Robust Fundamental Analysis

Thesis

M. Cavazza, 433085

Econometics, Quantitative Finance

October 30, 2017

Supervisors

dr. B. Diris	Erasmus University
dr. H.J.W.G. Kole	Erasmus University
prof. D. Van Dijk	Erasmus University
dr. M. Della Seta	APG

Abstract

According to market efficiency, it would not be possible to profit from mispricing in stock prices. In this scenario, in case investors are able to correctly price the outstanding information, there would not be mispricing signals to invest on. If instead, investors would not always value prices correctly, by estimating how, on average, investors reflect information into prices,it would be possible to detect firms where certain information have been over (under) weighted at a specific point in time. In order to estimate the weights that investors give to the information it is necessary to run a cross-sectional regression at company level with on the LHS the market cap and on the RHS the chosen explanatory variable. However, this analysis does not consider periods where the whole market has an irrational behavior (e.g. speculative bubble). I give evidence in this paper that it is essential to consider the time dimension to correctly estimate the full deviation of the market price from its fundamental. I achieve this by applying two different panel estimation techniques: Fama-MacBeth and Cointegration panel estimation. Moreover, I show what happens to the cross-sectional mispricing estimation when I add dummy variables to the model (Dummy for sectors, Dummy for Book to Market), in addition to the publicly available information, in order to control for differences in the price level for sub-groups of companies. I also give evidence that fundamental analysis actually works, by building a trading strategy that relies on price deviations from their fundamental values.

1 Introduction

The basic idea of Fundamental Analysis is that it is possible to estimate companies intrinsic fair values from publicly available information (e.g., Accounting Statement Variables) using simple statistical techniques. According to the efficient market hypothesis, Fundamental Analysis should not work, in the sense that it should not provide additional information over stock market valuations. In fact, if investors are able to correctly incorporate publicly available information into prices, mispricing would not exist and a company fair value would be fully reflected in its market valuation. Advocates of Fundamental Analysis support, often implicitly, a partially alternative view. Accord- ing to this view, the market portfolio is fairly priced on average and at each point in time. However, at company-level large deviations from a fair value can occur, and a Fundamental Analysis can help to identify them. Following this view, Fundamental Analysis can deliver a profitable strategy. If mispriced companies can be identified, and valuations tend to converge to their fair values, a strategy that goes long undervalued companies and short overvalued ones can earn abnormal returns. Bartram and Grinblatt (2014) supports this view. They show that a simple statistical analysis of publicly available information leads to a profitable mispricing signal. The purpose of this work is not to settle the dispute about market efficiency. The less ambitious goal is to propose a more sophisticated Fundamental analysis by relaxing the assumption that the market portfolio is fairly priced at each point in time. Indeed, there can be periods in which markets are mainly driven by sentiment (an example could be the so called DotCom bubble at the end of the nineties) and average valuations can deviate from their fundamentals. For this reason, I will assume that markets provide, on average, fair valuations but this occurs over a sufficiently long time horizon. Relaxing the assumption that the market portfolio is fairly priced at each point in time has direct implications on the estimation techniques. In a standard Fundamental Analysis, fair valuations could be obtained by fitting a series of cross-sectional regressions where the estimated coefficients represent, at each point in time, the fair marginal contributions to the firm value. If one assumes, as I do in this work, that the market can deviate from fundamentals in the short-run, the

time-series dimension of the data also becomes relevant. Fair marginal contributions can only be estimated taking into account a number of periods sufficiently large to cancel out the noise in the mispricing. For this reason, I estimate fair values through panel data techniques. In this way, the mispricing signals will have two dimensions, the cross-sectional dimension, and the time- series dimension. Therefore, the mispricing signals will be made by two components, one that takes into account price deviation from the intrinsic values at a company level (cross-sectional dimension), and one that takes into account the deviation of the whole market from its intrinsic value (time-series dimension). Panel data can be analyzed with different techniques, depending on the data type and the sample that we are dealing with. In my case, I have selected Fama-Macbeth cross-sectional regression due to its higher robustness in estimating the coefficients when time effects are present. As is understandable, fixed effects are not taken into account in the estimation. Instead, I add different dummies to the model (GICS sector dummie, book to market ratios dummie), in order to analyse if differences in sector or book to market have an impact on the company valuation. As is shown later in the paper, I find that the increase in the robustness of the analysis leads to better mispricing estimation. Indeed, when the time dimension is taken into account and the dummies are added to the model the portfolios perform in general better, leading to higher and less volatile positive returns. Another issue common to most of the Financial variables is the non-stationarity. From 1990 a lot of researchers have studied the effect of this issue on panel estimation, and they came up with different solutions to overcome the problem. I have tested for the presence of unit roots in my dataset and it resulted that most of my variables are indeed non-stationary. However, nobody ever tested the effect of this particular problem when Fama MacBeth is used as estimation technique. Therefore, as an extension of this work I will use a cointegration panel approach in order to estimate the long term relation between the variables. This will lead to more robust results, but it will also complicate the approach, losing the idea of simplicity implied by the Fundamental Analysis. In this framework, it will not only change the technique used, but also the interpretation of the mispricing. The cointegration approach, indeed, looks at the dynamic of the time series co-

movements; therefore, even the dynamic of the fundamentals (variables on the RHS) is modelled. In this way, the deviation of the price from the estimated fair value can be caused by movement on the RHS or the LHS of the model. This double interpretation of the mispricing is more realistic, and therefore more robust with respect to the Fama-MacBeth mispricing interpretation, where the deviation was simply due to the gravitation of the market cap around the fundamental value, where the latter had no dynamics. For both the techniques, the estimated mispricing signals will be used to divide the companies in different quintiles. Q5 will be identified with the most overpriced companies and Q1 as the most underpriced. Following this company division, I will base my trading strategy. I will set a long - short portfolio, that goes Long in Q1 and short in Q5. This procedure will be done every month, and the portfolio will be rebalanced. I evaluate the performances of the Fama-MacBeth approach for several averaging horizons (cross-sectional, 12 months, 24 months, and 60 months), and I obtain the best performances with the 12 months horizon. I also evaluate the performances of the cointegration model with an estimation horizon of 5 years, 60 months, and find out that it also leads to a profitable strategy, but not superior with respect to the Fama-MacBeth approach.

This paper will be an addition to the literature for different reasons. Firstly, it will shed some light on the market efficiency debate, investigating whether the markets are indeed efficient or at least they correct themselves in the long-run. Secondly, it will give more insight on the relation between Prices and Fundamentals. Thirdly, it will give some hints on whether it is possible to compute the Fundamental Analysis in a more robust way, in order to extend the findings of Bartram and Grinblaat (2014). All these goals will be reached by providing insight on the performance on an investing strategy base on the Fundamental Analysis. The Paper is divided into five sections. The first section is a Literature review, the second section is Data, the third section is the Methodology, the fourth section is Empirical Results, and the last one is the Conclusion.

2 Literature Review

2.1 Company Valuation

Company valuation is one of the main interests in the Financial Market. The ability of an investor to evaluate firms can be translated into higher profits. Indeed, researchers have put much effort to understand which model is better suited for the job. The direct consequence of all this effort is that we now have multiple ways to evaluate firms. Richardson et al (2010) reviewed, in their paper, all the models that are used to evaluate firms, trying to connect all the works done until that time on the subject.

Discounted Cash Flow The most common techniques of firm valuation are based on discounted cash flow models, where the assumption is that the price that we have now is simply the projection of future income. Frankel and Lee (1998), for example, use the EBO model (Edward Bell Ohlson model) to estimate firm fundamental values and use the ratio between this value and the actual price to predict long term cross-sectional returns. They find that this ratio better predicts cross-sectional returns in a long-term horizon. In addition, they have also shown that is possible to predict the forecast errors of future earnings prediction made by the analyst, and use this information inside the residual income model. This research is one of the few that studies if Fundamental analysis works because it tries to analyse the profitability caused by the price deviation from their intrinsic values. Indeed, it is rare to find literature that tries to prove whether the estimation of the fair value leads to abnormal returns, as I do in this work.

Multiples valuation Another way to valuate companies is by using multiples. Indeed, Baker and Ruback (1999) use this technique for the valuation, and argue that it has several advantages on the discounted method. It avoids all the estimation of future cash flows, and uses current account measures. Liu et al (2002) examine different value drivers to identify the best multiple, and find that forward earnings explain stock prices remarkably well. However, this methodology has some drawbacks: the basis of substitutability is chosen arbitrarily within the financial measures, choosing the multiple estimation technique is difficult, and finding a method to compare firms is also difficult. Even though these papers do not try to understand the profitability behind the company valuation, they indirectly study if Fundamental analysis works by researching how to explain stock prices. This is important information, also for my paper, because it gives hints on the best information to use to estimate the fair value.

Accounting variables Several accounting variables have been studied to understand their informative power in estimating the fair value or to explain stock returns. For example, Lev and Thiagarajan (1993) examine the value relevance of fundamentals[1] over earnings in explaining stock returns. Another example is, Jeffery et al. (1997), who investigate how informative fundamental signals are in respect of subsequent earnings. [1]

[1] the fundamentals are accounting variables that are used to estimate the fundamental values of companies.

The techniques that I use in my paper share the same goal as the methodologies cited before, estimating the intrinsic value of a company. However, the perspective of this analysis is different. Instead of using the basis of substitutability or future cash flow, this approach is based on financial statement analysis. The idea is that present accounting variables are connected with the current stock price, and by estimating the average link between the two, it is possible to back-out the fair value. This intrinsic value will then be indicated as a benchmark, and the stock price will move around it. This way of using the financial statement variables can be seen in the paper of (Bartram & Grinblatt, 2014), where they use the financial statement variable to fit the market cap at each point in time through a cross-sectional regression. In this way, the residuals can be interpreted as the mispricing signals, everything that is not explained by the information that can be found in the financial statement. Similarly to (Frankel & Lee, 1998), they also show the profits that a trading strategy based on the mispricing signals can generate, and find that this technique can lead to a significant abnormal return. I will as well estimate a model to evaluate over-under valuation of stocks, but I will not use the market cap in order to estimate the intrinsic value of the firms like (Bartram & Grinblatt, 2014). Instead, I will use the cumulative return, that still contains the time dimension, but they are not affected by the different size levels. Then I will explore the profitability of an investment strategy based on the mispricing signal, like (Frankel & Lee, 1998).

In my paper, I use the literature that analyse how informative are certain fundamentals (Lev & Thiagarajan, 1993), (Abarbanell & J.Bushee, 1997), to make a choice driven by theory when it is needed to make a variable selection like I do to estimate the intrinsic value. Like in most of the literature regarding the anomalies, a long - short portfolio is created in order to prove that Fundamental analysis works. This approach is, indeed, similar to the one used by Fama and French (1993), where they create the small-minus-big and the high-minus-low factor to capture the variation of stock returns together with the return on the market.

3 Data

Before describing the methodology, it is better to look at the data used in this work first. As stated before, the information used in the fundamental analysis is taken from the publicly available company Statements: balance sheet, income Statement, and cash flow statement. The period that I am considering is from 31/02/1999 till 31/08/2016, where I observe the monthly market cap for each company belonging to the MSCI US. In this way, I will have 211 trading months in my sample, and a number of firms that change over time, from 208 to 607. The variables selected to fit the cumulative returns are the most used financial statement variables for company valuation. I have also selected the variables depending on the availability, in order to have the largest sample of observation at each point in time, and the longest time series. The selected variables can be found in table 1. These variables do not have the same frequency as the market value, they indeed have a quarterly basis. I collected the data from MSCI US for the same interval of time that I have considered for the market cap. The variables that I am using in this paper I have selected them based on mainly three papers: (Bartram & Grinblatt, 2014), (Lev & Thiagarajan, 1993), and (Abarbanell & J.Bushee, 1997). In addition to the those variables I have added the "Past year return volatility" and "Beta" to see how the volatility and the exposure to the systematic risk are related with market cup. This is done to limit the problems of parameter uncertainty and multicollinearity. The following table shows the variables [1] that I considered.

Table 1: This table shows the different variables that I considered. The table is divided into four columns depending in which Company Statement is possible to find those variables. "D" is the abbreviation for Dummy

Balance-Sheet	**Income Statement**	**Cash Flow**	**Others**
Total Assets	Net Income	Cash and Short term Investment	Past year Return Volatility [2]
Current Assets	Sales		Beta [3]
Liabilities	Income Taxes		D for sectors
Stockholders Equity	Pre-Taxes Income		D for Book to Market
Dividends	Accounting Receivable		D for Dividend Yield
PPE	Operational Income		
Total Debt	Inventory		
Long-Term Debt			

In this paper, I did not insert inside the model all the variables showed in the table due to estimation uncertainty. More explanation of this is in the findings section.

When I proceed with the cointegration approach I further restrict my sample in order to obtain a balanced panel. This choice was taken in order to not complicate even further the cointegration analysis. It is also important to say that, the techniques that are used in this paper have only been

[1] A more detailed description of the variable can be found in table 35

[2] To calculate this variable I have taken the return on the past 12 trading months and calculated the standard deviation.

[3] This variable quantify the exposure to beta factor of the firm. If a firm has Beta equal to 100 it means that it has the same systematic risk as the market, and in case the variable has value > 100 (< 100) it means that its systematic risk is higher (lower) compared to the market. The summary statistics after in the paper, will show an average value of this variable higher then 100. This is due to the fact that in APG database the variable is calculated using a bigger sample of firms, and not only MSCI US.

used in the context of simulated data.

Characteristics of the data:

1. Monthly and Quarterly Frequency.
2. MSCI US DataBase.
3. The firm needs to have a Price per Share higher then 5$.

All the variables that I use, except for the market cap, are lagged 3 months to take into account the real availability of the data.

3.1 Dummy Variables

To make the analysis more robust, I will try to understand if there are valuation differences between sub-groups of firms. Therefore, I add dummy variables to my model in order to evaluate this differences.

The dummies that are used in this work are:

- GICS Sectors
- Book to Market

The reason why I chose the dummy for sectors is because firms that are active in different environments, might get priced in a different way by the market.

The other dummy variable follows from the findings of Bartram and Grinblatt (2014), where they notice a difference between the book to market of the underpriced and overpriced firms. Overvalued companies had lower B/M compared to the Undervalued ones. I created the dummy variables by firstly sorting the companies in five quintiles based on the B/M, from the lowest to the highest, and then I created the variables with value 1 if the company belong to that quintile and 0 elsewhere. This means that in the first quintile there will be the variable with the lowest book to market value and the fifth quintile there will be the companies with the highest book to market ratio.

Since I am using the dividend yield as my independent variable, I have the problem that some firms do not share any dividend. To overcome the issue, I have set their dividend yield equal to 0,and I have added a dummy variable with value 1 in case the company does not give any dividend.

3.2 Data Analysis

In this section I will analyse the data just described and the sample used. Table 2 shows the summary statistics of the variables that I am using, and table 3 shows the correlation matrix.

Table 2: This table shows the summary statistics of all the variables used in the regression. In the first row we can see, in order: Minimum value, Maximum value, Mean, Median, Standard Deviation, Correlation with Market Cap, Skewness, Kurtosis, Minimum number of observation, Maximum of observation, number of times that the variable is available and the percentage of non rejection of the adf test. Standard Deviation and Correlation, are calculated by taking for each variable the mean across time for each observation, and then I have made the calculation with the resulting vectors. Differently Mean, Median, Skewness and Kurtosis are calculated by first calculating the measure at each point in time across the observations and then average the resulting values across time. The percentage of non rejected adf test is calculated by estimating the adf test per individual and then divide the number of companies that do not reject the unit root presence hypothesis with the total number of companies.

	min	max	mean	median	std	corr	Skewness	Kurtosis	Min Obs	Max Obs	Time	%ADF
1.Total Assets	0.01	2577.15	41.22	8.21	11.19	-0.1	7.15	67.76	30	632	225	0.94
2.Current Assets	0	3692.08	61.12	19.74	12.55	0.22	7.51	83.5	32	526	224	0.66
3.Dividend Yield	0	95.37	1.7	1.04	0.36	0.04	2.32	17.65	30	632	217	1.00
4.Income Pre Tax	-60556	24573	383.06	119.85	148.97	0.17	2.5	55.16	30	631	225	0.11
5.Income Tax	-35645	17011	128.72	38.37	39.31	0.27	4.39	77.01	30	628	225	0.10
6.Inventories	0	515.01	14.4	2.61	4.75	0.05	5.98	59.69	31	561	224	0.82
7.Liabilities	-0.15	2341.28	32.36	4.82	9.22	-0.15	7.38	70.41	30	632	225	0.89
8.Long Term Debt	0	6281.48	78.77	19.56	20.33	-0.24	9.06	106.05	37	631	223	0.94
9.Net Income	-61.66	18.99	0.26	0.08	0.12	0.11	1.86	59.88	38	631	223	0.09
10.Operational Income	-6333.3	2582.1	47.63	15.77	12.38	0.11	3.48	48.3	30	631	225	0.24
11.Sales	-6.4	1315.65	40.03	11.74	6.59	0.13	5.74	51.97	30	631	225	0.54
12.Shareholder Equities	-911.42	2670.69	83.27	29.96	21.38	0.06	4.95	36.02	30	632	225	0.97
13.Cash	0	18922.8	216.27	56.51	92.03	-0.08	8.99	106.05	30	573	225	0.51
14.Other Assets	-1826.14	11253.73	113.47	19.96	28.48	-0.34	8.5	93.02	30	629	224	0.82
15.Total Debt	0	9991.99	122.61	22.21	30.55	-0.29	8.82	95.54	30	632	224	0.93
16.PPE	-4.24	2526.68	56.88	11.09	11.7	0.3	5.5	47.18	303	623	258	0.94
17.Betas	-164.36	400.11	109.79	109.21	8.92	-0.65	0.57	3.38	315	640	237	0.89
18.Return Volatility	0	1755.99	11.27	8.19	7.88	0.23	5.4	58.13	333	708	261	0.99
19.Price	0.61	6969.17	181.1	60.53	69.45	1	4.88	36.05	311	638	261	1.00

The variables are scaled in order get better comparable coefficients after the estimation. Looking at the higher moments we see that all the distributions are seriously skewed and not normal. This would suggest that a transformation of the variable could be necessary, however I have only applied the log transformation to the variables: ”Total Assets” and ”Total Debt” [1]. The correlation are all lower then 0.5, in absolute terms, apart for the variable ”Betas” that has the highest level. I have also run single ADF tests at the company level and calculated the percentage of acceptance of

[1] The decision to make the log transformation only for those two variables depends on the relation between them and the cumulative return, where I noticed from a series of scatter plot, where on the x axis I was plotting one independent variable and on the y axis the cumulative return, that those specific two variables needed a log transformation.

Table 3: This table shows the correlations between the variables that I use. In order to calculate them I have firstly estimated the correlations at each point in time, and then averaged the values across the period.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1.Total Asset	1																		
2.Current Assets	0.86	1																	
3.Dividend Yield	0.24	0.17	1																
4.Pre-Tax Income	0.44	0.48	-0.37	1															
5.Income Tax	0.22	0.25	-0.31	0.83	1														
6.Inventories	0.83	0.93	0.19	0.52	0.31	1													
7.Liabilities	0.99	0.79	0.25	0.38	0.19	0.76	1												
8.Long Term Debt	0.89	0.6	0.27	0.26	0.14	0.59	0.93	1											
9.Net Income	0.44	0.48	-0.36	0.98	0.73	0.52	0.38	0.25	1										
10.Operational Income	0.66	0.7	-0.18	0.92	0.73	0.74	0.59	0.44	0.89	1									
11.Sales	0.87	0.9	0.25	0.49	0.3	0.91	0.81	0.65	0.48	0.71	1								
12.ShHol Equity	0.91	0.95	0.2	0.55	0.28	0.95	0.84	0.68	0.56	0.76	0.93	1							
13.Cash	0.91	0.91	0.16	0.49	0.19	0.86	0.87	0.71	0.52	0.7	0.82	0.94	1						
14.Other Assets	0.72	0.35	0.18	0.18	0.04	0.31	0.79	0.84	0.19	0.29	0.44	0.44	0.53	1					
15.ToT Debt	0.68	0.28	0.17	0.06	0.06	0.23	0.76	0.87	0.04	0.17	0.4	0.34	0.4	0.87	1				
16.ppe	0.78	0.96	0.27	0.41	0.16	0.92	0.69	0.51	0.42	0.62	0.88	0.93	0.86	0.23	0.14	1			
17. Betas	0.34	0.17	0.27	0	-0.19	0.1	0.34	0.38	0.04	0.13	0.22	0.3	0.4	0.38	0.31	0.14	1		
18.Return Volatility	-0.41	-0.36	0.52	-0.65	-0.4	-0.43	-0.38	-0.27	-0.7	-0.67	-0.33	-0.45	-0.51	-0.22	-0.12	-0.24	-0.12	1	
19.Market Cap	-0.11	0.22	-0.11	0.16	0.26	0.16	-0.16	-0.3	0.1	0.11	0.12	0.05	-0.1	-0.34	-0.33	0.27	-0.65	0.23	1

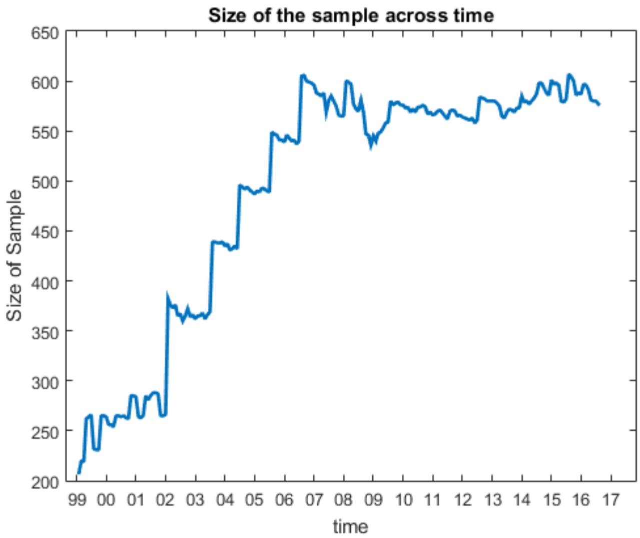
the unit - root presence. I have added an intercept as deterministic term in the adf test. However, results of a more robust way of testing for this hypothesis will be shown later in the paper.

Table 3 shows the correlation between the variables including the market cap. As shown in the previous table the correlation between the market cap and the other variables are all quite low. On the contrary, between the other variables the correlations are higher suggesting that adding all the variables in the model might cause multicollinearity. Indeed, the number of variables that I select to enter in the model are just a sub selection of the ones listed in the first column of this table.

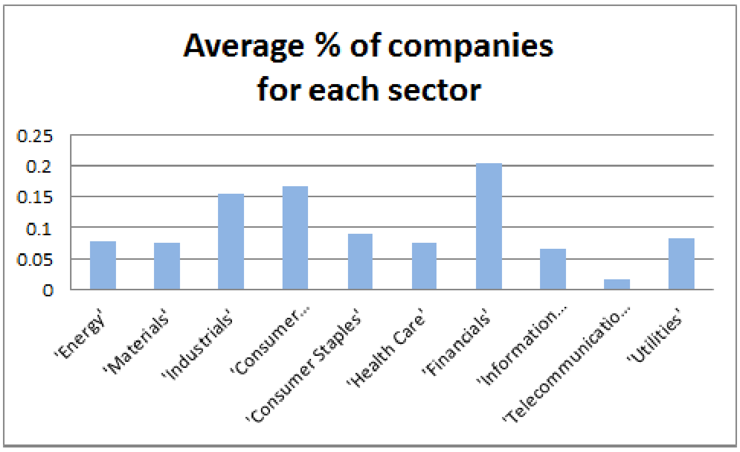
In figure 1a it is represented the sample size over time. I start the analysis only in the case that the number of observations in my sample is over 150, due to the estimation uncertainty of the parameters. The number of companies in the sample is not monotonically increasing due to companies that go out from the MSCI US Index.

In figure 1b, it is represented an overview of the average percentage of companies per sector that I have in my sample over the entire period

Figure 1: On the left figure, it is plotted the sample across time. At each point in time I calculate the number of companies that enter in the model estimation. On the right figure, I plot the histogram representing the average percentage of company belonging to the different sectors. The average is taken over the entire period that I consider to estimate the model



(a) Sample across time



(b) Average percentage companies per sector.

4 Methodology

Fundamental analysis is based on the concept that with publicly available information it is possible to back-out the fair value of a company. In this paper, I try two different techniques to estimate the intrinsic value of companies: the Fama-MacBeth approach and the panel cointegration approach. I discuss both techniques in the following paragraphs.

In most of the literature the focus is directed towards returns, where fundamentals signals or other variables are used in order to explain or forecast the returns. A possible model would be:

$$r_{it} = x_{it}\beta_{xt} + \varepsilon_{it}, \tag{1}$$

where r_{it} is the return of company i at time t, and x_{it} are the explanatory variables.

Using returns is common practice in all the Financial Literature, due to their better statistical characteristics. Prices are difficult to handle and therefore difficult to find in the literature. There are some exceptions, for instance Bartram and Grinblatt (2014), where they use market cap on a cross-sectional base. The model used in their paper is:

$$P_{it} = \boldsymbol{X}_{it}\beta_{xt} + \varepsilon_{it}, \tag{2}$$

where P_{it} is the market cap of company i at time t, and $\boldsymbol{X}_{it}$ are the explanatory variables. As explanatory variables they have selected variables from the financial statements. Using this model they have estimated the coefficients to then calculate the residuals, measures that they have used as mispricing signals.

This approach is problematic in the way that market cap scale with size and therefore the model needs to account for these differences, otherwise there would be a problem of heteroscedasticity. Even though I am following most of the work done by Bartram and Grinblatt (2014), in order to take into account this issue I use cumulative returns as dependent variable.

The general model representation of my thesis will be :

$$R_{it} = \frac{P_{it}}{P_{i0}} = \frac{\alpha_t t_i + \boldsymbol{X}_{it}\beta_{xt} + \varepsilon_{it}}{P_{i0}}, \tag{3}$$

where R_{it} is the cumulative return of the company i at time t and $\boldsymbol{X}_{it}$ are the explanatory variables and α_t is the coefficient of the individual time trend t_i at time t that I have added to model the effect of time on the cumulative returns. Similar to Bartram and Grinblatt (2014) I use finanicial statement variables on the right hand side.

The individual time trend is a variable that take value 1 in case it is the first time that the company enter in my sample, and will increase its value by 1 for every month that the company is in the sample. This avoids that companies that have been in my sample for a longer period are detected as overvalued or undervalued with respect to the ones that just entered in my sample.

Using this model I estimate the mispricing signals that I will use to detect undervaluation (overvaluation) of firms. In case all the relevant explanatory variables are inside the model,the mispricing signals would just be the residuals that I use as guidance to forecast future stock returns, as in Bartram and Grinblatt (2014). To test whether this is indeed the case, in the second part of the thesis the mispricing signals are used to create a trading strategy.

In the following paragraph I will explain two methodology that I have applied in order to estimate the coefficients, and therefore the mispricing signals.

4.1 Fama-Macbeth Regression

The assumption behind this thesis is: the market prices are, on average, correct over a long horizon. To work in this framework, the statistical technique needs to be more complicated than the cross-sectional regression used by Bartram and Grinblatt (2014). The coefficients of the estimated model need to be the proper weights of the value drivers over a certain period. Therefore, I need to use panel analysis instead of cross-sectional regressions.

I have decided to use the Fama-MacBeth[1] approach that is a widely-applied technique in Econometrics to model panel data. It has the advantage of being more robust to time effects with respect to the OLS estimator, and it also leads to consistent and asymptotically normal estimates under certain regularity conditions, Skoulakis (2008).

This approach it is still not statistically complicated. Indeed, I have decided to choose as a starting point a simple technique. The reason is that, since the relation between the price and the publicly available information depends only on how investors decide to react to certain news, if they incorporate the information in a simple way, then there is no need to use a highly sophisticated model.

The way the coefficient are estimated it is shown here:

$$\alpha = \frac{1}{T}\sum_{t=1}^{T}\alpha_t \qquad\qquad \beta_x = \frac{1}{T}\sum_{t=1}^{T}\beta_{xt}, \tag{4}$$

[1]The motivation that pushed me to use the Fama-MacBeth approach can be found in table 8. By looking at the ADF tests made on each beta time-series, we see that the Hypothesis of unit root presence is rejected for all the variables. This means that all the beta time series revert to an unconditional mean. Therefore, by averaging the betas across a period of time it is possible to get rid of all the noise around the estimation, and end up with a smoothed estimate of the relation between the variables.

$$R^*_{it} = \frac{\alpha t_i + \boldsymbol{X}_{it}\beta_x}{P_{i0}}, \tag{5}$$

R^*_{it} is the estimated fair value of the company, α is just the average across the cross sectional intercepts, and β_x is the average across the cross sectional vector of the explanatory variables coefficients.

This estimation technique is based on cross-sectional regressions where at each point in time, I estimate a cross-sectional model, equation 3. Then, by simply taking the average across these coefficients, like we can see in equation 4, it is possible to estimate the coefficients for the period of interest. The coefficients are averaged using a rolling window of length depending on the selected horizon, while the cumulative returns are calculated using an expanding window. Meaning that the cumulative return are always based on P_{i0}, and the time trend starts the counting in the first moment the company enter the MSCI US index. The periods that I have considered to average the cross-sectional coefficients are: 1 month horizon, 12 months horizon, 24 months horizon and 60 months horizon.

When I have estimated the coefficients I use them to fit the model and calculate the intrinsic values of the companies, as it is shown in equation 5.

Since on the left hand side of the regression I have divided P_{it} by P_{i0}, I will also need to adjust the financial statement variables that I use in order to fit it. Therefore I divide the right hand side elements also by the P_{i0}, in this way I maintain the proportion between the value of the fundamentals and the market cap. Regarding the dividend yield, since it is already divided by the market cap, I before multiply it by P_{it} and then I divide it by P_{i0}.

In addition to these variables I add two more in order to overcome two issues. The first issue is the changing in the number of outstanding shares due to SEO or buy-backs, that increases and decreases the market value of the companies, therefore I add the variable $\frac{\boldsymbol{NofShares}_{it}}{\boldsymbol{NofShares}_{i0}}$.

Even if on the right hand side of the model we have quarterly information and the frequency of the estimation is monthly, the coefficients change because the dependent variable change each month. The fluctuations of the slopes can be interpreted as the changes in the way investors evaluate companies. Across time, investors give different weights to the financial information, and in certain periods they might overweigh or underweigh certain characteristics. Therefore, in certain intervals of time the coefficients can deviate from their true mean due to investors irrational behaviour. Averaging the coefficients is therefore the perfect solution to smooth out all the noise caused by the market irrationality.

Once I have estimated the intrinsic value I will also be able to derive the mispricing signals just by looking at the model residuals, that is the difference between R_{it} and R^*_{it}.

$$\boldsymbol{\kappa}^*_{it} = R_{it} - R^*_{it} = \beta_{0t} + \boldsymbol{X}_{it}\beta_{xt} + \varepsilon_{it} - \beta_0 - \boldsymbol{X}_{it}\beta_x = \tag{6}$$

$$(\beta_{0t} + \boldsymbol{X}_{it}\beta_{xt} - \beta_0 - \boldsymbol{X}_{it}\beta_x) + \varepsilon_{it} \tag{7}$$

Equation 7 shows how the mispricing signal, $\boldsymbol{\kappa}^*_{it}$, is calculated and of what is composed: a cross-sectional component ε_{it}, that measures the deviation from the fair value at a company-level, and a time component (the difference inside the parenthesis) that measures deviation from the fair value of the entire market.

4.2 Cointegration Panel Model

However, the real problems of using cumulative returns in the model arise when the analysis moves from a cross-sectional towards a time series framework. The presence of a unit root in the time series is common, and the estimation techniques needs to account for this issue.

Non-stationarity is a problem of the dataset that can lead to the estimation of a spurious relation. This phenomenon depends on non-stationary variables having trends, and if variables that are not related share a similar trend, this can cause the model to inflate (deflate) their relation. In this case it is important to see if the variables are cointegrated.

To explain the meaning of cointegration, I will take the easiest example. In time series two I(1) (non-stationary) series are said to be cointegrated if there is a linear combination of the two that leads toward a series that is I(0), meaning that it is stationary. When we are dealing with linear models, we can say that the variables are cointegrated in case the residuals of the model are I(0). As a consequence, the two series will converge after a divergence. This is translatable as a true relation between them.

This means that if I have a matrix

$$\mathbf{Y_t} = [y_{1t}, y_{2t}, ..., y_{nt}], \tag{8}$$

where n is the number of time series. The time series will be cointegrated, if exist a vector $\boldsymbol{\beta}$ such that:

$$\boldsymbol{\beta Y_t} \sim I(0) \tag{9}$$

This vector of cointegrative coefficients, $\boldsymbol{\beta}$, is not unique because it can be multiplied by an infinite number of constants, $c\beta$. We can indeed normalize the vector in order to get:

$$y_{1t} + ... + \beta^* y_{rt} \sim I(0) \tag{10}$$

$$\boldsymbol{\beta^*} = c\boldsymbol{\beta} = \left[1, -\beta_2, ... - \beta_r\right], \tag{11}$$

we can rewrite this system as:

$$y_{1t} = \beta_2 y_{2t} ... + \beta_r y_{rt} + u_t, \tag{12}$$

where u_t is the stationary series. In this case the cointegrative relation is written as a linear relation between the y_{1t} and the other variables.

The reason why I have decided to estimate a panel cointegration model on my data is that when two variables are cointegrated you can identify two different dynamics in their relation, the long-term dynamics and the short-term dynamics. We can see the long-term dynamics as the equilibrium relation between the variables, and the short term dynamics as the fluctuations around the equilibrium relation. In my case it is important to account for these two components, since the financial market is characterized by movements around a long term trend that are also reflected in the estimated relation between the cumulative return and the financial statement variable inside my model.

With the Fama-MacBeth technique I try to control for these fluctuations by averaging the coefficients across-time, but the problem is that you never know for how long these deviation will continue to grow and when they will end.

Another reason why I have considered the cointegration model comes from the paper of Campbell and Shiller (1987) and the discounted dividend model, where they say that, if present value theory is correct, market prices and dividends are cointegrated. Therefore, changes in the market prices are caused by changes in the future dividend or changes in the discount rates, or both. Indeed, in their paper, they give evidence for this relation. They find a weak cointegration relation between prices and dividends, but also that the deviation from the present value model tend to remain for a long time and that the "spread", difference between the price and a multiple of the dividend, is too much volatile. I indeed will try to incorporate this cointegrative relation between the dividend and the price, while I will also add extra variables to try to better estimate the present value of the companies.

In this framework the regressors in equation 3 are integrated processes, this means that each variable can be written as:

$$x_{it} = x_{i,t-1} + u_{it} \tag{13}$$

when the regression errors of equation 3 that are stationary, the model represents a cointegrated relation between R_{it} and x_{it}.

The Bartram and Grinblatt (2014) assumption that deviations of the market caps from the fair value are more likely to decrease rather than increase, can be caused by the stationarity of the residuals. Therefore, it would mean that the residual of the model will revert toward a mean, and that the difference between the company value would decrease. This is the perfect example of cointegration, where there is a combination of the market caps and the explanatory variables that lead towards a stationary series. Indeed, if Bartram and Grinblaat assumption reflect the stationarity of the residuals the cointegration model should lead me towards the estimation of long-run coefficients.

In equation 3, we can see that the dependent variable is the cumulative return at company level. Since companies interact in the same markets, it is hard to believe that there are no dependencies between them. Therefore it is important to check that the companies are not cross-sectionally dependent.

Cross Sectional Dependence Assuming Independence across the observations is a strict assumption. From the moment that in my sample I have a large N (cross-section), it is important to explore this issue. Therefore, I use a simple test to check if my observation are cross-sectionally dependent. The test is called the CD test and it was created by Pesaran (2004).

I hereby show the formulas to compute the test:

$$CD = \sqrt[2]{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{p}_{ij} \right), \tag{14}$$

where p_{ij} is the correlation coefficient of the disturbances and is calculated as:

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} \varepsilon_{it}\varepsilon_{jt}}{\sqrt[2]{\sum_{t=1}^{T} \varepsilon_{it}^2} \sqrt[2]{\sum_{t=1}^{T} \varepsilon_{jt}^2}}, \tag{15}$$

ε_{it} is the residuals from an OLS regression, $y_{it} = \beta_i' x_{it} + \varepsilon_{it}$ ran for each observation with $i = 1, ..., N$ for the whole time horizon T. Under the null Hypothesis of cross-sectional independence the CD test distributes as $N(0, 1)$.

This test is applicable to different types of panel models including heterogeneous panels with short T and large N.

This test is an alternative of the Breush and Pagan (1980) test, that is affected from a size distortion when the time dimension is lower then the cross sectional dimension. The test is just the sum of all the possible pair correlation between the observations, multiplied by the square root of T and by a term that allows the test statistic to have a normal distribution with mean 0 and variance 1 under the H0.

In order to account for the cross-sectional dependence in my model I assume that the error have a factor structure.

$$\varepsilon_{it} = \lambda_i' F_i + e_{it}, \tag{16}$$

where λ_i are the factor loadings for the individual i, F_i are the common factors and e_{it} are the independent residuals.

Bai and Kao (2005) have proven that, in this framework, the coefficients estimated with OLS present a bias and they have suggested the CUP-FM technique, in order to estimate unbiased coefficients.

The subset of variables that I will use in the model is much more restricted in respect of the previous framework. The reason, is that the cointegration relation is a characteristic of stochastic processes. I select two variables that I believe cointegrate with the cumulative return. The first one is dividends, I have selected this one following the findings of (Shiller & Cambell, 1987). The second variable is sales, I have included this variable to also include firms that do not share any dividends. I will also add a dummy variable to estimate an intercept for the non-sharing-dividend firms.

In the next section I will show how to make an unbiased estimate of the coefficients, β_{fm}, using the CUP-FM technique.

4.2.1 Continuous Updated Fully Modified Estimator

The Continuous Updated Fully Modified Estimator (CUP-FM), Bai and Kao (2005), accounts for cointegration in panel datasets that have cross-sectional dependencies. It is an extension of the better known Fully Modified Estimator proposed by Phillips and Hansen (1990).

It is important to get familiar with the estimation of the long-term covariance matrix, and the one-sided long-term covariance matrix which are both used in the estimation.

First let's define the long-term covariance matrix, calculated on v_{it}, where $v_{it} = \left[\, F_t' \ e_{it} \ \Delta x_{it}' \,\right]'$, that is a vector formed by the estimated factors of the residuals, F, the cross-sectional independent residuals, e, and the change in the independent variable, Δx, with i describing the individual and t describing the time.

$$\Omega_i = \sum_{t=-\infty}^{+\infty} E(v_{i0}v_{it}') = \Sigma_i + \Gamma_i + \Gamma_i' = \begin{bmatrix} \Omega_{Fi} & \Omega_{Fei} & \Omega_{Fui} \\ \Omega_{eFi} & \Omega_{ei} & \Omega_{eui} \\ \Omega_{uFi} & \Omega_{uei} & \Omega_{ui} \end{bmatrix}, \tag{17}$$

where

$$\Sigma_i = E(v_{i0}v_{it}') = \begin{bmatrix} \Sigma_{Fi} & \Sigma_{ei} & \Sigma_{Fui} \\ \Sigma_{eFi} & \Sigma_{ei} & \Sigma_{eui} \\ \Sigma_{uFi} & \Sigma_{uei} & \Sigma_{ui} \end{bmatrix}, \tag{18}$$

and

$$\Gamma_i = \sum_{t=1}^{+\infty} E(v_{i0}v_{it}') = \begin{bmatrix} \Gamma_{Fi} & \Gamma_{Fei} & \Gamma_{Fui} \\ \Gamma_{eFi} & \Gamma_{ei} & \Gamma_{eui} \\ \Gamma_{uFi} & \Gamma_{uei} & \Gamma_{ui} \end{bmatrix} \tag{19}$$

The one-side long-term covariance matrix is estimated as follows.

$$\Delta_i = \Sigma_i + \Gamma_i \tag{20}$$

Let's also define two other matrices that will be useful in the estimation

$$\Omega_{bi} = \begin{bmatrix} \Omega_{Fi} & \Omega_{Fei} \\ \Omega_{eFi} & \Omega_{eFi} \end{bmatrix} \qquad \Omega_{bui} = \begin{bmatrix} \Omega_{Fui} \\ \Omega_{eui} \end{bmatrix} \tag{21}$$

Now I will show how to estimate the unbiased vector of coefficients.

The whole point of the Fully modified estimator is to correct the data for endogeneity and serial correlation, and re-estimate the regression after this correction.

let's just take 2 series from the example in 12.

To achieve the endogeneity correction we will need to correct the dependent variable:

$$R_{it}^* = (R_{it} - \bar{R}_{it}) - (\lambda_i'\Omega_{Fui} + \Omega_{eui})\Omega_{ui}^{-1}\Delta x_{it}, \tag{22}$$

and in order to correct for the serial-correlation I will need to estimate the following matrix:

$$\Delta_{bui}^{*} = \Delta_{bui} - \Omega_{bui}\Omega_{ui}^{-1}\Delta_{ui} \tag{23}$$

$$\Delta_{bui}^{*} = \begin{bmatrix} \Delta_{Fui}^{*} \\ \Delta_{eui}^{*} \end{bmatrix}, \tag{24}$$

with this two correction I am able to estimate the β_{fm}:

$$\beta_{fm} = \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\hat{R}_{it}^{*}(x_{it}-\bar{x}_i)' - T(\hat{\lambda}_i'\hat{\Delta}_{Fui}^{*} + \hat{\Delta}_{eui}^{*})\right)\right]\left[\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'\right]^{-1} \tag{25}$$

These corrections are made in order to clean the model from any short-term dynamics that can be caused by endogenous variables not included in the model and from a feedback dynamic of prices. This correction will only maintain the long-term dynamic.

Feasible estimation Following Bai and Kao (2005) I will show how to feasibly estimate the different matrix. First of all I will explain how to obtain $\hat{v}_{it}$.

In order to obtain $\hat{v}_{it}$ I will need to estimate the regression errors, e_{it}.

$$\hat{e}_{it} = (R_{it} - \bar{R}_i) - \hat{\beta}(x_{it} - \bar{x}_i), \tag{26}$$

where the $\hat{\beta}$ is estimated through an OLS technique.

I then estimate the factor structure of the residual using a PCA (Principal Component Analysis). This will produce a number R of factors, that are perfectly suited to model the co-movement of the series. Once I have all the component of the w_{it}, I estimate the long-term covariance matrix and the one-side long-term covariance matrix using the formula :

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=1}^{T}\hat{v}_{it}\hat{v}_{it}' + \frac{1}{T}\sum_{\tau=1}^{l}\omega_{\tau l}\sum_{t=\tau+1}^{T}(\hat{v}_{it}\hat{v}_{it-\tau}' + \hat{v}_{it-\tau}\hat{v}_{it}')\right\}, \tag{27}$$

where ω is a kernel function, in this work I have selected the Bartlett kernel. I assume in this paper that $\Omega = \Omega_i$. In order to estimate the one-sided long-term covariance matrix a similar formula is used. Again the assumption is that $\Delta = \Delta_i$.

Once I have estimated all the matrix that I need, then I am able to estimate all the corrections and therefore also eq. 25.

However in this thesis I am not directly interested in the relation between the dependent variable (cumulative return) and the independent variables. Instead, I am interested in which one are the mispricing signals. Following Bai and Kao (2005) the residuals of the CUP-FM will look like:

$$\hat{e}^*_{it} = \hat{e}_{it} - (\hat{\lambda}'_i\hat{\Omega}_{Fui} + \hat{\Omega}_{eui})\hat{\Omega}^{-1}_{ui}\hat{\Delta}x_{it}, \tag{28}$$

and this ones will be the residuals of the model that looks like:

$$R^*_{it} = \beta_{fm}(x_{it} - \bar{x_i}) + e^*_{it}, \tag{29}$$

where the R^*_{it} is the dependent variable corrected for endogeneity, $\beta_{fm,p}$ are the coefficients, and e^*_{it} is the residual, the mispricing signal that I will use.

I use these residuals because they describe the unexplained variance of the cumulative return, when I consider the long-term dynamics and the short-term dynamics. Therefore, we can consider these residuals as the unexplained component of cumulative return that I assume to contract rather then expand.

This estimation is done iteratively, because as Bai and Kao show, this improves the small sample properties compared to the Fully Modified Estimator. This means that at each iteration I estimate the residuals and I use them to run the next iteration. In order to do so, at the end of every iteration, the coefficients are used to estimate the residuals $\hat{e}_{it}$, as follows:

$$\hat{e}_{it} = (R_{it} - \bar{R}_i) + \hat{\beta_{fm}}(x_{it} - \bar{x}_i) \tag{30}$$

4.2.2 Unit Root Test

Before proceeding with the cointegration estimation it is important to test the non-stationarity of the variables. To test it, I need to check the presence of a unit root in the data. This is a normal procedure in time series analysis, where tests like the Augmented Dikey Fuller are implemented.

Lots of literature has been produced regarding unit root tests, cointegration tests and cointegration panel estimation in the past 25 years. Levin, Lin and Chu (2002) is an example, where they proposed and studied the LLC test, with the purpose to check for unit root presence in a panel dataset. By studying the asymptotic proprieties of this test, they found that it has a higher power compared to individual ones.

The LLC test together with most of the panel unit root tests makes an assumption that helps to simplify the analysis, they assume cross-sectional independence across the observations. This assumption can become too restrictive, especially when the number of cross-sections increases. Likely, unit root tests that account for this issue have been studied like in Moon and Perron (2004c), where they use a dynamic factor model to capture the error dependence.

In this thesis I am dealing with panel data, therefore I am not dealing with one time series but for each variable I have N (number of observation) time series. This leads to complications in the test estimation, but also to an increase of its robustness. It is indeed known that this type of test has a low power, and even the not refusing of the H0 cannot be automatically translated in the presence of unit root.

The test that I have selected is the one suggested by Moon and Perron (2004c), because it accounts for cross-sectional dependence and it is particularly adapt in case you have a large cross-section. This test is based on a dynamic factor model:

$$y_{it} = \alpha_i + y_{it}^0 \tag{31}$$

$$y_{it}^0 = \rho_i y_{i,t-1}^0 + \epsilon_{it}, \tag{32}$$

where ϵ_i are unobservable error terms with a factor structure and α_i are fixed effects. The unobserved errors are generated by $\boldsymbol{R}$ random factors and from idiosyncratic shocks, like:

$$\epsilon_{it} = \lambda_i^{'} f_t + e_{it}, \tag{33}$$

where λ_i are the non-random factor loadings for observation i, f_t are the random factors, e_{it} are the idiosyncratic components. The number of factors that generates the unobservable errors is unknown. The correlation that is between the cross-sectional units it comes from this unknown factors, and the extent of the correlation is determined by the factor loadings coefficients. By defining $Q_\lambda = I - \lambda(\lambda'\lambda)^{-1}\lambda'$ as the matrix projection onto the orthogonal space with respect to the factor loadings, is possible to calculate the de-factored data and residuals as, YQ_λ and eQ_λ, that do not have any cross-sectional dependence. Y is the matrix of data containing N observation, observed for T times. Moon and Perron (2004c) estimate the pooled autoregressive coefficient defined as, ρ^+_{pool}. To estimate the factor structure of the residuals I use the most known PCA.

We then have:

$$Y^0 = Y^0_{-1} + \lambda^{'} f + e \tag{34}$$

To get the de-factored model we multiply per Q_λ

$$Y^0 Q_\lambda = Y^0_{-1} Q_\lambda + e \tag{35}$$

In this setting the pooled estimate is

$$\hat{\rho}^+_{pool} = \frac{tr(Y_{-1}Q_\lambda Y') - NT\tau^N_e}{tr(Y_{-1}Q_\lambda Y'_{-1})} \tag{36}$$

To test our hypothesis of $H0 : \rho^+_{pool} = 1$ it is possible to apply two different t-tests:

$$t_a = \frac{\sqrt[2]{N}T(\hat{\rho}^+_{pool} - 1)}{\sqrt[2]{\frac{2\phi^4_e}{\omega^4_e}}} \tag{37}$$

$$t_b = \sqrt[2]{N}T(\hat{\rho}^+_{pool} - 1)\sqrt[2]{\frac{1}{NT^2} tr(Y_{-1}Q_\lambda Y'_{-1})}\frac{\omega^2_e}{\phi^4_e} \tag{38}$$

$$t_a, t_b \Rightarrow \mathcal{N}(-\mu_\theta \sqrt[2]{\frac{\omega^4_e}{2\phi^4_e}}, 1) \tag{39}$$

Under the null Hypothesis $H0 : \rho^+_{pool} = 1$, μ_θ will also be equal to 0. Therefore the test distributes as $N(0, 1)$.

To estimate the pooled coefficient and to test the Hypothesis, the parameter τ^N_e, ϕ^4_e, ω^2_e, need to be estimated. Let's define $\hat{\Gamma}_i(j) = \frac{1}{T}\sum_t \hat{e}_{it}\hat{e}_{i,t+j}$ as the sample covariance of the i^{th} observation.

To continue with the estimations we need to use the following kernel estimations

$$\hat{\tau}_{e,i} = \sum_{j=1}^{T-1} w(\frac{j}{h})\hat{\Gamma}_i(j) \tag{40}$$

$$\hat{\omega}_{e,i}^2 = \sum_{j=-T+1}^{T-1} w(\frac{j}{h})\hat{\Gamma}_i(j) \tag{41}$$

We then get τ_e^N, ω_e^2 by taking the mean of all the estimation for each observation, with $i = 1, ..., N$. The estimation of ϕ_e^4 is $\frac{1}{N}\sum_{i=1}^{N} \hat{\omega}_e^4$. To estimate the parameters I use the Bartlett kernel.

In addition, I compare the result of this test with one that does not take into consideration the cross-sectional dependence. I have chosen the LLC test (Levin Lin and Chu (2002)), that in case the hypothesis of cross-sectional independence is violated the test suffers from size distortion.

4.2.3 Cointegration test

After the presence of a unit root inside the variables has been proved, it is important to understand the relation between the dependent variable (cumulative return) and the independent variables (financial statement variables). When we have non-stationary variables in our model the estimated relation can be: a spurious relation, a cointegrated relation or no relation.

Only the second one is important to study, and indeed it is essential to test if there is cointegration.

Kao (1999) created a test called Residual-Based DF and ADF tests that have a null hypothesis of no cointegration. It is based on DF or ADF tests, that are used to check if the residuals of the pooled panel model are stationary (I(0)).

Even here the cross-sectional dependence is a key factor. Gengenbach, Palm and Urbain have created a test based on the Kao (1999) and the Johansen tests that account for this problem.

What they proposed is just an extension of the most known test for no cointegration (Kao Test), that is applied to the de-factored data. The assumption behind the test is that each variable has a factor structure, that leads to the cross-sectional dependence. It is important then to check whether both the factors and the de-factored data have an I(1) process. The rejection of the no cointegration hypothesis is reached when the residuals of the linear regression of the de-factored data are stationary (meaning that there is cointegration between the dependent and independent variables) and the factors of the different variables are cointegrated.

This technique follows two steps:

- Step 1:Apply a unit root analysis on the de-factored variables and on the common factors in order to identify if the non-stationarity is driven only by a limited number of factors, or if there is an idiosyncratic non-stationary component.
- Step 2: Once you have discovered the source of the non-stationarity proceed to identify the cointegration between the common factors and between the de-factored variables

As a unit root test in step one I used the Moon and Perron type of test for the de-factored data and the Augmented Dickey Fuller test for the factor structure. To test the cointegration between the de-factored variables I used the Residual-Based DF Test (Kao (1999)). The null hypothesis of this test is the presence of no cointegration.

The Kao test or Residual-Based DF Test is based on the model:

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + e_{it}, \tag{42}$$

where either y and x are $I(1)$ and non-cointegrated . In case $z_{it} = \mu_i$, Kao proposed to make a DF test on the residuals of the regression, in order to identify the presence of no cointegration. To estimate the residuals we can use a fixed effect model, $\hat{e}_{it} = \dot{y}_{it} - \dot{x}_{it}$ where $\dot{y}_{it}$ and $\dot{x}_{it}$ are the demeaned observation. We can then estimate the autoregressive model.

$$\hat{e}_{it} = \rho \hat{e}_{i,t-1} + \nu_{it} \tag{43}$$

$$\hat{\rho} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e_{it}} \hat{e_{i,t-1}}}{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e_{it}^2}} \tag{44}$$

$$t_\rho = \frac{(\hat{\rho} - 1) \sqrt[2]{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e_{it}^2}}}{s_e}, \tag{45}$$

where t_ρ is the t statistic of the coefficient and $s_e = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} (\hat{e}_{it} - \hat{\rho}\hat{e}_{i,t-1})^2}{NT}$. The DF test that have been proposed are:

$$DF_\rho^* = \frac{\sqrt[2]{N}T(\hat{\rho} - 1) + \frac{3\sqrt[2]{N}\hat{\sigma}_\nu^2}{\hat{\sigma}_{0\nu}^2}}{\sqrt[2]{3 + \frac{36\hat{\sigma}_\nu^4}{5\hat{\sigma}_{0\nu}^4}}} \tag{46}$$

$$DF_t^* = \frac{t_\rho + \frac{\sqrt[2]{6N}\hat{\sigma}_\nu}{2\hat{\sigma}_{0\nu}^2}}{\sqrt{\frac{\hat{\sigma}_{0\nu}^2}{2\hat{\sigma}_\nu^2} + \frac{3\hat{\sigma}_\nu^2}{10\hat{\sigma}_{0\nu}^2}}}, \tag{47}$$

where $\hat{\sigma}_\nu$ is just the estimated variance of the ν_{it}, the residuals of the autoregressive model, and $\hat{\sigma}_{0\nu}$ is the estimated long-term variance, that I have estimated by applying the formula shown at equation 27.

The asymptotic distribution of the test converges to a $N(0, 1)$.

In the Kao(1999) paper an ADF test was also proposed. I have estimated it but decided not to show the results because similar to the ones obtained with the DF test, even for different lags.

To test the cointegration between the common factors I use the more standard Johansen test.

4.3 Portfolio Formation

To test if my assumption is correct, and if fundamental analysis works, I need to see if future cumulative returns converge towards the estimated fair value.

To do so, I build up an out-of-sample trading strategy based on the estimated mispricing signals.

Indeed, at each month I will have a vector containing the mispricing signal for each company, that I order from the lowest to the highest value. Those signals will be used to split the sample in five quintiles, where Q1 are the underpriced firms (with negative mispricing signal), and Q5 are the overpriced ones (with positive mispricing signal). The basic idea behind the trading strategy is to go long in Q1 and short in Q5.

If this investment strategy ends up to be profitable, I can say that fundamental analysis works. In order to calculate the profitability of the portfolios, and to compare them, I will calculate several statistics: annualized return, Sharpe ratio, standard deviation, min (max) return and percentage of positive returns. This portfolio statistics are calculated taking into account the same interval of time. It also becomes important to understand why the strategies are profitable, and whether the mispricing signals simply reflect some of the most known anomalies. This is verified by running this regression:

$$\boldsymbol{r}_{qt} = \boldsymbol{\alpha}_q + \boldsymbol{r}_{mt}\boldsymbol{\beta}_m + \sum_{j=1}^{p} \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_{qt}, \tag{48}$$

where the $\boldsymbol{r}_{qt}$ is the return of the q^{th} quintile at each point in time, $\boldsymbol{\alpha}_q$ is the intercept, the $\boldsymbol{r}_m$ is the return on the market and $\boldsymbol{X}_j$ are the p anomalies (like SMB and HML) that I take into account in order to explain the return and $\boldsymbol{\beta}_j$ are their coefficients.

Then by looking at the significance level of the variables, I can test if there is a significant performance not explained by the factors (significant alpha), and which ones are the relations of the portfolios with the anomalies taken into account.

In the case that the trading strategies based on mispricing signals estimated with longer horizons end up being profitable, it can give me some evidences on whether the market tends to revert itself toward a long term average. The reason is that in case the coefficients would be the true one at each point in time, then I would only be able to profit from the divergences of companies from their “fundamentals”, but not to profit from correction of the market as a whole. The coefficients that I have when I average the weights across time can be considered as a conditional mean of the weights over a certain interval of time. In case the dynamics of this coefficients tend to be mean reverting, then I am able to profit from this behaviour of the market. Therefore, I believe that periods when the value drivers are not in line with the conditional means are periods where the market cap are not on average correct.

Moreover, I will study the dynamics of the mispricing signals, in order to understand if there is a correction after a deviation, and how long does it take to happen. The information that I derive from these analysis are then incorporated into the portfolio strategy, to see if it leads to higher performances.

5 Empirical Results

In this section, I give an overview of the different results that I obtain for each analysis. In the first section I show the results regarding the non-stationarity and cointegration tests, in the second section I discuss the portfolio findings for each model, in the third section I analyse the convergence of the cointegration signals, and as a final section I make a summary of statistics per quintile for each approach.

I will show the results for:

- Method A = LHS, Cumulative Return, RHS, All the Financial Variables
- Method B = Method A + Dummy Variables
- Cointegration Model = RHS, Cumulative return, LHS, Dividend, Sales, Dummy dividend.

Since the number of parameters increases drastically with the addition of the dummies, I have decided to decrease the number of variables by eliminating highly correlated ones.[1][2] This will decrease the noise in my estimation.The variable that are left after the selection are listed in table 4.

For each method I estimate the coefficients with four different time horizons: Cross-sectional at each t, 12 Months Average, 24 Months Average and 60 Months Average.

[1]The results of the model with all the variables are not showed in this paper, because due to the estimation uncertainty I could not rely on the mispricing signals to create a portfolio strategy, especially during the first period when the sample is small. The correlation Matrix used in order to select the variables is in table 12 in the appendix

[2]Motivation of the decrease in the number of independent variables can also be found in Bartram and Grinblaat (2014), where they say that due to estimation uncertainty they can have maximum 28 variables. In addition, they say that 6 of this variables are perfectly spanned by the other 22 and that half of the latter explains 98% of the variance.

Table 4: This table shows the different variables that are used in the model estimation. The table is divided into three columns depending in which Company Statement the variables are. The * means that I have takeen the log of that variable, this is justified from the relation of the variable with the cumulative return

Balance-Sheet	Income Statement	Others
Total Assets*	Net Income	Past year Return Volatility
Dividends Yield	Sales	Beta
Total Debt*	Income Taxes	dummy for sectors
Number of Shares	Operational Income	dummy for Book to Market
		dummy for Dividend Yield

5.1 Non-stationarity and cointegration test

As it has been stated previously, before estimating the relation between some variables it is important to test the presence of unit roots. In case we find that the variables are indeed non-stationary, it is necessary to investigate which type of relation there is between them. Therefore, the next step is to apply a cointegration test that checks the presence of a cointegrative relation. The last step, if the variables are actually cointegrated, is to apply a panel cointegration estimation techniques.

In this section I will show the results of the unit root test and the cointegration test. The variables that I have selected to enter in the cointegration model are: dividend, net income, operational income, and sales. Therefore, I run the panel unit root test and the cointegration test only on this subset of variables. It is important to notice that now I do not use the dividend yield but the dividend itself. I do this in order to follow the findings of (Shiller & Cambell, 1987).

5.1.1 Panel Unit Root

As a panel unit root test, I have used the Moon and Perron one that I have explained in the Methodology. This test is then compared with the percentage of non-rejection of singular ADF statistics made on each individual, and the LLC test (Levin Lin and Chu(2002)). In this way, I can see the robustness of the panel approach compared to the individual approach, and the one that does not consider cross-sectional dependence. Indeed, if we look at table 5 we can see that the singular ADF and the LLC test are close to each other, and we can see that the Moon and Perron test statistics give a different result especially regarding the income variables. We can notice the extreme results of the LLC test, where the p-values are equal to 1. This can be due to the fact that LLC test is affected by size distortion in the presence of cross-sectional dependent observation.

Table 5: This table shows the outcome of the unit root tests. In the first row there are the percentages ADF test, where I calculate the percentage of not rejection of the $H0 : \rho = 1$, where ρ is the coefficient for the ADF test, the values in the bracket are the percentage of rejection with p-value 0.1, the one not in the bracket have p-value 0.05. The second line shows the Z-Scores of the Moon and Perron tests where rejection of the $H0$ means stationarity of the variable, under it there are the p-value related to the Z-scores. In the 4th line, there are the T-statistics of the LLC test, and the respective p-values underneath. The percentage adf is calculated on 125 companies, that are the ones that stay inside my sample for the whole period.

	Dividend	Net Income	Operational Income	Sales	Cumulative return
Singular ADF	0.99(0.98)	0.46(0.39)	0.74(0.66)	0.96(0.95)	0.98(0.98)
Z-score	-1.35	-10.46	-4.76	1.76	-0.80
Moon and Perron	0.09	0.00	0.00	0.96	0.21
T - stat	25.46	-1.38	-5.68	-1.49	5.88
LLC Test	1	0.28	1	1	1

In the Moon and Perron test a number of factors had to be chosen in order to eliminate the cross-sectional dependence. Bai and Ng (2002) proposed a technique in order to estimate an optimal number of factors. However, in this paper the number of factors is treated as known and I have selected three factors[1].

5.1.2 Panel Cointegration Test

To test for cointegration I use the Gengenbach, Palm and Urbain Test that is based on the Kao test and the Johansen test. It allows me to understand if there is cointegration when cross-sectional dependence is present in the data. It also gives intuition on whether the non-stationarity of the variables is due to a limited number of components or due to idiosyncratic components. In the first step, I estimate the factor structure that explains the cross-sectional dependence. Then, I check if the de-factored panel data and the factor structure present unit roots. To do so, I use the Moon and Perron test for the de-factored data and the ADF test for the factors. The outcome is that all the common factors are non-stationary and that the only variables that present a unit-root are: Dividends and Sales. Then as the last step, I check if there is cointegration between the non-stationary de-factored variables, using the Kao(1999) test, and if there is cointegration between the factors, using the Johansen test. In this way, it is also possible to understand if the variables are cointegrated only due to the factors or also at an individual level. Table 6 shows the result of the Kao Test on these variables.

We can see that both of the tests reject the hypothesis of no-cointegration.

Table 7 shows the output of the Johansen test. The test has rejected the presence of 0 cointegrative

[1]Other numbers of factors have been used and similar results were found.

Table 6: This table shows the outcome of the cointegration test, where the $H0$ is no cointegration. In the header, from left to right, there are the coefficient estimate ρ, the DF_ρ and its p-value. Then there are the DF_t and its p-value. I have estimated this test on the model that has on the LHS the cumulative return, and on the RHS the 2 non-stationary variables listed before.

	ρ	DF_ρ	p value ρ	DF_t	p value
Kao test	0.97	-8.29	0.00	31.33	1

Table 7: This table shows the result of the Johansen test. The first column is showing is the number of cointegrative relation. Following there are: the outcome of the test (0 not rejecting of H0 and 1 rejecting of H0), test statistic for that specific number of cointegrative relation, critical value for the right tail probabilities, pvalue of rejecting or not rejecting the H0, eigenvalue associated with H(n of cointegrative relation +1).

r	h	stat	cValue	pValue	eigVal
0	1	215.68	208.44	0.02	0.31
1	0	148.92	169.60	0.38	0.20
2	0	108.33	134.68	0.60	0.16
3	0	77.62	103.85	0.70	0.15
4	0	49.22	76.97	0.87	0.11
5	0	27.41	54.08	0.96	0.07
6	0	14.99	35.19	0.95	0.04
7	0	8.19	20.26	0.81	0.03
8	0	3.38	9.16	0.58	0.02

relations in favour of one cointegrative relation. This is a positive result because the rejection of the no-cointegration Hypothesis is reached only in case both, the common factors and the de-factored variables cointegrate.

Even for this test a number of factors had to be chosen, as before I have chosen three factors [1].

Model coefficients

Before discussing the portfolio results for each methodology I show the coefficient for each method. The following table shows the coefficients and the adf test statistic applied on each coefficients time series. I have done this test in order to check that the series of the coefficients are mean reverting.

Method A The coefficient for this method are all significant a part for the variable other assets, and all of them are mean reverting. The variable total debt is the only one with a negative coefficient, all the other ones have a positive relation with the cumulative return. This result is in line with my expectations, since a high debt is usually seen as negative and, on the contrary, having

[1] Other numbers of factors have been used and similar results were found.

Table 8: This table shows the coefficient for all the methods. The table is divided into three parts, one for each model. For methods A and B there are 5 sub-columns, for the cointegration model there are only the first three columns. In the first three column there are the estimated coefficient (column name "Estimates") with the respective T-statistics (column name "T stat") and significance level (***, pvalue ¡ 0.01, **, pvalue ¡ 0.05, *, pvalue ¡ 0.1). In the last two columns are displayed the results of the ADF test (column name "Adf"), applied on the time series of the coefficients. The coefficients that are displayed here are the average across the whole coefficients estimated at each point in time (211 trading months for methods A and B and 151 for the cointegration model). In order to make the estimation feasible, I had to exclude one column from each of the dummy matrices. For sector I have excluded utilities and for B/M the

Total Assets	0.278	17.21	***	-2.68	**	0.288	16.61	***	-2.47	**			
Dividend Yield	0.014	6.54	***	-2.81	**	0.035	18.13	***	-2.64	**			
Dividend											0.14	1.11	
Income Tax	0.055	5.11	***	-5.59	***	0.056	5.14	***	-5.37	***			
Net Income	0.117	15.38	***	-4.8	***	0.095	13.92	***	-5.16	***			
Operational Income	0.067	14.66	***	-4.7	***	0.069	15.15	***	-4.53	***			
Shareholder Equities	0.023	10.3	***	-4.1	***	0.041	17.86	***	-3.52	***			
Sales	0.257	14.05	***	-4.48	***	0.274	13.92	***	-4.04	***	0.10	18.01	***
Other Assets	0.019	3.02	***	-4.23	***	0	-0.05		-5.44	***			
Total Debt	-0.236	-17.58	***	-2.48	**	-0.208	-17.84	***	-2.62	**			
Betas	0.271	8.85	***	-3.04	**	0.321	10.05	***	-3.23	**			
Return volatility	0.06	3.79	***	-3.9	***	0.086	5.62	***	-3.77	***	-0.04	-13.4	***
Dummy Dividend	0.007	13.11	***	-3.43	***	0.007	12.42	***	-2.68	**			
Trend	0.014	16.05	***	-2.3	**	0.029	14.23	***	-2.13	**			
Energy						0	0.55		-3.91	***			
Materials						-0.005	-7.43	***	-3.38	***			
Industrials						-0.007	-9.51	***	-2.76	**			
Consumer Discretionary						-0.009	-10.96	***	-2.66	**			
Consumer Staples						-0.01	-13.23	***	-2.74	**			
Health						-0.003	-4.54	***	-3.16	**			
Financials						0.001	0.91		-3.14	**			
Information Technology						0.003	3.62	***	-4.09	***			
Telecomunuication Service						-0.002	-2.43	**	-3.73	***			
Q1 B/M						-0.022	-16.42	***	-1.74	*			
Q2 B/M						-0.017	-14.93	***	-1.89	*			
Q3 B/M						-0.012	-13.21	***	-2.09	**			
Q4 B/M						-0.006	-8.83	***	-2.91	**			
	min R2	max R2	Avg R2			min R2	max R2	Avg R2			min R2	max R2	Avg R2
	0.24	0.88	0.61			0.33	0.89	0.67			0	0.23	0.03

a lot of assets or a high income or a high dividend are seen as positive information. Moreover, "Shareholder equity" has also a positive relation due to the SEO, and I also expected "Betas" and "Return volatility" to be positive due to the theory that higher risk bring to a higher expected return.

Method B Even here most of the coefficient are significant. The variables other assets become not significant. Apart from that, the sign of the variables that belong also to method A and their significance did not change. The dummies per sector are almost all significant with slightly different effects. The dummies on book to market are all significant with an increasing pattern suggesting that more the ratio between the book value and the market cap is higher and higher is the cumulative return level. The average R squared is only slightly different from the average R squared of method A, suggesting that the additional variables do not add explanatory power to the model.

Cointegration model This model is composed of only three variables, where only two are significant. The only one that is not significant is dividend. This suggests that the cointegrative relation, underlined by (Shiller & Cambell, 1987), between the dividends and prices comes from the discount rates and not the future dividends. The average R squared is in this model really low, suggesting that the model is not able to properly explain the dependent variable.

5.2 Portfolio Findings

In this section I will show the results of the portfolio strategies based on the mispricing signals. At each point in time I sort the companies in different quintiles depending on their mispricing signals. The portfolio is then created by going long in the companies belonging to Q1, undervalued, and short in firms belonging to Q5, overvalued. The sorting is done at each trading month.

5.2.1 Method A

Table 9 summarizes all the information of the portfolios built using the signals estimated with method A, for different horizons, and for different weighting schemes.

Equally Weighted By looking at the annualized return across the horizon, it is clear that investing on the mispricing signals leads to a profitable strategy. In addition, I find that averaging the coefficient across time leads to more robust mispricing estimations. Indeed, I obtain Sharpe ratios that increase with the horizons. Regarding the volatility of the portfolios, we can see that it increase in the 12 months horizon and then it decreases.

Value Weighted Similar as before the value weighted portfolios created by sorting the stocks on the mispricing signals give positive results. In this case, the performances are on average better with respect to the other weighting method, with higher annualized returns and a lower standard deviation. Here the Sharpe ratios do not increase with the horizon, but they reach their maximum with the 24 months horizon.

The difference in performances of the two weighting scheme can also be seen from figures 5 and 6 in the appendix.

Table 9: The table shows the summary statistics of the long - short portfolio strategy based on method A. The table is divided into two depending on the weighting criteria. The annualized return is the annualized return of the portfolio. The Sharpe Ratio it has been calculated using $\frac{(Rt-Rf)}{\sigma_r}$ where Rf is the risk free return. % Positive return is the percentage of positive return in the series

	Equally Weighted				Value Weighted			
	0 Months	12 Months	24 Months	60 Months	0 Months	12 Months	24 Months	60 Months
Annualized return	0.033	0.047	0.043	0.042	0.046	0.060	0.064	0.049
Std Dev	0.069	0.086	0.073	0.069	0.067	0.075	0.071	0.065
T-stat	1.685	1.879	1.999	2.129	2.331	2.699	3.070	2.595
Sharpe	0.487	0.547	0.580	0.618	0.678	0.789	0.900	0.755
Min Ret	-0.077	-0.076	-0.056	-0.044	-0.046	-0.054	-0.054	-0.039
Max Ret	0.091	0.132	0.083	0.077	0.078	0.089	0.091	0.077
% Pos Ret	0.588	0.588	0.581	0.588	0.574	0.574	0.588	0.615

5.2.2 Method B

Table 10 shows the performance of the portfolios created using the signals estimated with method B, for all the four different horizons and with equal and value weights.

The results regarding method B are worst, in terms of Sharpe ratios, then method A. The reason is a general decrease in the portfolios annualized return. However, even here we can see that accounting for the time dimension improve the mispricing signals estimation.

The fact that method B gives worst performances in comparison to method A suggests that taking into account sector differences and B/M differences lead to worst mispricing signal estimation.

Table 10: This is the summary statistics of the long - short portfolo based on method B

	Equally Weighted				Value Weighted			
	0 Months	12 Months	24 Months	60 Months	0 Months	12 Months	24 Months	60 Months
Annualized return	0.022	0.033	0.032	0.025	0.025	0.034	0.036	0.025
Std Dev	0.059	0.072	0.068	0.069	0.054	0.065	0.060	0.059
T-stat	1.295	1.559	1.632	1.279	1.618	1.789	2.090	1.453
Sharpe	0.372	0.450	0.471	0.368	0.466	0.517	0.605	0.418
Min Ret	-0.058	-0.064	-0.059	-0.052	-0.033	-0.053	-0.044	-0.033
Max Ret	0.094	0.086	0.101	0.095	0.056	0.062	0.070	0.060
% Pos Ret	0.581	0.595	0.574	0.534	0.547	0.561	0.561	0.547

5.2.3 Cointegration Model

Table 11 summarizes all the portfolio statistics relative to the cointegration model. I have estimated the signals using a rolling window of 60 months, this is the maximum length that I could use in order to get a long enough time series to incorporate different periods in the financial market (e.g. Pre-Crisis(2004 - 2008), Crisis(2008 - 2010)..). The variables that I have used are the same variables that I have used for the Kao test in table 6, and are those that after being de-factored present a unit root. By looking at the portfolios in table 11 we can see that the use of a more sophisticated technique does not lead to higher performances. For both the weighting scheme the portfolios have negative performances. These results suggest that the residual corrected to account for the short dynamics are not a good indicator of mispricing.

This poor result can also be caused by different reasons. Firstly, the variables used in the model are just a few and it is possible that some variables are omitted from the model, secondly, it can be that the rolling window that I am using for the estimation is not optimal, thirdly, the selection bias that I have created by adjusting the sample in order to have a balanced panel creates a distortion in the estimation, or it is possible that the technique leads to a lot of parameter uncertainty.

Table 11: This table shows the returns of the Q1-Q5 portfolio, created using the mispricing signal estimated with the CUP-FM method. Read table 5 for more description about the statistics.

	Equally Weighted	Value Weighted
Annualized return	-0.018	0.001
Std Dev	0.127	0.114
T-stat	-0.512	0.031
Sharpe	-0.144	0.009
Min Ret	-0.109	-0.083
Max Ret	0.170	0.121
% Pos Ret	0.466	0.480

Conclusion From the tables above we could observe that averaging the beta over time leads to better signal estimation. In addition, adding group specific effect for sector and B/M to the model does not improve the results, and it decreases the general performances. Moreover, when the time dimension is extended to estimate the long-run relationship between the variables it leads to an apparent randomness in the portfolio performances.

5.3 Signal convergence

In this section I will analyse the behaviour of the mispricing signals, to understand if they converge and how long does it take to happen. This information will then be incorporated in the portfolio strategy and see if it leads to better performances. Figure 3 gives a full overview of the speed of convergence of the mispricing signals for each portfolio. In the graphs I show the percentage of the difference between the delayed signals and the contemporaneous ones for each portfolio. In order to calculate the percentage I have used the formula:

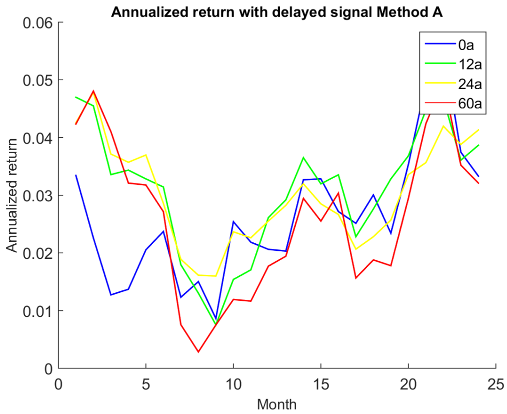
$$\delta_{it,j} = (\frac{\boldsymbol{\kappa}^*_{i,t+j}}{\boldsymbol{\kappa}^*_{i,t}}) - 1, \tag{49}$$

where δ_{it} is the percentage difference between the delayed mispricing signal, delayed of j months, and the contemporaneous mispricing at time t for company i, $\boldsymbol{\kappa}^*$ is the estimated mispricing signal of the company i at time t with delay j. δ_{it} will have wither positive or negative values, where positive means that the deviation is increasing, and negative that the deviation is contracting. The graph shows the average of this measure across time and individuals. The maximum delay that I have chosen is 24 months.

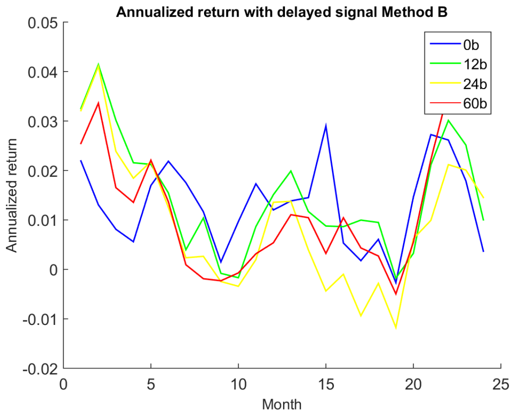
The first thing that is easy to notice is that, on average, the deviation for method A and B tend to contract rather than expand. Even for the cointegration model, where we have seen negative performances in the portfolio based on its mispricing signals, the deviations contract in the following months. We can see that differences are present in the mispricing dynamic across horizons, but only slight changes are present across models. This phenomenon is strange since we see different portfolio performances between the two models. The time dimension, instead, effectively affects the deviation behaviour and indeed we see different patterns across the horizons.

For all the horizon the difference between the cumulative return and the intrinsic value tend to decrease monotonically. I, therefore, build several trading strategies using delayed signals, starting from using contemporaneous signals to using two years delayed signals. I do this only for method A and B since are the only one that ended up being profitable. In figure 2 I have plotted the annualized return of the strategies based on the different delayed signals for both, equally and value weighted portfolios. In most cases, the portfolio performances drastically decrease in the

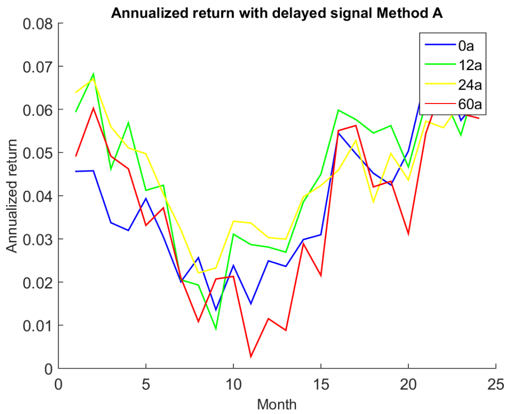
Figure 2: In these figures I have plotted the annualized return of the portfolios build using delayed signals. The y axes represent the annualized return and the x axes the months of delay.



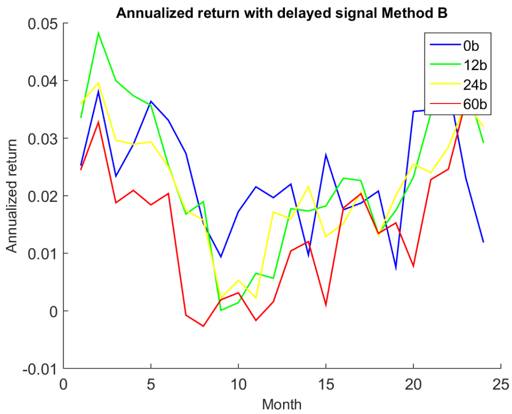
(a) Equally weighted



(b) Equally Weighted



(c) Value Weighted

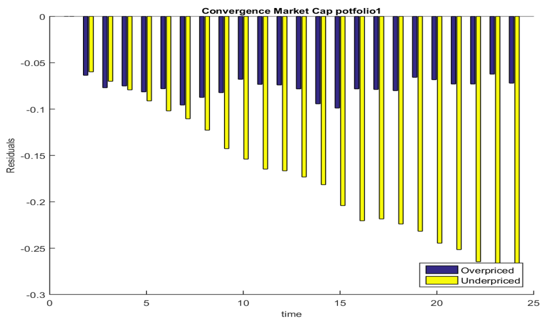


(d) Value Weighted

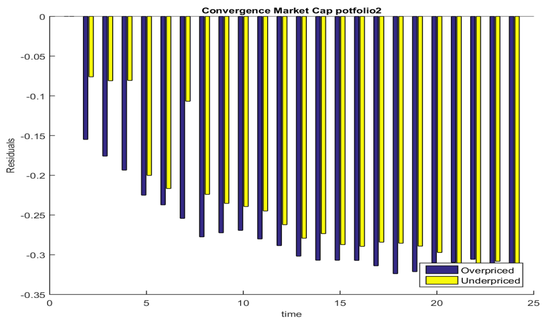
first 5 to 10 months. In only a few cases, especially for the 1 month horizon, we can find an increase in the performance in the first months.

Starting from the fact that the best performance is reached, on average, around the first 4 months, I have created the same Long-short portfolios as in the previous section but with different rebalance horizons. I then try to understand if a buy and hold portfolio can, therefore, be profitable. In figure 4 in the appendix there are plotted the portfolio performances when I change the rebalance frequency. I have tested the frequencies from once a month to once every six months. I have plotted two portfolio statistics: Annualized return and the Sharpe ratio. An important finding is that for most of the portfolios the performance increases when the rebalance frequency is less. For the strategies that use averaged coefficient across time, the improvements happen already if we rebalance every two months. For the 1 months horizon strategy, there is no clear pattern.

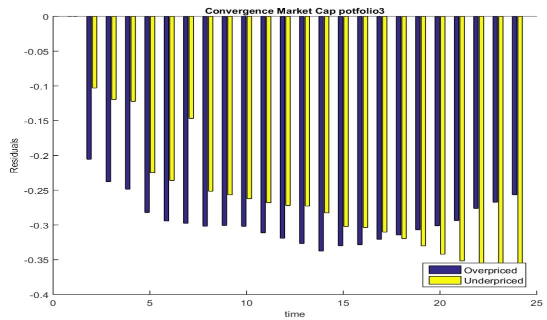
Figure 3: In this set of figures are shown the convergence rates of the cumulative return toward their fundamental values, of the overpriced and underpriced companies. The Y axes represent the percentage of convergence toward the intrinsic value, the X axes represent the time interval. In order to understand if the deviation of the Market Cap from the estimated Fundamental value contract, I have taken, for each point in time, the ratio of the difference between the Fundamental value at time t and the cumulative return for each point in time from t+1 till t+24 and the difference between the same fundamental value and the cumulative return at time t, for each individual, and then I subtract 1. In this way I am able to see if the cumulative return in the future move toward the estimated intrinsic value. I do this for each individual at each point in time, and what is shown in the graph is simply the average across time and individuals of the several patterns.



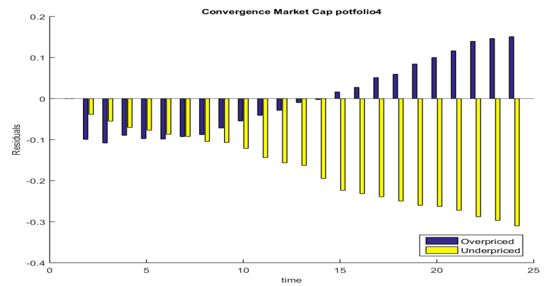
(a) horizon 1, method A



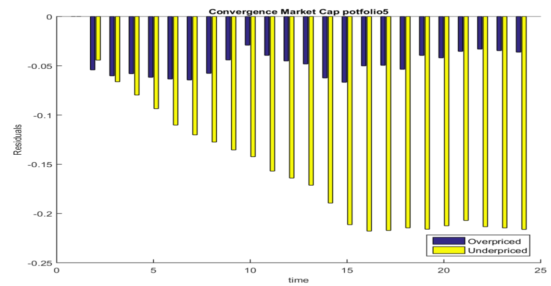
(b) horizon 12, method A



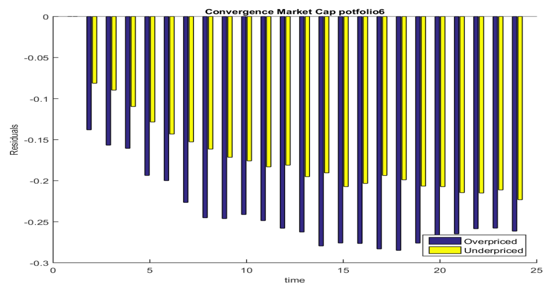
(c) horizon 24, method A



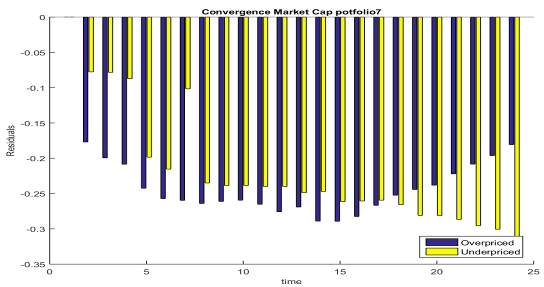
(d) horizon 60, method A



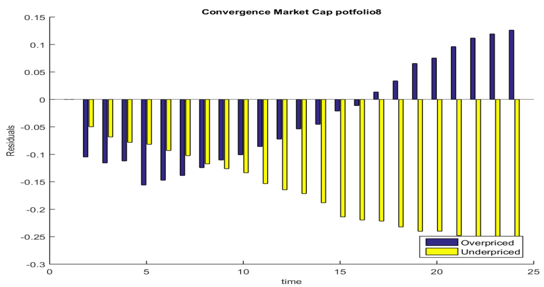
(e) horizon 1, method B



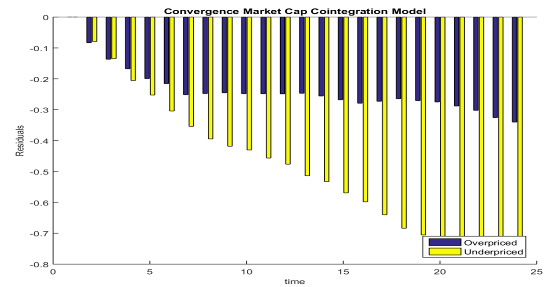
(f) horizon 12, method B



(g) horizon 24, method B

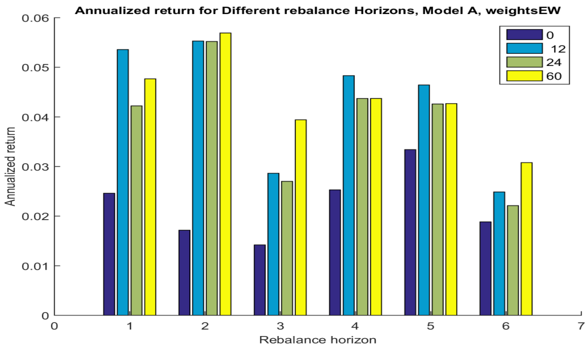


(h) horizon 60, method B

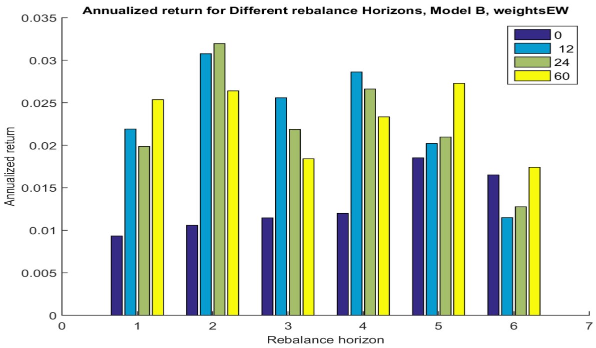


(i) horizon 60, Convergence Model

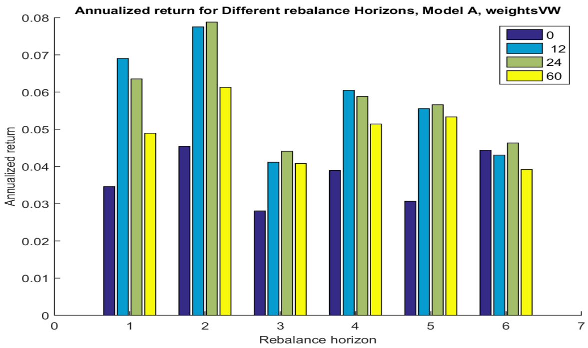
Figure 4: In this figure are shown the Annualized returns and the Sharpe ratios of the different portfolios, for six rebalance frequencies. For each model I have tested the performances in case I rebalance my portfolio from once every month to once every six months. The results are shown for the Equally weighted and the Value weighted portfolios.



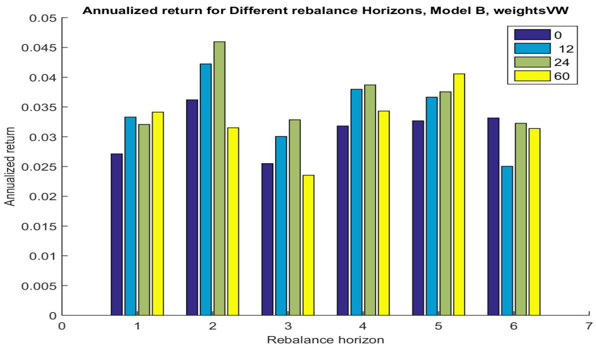
(a) Return method A, Equally Weighted



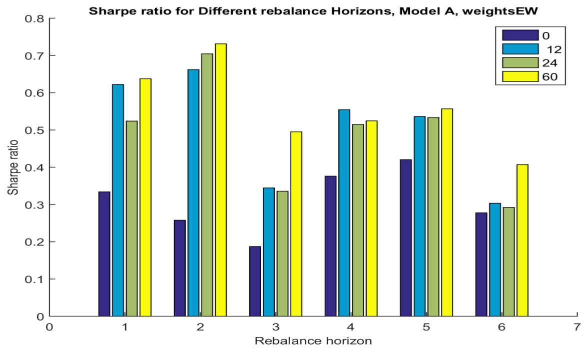
(b) Return method B, Equally Weighted



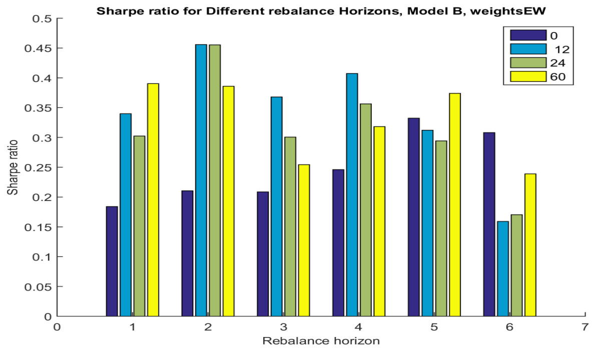
(c) Return method A, Value Weighted



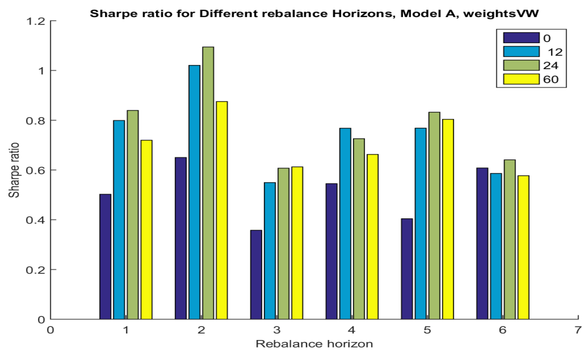
(d) Return method B, Value Weighted



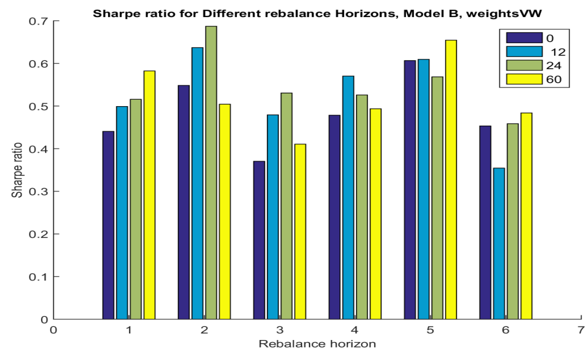
(e) Sharpe method A, Equally Weighted



(f) Sharpe method B, Equally Weighted



(g) Sharpe method A, Value Weighted



(h) Sharpe method B, Value Weighted

5.4 Alpha Significance

In this section I will test whether the returns of the portfolios made using the estimated signals give significant alphas or if they are entirely explained by the most known anomalies.

I estimate the model (equation 48) for each quintile, and for the long - short portfolio created by subtracting the return in the quintile with the most underpriced firms from the one in the quintile with the most overpriced firms. I do this in order to understand which quintiles are more exposed to the factors and if there is a pattern in the estimated alphas. I estimate the model for both, value and equally weighted portfolios. The anomalies that I use are: Size (SMB), Value (HML), Momentum (MoM) and Low Volatility (BaB)[1]. I have selected these because in the summary statistics I could observe some differences across the quintiles for: size, value, past return and exposure to beta.

In this section I show the results for the 1 and 12 months horizon. All the remaining results can be found in the tables from number 27 to number 30 that can be found in the appendix.

Again, I have decided to show only the results for the cross-sectional regression model, and the 12 month horizon average model. I have made this choice because the most interesting result can be found in these tables.

Method A and B The most important thing to notice from these tables are the significant alphas for both the weighting scheme for method A. For method B ,in general, all the alphas are positive, even though no one is significant.

Something unexpected is the negative relation between the portfolio returns and the HML factor, because, as we will see later in the paper, underpriced companies tend to have a higher B/M compared to the overpriced ones.

For both the methods it is important to see that: SMB has, in general, a positive relation with the long - short portfolio return, and that MoM and BaB factors have a negative relation with it. This is what I would have expected by simply looking at the summary statistics.

[1] SMB, HML are built following the Fama and French (1993) paper. MoM is built following the Kenneth R. French data library. BaB is built following the Frazzini and Pedersen (2014) paper.

Cointegration Model The results are shown in table 20-21 in the appendix. In this case the alphas estimated in the long - short strategy are positive but not significant. This is an expected result since the portfolios annualized returns are close to 0. The portfolio has a positive and significant relation with the SMB factor and a negative relation with MOM.

Table 12: This table shows the coefficients, and the relative t statistics, of the model that I estimate in order to check the presence of a significant alpha. The table is divided into 6 columns, one for each quintile plus one for the Q1 minus Q5 portfolio. For each column I show the coefficients for the models where I consider the anomalies: SMB, HML, MOM and BaB. In addition at the bottom of every column there is the information ratio of the portfolio, the R-squared and the number of observation used to estimate the model. To estimate the Information ratio I have followed the technique used in Bartram and Grinblatt (2014) paper, I have multiplied the t-statistic of the intercept by, $\sqrt{\frac{12}{T}}$. SMB, HML and MOM factors are build following the Kenneth and French dataset approach, and the BaB factor following Frazzini and Pedersen (2014). For this table the portfolio quintiles are made using the cross-sectional beta estimates. The portfolio are Equally Weighted. The following table represent the results for method A.

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**
Alpha	0.001	0.937	-0.001	-0.838	0	0.411	0.001	1.438	-0.002	-2.168 **	0.002	1.915 *
Ret_Mkt	1.047	31.578 ***	0.976	33.471 ***	0.983	38.499 ***	0.941	27.221 ***	0.975	30.146 ***	0.072	1.356
SMB	0.214	3.073 **	0.178	2.905 **	0.109	2.026 **	-0.065	-0.891	-0.353	-5.202 ***	0.567	5.113 ***
HML	-0.258	-3.622 ***	-0.267	-4.264 ***	-0.056	-1.02	0.124	1.667 *	0.24	3.463 ***	-0.498	-4.394 ***
MoM	-0.151	-5.958 ***	-0.064	-2.874 **	-0.041	-2.076 **	0.115	4.331 ***	0.127	5.122 ***	-0.278	-6.875 ***
BaB	-0.049	-1.106	0.052	1.343	0.048	1.42	0.048	1.037	-0.033	-0.761	-0.016	-0.228
Information Ratio	0.272		-0.244		0.119		0.418		-0.63		0.557	
R Squared	0.96		0.97		0.98		0.95		0.95		0.48	
Observations	142		142		142		142		142		142	

Table 13: This table shows the same results as in the table above, but in this case the portfolios are value weighted. The signals are estimated using method A.

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**
Alpha	0.001	1.293	0.001	1.16	0.001	1.093	0.002	1.735 *	-0.002	-2.881 **	0.003	2.287 **
Ret_Mkt	1.103	24.893 ***	0.966	27.721 ***	0.944	28.833 ***	0.953	23.568 ***	1.011	33.321 ***	0.092	1.457
SMB	0.253	2.984 **	0.165	2.478 **	0.133	2.121 **	-0.072	-0.927	-0.193	-3.324 **	0.445	3.684 ***
HML	-0.062	-0.689	-0.162	-2.296 **	0.158	2.38 **	0.055	0.672	0.062	1.014	-0.124	-0.969
MoM	-0.069	-2.074 **	-0.09	-3.421 ***	-0.083	-3.351 **	0.048	1.566	0.069	3.017 **	-0.139	-2.899 **
BaB	-0.136	-2.315 **	0.067	1.443	0.1	2.303 **	0.031	0.585	-0.026	-0.653	-0.11	-1.308
Information Ratio	0.376		0.337		0.318		0.504		-0.838		0.665	
R Squared	0.92		0.94		0.95		0.92		0.95		0.27	
Observations	142		142		142		142		142		142	

Table 14: 12 Months horizon method A, Equally Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**	
Alpha	0.002	1.781	*	0	-0.228		-0.001	-0.88		0	0.223		-0.001	-1.563		0.003	1.937	*
Ret_Mkt	1.107	26.318	***	0.992	33.765	***	0.985	36.18	***	0.878	27.593	***	0.959	27.948	***	0.148	2.233	**
SMB	0.253	2.869	**	0.105	1.705	*	0.085	1.492		0.045	0.674		-0.407	-5.647	***	0.66	4.738	***
HML	-0.16	-1.776	*	-0.245	-3.892	***	-0.154	-2.64	**	0.144	2.11	**	0.2	2.719	**	-0.36	-2.532	**
MoM	-0.159	-4.945	***	-0.07	-3.098	**	-0.005	-0.23		0.108	4.42	***	0.112	4.277	***	-0.272	-5.346	***
BaB	-0.12	-2.139	**	0.03	0.767		0.071	1.973	*	0.111	2.619	**	-0.026	-0.573		-0.094	-1.059	
Information Ratio	0.518			-0.066			-0.256			0.065			-0.454			0.563		
R Squared	0.95			0.97			0.97			0.96			0.95			0.48		
Observations	142			142			142			142			142			142		

Table 15: 12 Months horizon method A , Value Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**	
Alpha	0.002	1.763	*	0.001	0.81		0	-0.066		0.002	1.445		-0.002	-2.572	**	0.004	2.409	**
Ret_Mkt	1.147	24.559	***	0.988	29.575	***	0.898	25.28	***	0.923	20.593	***	1.006	33.328	***	0.141	2.122	**
SMB	0.258	2.896	**	0.149	2.341	**	0.185	2.721	**	0.085	0.991		-0.306	-5.314	***	0.564	4.452	***
HML	0.052	0.549		-0.13	-1.923	*	-0.106	-1.47		0.064	0.709		0.108	1.769	*	-0.056	-0.418	
MoM	-0.049	-1.386		-0.103	-4.073	***	-0.088	-3.292	**	0.107	3.175	**	0.05	2.187	**	-0.099	-1.969	*
BaB	-0.175	-2.83	**	0.08	1.809	*	0.15	3.195	**	0.034	0.578		-0.031	-0.78		-0.144	-1.636	
Information Ratio	0.512			0.236			-0.019			0.42			-0.748			0.7		
R Squared	0.92			0.95			0.94			0.9			0.95			0.35		
Observations	142			142			142			142			142			142		

Table 16: Cross-sectional regression method B, Equally Weighted Portfolios

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**
Alpha	0.001	0.809	0	-0.727	-0.001	-1.08	0.002	2.188 **	-0.001	-1.461	0.002	1.352
Ret_Mkt	1.064	36.024 ***	0.997	35.577 ***	0.975	36.299 ***	0.899	30.834 ***	0.985	35.273 ***	0.078	1.639
SMB	0.101	1.637	0.102	1.738 *	0.151	2.682 **	-0.014	-0.222	-0.26	-4.437 ***	0.362	3.601 ***
HML	-0.17	-2.682 **	-0.155	-2.575 **	-0.12	-2.085 **	0.174	2.777 **	0.059	0.981	-0.229	-2.228 **
MoM	-0.097	-4.297 ***	-0.088	-4.081 ***	-0.014	-0.7	0.061	2.737 **	0.124	5.799 ***	-0.221	-6.037 ***
BaB	-0.106	-2.709 **	0.012	0.313	0.061	1.693 *	0.131	3.385 ***	-0.03	-0.814	-0.076	-1.197
Information Ratio	0.235		-0.211		-0.314		0.636		-0.425		0.393	
R Squared	0.97		0.97		0.97		0.97		0.97		0.43	
Observations	142		142		142		142		142		142	

Table 17: Cross-sectional regression method B, Value Weighted Portfolios

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**
Alpha	0.001	0.809	0	-0.727	-0.001	-1.08	0.002	2.188 **	-0.001	-1.461	0.002	1.352
Ret_Mkt	1.064	36.024 ***	0.997	35.577 ***	0.975	36.299 ***	0.899	30.834 ***	0.985	35.273 ***	0.078	1.639
SMB	0.101	1.637	0.102	1.738 *	0.151	2.682 **	-0.014	-0.222	-0.26	-4.437 ***	0.362	3.601 ***
HML	-0.17	-2.682 **	-0.155	-2.575 **	-0.12	-2.085 **	0.174	2.777 **	0.059	0.981	-0.229	-2.228 **
MoM	-0.097	-4.297 ***	-0.088	-4.081 ***	-0.014	-0.7	0.061	2.737 **	0.124	5.799 ***	-0.221	-6.037 ***
BaB	-0.106	-2.709 **	0.012	0.313	0.061	1.693 *	0.131	3.385 ***	-0.03	-0.814	-0.076	-1.197
Information Ratio	0.235		-0.211		-0.314		0.636		-0.425		0.393	
R Squared	0.97		0.97		0.97		0.97		0.97		0.43	
Observations	142		142		142		142		142		142	

Table 18: 12 Months horizon method B, Equally Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat	
Alpha	0.001	0.918		0	0.397		-0.001	-0.984		0.001	0.894		-0.001	-1.394		0.002	1.308	
Ret_Mkt	1.122	30.015	***	1.048	33.612	***	0.961	33.865	***	0.875	28.117	***	0.915	26.03	***	0.206	3.235	**
SMB	0.008	0.106		0.131	2.004	**	0.142	2.388	**	0.079	1.211		-0.279	-3.779	***	0.287	2.148	**
HML	-0.045	-0.566		-0.196	-2.933	**	-0.153	-2.517	**	0.056	0.845		0.127	1.678	*	-0.172	-1.258	
MoM	-0.06	-2.091	**	-0.087	-3.651	***	-0.031	-1.446		0.082	3.457	***	0.083	3.071	**	-0.143	-2.921	**
BaB	-0.11	-2.204	**	-0.103	-2.478	**	0.106	2.818	**	0.112	2.711	**	0.06	1.29		-0.17	-2.004	**
Information Ratio	0.267			0.115			-0.286			0.26			-0.405			0.38		
R Squared	0.96			0.97			0.97			0.96			0.94			0.33		
Observations	142			142			142			142			142			142		

Table 19: 12 Months horizon method B, Value Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat	
Alpha	0	0.324		0.002	2.476	**	0	-0.427		0.001	1.104		-0.002	-2.199	**	0.002	1.24	
Ret_Mkt	1.072	25.473	***	1.108	32.115	***	0.947	26.288	***	0.839	21.001	***	0.979	31.564	***	0.092	1.401	
SMB	0.086	1.067		0.083	1.264		0.082	1.197		0.172	2.26	**	-0.213	-3.603	***	0.299	2.373	**
HML	0.127	1.489		-0.119	-1.705	*	0.021	0.287		0.044	0.545		-0.017	-0.274		0.144	1.078	
MoM	0.068	2.153	**	-0.045	-1.744	*	-0.066	-2.42	**	-0.002	-0.053		0.008	0.348		0.06	1.208	
BaB	-0.071	-1.269		-0.179	-3.905	***	0.07	1.46		0.204	3.845	***	0.038	0.916		-0.108	-1.239	
Information Ratio	0.094			0.72			-0.124			0.321			-0.639			0.361		
R Squared	0.93			0.95			0.94			0.92			0.95			0.15		
Observations	142			142			142			142			142			142		

Table 20: Cointegration Technique, Equally Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat	
Alpha	0	-0.274		0.001	1.217		0.001	0.577		0	0.459		-0.002	-1.807	*	0.001	0.517	
Ret_Mkt	1.15	19.493	***	0.979	25.283	***	0.907	25.664	***	0.888	30.272	***	0.898	28.651	***	0.253	3.231	**
SMB	0.023	0.175		0.134	1.534		0.062	0.773		-0.021	-0.312		-0.341	-4.839	***	0.365	2.07	**
HML	0.003	0.022		-0.167	-1.665	*	-0.17	-1.854	*	-0.154	-2.031	**	0.131	1.622		-0.128	-0.633	
MoM	-0.322	-5.305	***	-0.012	-0.311		0.033	0.909		0.15	4.977	***	0.23	7.157	***	-0.552	-6.869	***
BaB	-0.151	-2.201	**	0.088	1.961	*	0.152	3.704	***	0.126	3.681	***	0.028	0.78		-0.18	-1.973	*
Information Ratio	-0.104			0.463			0.219			0.175			-0.687			0.197		
R Squared	0.9			0.95			0.95			0.96			0.96			0.48		
Observations	83			83			83			83			83			83		

Table 21: Cointegration Technique, Value Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat	
Alpha	0.001	0.34		0.002	1.786	*	0	0.34		0	0.097		-0.002	-1.716	*	0.002	0.945	
Ret_Mkt	1.154	19.103	***	0.993	18.932	***	0.941	22.385	***	0.945	24.802	***	0.998	29.262	***	0.157	1.871	*
SMB	0.163	1.313		0.098	0.909		0.124	1.435		0.137	1.749	*	-0.212	-3.025	**	0.374	2.181	**
HML	-0.097	-0.675		-0.131	-1.047		-0.019	-0.189		-0.106	-1.163		0.208	2.557	**	-0.305	-1.529	
MoM	-0.402	-6.84	***	-0.06	-1.179		-0.096	-2.339	**	0.071	1.926	*	0.174	5.257	***	-0.576	-7.082	***
BaB	-0.146	-2.119	**	0.02	0.334		0.123	2.551	**	0.056	1.277		-0.038	-0.969		-0.109	-1.136	
Information Ratio	0.129			0.679			0.129			0.037			-0.653			0.359		
R Squared	0.88			0.89			0.93			0.94			0.95			0.4		
Observations	83			83			83			83			83			83		

5.5 Quantile Descriptive Statistics

In this paragraph I will dig more into the quintiles firm differences, and how my models detect Undervaluation or Overvaluation of firms.

I then compare the results between the quintiles and differences in the characteristics and the performances.

I only show the first two estimation horizons, because similar conclusions can be found when the period taken into account is longer.[1]

5.5.1 Method A

In table 22 there is a summarized description of the companies that enter in the different quintiles. The quintiles are divided depending on the misprices, where Q1 are the underpriced firms and Q5 are the overpriced ones.

Horizon 1 What it is clear by looking at the table, is that the underpriced companies and the overpriced companies are quite similar compared to those that enter in the middle quintiles. Indeed, if we look at the Total Asset we can see that the entity of this variable in Q1 and Q5 is a lot higher compared to the companies that belong to the Q2, Q3 and Q4.

By looking at the independent variables we see that the underpriced companies have a higher income, more assets, and a lower debt. Other interesting aspects are that the underpriced companies are indeed smaller compared to the overpriced, with higher Book to Market, and also higher Roe. However, the smallest and more value firms are those in the middle quintiles.

The next statistics are about the future and past return, where we see a difference in the average future return between Overvalued and Undervalued firms, but the returns are not monotonically decreasing, where indeed the firms with higher future return are clustered in the middle quintiles. Looking at the past return we see that the underpriced companies are past losers compared to the overpriced.

[1]The quintile statistics for the 24 and 60 months horizon can be seen in the appendix.

Horizon 12 By increasing the horizon of the estimation, the results in the different quintiles for the independent variables do not change that much. The behaviours are just a little bit less or more emphasized, depending on the variable considered. There is indeed a larger difference in the income and in the return volatility and the systematic risk of the companies that belong to Q1 in comparison to the ones that belong to Q5. The Book to Market is still higher for Q1 in respect of Q5.

By looking at the return section, we can see that there is a larger gap between the extreme quintiles. We can also see that the differences between the past return, with respect to the horizon 1, is much higher when we look at the past trading month, but it becomes of the same entity when we look at the past 60 trading months.

5.5.2 Method B

The results that come out from method B are not far from those that are observed in method A, especially when we compare the independent variables. The results can be found in table 23.

Horizon 1 Here the difference between the quintiles in all the independent variables is more emphasized. The total assets difference is still present when we compare the Q1 and Q5 with Q2, Q3 and Q4.

Right under the independent variables we can see that the companies belonging to Q1 are really similar in size compared to the ones that belong to Q5. This is not a good result, meaning that the model is driven by the company size.

Regarding the B/M we now have that the overpriced companies have a higher B/M compared to the underpriced. However, like the previous model, the quintiles with higher B/M are those in the middle.

The difference in the Roe is now much more emphasized, meaning that the underpriced companies are more profitable than before with respect to the overpriced ones.

Even in the return section we have similar results as before. In this case we have a smaller gap for the future return and a lower difference in the past return.

Horizon 12 The consequences of averaging the coefficient have a similar outcome with respect to method A. Indeed, the variables have similar behaviour as in the horizon 1, just a bit more accentuated.

With this model we see minimal changes for the dependent and independent variables,similar to the one already described in method A.

Even in the return section we see similarities, with higher return in the underpriced quintile, and a much higher difference in the past return.

5.5.3 Cointegration Model

As for methodology A and B I have summarized the results per quintiles.

What we see in table 24 is a really similar behaviour in the quintile formation in respect of the Fama-MacBeth approach. Even here we can see a polarization of firms with high Total assets in the extreme quintiles. What is also noticeable is that now we have a significant difference in the size of the overpriced and the underpriced companies. On average the underpriced firms have lower Assets, higher Beta exposure and have less income. Regarding the Beta exposure we see that the extreme quintiles are, on average, riskier with respect to the one in the middle. Another thing to underline is that the underpriced quintile (Q1) gives on average higher dividend yield. The Book to Market ratio is high for the underpriced firms and it slowly decreases. What is also possible to observe is that the overpriced quintile has a higher average future return with respect to the underpriced quintile, and that the firms with higher return are the one in the middle. Regarding the average past return, the underpriced can be labeled as past loser.

In conclusion, we can summarize all the findings by saying that, in methodology A and B the underpriced firms are on average smaller with higher value with respect to the overpriced ones. They are also more profitable, with higher future returns and lower past returns. For methodology A and B the companies belonging to Q1 are on average riskier. Regarding the cointegration model the underpriced companies are also smaller with lower assets, higher dividend yield, lower profitability, and with lower future returns compared to the overvalued ones.

Table 22: This table shows the summary statistics for each quintile. The estimation of the quintiles has been done using methodology A. It is divided into two blocks, horizon 1 and horizon 12, depending on the estimation horizon of the coefficients. On the left column there are the characteristics that I have taken into consideration. The first twelve are the variables that I use in my model (missing the dummy on dividend yield), the 13th is the dependent variable of my model. Underneath, there is the average Book to Market and the Average Return on Equity. Then there is a little summary on the quintile return, where I look at the average return in the next trading month, the percentage of positive return during all the trading months and the average past return of: previous month, previous 12 months and previous 60 months. All the values are calculated by first taking a cross-sectional average and then an average on the time series. All the variables are multiplied by 10^x in order to have the similar level

	horizon 1					**horizon 12**				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**
Total Assets	10.08	5.09	4.28	4.05	10.97	10.17	5.11	4.23	4.04	10.92
Dividend Yield	5.37	4.24	3.87	3.69	4.19	5.33	4.15	3.88	3.76	4.23
Income Tax	3.31	1.71	1.25	1.52	2.63	3.43	1.74	1.51	1.49	2.25
Net Income	8.51	3.87	3.52	3.41	5.69	9.06	3.97	3.45	3.49	5.03
Operational Income	14.14	6.83	6.05	6.06	10.85	14.34	6.82	5.95	6.15	10.68
Sales	11.61	5.63	4.79	4.77	9.58	11.6	5.7	4.69	4.69	9.69
Shareholder Equities	3.77	2.16	1.9	1.95	2.99	3.75	2.16	1.88	1.98	2.99
Other Assets	2.31	1.27	1.12	1.11	2.99	2.33	1.23	1.08	1.12	3.02
Total Debt	2.4	1.27	1.13	1.18	3	2.39	1.22	1.13	1.2	3.04
Betas	1.15	1.11	1.11	1.1	1.11	1.16	1.11	1.12	1.08	1.1
Return Volatility	0.79	0.69	0.72	0.72	0.8	0.91	0.69	0.7	0.7	0.73
Trend	130.05	111.64	91.58	82.36	110.87	130.41	111.66	91.73	82.95	109.76
Market Cap	16.13	14.07	14.62	16.11	28.19	15.77	13.96	14.67	16.56	28.17
Book to Market	0.3	0.32	0.33	0.3	0.2	0.31	0.32	0.34	0.29	0.2
Roe	0.23	0.24	0.31	0.3	0.15	0.22	0.26	0.32	0.31	0.12
Quantile Return t+1	0.97	0.84	0.92	0.98	0.68	1.09	0.87	0.86	0.89	0.68
% Positive Return	0.57	0.57	0.56	0.57	0.55	0.57	0.57	0.56	0.57	0.55
Return t	0.29	0.49	0.82	1.25	1.78	0.25	0.45	0.84	1.25	1.84
Return (t-12) - t	4.87	5.31	7.04	11.42	19.86	4.1	4.42	7.4	12.24	20.49
Return (t-60) - t	24.85	22.1	28.09	41.19	66.43	25.04	20.55	29.44	42.01	67.23

Table 23: This table shows the same quintile summary statistics of the previous table for method B.

	horizon 1					**horizon 12**				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**
Total Assets	8.71	5.29	4.56	4.52	11.38	8.86	5.32	4.48	4.42	11.39
Dividend Yield	4.99	4.48	4.09	3.75	4.05	5.01	4.35	4.15	3.69	4.16
Income Tax	3.15	1.84	1.36	1.56	2.52	3.38	1.84	1.6	1.53	2.07
Net Income	8.21	4.23	3.77	3.4	5.39	8.93	4.26	3.6	3.48	4.72
Operational Income	13.6	7.21	6.33	6.23	10.57	14.05	7.22	6.22	6.21	10.24
Sales	10.97	5.9	5.05	4.88	9.57	11.14	5.84	4.84	4.93	9.63
Shareholder Equities	3.6	2.18	1.98	2.03	2.98	3.59	2.2	1.96	2.05	2.96
Other Assets	2.1	1.32	1.15	1.13	3.09	2.14	1.26	1.08	1.19	3.12
Total Debt	2.21	1.3	1.14	1.22	3.11	2.22	1.22	1.14	1.24	3.16
Betas	1.15	1.12	1.12	1.08	1.1	1.16	1.14	1.11	1.08	1.09
Return Volatility	0.78	0.73	0.72	0.74	0.76	0.87	0.72	0.69	0.7	0.74
Trend	128.95	108.89	92.01	85.46	111.19	130.19	107.57	92.11	86.43	110.21
Market Cap	19.83	14.49	13.16	14.58	27.06	19.41	14.43	13.23	15.31	26.73
Book to Market	0.27	0.32	0.33	0.31	0.22	0.27	0.32	0.34	0.31	0.21
Roe	0.36	0.3	0.22	0.23	0.12	0.34	0.32	0.24	0.24	0.1
Quantile Return t+1	0.94	0.84	0.83	1.03	0.75	1	0.9	0.85	0.94	0.71
% Positive Return	0.56	0.56	0.56	0.57	0.55	0.57	0.56	0.56	0.57	0.55
Return t	0.38	0.52	0.71	1.21	1.81	0.17	0.43	0.78	1.32	1.93
Return (t-12) - t	5.64	5.74	7.22	10.88	19.02	5.08	4.89	7.03	11.73	19.94
Return (t-60) - t	25.69	24.88	31.6	36.94	62.4	26.19	25.93	28.86	37.27	64.06

Table 24: In this table we can see the same summary statistics of the tables above for the cointegration model.

	Q1	Q2	Q3	Q4	Q5
Total Assets	10.4	5.19	5.01	6.66	13.88
Dividend Yield	8.97	8.79	8.49	8.57	8.33
Income Tax	2.31	1.53	1.68	2.37	5.29
Net Income	4.41	3.65	3.9	5.55	12.66
Operational Income	9.58	6.47	6.71	9.29	19.56
Shareholder Equities	2.74	2.2	2.45	2.81	4.73
Other Assets	2.45	1.35	1.33	1.66	3.31
Total Debt	2.66	1.28	1.26	1.51	3.14
Betas	1.15	1.04	1.03	1.03	1.06
Return Volatility	0.81	0.61	0.56	0.52	0.54
Trend	143.78	134.03	133.55	140.32	146.68
Market Cap	24.75	18.41	19.2	23.52	34.84
Book to Market	0.29	0.26	0.25	0.2	0.13
Roe	0.22	0.26	0.32	0.28	0.21
Quantile Return t+1	0.75	1.03	0.96	0.98	0.77
% Positive Return	0.55	0.59	0.58	0.58	0.57
Return t	-0.6	0.48	1.11	1.58	2.09
Return (t-12) - t	-6.05	2.81	8.22	14.19	22.22
Return (t-60) - t	-0.96	11.89	28.08	49.08	88.53

6 Conclusion

In this work, we have seen how a strategy based on mispricing signals would perform. We have seen two different techniques to estimate these signals: one based on cross-sectional regression and one based on cointegration panel estimation. The results from the first approach suggest that fundamental analysis actually works. By investing on the mispricing signals the portfolio leads to a profitable strategy. We have also seen that the profitability increases when the signal estimation becomes more robust. Indeed, when the coefficients are averaged across time the profitability increases. Not the same can be said when we add dummies for sector and book to market, that decrease the performances.

The fact that Fundamental Analysis works, it is in contradiction with the market efficiency theory because if all the information are already reflected in the price, mispricing signals should therefore only be noise. The positive return in the estimated portfolios and the better performance when the coefficients are averaged across time suggest that the market prices are not correct at each point in time but tend to correct on a longer horizon and that these prices converge on what we think to be the fair value.

The mispricing can ,therefore, be divided into a cross-sectional component and a time series component. The measurement of the latter resulted in this paper to be a key factor in the portfolios higher returns, suggesting the bigger importance of detecting period where investors over(under) weight certain firm characteristics. The cross-sectional dimension alone assumes that the market in every moment is correct, not allowing the possibility of investors irrational exuberance for certain types of information.

We have also seen that for method A, the portfolios based on its mispricing signals lead to significant alphas. The fact that for method B we did not get the same results suggest that adding group differences add information that lower the power of the mispricing signal.

Even for the portfolio made using the mispricing signal estimated with the CUP-FM technique, the performances decrease.

It is known that in this case, with the presence of non-stationary variables, cointegration estimation techniques have to be used. However, these techniques rely on several heavy assumptions and have more parameter uncertainty with respect to a normal OLS. These models are indeed difficult to handle and difficult to estimate. This is the reason why these models are not really used in practice, and are replaced by easier ones. However nowadays, where panel techniques are used on datasets with long time series, it is important to better estimate the long run relation between the variables. Therefore, these techniques will need to be applied more often, and they will grow in importance in the Financial Industry.

References

Abarbanell, J. S., & J.Bushee, B. (1997). Fundamental analysis, future earnings, and stock prices. *Journal of Accounting Research*, *35(1)*, 1-24.

Bai, J., & Kao, C. (2005). On the estimation and inference of panel cointegratio model with cross-sectional depencence. *Centre for Policy Research*, *Paper 89*.

Bai, J., & Ng, S. (2002). Determining the number of factors in approxmate factor model. *Econometrica*, *70*, 191-221.

Bai, J., & Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, *72*, 1127-1177.

Baker, M., & Ruback, R. S. (1999). Estimating industry multiples. *Working Paper*.

Baltagi, B. H. (2005). *Econometric analysis of panel data*. John Wiley and Sons, Ltd.

Bartram, S. M., & Grinblatt, M. (2014). Fundamental analysis works. *Working Paper*.

Breitung, J., & Pesaran, M. H. (2005). Unit roots and cointegration in panels. *Working Paper*.

Cooper, I. A., & Lambertides, N. (2014). Is there a limit to the accuracy of equity valuation using multiples?

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

Frankel, R., & Lee, C. M. (1998). Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics*, *25*, 283-319.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*, 1-25.

Gengenbach, C., Palm, F. C., & Urbain, J.-P. (2005). Panel cointegration testing in the presence of common factors. *Working paper*.

Greig, A. C. (1992). Fundamental analysis and subsequent stock returns. *Journal of Accounting and Economics*.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, *90*, 1-44.

Kao, C., & Chiang, M.-H. (1999). On the estimation and inference of a cointegrated regression in panel data. *Centre for Policy Research*, *Paper 145*.

Lev, B., & Thiagarajan, S. R. (1993). Fundamental information analysis. *Journal of Accounting Research*, *31(2)*, 190-215.

Levin, A., Lin, C.-F., & Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample proprieties. *Journal of Econometrics*, *108*, 1-24.

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal of Accounting Research*, *40(1)*, 135-172.

Moon, H., B, P., & Phillips, P. (2006). On the breitung test for panel unit roots and local asymptotic power. *Cambridge University Press*, *22*, 1179-1190.

Moon, H., & Phillips, P. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, *67*, 1057-1111.

Moon, H., & Phillips, P. (2000). Estimation of autoregressive roots near unity using panel data. *Cambridge University Press*, *16*, 927-997.

Moon, H. R., & Perron, B. (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics*, *122*, 81-126.

OU, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*.

Pesaran, M. (2004). General diagnostic tests for cross section dependence in panels. *Trinity Collage, Cambridge, Working paper*.

Phillips, P. C. (1993). Fully modified least squares. *Cowles Foundation for Research in Economics, Yale University*.

Richardson, S., Tuna, I., & Wysocki, P. (2010). Accounting anomalies and fundamental analysis: A review of recent research advances. *Journal of Accounting and Economics*, *50*, 410-454.

Shiller, R. J., & Cambell, J. K. (1987). Cointegration and tests for present value models. *Journal of Political Economic*, *95*, 1062-1088.

Skoulakis, G. (2008). Panel data inference in finance: Least-squares vs fama-macbeth. *SSRN Electronic Journal*.

Appendix

Summary statistic per Quantile

Table 25: This table shows the summary statistics of each quintile. The estimation of the quintiles has been done using method A. It is divided into two blocks, horizon 24 and horizon 60, depending on the Estimation horizon of the Beta estimates. On the column on the left there are all the characteristics that I have taken into consideration. The first nine are the variables that I use in my model, the 10th is the dependent variable of my model. Underneath, there is the average Book to Market and the Average Return on Equity. Then there is a little summary on the quintiles return, where I look at the average return in the next trading month, the percentage of positive return during all the trading months and the average past return: previous month, previous 12 months and previous 60 months. All the variables are multiplied by 10^x n order to have the same level

	horizon 24					horizon 60				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**
Total Assets	10	5.16	4.19	4.06	11.06	9.46	5.1	4.46	4.37	11.08
Dividend Yield	5.09	4.1	3.95	3.86	4.37	4.63	4.09	3.89	3.86	4.89
Income Tax	3.31	1.77	1.53	1.56	2.25	2.99	1.82	1.62	1.68	2.31
Net Income	9.22	4.05	3.44	3.51	4.78	9.13	4.21	3.66	3.7	4.29
Operational Income	14.24	6.85	6	6.21	10.64	13.8	7.06	6.3	6.55	10.23
Sales	11.61	5.59	4.73	4.76	9.68	11.52	5.82	4.91	4.94	9.19
Shareholder Equities	3.75	2.15	1.88	2.01	2.98	3.78	2.11	1.92	2.1	2.87
Other Assets	2.26	1.27	1.12	1.1	3.04	2.19	1.34	1.24	1.13	2.89
Total Debt	2.37	1.19	1.12	1.21	3.09	2.23	1.24	1.22	1.23	3.06
Betas	1.17	1.12	1.12	1.08	1.1	1.19	1.12	1.13	1.06	1.08
Return Volatility	0.84	0.69	0.71	0.69	0.79	0.82	0.69	0.73	0.71	0.77
Trend	130.89	111.66	90.98	83.54	109.43	135.87	114.83	91.85	82.13	101.83
Market Cap	15.47	14.07	14.58	16.71	28.29	15.79	14.5	15.74	17.06	26.02
Book to Market	0.3	0.32	0.34	0.29	0.2	0.29	0.32	0.33	0.3	0.22
Roe	0.24	0.25	0.33	0.29	0.12	0.21	0.27	0.36	0.25	0.14
Quantile Return t+1	1.05	0.91	0.88	0.86	0.69	1.04	0.92	0.86	0.86	0.71
% Positive Return	0.57	0.57	0.56	0.56	0.55	0.57	0.57	0.56	0.56	0.55
Return t	0.09	0.49	0.86	1.28	1.91	0.16	0.51	0.84	1.24	1.87
Return (t-12) - t	4.37	4.38	7.7	11.83	20.39	4.44	4.98	7.99	11.34	20.23
Return (t-60) - t	24.34	20.27	30.57	42.39	67.29	23.85	19.47	32.09	45.95	69.33

Table 26: This table shows the same quintile summary statistics of the previous table, but I have used method B in order to estimate the mispricing signals

	horizon 24					horizon 60				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**
Total Assets	8.81	5.31	4.58	4.26	11.51	8.18	5.64	4.7	4.46	11.5
Dividend Yield	4.87	4.26	4.16	3.75	4.31	4.43	4.15	4.13	4.07	4.58
Income Tax	3.3	1.88	1.6	1.53	2.12	2.95	1.96	1.67	1.58	2.26
Net Income	9.12	4.4	3.59	3.43	4.45	9.08	4.47	3.81	3.49	4.15
Operational Income	14	7.28	6.24	6.13	10.3	13.6	7.4	6.52	6.32	10.1
Sales	11.17	5.76	4.9	4.85	9.69	11.09	5.85	5.11	4.98	9.35
Shareholder Equities	3.62	2.16	1.98	2.03	2.97	3.6	2.23	2.02	2.04	2.88
Other Assets	2.1	1.28	1.1	1.16	3.14	1.98	1.4	1.2	1.19	3.02
Total Debt	2.21	1.23	1.14	1.2	3.19	2.07	1.32	1.2	1.26	3.13
Betas	1.16	1.14	1.11	1.08	1.09	1.17	1.15	1.12	1.07	1.07
Return Volatility	0.79	0.7	0.71	0.7	0.81	0.8	0.71	0.71	0.72	0.77
Trend	131.15	107.04	91.17	87.21	109.94	135.43	109.63	92.26	85.6	103.59
Market Cap	19.37	14.62	13.04	15.32	26.77	20.38	15.16	13.05	14.85	25.68
Book to Market	0.27	0.31	0.35	0.31	0.22	0.25	0.31	0.34	0.32	0.24
Roe	0.37	0.3	0.23	0.23	0.1	0.38	0.29	0.23	0.21	0.11
Quantile Return t+1	0.98	0.89	0.85	0.95	0.71	0.96	0.9	0.81	0.97	0.75
% Positive Return	0.57	0.57	0.56	0.57	0.55	0.56	0.57	0.56	0.57	0.56
Return t	0.05	0.41	0.72	1.41	2.03	0.14	0.34	0.81	1.32	2.01
Return (t-12) - t	4.92	4.77	6.87	12.14	19.99	4.92	4.82	7.38	11.61	20.29
Return (t-60) - t	25.24	26.15	29.55	37.89	64.15	25.77	25.24	31.43	38.42	64.39

Table 27: This table shows the coefficients, and the relative t statistics, of the model that I estimate in order to check the presence of a significant alpha. The table is divided into 6 columns, one for each quintile plus one for the Q1 minus Q5 portfolio. For each column I show the coefficients for the models where I consider the anomalies: SMB, HML, MOM and BaB. In addition at the bottom of every column there is the information ratio of the portfolio, the R-squared and the number of observation used to estimate the model. To estimate the Information ratio I have followed the technique used in Bartram and Grinblatt (2014) paper, I have multiplied the t-statistic of the intercept by, $\sqrt{\frac{12}{T}}$. SMB, HML and MOM factors are build following the Kenneth and French dataset approach, and the BaB factor following Frazzini and Pedersen (2014). For this table the portfolio quintiles are made using the 24 months averaged beta estimates. The portfolio is Equally Weighted. The following table represents the results for method A.

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**
Alpha	0.001	1.584	0	0.328	0	-0.306	0	-0.172	-0.001	-1.543	0.003	1.818 *
Ret_Mkt	1.102	31.185 ***	1.005	35.162 ***	0.97	36.824 ***	0.866	28.581 ***	0.978	27.607 ***	0.124	2.036 **
SMB	0.224	3.017 **	0.042	0.707	0.175	3.165 **	0.031	0.495	-0.391	-5.26 ***	0.615	4.815 ***
HML	-0.164	-2.158 **	-0.27	-4.403 ***	-0.141	-2.503 **	0.132	2.024 **	0.229	3.014 **	-0.393	-3.008 **
MoM	-0.119	-4.4 ***	-0.062	-2.826 **	-0.023	-1.12	0.128	5.499 ***	0.062	2.302 **	-0.182	-3.895 ***
BaB	-0.12	-2.558 **	0.046	1.203	0.044	1.244	0.119	2.94 **	-0.021	-0.449	-0.099	-1.224
Information Ratio	0.46		0.095		-0.089		-0.05		-0.449		0.529	
R Squared	0.96		0.97		0.97		0.96		0.95		0.41	
Observations	142		142		142		142		142		142	

Table 28: This table shows the same results as in the table above, but in this case the portfolios are value weighted. The signals are estimated using method A.

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**	β	**T - stat**
Alpha	0.002	1.587	0.002	1.955 *	0	0.292	0.001	1.257	-0.002	-3.333 **	0.004	2.663 **
Ret_Mkt	1.134	25.011 ***	1.016	28.263 ***	0.886	25.346 ***	0.928	23.294 ***	1.002	32.984 ***	0.131	2.021 **
SMB	0.208	2.402 **	0.087	1.269	0.205	3.078 **	0.111	1.458	-0.311	-5.363 ***	0.519	4.179 ***
HML	0.115	1.25	-0.208	-2.862 **	0.021	0.299	-0.164	-2.029 **	0.173	2.809 **	-0.058	-0.441
MoM	-0.043	-1.27	-0.091	-3.364 ***	-0.031	-1.167	0.087	2.895 **	0.032	1.402	-0.076	-1.54
BaB	-0.148	-2.463 **	0.061	1.284	0.109	2.362 **	0.023	0.444	-0.003	-0.082	-0.145	-1.678 *
Information Ratio	0.461		0.568		0.085		0.365		-0.969		0.774	
R Squared	0.92		0.94		0.94		0.91		0.95		0.31	
Observations	142		142		142		142		142		142	

Table 29: 60 months horizon method A, Equally Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**	
Alpha	0.002	1.983	**	0	0.371		0	-0.48		0	-0.391		-0.001	-1.428		0.003	1.952	*
Ret_Mkt	1.094	32.038	***	0.991	35.584	***	0.993	34.375	***	0.89	27.98	***	0.953	25.273	***	0.141	2.271	**
SMB	0.058	0.805		0.048	0.819		0.21	3.471	***	0.021	0.321		-0.256	-3.235	**	0.314	2.399	**
HML	-0.1	-1.364		-0.213	-3.557	***	-0.289	-4.67	***	0.178	2.61	**	0.21	2.602	**	-0.31	-2.322	**
MoM	-0.116	-4.426	***	-0.064	-3.012	**	0.022	1.012		0.094	3.863	***	0.05	1.746	*	-0.166	-3.483	***
BaB	-0.131	-2.877	**	0.081	2.188	**	-0.01	-0.26		0.11	2.598	**	0.017	0.33		-0.147	-1.777	*
Information Ratio	0.577			0.108			-0.14			-0.114			-0.415			0.567		
R Squared	0.96			0.97			0.97			0.96			0.94			0.26		
Observations	142			142			142			142			142			142		

Table 30: 60 months horizon method A , Value Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**		β	**T - stat**	
Alpha	0.001	1.344		0.002	2.519	**	0	-0.488		0.001	0.631		-0.002	-2.417	**	0.003	2.036	**
Ret_Mkt	1.106	23.979	***	0.986	29.334	***	0.959	26.873	***	0.897	22.459	***	1.016	35.098	***	0.091	1.4	
SMB	0.08	0.911		0.114	1.769	*	0.099	1.446		0.067	0.879		-0.213	-3.86	***	0.294	2.372	**
HML	0.126	1.355		-0.21	-3.089	**	-0.11	-1.52		0.091	1.127		0.113	1.924	*	0.014	0.105	
MoM	-0.023	-0.666		-0.085	-3.342	**	-0.004	-0.145		0.074	2.441	**	0.027	1.236		-0.05	-1.026	
BaB	-0.146	-2.393	**	0.107	2.41	**	0.081	1.723	*	0.015	0.277		-0.027	-0.704		-0.119	-1.389	
Information Ratio	0.391			0.732			-0.142			0.184			-0.703			0.592		
R Squared	0.91			0.95			0.94			0.91			0.96			0.15		
Observations	142			142			142			142			142			142		

Table 31: 24 months horizon method B, Equally Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat	
Alpha	0.001	0.981		0	0.21		0	-0.531		0.001	1.025		-0.001	-1.448		0.002	1.39	
Ret_Mkt	1.13	33.61	***	1.03	31.939	***	0.985	32.627	***	0.846	28.868	***	0.929	23.878	***	0.201	3.123	**
SMB	-0.038	-0.535		0.072	1.064		0.208	3.281	**	0.101	1.644		-0.261	-3.199	**	0.224	1.658	*
HML	-0.071	-0.983		-0.17	-2.462	**	-0.26	-4.012	***	0.08	1.27		0.208	2.494	**	-0.279	-2.025	**
MoM	-0.053	-2.075	**	-0.065	-2.622	**	-0.031	-1.348		0.081	3.622	***	0.054	1.82	*	-0.108	-2.188	**
BaB	-0.145	-3.235	**	-0.045	-1.045		0.015	0.367		0.174	4.45	***	0.068	1.318		-0.213	-2.492	**
Information Ratio	0.285			0.061			-0.154			0.298			-0.421			0.404		
R Squared	0.96			0.96			0.97			0.96			0.94			0.23		
Observations	142			142			142			142			142			142		

Table 32: 24 months horizon method B, Equally Weighted Portfolios

	Q1			Q2			Q3			Q4			Q5			Q1 - Q5		
	β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat		β	T - stat	
Alpha	0.001	0.678		0.002	1.719	*	0	0.113		0.001	0.987		-0.002	-2.289	**	0.002	1.557	
Ret_Mkt	1.102	28.151	***	1.04	25.547	***	0.99	26.626	***	0.802	23.693	***	0.984	31.056	***	0.118	1.86	*
SMB	0.024	0.325		0.009	0.117		0.187	2.634	**	0.107	1.656		-0.184	-3.034	**	0.208	1.71	*
HML	0.013	0.16		0.033	0.4		-0.201	-2.677	**	0.018	0.264		0.101	1.578		-0.088	-0.687	
MoM	0.074	2.502	**	-0.079	-2.564	**	-0.046	-1.644		0.059	2.301	**	-0.033	-1.376		0.107	2.224	**
BaB	-0.147	-2.824	**	-0.103	-1.916	*	0.008	0.154		0.256	5.712	***	0.048	1.152		-0.195	-2.311	**
Information Ratio	0.197			0.5			0.033			0.287			-0.665			0.453		
R Squared	0.93			0.92			0.93			0.94			0.95			0.06		
Observations	142			142			142			142			142			142		

Table 33: 60 Months horizon method B, Equally Weighted Portfolios

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	**β**	**T - stat**	**β**	**T - stat**	**β**	**T - stat**	**β**	**T - stat**	**β**	**T - stat**
Alpha	0.001	1.142	0	-0.083	-0.001	-0.968	0.001	1.231	-0.001	-1.163	0.002	1.292
Ret_Mkt	1.097	32.257 ***	1.077	37.363 ***	0.973	33.608 ***	0.861	28.665 ***	0.912	22.156 ***	0.185	2.752 **
SMB	-0.138	-1.932 *	0.071	1.165	0.223	3.67 ***	0.12	1.897 *	-0.193	-2.236 **	0.055	0.393
HML	-0.039	-0.53	-0.241	-3.889 ***	-0.219	-3.52 ***	0.047	0.724	0.239	2.701 **	-0.277	-1.925 *
MoM	-0.08	-3.073 **	-0.08	-3.621 ***	0.03	1.373	0.065	2.841 **	0.051	1.616	-0.131	-2.547 **
BaB	-0.121	-2.676 **	-0.054	-1.417	0.001	0.038	0.137	3.418 ***	0.104	1.902 *	-0.225	-2.522 **
Information Ratio	0.332		-0.024		-0.281		0.358		-0.338		0.376	
R Squared	0.96		0.97		0.97		0.96		0.93		0.14	
Observations	142		142		142		142		142		142	

Table 34: 60 Months horizon method B, Value Weighted Portfolios

	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	β	**T - stat**	**β**	**T - stat**	**β**	**T - stat**	**β**	**T - stat**	**β**	**T - stat**	**β**	**T - stat**
Alpha	0	0.532	0	0.53	0	0.115	0.002	2.532 **	-0.001	-1.675 *	0.002	1.234
Ret_Mkt	1.046	29.218 ***	1.101	32.515 ***	0.96	30.802 ***	0.854	23.624 ***	0.975	28.492 ***	0.071	1.145
SMB	-0.047	-0.693	0.03	0.465	0.132	2.22 **	0.145	2.105 **	-0.138	-2.113 **	0.091	0.767
HML	0.031	0.432	-0.043	-0.623	-0.097	-1.53	-0.042	-0.57	0.152	2.193 **	-0.121	-0.962
MoM	-0.004	-0.153	-0.094	-3.688 ***	0.053	2.269 **	0.044	1.601	-0.017	-0.653	0.013	0.272
BaB	-0.06	-1.258	-0.113	-2.522 **	-0.006	-0.134	0.165	3.446 ***	0.057	1.257	-0.117	-1.422
Information Ratio	0.155		0.154		0.033		0.736		-0.487		0.359	
R Squared	0.94		0.95		0.95		0.93		0.94		-0.01	
Observations	142		142		142		142		142		142	

Figure 5: This figure shows the performance of the Equally weighted long - short portfolios created using the signals estimated with method A. The plot is the comulative return of the strategy across time, from the beginning of 2004 till August 2016. The colors are referred to the different time horizon that I use to average the coefficients. On the North-west part of the figure there is the horizon's length associated with the colors.

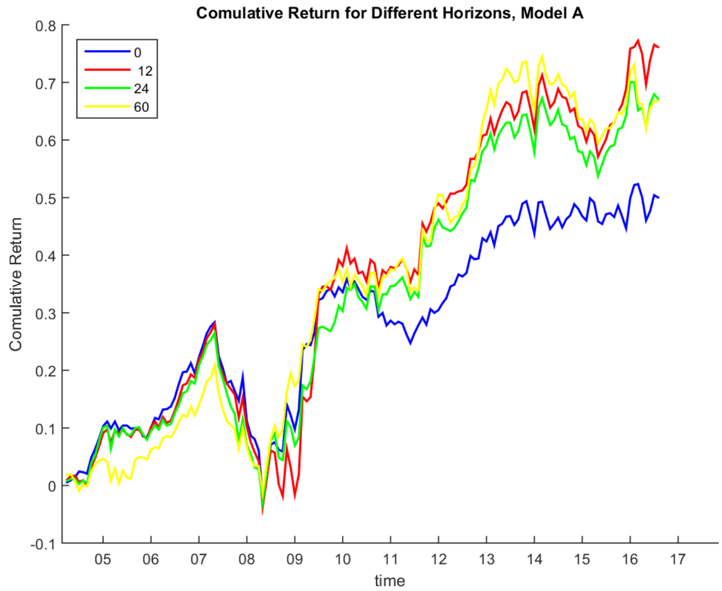


Figure 6: This figure shows the performance of the Value weighted long - short portfolios created using the signals estimated with method A. For more description on the graph look at figure 5

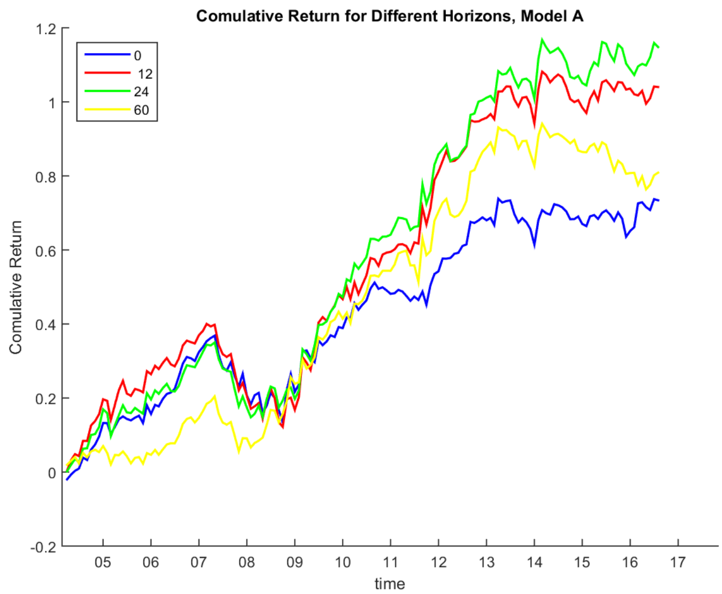


Figure 7: This figure shows the performance of the Equally weighted long - short portfolios created using the signals estimated with method B. For more description on the graph look at figure 5

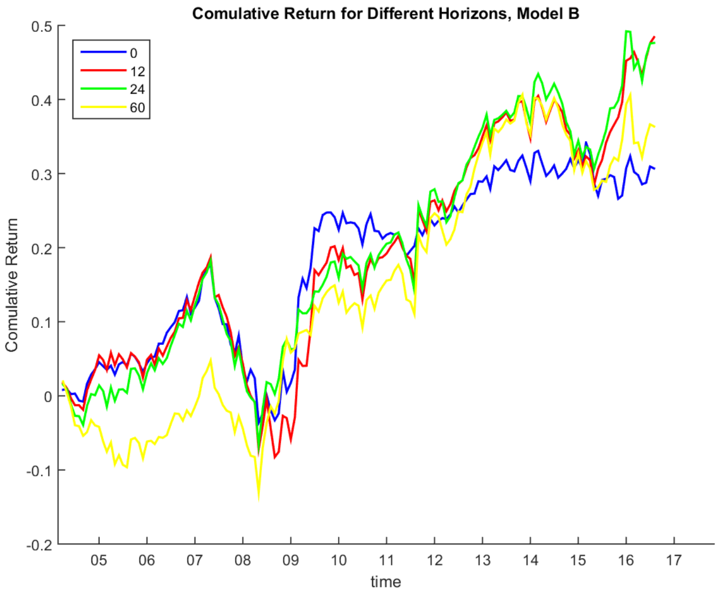


Figure 8: This figure shows the performance of the Value weighted long - short portfolios created using the signals estimated with method B. For more description on the graph look at figure 5

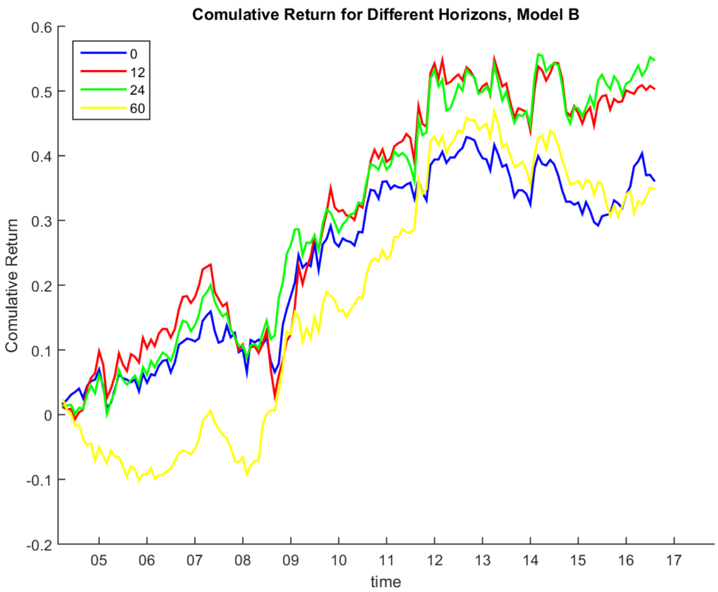


Figure 9: This figure shows the performance of the Equally weighted long - short portfolios created using the signals estimated with the cointegration technique. For more description on the graph look at figure 5

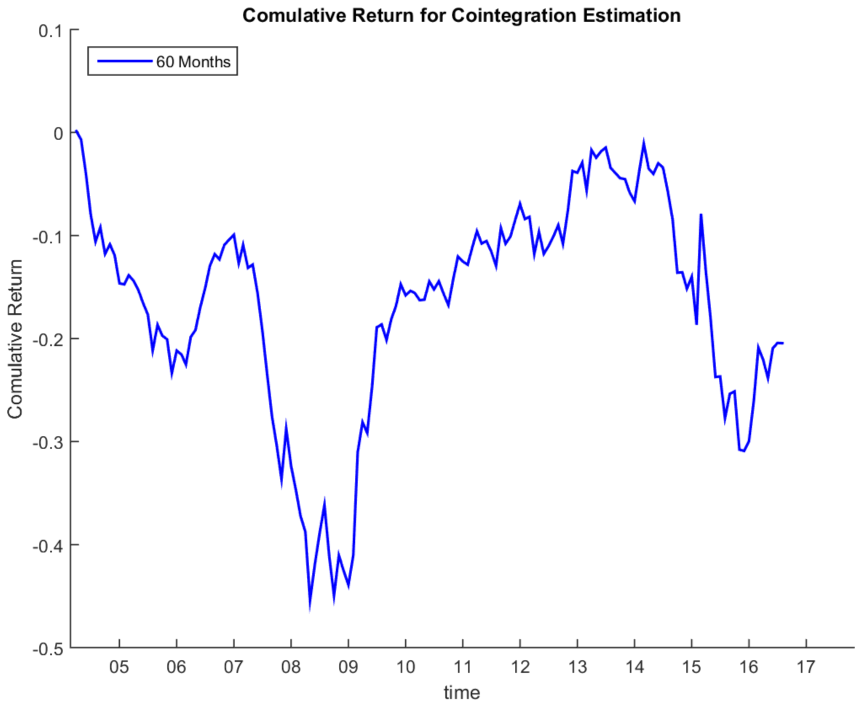


Figure 10: This figure shows the performance of the Value weighted long - short portfolios created using the signals estimated with the cointegration technique. For more description on the graph look at figure 5

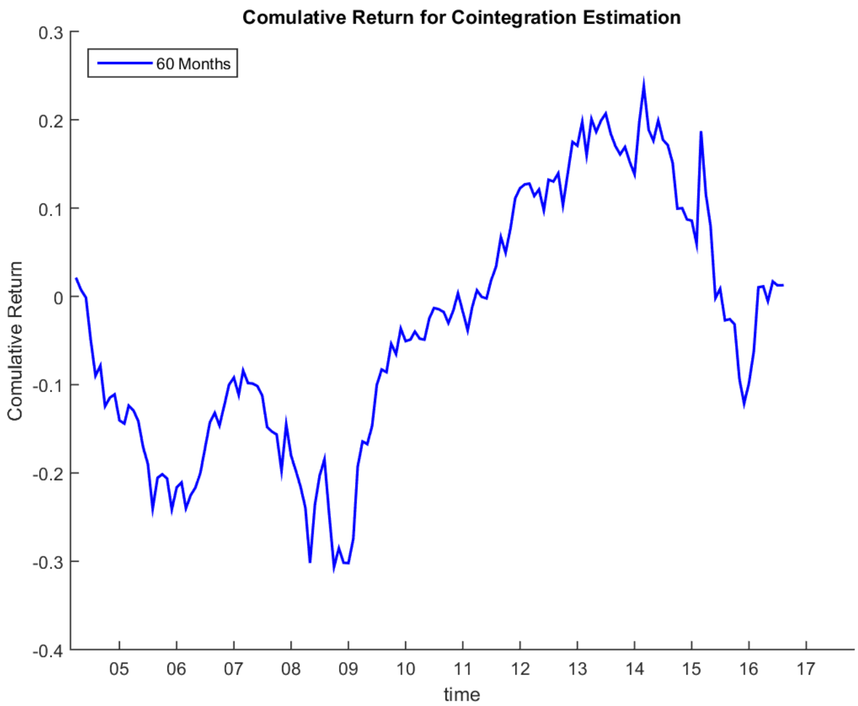


Table 35: In this table are available all the description of the variables used in this work. All the information are taken from the Datatypes definition guide of Worldscope Database.

Variables	**Descrption**
Accounting Receivable	represent the amounts due to the company resulting from the sale of goods and services on credit to customers (after applicable reserves). These assets should reasonably be expected to be collected within a year or within the normal operating cycle of a business.
Beta	A measure of market risk which shows the relationship between the volatility of the stock and the volatility of the market. This coefficient is based on between 23 and 35 consecutive month end price percent changes and their relativity to a local market index.
Cash and Short term Investment	represents the sum of cash and short term investments.
Current Assets	represents cash and other assets that are reasonably expected to be realized in cash, sold or consumed within one year or one operating cycle. Generally, it is the sum of cash and equivalents, receivables, inventories, prepaid expenses and other current assets. For non-U.S. corporations, long term receivables are excluded from current assets even though included in net receivables.
Dvidends	Dividends Per Share-Last 12 Months / Market Price-Current * 100
Income Taxes	represent all income taxes levied on the income of a company by federal, state and foreign governments.
Pre-Taxes Income	represents all income/loss before any federal, state or local taxes. Extraordinary items reported net of taxes are excluded.
Inventory	represent tangible items or merchandise net of advances and obsolescence acquired for either (1) resale directly or (2) included in the production of finished goods manufactured for sale in the normal course of operation.
Liabilities	represent all short and long term obligations expected to be satisfied by the company.
Long-Term Debt	represents all interest bearing financial obligations, excluding amounts due within one year. It is shown net of premium or discount.
Net Income	represents the net income of the company converted to U.S. dollars using the fiscal year end exchange rate. See the definition for NET INCOME BEFORE PREFERRED DIVIDENDS for information regarding the net income used in this calculation.
Operational Income	represents the difference between sales and total operating expenses
Past year Return Volatility	sqrt((sum of the past 12 months return)^2/12)
PPE	represents tangible assets with an expected useful life of over one year which are expected to be used to produce goods for sale or for distribution of services.
Sales	represent gross sales and other operating revenue less discounts, returns and allowances.
Stockholders Equity	represents the sum of Preferred Stock and Common Shareholders Equity.
Total Assets	represent the sum of total current assets, long term receivables, investment in unconsolidated subsidiaries, other investments, net property plant and equipment and other assets.
Total Debt	represents all interest bearing and capitalized lease obligations. It is the sum of long and short term debt.