COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

A FIRM-FUNDAMENTALS BASED CORPORATE BOND INVESTMENT STRATEGY

MASTER THESIS

2016

Bc. Michaela Floriánová

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

A firm-fundamentals based corporate bond investment strategy

MASTER THESIS

Study programme:	Economic and Financial Mathematics
Field of Study:	1114 Applied Mathematics
Department:	FMFI KAMŠ -Department of Applied Mathematics and Statistics
Supervisor:	Mgr. Juraj Katriak

Bratislava 2016

Michaela Floriánová

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Investičná stratégia pre firemné dlhopisy založená na účtovných dátach firiem

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	1114 Aplikovaná matematika
Katedra:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	Mgr. Juraj Katriak

Bratislava 2016

Michaela Floriánová

43960171

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Bc. Michaela Floriánová
Študijný program:	ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor:	aplikovaná matematika
Typ záverečnej práce:	diplomová
Jazyk záverečnej práce:	anglický
Sekundárny jazyk:	slovenský

Názov: A firm-fundamentals based corporate bond investment strategy.
Investičná stratégia pre firemné dlhopisy založená na účtovných dátach firiem.

Cieľ: A study of Goyal et al. (Goyal et al.: Is the Cross-Section of Expected Bond Returns Influenced by Equity Return Predictors?) suggests that common equity return predictors also have forecasting power for US corporate bonds. The aim of the thesis is to investigate if similar results can be obtained for an investable portfolio of European or US corporate bonds on a more recent dataset and subject to data availability. The thesis should consist of three parts. In the first part a dataset of corporate bond returns and corresponding firm data should be collected from available sources (mainly DataStream and Bloomberg). In the second part it should be investigated which of the available firm data have predictive power for corporate bond returns. In the third part some investment strategies should be analyzed. The thesis should be written in English. The preferred programming language is R.

Vedúci:	Mgr. Juraj Katriak
Katedra:	FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry:	prof. RNDr. Daniel Ševčovič, CSc.
Dátum zadania:	10.02.2015

Dátum schválenia: 11.02.2015

prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

....
študent

....
vedúci práce

Acknowledgements I would like to express my deep sense of gratitude to my supervisor Mgr. Juraj Katriak for his systematic support and expert guidance during the elaboration of this thesis. Thanks goes to the firm Spängler IQAM Invest for provision of access to the necessary data. In this way I would also like to thank to my family and Matúš for all their support during my studies.

Abstract

Floriánová, Michaela: *A firm-fundamentals based corporate bond investment strategy* [Master thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics a Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Juraj Katriak, Bratislava, 2016, 60 p.

The aim of this thesis is threefold. We build a joint database of single issues of corporate bonds and firm-fundamental data for European and US firms. Further, using the database we investigate along the lines of Goyal *et al.* study [6], whether firm-fundamentals have predictive power for corporate bond returns of the respective firms. Based on the results of the regression study we devise simple investment strategies and assess their economic significance.

Keywords: corporate bond, firm-fundamentals, panel data, investment strategy

Abstrakt

Floriánová, Michaela: *Ivestičná stratégia pre firemné dlhopisy založená na účtovných dátach firiem* [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Juraj Katriak, Bratislava, 2016, 60 s.

Cieľ tejto diplomovej práce je trojaký. Zostavíme spoločnú databázu pozostávajúcu z jednotlivých emisií korporátnych dlhopisov a účtovných dát Európskych a Amerických firiem. S použitím tejto databázy a v súlade so zisteniami štúdie od Goyala *a spol* [6] zistíme, či vybrané účtovné dáta firiem signifikantne predikujú výnosy korporátnych dlhopisov. Na základe výsledkov z regresie navrhneme jednoduché investičné stratégie a zhodnotíme ich ekonomickú významnosť.

Kľúčové slová: firemný dlhopis, účtovný ukazovateľ firmy, panelové dáta, investičná stratégia

Contents

List of Figures

List of Tables

Introduction

The low government interest rates and high volatility of equity investments in the recent years makes the portfolio managers search for investment strategies with reasonable risk/return characteristics ever harder. In this thesis we focus on the construction of an investment strategy for corporate bonds.

Our starting point is the paper by Goyal *et al.* [6], where the authors investigate the predicting power of typical equity return predictors on corporate bond returns in an US dataset. The aim of this thesis is threefold. First we build a joint database of single issues of corporate bonds and firm fundamental data. The database comprises EUR denominated as well as USD denominated corporate bond issues and the corresponding firm data in the time span between 1999 and 2014.

Second we use this database in a regression framework to test whether selected firm fundamentals have predictive power on the return of corporate bonds issued by the respective firms. While we found some significant factors, there were no such strong results as in the Goyal *et al.* study which would be consistent for both the European and US region on the whole time span of the dataset. Nonetheless in the post-crisis period we found some significant factors similar to those of Goyal *et al.*

And third we use the findings from the regression study to devise a simple investment strategy. The strategy is backtested in an out-of-sample backtest and compared to benchmarks.

Accordingly, this thesis is divided into three parts. The first part describes the data collection process and the corporate bond/firm fundamentals database. In the second part we perform our regression analysis and in the last part we analyze the investment strategies.

1 Data

In our analysis we were limited by data availability. While we believe that our historical universe of single issues of corporate bonds provides a good gauge of the liquid part of the market, we think that our analysis would benefit from a larger dataset of firm-fundamental data.

The fundamental firm data was available for firms in the compositions of two equity indices – S&P 500 for US and STOXX 600 for Europe. The US data was available from August 1989 on and the EU data from August 1999 on. Both dataset end in December 2014. The firm data stems mainly from the Worldscope database and was obtained through Datastream. Also from a practical point of view, the narrowing down to large S&P 500 and STOXX 600 firms brings our analysis nearer to the possibilities of investors who prefer liquid issues from well-know firms.

The return data for corporate bonds was obtained from the compositions of two Bank of America Merrill Lynch Indices – The BofA Merrill Lynch US Corporate Index (C0A0) for USD denominated issues and The BofA Merrill Lynch Euro Corporate Index (ER00) for EUR denominated issues. The data was obtained through Bloomberg. The compositions were available from December 1972 for C0A0 and December 1995 for ER00 on until December 2014.

Unfortunately, the two databases do not have a common firm identifier and we had to rely mostly on the firm descriptions (names) to merge the databases. Internet search had to be used for firms that had merged or were acquired by other firms, or changed their name etc. Since the merging procedure was difficult and since we want to concentrate our attention to recent data, we decided to use only data after December 1999.

We also paid attention not to introduce any forward-looking biases into our data. Both databases as well as the merging process take into account the historical availability of the data.

In this chapter, we describe both datasets, provide various statistical information and describe the merging algorithm.

1.1 Corporate Bond Data

As mentioned above, we collected the corporate bond data for the USD and EUR market from two large corporate bond indices, the ER00 Index and C0A0 Index. The dataset consists of montly collections of descriptive and return data for single bond issues. The ER00 contains EUR denominated issues from companies not necessarily based in the European Economic and Monetary Union (EMU) and the C0A0 index contains USD denominated corporate bond issues.

1.1.1 ER00 and C0A0

The ER00 index tracks the performance of Euro denominated investment grade corporate debt publicly issued in the Eurobond or Euro member domestic markets, while the C0A0 index tracks the performance of US dollar denominated investment grade corporate debt publicly issued in the US domestic market. Furthermore, there are the following requirements for the inclusion of a security in the indices:

1. an investment grade rating,
2. at least 18 months to final maturity at the time of issuance,
3. at least one year remaining term to final maturity,
4. fixed coupon schedule,
5. minimum amount outstanding of €100 million / $150 million before 2005 and €/$250 million after 2005 for the ER00 and C0A0 index respectively.

Index constituents are capitalization-weighted. The weight is based on their current amount outstanding times the market price plus accrued interest. Accrued interest is in both indices calculated assuming next-day settlement. Cash flows from bond payments that are received during the month are retained in the index until the end of the month and then are removed as part of the rebalancing. Cash does not earn any reinvestment income while it is held in the index. Both indices are rebalanced on the last calendar day of the month, based on information available up to and including the third business day before the last business day of the month [2].

1.1.2 Bond returns

The monthly total return of a (corporate) bond i is defined as:

$$R_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1, \tag{1}$$

where $P_{i,t}$ is the price of corporate bond i at time t, $AI_{i,t}$ represents the accrued interest at time t and $C_{i,t}$ is the paid coupon during month t.

The total return of a corporate bond can be decomposed into two components: the risk-free return of a matching government bond and the firm-specific return. Such decomposition makes also sense from a practical point-of-view because the risk-free returns can be efficiently managed (hedged) by liquid interest futures.

Hence we would like to concentrate our attention to the explanation of the firm specific returns. Therefore, we will use excess returns, which are defined as the monthly total return of a corporate bond minus (in excess of) the monthly risk-free return. The excess returns are provided in the data, and as described in [2] are calculated as the monthly total return of a bond minus the monthly total return of a duration-matched basket of government bonds:

$$R^{e}_{i,t} = R_{i,t} - R^{rf}_{i,t}, \tag{2}$$

where $R^{rf}_{i,t}$ is the monthly return of such a basket. The corporate bond excess returns are immune to parallel shifts in the government yield curve.

1.1.3 Firm returns

In this thesis we want to investigate the predictive power of firm fundamental data on the (excess) returns of corporate bonds issued by the firm. A firm can issue many bonds which differ in maturity and other features. Thus in practice when investing into the debt of a firm one has to decide which particular bonds to buy. These decisions are not trivial and are based on many criteria, most prominent of which is the liquidity of the issues. For an overview of the number of issues by firms in our indices see Table 1 and Table 2.

number of issues	1	2	3-10	11-50	51-100	101-
percentage of firms	33.8%	18.7%	32.8%	13.5%	1.0%	0.2%

Table 1: Distribution of number of bond issues by firms in the ER00 index

number of issues	1	2	3-10	11-50	51-100	101-
percentage of firms	25.5%	15.2%	37.7%	19.3%	1.7%	0.6%

Table 2: Distribution of number of bond issues by firms in the C0A0 index

However for the purpose of our analysis we do not make any such decisions and use the weighed excess return of all firms bonds which are included in our indices in a particular month:

$$R_{i,t}^{fe} = \frac{\sum_{i=1}^{W_{i,t}} w_{i,t} R_{j,t}^{e}}{W_{i,t}}, \tag{3}$$

where $R_{i,t}$ is the price of corporate bond i at time t, $w_{i,t}$ is the index weight of corporate bond i at the beginning of month t and $W_{i,t} = \sum w_{i,t}$ is the total weight of firm's i issues in the index at time t.

1.1.4 Descriptive Statistics

As pointed out above, we use the bond data from both indices in the timespan from December 1999 on until December 2014. In this section we provide some descriptive statistics on the corporate bond database.

- **Basic summary**

Following table includes the basic statistical information for both indices ER00 and C0A0:

Statistics	ER00	C0A0
total number of bond issues	6732	18480
minimum of bond issues per month	941	3031
average number of bond issues per month	1420	4065
maximum of bond issues per month	1946	6455
total number of issuers	1679	3866
minimum of issuers per month	357	1045
average number of issuers per month	533	1260
maximum of issuers per month	605	1610

Table 3: Corporate bond dataset descriptive statistics

- **Total number of bonds and bonds issuers over time**

In the Figures 1 and 2 we plot the size of indices, i.e. number of different bond issues over time:

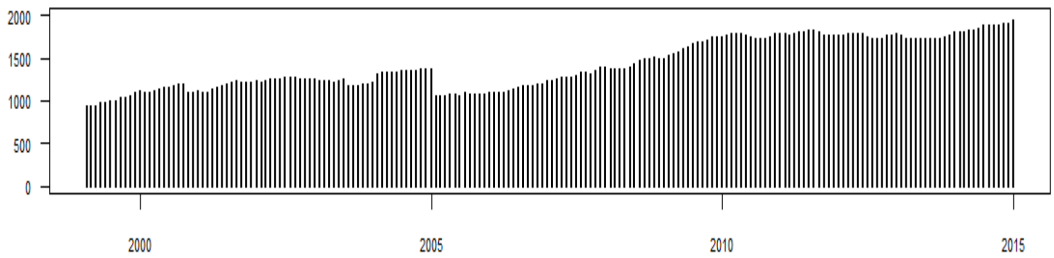


Figure 1: Number of bond issues in the ER00 index over time

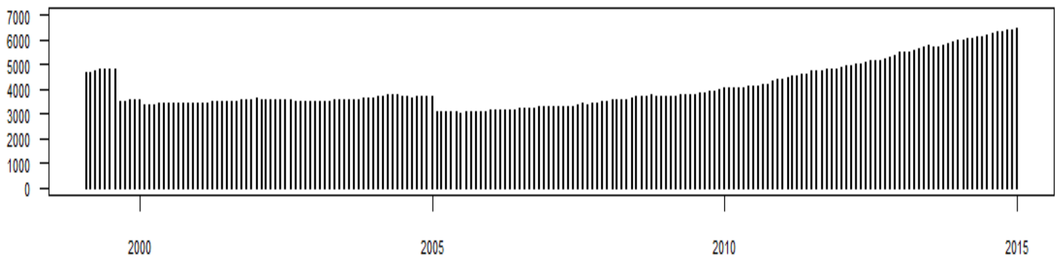


Figure 2: Number of bond issues in the C0A0 index over time

Both indices grow in the total number of bond issues over time. We observe a drop in the number of issues at the beginning of the year 2005 in both indices – the number of issues dropped by 321 and 660 for the ER00 and C0A0 index respectively between December 2004 and January 2005. This was caused by change in the eligibility criteria for both indices [2]. The minimum size requirement increased from €100 million to €250 million for the EUR denominated bond and from \$150 million to \$250 million for the USD-denominated bonds. The percentage decline lower for the C0A0 index, because at the same time special 144a securities were qualified for the inclusion into US index.

The trend in the number of issuers over time is less pronounced than in the case of issues, see Figures 3 and 4.

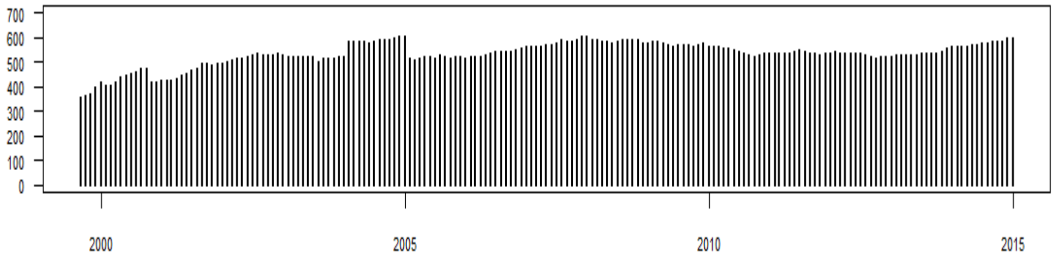


Figure 3: Number of bond issuers in the ER00 index over time

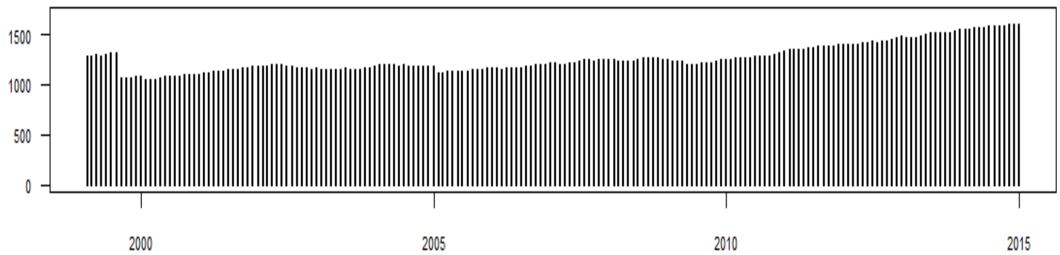


Figure 4: Number of bonds issuers in the C0A0 index over time

We notice that in January 2005 not only the number of bond issues declined, but also the number of bond issuers, what was probably caused by the fact that some issuers didn't have bonds outstanding that would meet the minimum size requirements.

- **Returns**

Here we provide a quick graphical overview of returns of both corporate indices in comparison with equities and government bonds of similar duration. More detailed analysis is provided in chapter 3. In the figures 5 and 6 we can see that the corporate return have interesting risk/return profiles, especially in the most recent period after 2009.

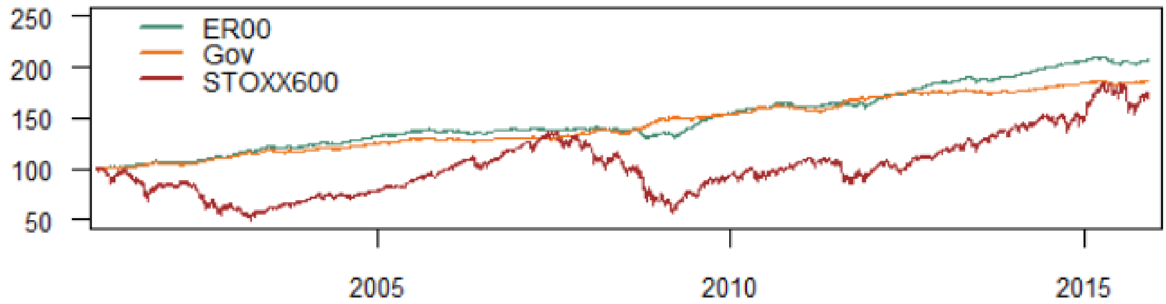


Figure 5: Total returns of the ER00 index in comparison with STOXX 600 and German government index G5D0

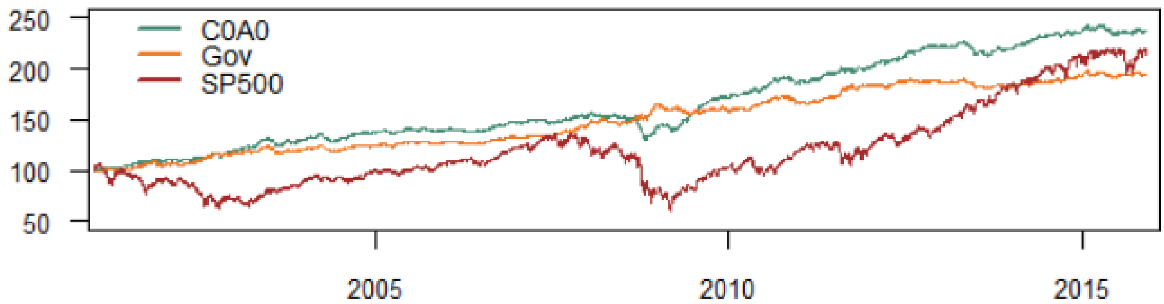


Figure 6: Total returns of the C0A0 index in comparison with S&P 500 and US government index G0Q0

- **Corporate bond ratings**

A bond rating refers to the grade given to bonds which mirrors the credit quality and informs investors of creditworthiness of the corporate bond and about the risk to default. Bond credit rating is determined by private and independent rating companies (e.g. Standard & Poor's, Fitch or Moody's). Rating agencies use similar methodologies, but they differentiate themselves in various combinations of letters. Composite rating for bonds from ER00 and C0A0 indices is the average of the Moody's, Standard & Poor's and Fitch bond ratings, for details refer to [2].

There are several main groups of ratings:

- AAA and AA: High credit-quality investment grade (IG)
- A and BBB: Medium credit-quality IG
- BB, B, CCC, CC, C: Low credit-quality, non-IG known as junk bonds
- D: Bonds in default for non-payment
- NR: No rating due to insufficient information on which to base a rating

In tables 4 and 5 below we calculated the distribution of rating categories in both indices.

	AAA	AA	A	BBB	NR
percentage	10 %	23 %	41 %	25 %	1 %

Table 4: Distribution of ratings in the ER00 index

	AAA	AA	A	BBB
percentage	2 %	11 %	41 %	46 %

Table 5: Distribution of ratings in the C0A0 index

As we mentioned above both ER00 and C0A0 are investment grade indices. Historically there were some non-rated issues in the ER00 index at the beginning of our timespan. The European index contains almost three quarters of A and higher graded bonds, while for the US index the ratio is only about one half.

- **Duration**

Duration of a bond measures the sensitivity of bond price to a change in interest rates. There are several types of duration including Effective duration, Modified Duration and Macaulay duration.

The duration measure we use is the Effective duration which measures the % change in the price of a bond given a parallel shift in the government yield curve while keeping the corporate spread constant. A theoretical price is calculated by discounting the bond's cash flows using the shifted yield curve.

In Table 6 we provide short summary of Effective duration for both indices:

	Min	1st Q	Median	Mean	3rd Q	Max
ER00	3.763	4.206	4.502	4.479	4.725	5.227
C0A0	5.317	5.720	5.938	6.052	6.410	7.161

Table 6: Summary statistics of Effective duration for ER00 and C0A0 indices

The average duration of the ER00 and C0A0 indices is similar to the durations of common government bond indices, German government index G5D0 and US government index G0Q0:

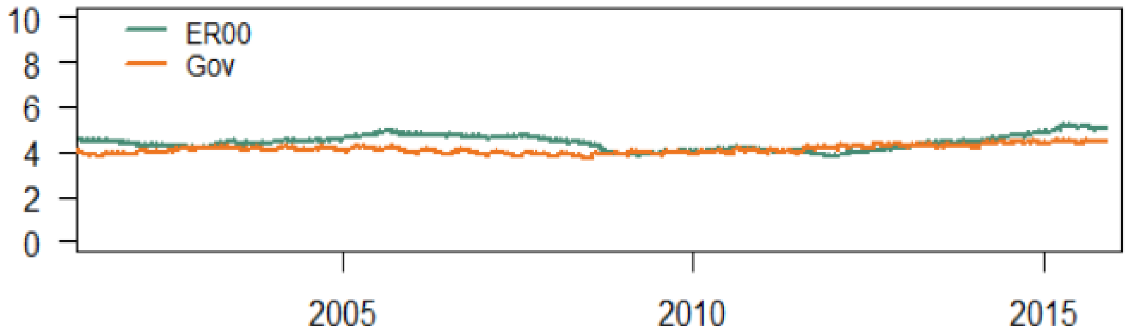


Figure 7: Effective duration of the ER00 index vs. German government index G5D0

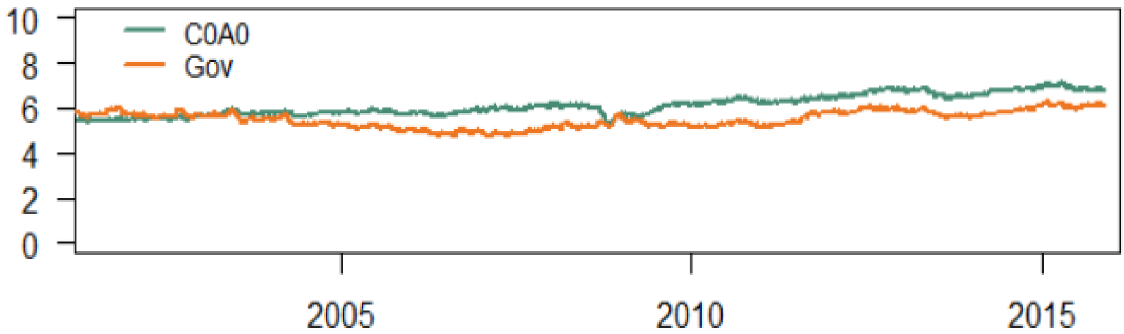


Figure 8: Effective duration of the C0A0 index vs. US government index G0Q0

1.2 Firm data

As we mentioned above, in our analysis we were limited by data availability. In the case of fundamental firm data it was the intersection of the compositions of the STOXX 600 and S&P 500 Indices with the firms available in the Worldscope database.

1.2.1 STOXX 600

The STOXX Europe 600 Index belongs to the index provider STOXX Limited. The company was founded in 1997 and today calculates more than 7000 indices. The STOXX Index family represents a wide range of stocks covering different market segments and different investment strategies.
The STOXX Europe 600 Index is derived from the STOXX Europe Total Market Index and is a subset of the STOXX Global 1800 Index (see [8]). With a fixed number of 600 components, the STOXX Europe 600 Index represents large, mid and small capitalization companies across 18 countries of the European region: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom. The inception date of the STOXX Europe 600 Index is 31.12.1998.

1.2.2 S&P 500

The S&P 500 is an US stock market index based on the market capitalization of 500 largest companies which have common stock listed on the NYSE or NASDAQ exchanges. The S&P 500 index components and their weightings are determined by S&P Dow Jones Indices. S&P differs from other US stock market indices, because of its diverse constituency and weighting methodology as this index has been traditionally capitalization-weighted; that is, movements in the prices of stocks with higher market capitalization (the share price times the number of shares outstanding) have a greater impact on the value of the index than do companies with smaller market capitalization. This index is considered to be one of the best representations of the US stock market. The inception date of the S&P 500 Index is 31.12.1963.

1.2.3 Worldscope Global Database

The Worldscope Global Database [10] offers detailed fundamental data on the world's leading public companies. Origin of this database roots in the international investment management activities of global management company Wright Investors' Service based in the United States. After a joint venture and establishment of several research and data collection centers around the world, the corporation was acquired by the Thomson Corporation in 2000. Today the database provides access to financial information of about 10 000 companies in more than 40 countries. Fundamental data are taken from three well-known financial statements: Balance Sheet, Income Statement and Cash Flow Statement. The base year for the Worldscope Database is 1980, although statistically significant company and data item representation is best represented from January 1985 forward. The Worldscope database is much broader than the compositions of the S&P 500 and STOXX 600 indices, and contains fundamental data for all firms in the historical compositions of the indices at least from December 1999 on.

1.2.4 Statistics - firm data

Compositions of our two firm indices STOXX 600 and S&P 500 are of monthly frequency. Despite the fact that firm-fundamentals collected from Worldscope are available, we are restricted to the presence of the firm in stock market indices. In the next table we include some information for datasets of firms from indices STOXX 600 and S&P 500 calculated for the timespan December 1999 - December 2014:

Statistics	STOXX 600	S&P 500
total number of firms in dataset	1323	964
number of firms in index during the whole timespan	225	240
firm's minimum duration in index (in months)	1	1
firm's average duration in index (in months)	82.1	93.9
firm's maximal duration in index (in months)	181	181

Table 7: Firm dataset descriptive statistics

1.3 Common corporate bond/firm-fundamentals database

In order to be able to conduct our analysis of the influence of the firm-fundamental data on the corporate bond returns we need a common database – the intersection of the firms in the composition of the corporate bond indices and the equity indices.

As we already mentioned, the two databases do not have a common firm identifier.

Our first try was to merge the databases by the firm-part of the International Securities Identification Number (ISIN). This unique number is available for all equities in the firm data and all single bond issues in the corporate data.

In recent US data (equity and most of the bonds) the ISIN numbers have a common firm-specific part. With this method we were able to match only some of the US firms, because for firm in older compositions this feature partly disappears. For the European firms this method doesn't work, because the corporate bond and equities ISINs do not have a common part. Therefore we devised an algorithm for merging of the two databases. At the end of each month, for all bonds and equities in the composition of the indices at that time, we merged the databases using the following steps:

1. Matching based on ISIN – firms and bonds in the US data where the first 8 characters of the ISIN matched were merged. This first step works only for US data and for companies where the ISIN didn't change (for example after a merger), since in Datastream ISINs in the historical compositions are overwritten with new ones, once there is a change.
2. Matching based on Description (Firm Name) – we collected all name description fields from Datastream and Bloomberg for the firms in the historical compositions of the S&P 500 and STOXX 600. We deleted common abbreviations, spaces, commas and so on from these fields, put them to lowercase and matched with the descriptions in the corporate bond data. In this way were able to match much more of the data, including European companies.
3. Matching based on Ticker – there is a "Ticker" field in the corporate dataset, which suffers from the some of the same problems as the ISIN. Some of tickers repeat in the history for unrelated firms, and from Bloomberg we were able to obtain only the ticker symbols as they are now, not the pre-merger histories.
4. Manual conflict resolving – For companies where there were conflicting matches from the three steps above we carried out an internet search and looked for possible mergers, acquisitions or bankruptcies. The list of manually matched companies is in the Appendix A.

After completing the merging procedure we have obtained two separate datasets.

The first one is the European dataset consisting of the merged STOXX 600 and ER00 compositions. And the second one is the US dataset with merge S&P 500 and C0A0 compositions. The European dataset contains EUR denominated returns of corporate bonds, and the returns in the US dataset are USD denominated.

Table 8 provides an overview of the complete database with our two datasets:

Statistics	ER00 dataset	C0A0 dataset
number of single issues	86333	421268
number of firms in the dataset	337	636
number of firm-month	19823	53072
average number of firms per month	256.2	662.4
average number of months for a firm	58.8	83.4

Table 8: Common corporate bond/firm-fundamentals database summary

2 Corporate Bonds Return Prediction

In this chapter we investigate the predictive power of some common equity return predictors on corporate bond returns in both datasets which were described in the previous chapter.

Goyal *et al.* [6] studied the same problem on a much larger (in both history length and number of firms at a given point in time) US dataset. In their study they used ten equity return predictors and three corporate bond related variables. The authors found that some equity return predictors have significant predictive power for corporate bond returns. More specifically, in their investment grade subset they found that out of the equity predictors the size and equity-momentum factors have predictive power, and out of the corporate bond related variables the bond momentum and distance to default.

Our study differs in the in the length and breadth of the data[1], and also in the firm and bond specific factors we used.

In the Worldscope database many historical variables for individual companies are available. Based on some criteria that we describe below, we chose eleven firm factors to include in our regression analysis. These factors are same or similar to those used in Goyal *et al.* As for the bond specific factors we used probability of default and bond momentum as in Goyal *et al.*, a liquidity measure wasn't available to us. Here we would like to point out that, firstly given our dataset we focus on the most liquid firms and secondly the liquidity factor was not found to be significant in the study by Goyal*et al.*

2.1 Firm fundamentals

Fundamental analysis [2] is a method used to determine the value of a stock by analyzing the financial data that is *fundamental* to the company. In other words, fundamental

[1]Goyal *et al.* have a dataset that spans from 1973 to 2011, our dataset is from December 1999 to 2014; we include only the most liquid investment grade firms from two of the most prominent stock/corporate bond indices, whereas Goyal *et al.* have very broad dataset which also include sub-investment grade bond. Also our analysis includes the European dataset whereas study of Goyal *et al.* focuses only on US data.

[2]source: Investopedia: Fundamental Analysis

analysis takes into consideration only those variables that are directly related to the company itself, such as its returns or dividends. Fundamental analysis focuses on the company's business in order to determine whether or not the stock should be bought or sold and it does not look at the overall state of the market. This analysis does not include behavioral variables in its methodology. Firm-fundamentals represent the various accounting and other variables related to the company, on which fundamental analysis is concentrated.

From the firm fundamental data, which is available in the Worldscope Database,[3] we selected eleven factors based on the following criteria:

- meaningful economic interpretation of the factor,
- correlation with other factors,
- data availability (the percentage of non-missing values).

2.1.1 Description of chosen factors

As described above we use the following equity return predictors. The factors were either downloaded directly from Datastream, or calculated based on the Datastream data[4].

1.) *Beta (Beta)*: is the coefficient from regression of the absolute simple returns of the equity on the benchmark absolute simple returns, where the benchmark is the STOXX 600 index for European and S&P 500 for US data. The regression is done with weekly data on a one year horizon. Calculated factor.

2.) *Book to Price (BTP)*: is calculated as the book value per share divided by the share price. Datastream field.

3.) *Dividend Yield (DY)*: expresses the dividend per share as a % of the share price. The underlying dividend calculation is based on an anticipated annual dividend and excludes special or once-off dividends. Datastream field.

4.) *Earnings Per Share Growth (EPS.growth)*: is calculated as (Current EPS/(Last Year's EPS $-1) * 100$. Earnings Per Share (EPS) represents

[3] There are about 150 historical fundamental variables available for each firm

[4] We indicate in the list of descriptions, whether the factor is a "calculated factor" or a "datastream field".

the earnings for the 12 months ended the last calendar quarter of the year for US corporations and the fiscal year for non-US corporations. Calculated factor.

5.) *Earnings Quality (EQ)*: is the beta coefficient from the regression of Net Income Before Extra Items and Preferred Dividend divided by Total Assets on its one week lagged value. Calculated factor.

6.) *Equity Momentum (Eq.mom)*: is the cumulative 1 year equity return. Calculated factor.

7.) *Gearing (Gearing)*: measures the extent to which the operations of the firm are funded by firm's lenders versus shareholders. Gearing is calculated as: (Long Term Debt+Short Term Debt&Current Portion of Long Term Debt) / Common Equity * 100. Datastream field.

8.) *Long Term Leverage (LTLev)*: is the t-value for the hypothesis that Long Term Leverage > 5. Leverage is the ratio of Total Liabilities to Common Equity and expresses the amount of debt used to finance the firm's assets. Calculated factor.

9.) *Long Term Return on Assets (LTROA)*: is a quality factor representing long term stable earnings. It is calculated as 10 years moving average of ROA, where ROA is calculated as: (Net Income – Bottom Line + ((Interest Expense on Debt-Interest Capitalized) * (1-Tax Rate))) / Average of Last Year's and Current Year's Total Assets * 100. Datastream field.

10.) *Market value (MV)*: is equal to the share price multiplied by the number of ordinary shares in issue and is displayed in millions of units of local currency. Datastream field.

11.) *Post Earnings Announcement Drift (PEAD)*: is the change in 1 quarter change in EPS (Earnings Per Share) divided by standard deviation of EPS from last 8 quarters. Calculated factor.

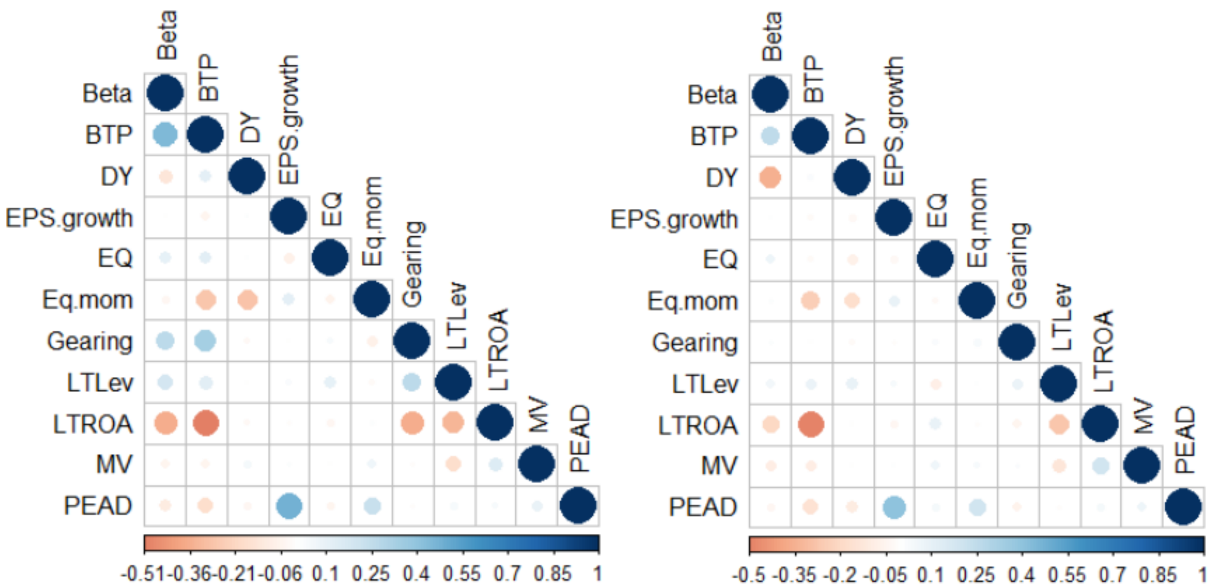
Following bond specific factors were used:

12.) *Probability to Default (Prob.Def)*: default probability from the Merton model, as calculated by Bloomberg.

13.) *Bond momentum (Bond.mom)*: last month's corporate return; lagged left hand side variable.

All of the accounting variables are shifted in order to account for the fact that the information in is published with a time delay. This time shift is important as we don't want to introduce a forward-looking bias into our regression.

In Figure 9 we plot a graphical representation of the correlation matrix of the chosen factors for European and US datasets. The bigger and darker the circle, the higher the correlation of the factors. Blue color indicates positive correlation while red color represents negative correlation. There are no extremely correlated factors which is a consequence of our factor selection method.



(a) European dataset

(b) US dataset

Figure 9: Correlation matrix of selected equity factors for European and US dataset

2.2 Panel data regression

We analyze our data in a panel regression. Since not all firms and factors have data at all points in time our panel is unbalanced. In financial literature often the Fama-Macbeth (see [5],[7]) two step regression is used to analyze such datasets. We decided against this method because firms often go in and out of our dataset and we would lose a large portion of our data. Instead we use a standard panel regression as described e.g. in [1] and implemented in the R-package *plm* [4].

2.2.1 Linear panel models

The general linear panel model used in econometrics is given by:

$$y_{it} = \alpha_{it} + \beta_{it}^T x_{it} + \epsilon_{it}, \tag{4}$$

where $i = 1, ..., n$ is the individual index, $t = 1, ..., T$ is the time index and ϵ_{it} is the unobservable random disturbance term. To decrease the complexity of the model, some assumptions have to be made.

After adding an assumption of parameter homogeneity, which says that $\alpha_{it} = \alpha$ for $\forall i \forall t$ and $\beta_{it} = \beta$ for $\forall i \forall t$, the model becomes a standard linear model pooling all the data across individual i and time t. The **pooling model** is defined as follows:

$$y_{it} = \alpha + \beta^T x_{it} + \epsilon_{it}, \tag{5}$$

with $E(\epsilon_{it}) = 0$ and $Var(\epsilon_{it}) = \sigma^2$. This model is usually unrealistic, because of the assumption that ϵ_{it} are independent across i and t.

The more realistic choice is to model the individual heterogeneity, where we can assume that the error term consists of two components, $\epsilon_{it} = \mu_i + \nu_{it}$. One part does not change over time only among units (individual error μ_i), and the other that varies both among time and groups (idiosyncratic error ν_{it}). The **unobserved effects model** is defined as:

$$y_{it} = \alpha + \beta^T x_{it} + \mu_i + \nu_{it}, \tag{6}$$

the idiosyncratic error ν_{it} is assumed to be independent of individual error component μ_i and regressors x_{it}, while the individual component μ_i can be either correlated with regressors or independent of regressors x_{it}.

In case that x_{it} and μ_i are correlated, the method of ordinary least squares (OLS) would lead to inconsistent estimators and therefore individual error μ_i is treated as another set of n parameters to be estimated and this is called the **fixed effects model (FE)**.

In the **Random effects model (RE)** the individual components μ_i (and therefore also the overall ϵ_{it}) are not correlated with the regressors x_{it} and so the OLS method is consistent. This model is estimated using method of generalized least squares estimators (GLS), because the OLS method is not efficient.

2.2.2 Estimation methods

The OLS estimation framework is used in estimation methods for all basic models introduced above. The pooling model is the simplest model, as it assumes parameter homogeneity, this model simply pools all the data and then estimates the parameters using the OLS method. The Fixed effects model is estimated by OLS on transformed data. The transformation is called time-demeaning and is done by subtracting the mean from each variable. The Random effects model is the most complex one, because GLS estimation methods have to cope with the serial correlation caused by invariance among groups. Again equivalence to the OLS estimation can be achieved by transforming the data. In this case the OLS has to be run on quasi-demeaned data. Quasi-demeaning is defined as:

$$y_{it} - \lambda \bar{y_{it}} = (X_{it} - \lambda \bar{X}_i)\beta + (\epsilon_{it} - \lambda \bar{\mu_i}), \tag{7}$$

where $\bar{y}$ and $\bar{X}$ are the time means of y and X and

$$\lambda = 1 - (\frac{\sigma_\mu^2}{\sigma_\mu^2 + T\sigma_\nu^2})^{1/2}. \tag{8}$$

As $T \to \infty$, $\lambda \to 1$ and the RE model becomes equivalent to the FE model. Again, we can see that the estimate for β using feasible RE estimator can be obtain also in a way of estimating $\bar{\lambda}$ and than using the basic OLS regression on transformed data. Moreover, for $\lambda = 0$ we would get the pooling model and for $\lambda = 1$ the estimator in FE model.

2.2.3 Tests of hypothesis in linear panel models

We introduce tests and methods which helps to detect the correct linear panel model specification for a given dataset.

- **Test of poolability**

Using the *pooltest* we can detect whether we accept or reject the null hypothesis saying that the same coefficients apply to each individual at all times. This test, whose t-statistics has an F-distribution (therefore it belongs to group of standard F-tests) compares model based on estimation for each individual-time combination and the model obtained for the full sample.

- **Hausman test**

Statistically, FE are always a reasonable way of modeling the panel data, as this model always give consistent results, however it may not be the most efficient model. RE model gives better *p-values* as this model provides more efficient estimator. Therefore we need to justify that run RE model is correct. *Hausman test* is a statistical hypothesis test and in panel data regression it is used for choice between the RE and the FE model. This tests evaluates the consistency of the RE model estimator when compared to the less efficient, but consistent alternative FE model estimator. We test the hypothesis:

H_0: the coefficients estimated by the efficient RE estimator are the same as the ones estimated by the consistent FE estimator.

If we get a significant *p-value* (less than 5% as we work with the 95% confidence interval), then we use the consistent FE fixed effect model, while a $p - value > 5\%$ can also lead to use of more efficient and also consistent RE model.

- **Robust covariance matrix estimation**

The linear panel model given by equation 5 assumes that the regression disturbances are homoscedastic with the same variance across individuals and time. In an unbalanced model, where the cross-sectional units vary in size the assumption of homoscedasticity is quite restrictive. Assuming homoscedastic disturbances when heteroscedasticity is present will still result in consistent estimates of the regression coefficients, but these estimates will not be efficient.

There are several possibilities for testing heteroskedacticity, discussed e.g in [3], the Breusch-Pagan test can be implemented in FE model as long as the sample size is large.

The main problem is that in case of present heteroscedasticity the standard errors of these estimates are biased and we need to compute robust standard errors corrections.

All versions of the robust covariance matrix estimation are assuming that there is no correlation between different units and thus heteroscedasticity can be present only across the firms. The common types of estimators in RE model case are the White estimators, which don't allow for serial correlation, only for general heteroskedaticity. On the other hand, demeaning process used by FE model induces serial correlations in errors and therefore, for this model would be the White estimator inconsistent and is replaced by version *"arrelano"* [9].

2.3 Our model specification

In our model specification we regress the firm corporate bonds excess return on the lagged equity factors and bond specific factors as described above.

For both datasets, the *Pooling tests* strongly rejects the hypothesis that the same coefficients apply across all firms and we only need to decide between FE or RE model. Although *Hausman test* has not directly rejected the hypotheses that RE estimate results could be consistent, we choose to work with the consistent and recommended FE model. Due to the present heteroscedasticity in the error terms, we use the robust covariance matrix estimator.

Our regression specification is:

$$R_{it}^{fe} = \alpha_i + \beta Xfund_{it-1} + \gamma Prob.Def_{it-1} + \delta Bond.mom_{it-1} + \epsilon_{it}, \tag{9}$$

where R_{it}^{fe} represent the excess bond returns for the firm i at the time t and $Xfund_{it-1}$ are the lagged firm-fundamentals. At last, ϵ_{it} is the random disturbance error.

Prior to running the regression, we normalize the factors by standard score, so that we can compare the magnitudes of the estimated coefficients.

2.4 Estimation results

Tables 10 and 11 show the estimation results for the European and US datasets.

	Estimate	Std. Error	t value	Pr(>|t|)	
Beta	0.0243698	0.0312143	0.7807	0.434978	
BTP	0.0281921	0.0303764	0.9281	0.353377	
DY	-0.0120333	0.0297326	-0.4047	0.685692	
EPS.growth	-0.0208596	0.0140531	-1.4843	0.137741	
EQ	-0.0153095	0.0137773	-1.1112	0.266498	
Eq.mom	0.0378024	0.0212222	1.7813	0.074892	.
Gearing	-0.0176924	0.0200357	-0.8830	0.377229	
LTLev	0.0352124	0.0306142	1.1502	0.250083	
LTROA	-0.0023032	0.0234609	-0.0982	0.921797	
MV	-0.0086491	0.0165872	-0.5214	0.602075	
PEAD	-0.0037740	0.0156073	-0.2418	0.808929	
Bond.mom	0.2157163	0.0799143	2.6993	0.006956	**
Prob.Def	-0.0246557	0.0343198	-0.7184	0.472518	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 10: Factor coefficients from regression on the European dataset

	Estimate	Std. Error	t value	Pr(>|t|)	
Beta	0.0120735	0.0212398	0.5684	0.56974	
BTP	0.0189679	0.0385087	0.4926	0.62233	
DY	-0.0266758	0.0225855	-1.1811	0.23757	
EPS.growth	-0.0087285	0.0165442	-0.5276	0.59779	
EQ	0.0119035	0.0112815	1.0551	0.29137	
Eq.mom	0.0353484	0.0179042	1.9743	0.04836	*
Gearing	-0.0027388	0.0148158	-0.1849	0.85334	
LTLev	0.0195931	0.0119531	1.6392	0.10119	
LTROA	-0.0079034	0.0175147	-0.4512	0.65182	
MV	-0.0086630	0.0173774	-0.4985	0.61812	
PEAD	0.0059109	0.0106263	0.5562	0.57804	
Bond.mom	-0.0380568	0.0395927	-0.9612	0.33645	
Prob.Def	0.0049983	0.0291215	0.1716	0.86373	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 11: Factor coefficients from regression on the US dataset

We observe that the only significant equity factor in both datasets is the *Equity Momentum* at the 10% significance level. The significance of the *Eq.mom* factor is consistent with the findings of Goyal *et al.* The coefficient of determination R^2 is very low for both regressions (at 0.014 for the European and dataset 0.0016 for the US dataset).

2.4.1 Impact of economical crisis

Since we observe such weak significance results and low R^2 values we split our dataset to pre-crisis, crisis and post-crisis horizons and we assume that the financial crisis might have a considerable impact on our results.

The three periods are defined as follows:

1.) the pre-crisis period: from December 1999 to July 2008,
2.) the crisis period: from August 2008 to March 2009,
3.) the post-crisis period: from April 2009 to December 2014.

2.4.2 Results for European data

Now we look at the results of regression for European dataset on our three periods. Following table shows the coefficients of determination R^2 from the regressions in the respective time periods.

European data	FE model R^2
Before Crisis	0.010978
During Crisis	0.098999
After Crisis	0.032976
Whole Dataset	0.01414

Table 9: Statistics from regression of European corporate bonds excess returns

Although the coefficients of determination are quite small in all three periods, we can see an improvement in comparison to the regression on the whole sample. For example in the crisis period, our model explains up to 9% of the variation in the data.

The Brausch-Pagan test (suggested in [3]) rejects the hypothesis of homoscedastic errors in all three periods. Therefore we again use a robust covariance matrix estimator of the "arrelano" type which is recommended for the FE model [9].

The Figures 12 - 14 present the results of the estimation for three periods. The significant factor are marked by ./*/** or ***, depending on the level of significance.

In the pre-crisis period the only significant factor is the *Equity Momentum*:

```
              Estimate   Std. Error t value Pr(>|t|)
Beta        0.00535324  0.01809382  0.2959  0.76735
BTP        -0.02384780  0.01951050 -1.2223  0.22164
DY          0.05199446  0.04401297  1.1813  0.23751
EPS.growth -0.00693328  0.00806204 -0.8600  0.38983
EQ         -0.01199081  0.01250230 -0.9591  0.33755
Eq.mom      0.03667690  0.01522258  2.4094  0.01601 *
Gearing    -0.01002674  0.01222132 -0.8204  0.41200
LTLev      -0.00817399  0.01042560 -0.7840  0.43305
LTROA      -0.01866954  0.01365890 -1.3668  0.17172
MV          0.01451792  0.01415304  1.0258  0.30503
PEAD       -0.00052966  0.01161378 -0.0456  0.96363
Bond.mom    0.10686225  0.10124072  1.0555  0.29123
Prob.Def   -0.03928829  0.04688323 -0.8380  0.40206
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 12: Estimation results from the pre-crisis period for the European dataset

During crisis factors *LTLev* and *LTROA* are significant at the 10% significance level:

```
            Estimate Std. Error t value Pr(>|t|)
Beta       -0.454205   0.296089 -1.5340  0.12551
BTP         0.015332   0.357927  0.0428  0.96585
DY         -0.020980   0.239825 -0.0875  0.93032
EPS.growth -0.204845   0.222146 -0.9221  0.35681
EQ          0.110350   0.100832  1.0944  0.27418
Eq.mom      0.157162   0.189855  0.8278  0.40808
Gearing     0.090601   0.349192  0.2595  0.79536
LTLev      -0.340945   0.199305 -1.7107  0.08762 .
LTROA       0.488422   0.270533  1.8054  0.07147 .
MV          0.240315   0.151551  1.5857  0.11329
PEAD        0.120172   0.167491  0.7175  0.47333
Bond.mom    1.024381   0.466239  2.1971  0.02836 *
Prob.Def   -0.251225   0.293703 -0.8554  0.39266
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 13: Estimation results from the crisis period for the European dataset

Results from post-crisis period show more significant factors:

	Estimate	Std. Error	t value	Pr(>|t|)	
Beta	0.121957	0.051586	2.3642	0.018100	*
BTP	0.075979	0.042316	1.7955	0.072619	.
DY	-0.015633	0.029815	-0.5243	0.600076	
EPS.growth	-0.032586	0.022969	-1.4187	0.156041	
EQ	-0.063647	0.025293	-2.5164	0.011879	*
Eq.mom	-0.016187	0.039858	-0.4061	0.684663	
Gearing	-0.072864	0.023550	-3.0940	0.001983	**
LTLev	0.129471	0.048111	2.6911	0.007139	**
LTROA	-0.046985	0.034428	-1.3647	0.172382	
MV	-0.050445	0.016160	-3.1216	0.001806	**
PEAD	-0.020035	0.024988	-0.8018	0.422704	
Bond.mom	0.183404	0.112422	1.6314	0.102855	
Prob.Def	-0.034112	0.040514	-0.8420	0.399830	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 14: Estimation results from the post-crisis period for the European dataset

We observe that there are no factors which would be significant in all three periods. However in the most recent, post-crisis period, the following factors seem to have significant forecasting power for corporate bond returns in the European dataset: *Beta*, *Earnings Quality*, *Gearing*, *Long Term Leverage*, *Market Value* and also *Book to Price*.

2.4.3 Results for US data

As our analysis is done separately for European and US, we repeat the same steps in US dataset. We run three regressions according to the time period: pre-crisis, crisis and post-crisis. The coefficients of determination R^2 are presented the following table:

US data	FE model R^2
Before Crisis	0.0070271
During Crisis	0.025946
After Crisis	0.025184
Whole Dataset	0.0015611

Table 10: Statistics from regression of US corporate bonds excess returns

Even though that the coefficient of determination of the model is even smaller than for European dataset, it is still reasonable, as we do not except the equity return predictors to explain too much variance in bond returns. There is also an improvement of R^2 for the three periods when compared to the regression on the whole sample.

Similarly as in the European case, the hypothesis of homoscedasticity is rejected in all time periods and we move on to robust covariance matrix estimator.

We can compare the results of robust regression analysis for all three periods in Figures 15 - 17. The significant codes are marked according to the level of significance, explained at the bottom of figure.

Before crisis we have one significant factor *Equity Momentum*:

```
              Estimate  Std. Error t value  Pr(>|t|)
Beta       -0.02630429 0.01754535 -1.4992    0.1338
BTP        -0.04122884 0.03555020 -1.1597    0.2462
DY         -0.02818895 0.02407383 -1.1709    0.2416
EPS.growth  0.00920865 0.01059602  0.8691    0.3848
EQ          0.00123220 0.01123701  0.1097    0.9127
Eq.mom      0.05722620 0.01422558  4.0228 5.776e-05 ***
Gearing     0.01293770 0.01838791  0.7036    0.4817
LTLev      -0.01204217 0.01228159 -0.9805    0.3268
LTROA      -0.02135464 0.01779742 -1.1999    0.2302
MV         -0.00289480 0.00862468 -0.3356    0.7371
PEAD       -0.00093303 0.01439121 -0.0648    0.9483
Bond.mom   -0.03002787 0.02639355 -1.1377    0.2553
Prob.Def   -0.03794992 0.03435551 -1.1046    0.2693
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 15: Estimation results from the pre-crisis period for the US dataset

During crisis we have also one significant factor *Long Term Return on Asset*:

```
             Estimate Std. Error t value  Pr(>|t|)
Beta       -0.336864   0.235292 -1.4317 0.1524548
BTP         0.777231   0.749180  1.0374 0.2997071
DY         -0.077319   0.466052 -0.1659 0.8682576
EPS.growth  0.019495   0.289657  0.0673 0.9463496
EQ          0.057337   0.175113  0.3274 0.7433907
Eq.mom     -0.040058   0.262605 -0.1525 0.8787825
Gearing     0.134856   0.254966  0.5289 0.5969458
LTLev       0.012240   0.108065  0.1133 0.9098394
LTROA       0.424376   0.119596  3.5484 0.0004002 ***
MV          0.134834   0.315421  0.4275 0.6690993
PEAD        0.077105   0.210652  0.3660 0.7143976
Bond.mom   -0.410853   0.594054 -0.6916 0.4892954
Prob.Def   -0.645642   0.527443 -1.2241 0.2211187
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 16: Estimation results from the crisis period for the US dataset

The post-crisis period brings again the highest number of significant factors:

	Estimate	Std. Error	t value	Pr(>|t|)	
Beta	0.0976047	0.0397266	2.4569	0.0140263	*
BTP	0.0353398	0.0389937	0.9063	0.3647954	
DY	0.0173275	0.0214741	0.8069	0.4197357	
EPS.growth	-0.0495518	0.0220893	-2.2432	0.0248971	*
EQ	0.0151427	0.0147219	1.0286	0.3036942	
Eq.mom	-0.0292313	0.0311164	-0.9394	0.3475346	
Gearing	0.0116362	0.0096779	1.2023	0.2292518	
LTLev	0.0398645	0.0110101	3.6207	0.0002949	***
LTROA	-0.0512529	0.0163158	-3.1413	0.0016856	**
MV	-0.0299352	0.0151412	-1.9771	0.0480548	*
PEAD	0.0049121	0.0106458	0.4614	0.6445077	
Bond.mom	-0.0349480	0.0580316	-0.6022	0.5470347	
Prob.Def	0.0715935	0.0360360	1.9867	0.0469735	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 17: Estimation results from the post-crisis period for the US dataset

From the results above we observe that we have again more significant factors in post-crisis period, some of which are the same as for European model. The factors that have some power to predict US corporate bond excess returns found from the post-crisis periods are: *Beta*, *Earnings per Share growth*, *Long Term Return on Assets*, *Long Term Leverage* and *Market Value*.

2.4.4 Conclusion

Using panel data regression analysis, we suggest that the only significant factor common for both European and US datasets based on the whole timespan 1999-2014 is the *Equity Momentum* factor.

The division of our timespan into three sub-periods due to the significant market turbulence during financial crisis has proved itself to be reasonable. Factors *Beta*, *Long Term Leverage* and *Market Value* are all significant in post-crisis period in both the European and US datasets.

3 Backtest

While in the previous chapter we focused on statistical significance of equity factors on corporate bond returns, we now turn to the question of economic significance.

We devise a simple investment strategy based on the findings from the previous chapter. We evaluate the historical performance in an out-of-sample backtest and calculate the break-even trading cost. The break-even costs in comparison with actual trading costs will provide a measure of economic significance. We run separate backtests for the European and US datasets. Below by investing "into a firm" we mean investing into the corporate bonds of the firm which are available at a given point in time (for the calculation of the firm returns see Section 1.1.3).

3.1 Backtesting framework

We will backtest our findings from the previous chapter with a simple investment strategy, which is based on scoring available firms and investing our portfolio to the firms with the best scores. The score of each firm is calculated as the mean of the z-scores of the significant factors we identified in the previous chapter.

In our backtest we pay special attention not to introduce any forward-looking bias into the analysis, which would render the backtest results useless. Hence, given the equity factors, at any point in time where an investment decision is being made we use only information available at that time.

3.1.1 Scoring

To choose the firms in which we will invest our portfolio we use the method of scoring. This method is widely used among portfolio managers due to its robustness and easy interpretation. In the scoring method each firm is assigned a score. This score is calculated as the mean of the normalized factor values of the respective firm. The normalization is done using the cross-sectional standard z-score of the respective factors. This normalization ensures that the values of the factors are comparable between each other.

The z-score for factor i is defined as:

$$z_{ij} = \frac{F_{ij} - \mu_i}{\sigma_i}, \tag{10}$$

where $\mu_i = \frac{\sum_{j=1}^{N} F_{ij}}{N}$ is the mean of the factor values F_{ij} and $\sigma_i = \sqrt{\frac{\sum_{j=1}^{N}(F_{ij} - \mu_i)^2}{N-1}}$ is the standard deviation of the factor i, while $j = 1, ..., N$. To take into account a possible negative direction of the factor we multiply the relevant z-score by -1.

3.1.2 Backtesting algorithm

Our backtesting algorithm is as follows. We start with a cash portfolio of 100 money units (e.g. EUR or USD) at 1999-12-31. At the end of each month, when the investment universe for the next month becomes available we:

- calculate the portfolio performance in the past month as the weighted performance of the firms in portfolio,
- sell the whole existing portfolio,
- score the available firms as described above,
- divide the portfolio value equally weighted between the 50 firms with the highest score,
- calculate the transaction costs by comparing the old and the new portfolio composition and taking into account the specified bid/ask spreads,
- advance to the end of next month.

3.1.3 Trading costs

A very important part of the analysis are the transaction costs, i.e. the costs associated with the transactions taking place in our investment strategy. There are two kinds of costs associated with each trade – the ticket fee and the bid/ask spreads. The ticket fee is the amount paid for a trade, regardless of the size of the trade. For reasonably sized portfolios these fees are negligible and we will not take them into account in or analyses. The bid/ask spreads are the more important costs of trading, because these are quoted as percentage of the price of the investment (bonds in our case) and thus can become huge for a portfolio with a high turnover. The bid/ask spread represents the percentage difference between the highest price a buyer is ready to pay and the lowest price that a seller is willing to accept. The actual price for which the asset is traded is usually somewhere between the bid and ask price. We assume that for each sold (bought) bond the actual prices is lower (higher) by the half of the bid/ask spread. The actual bid/ask spreads for the bonds in the ER00 and C0A0 indices are summarized in Table 11:

Statistics	ER00	C0A0
minimum bid/ask spread	0.11%	0.03%
average bid/ask spread	0.52%	0.96%
maximum bid/ask spread	1.87%	5.73%

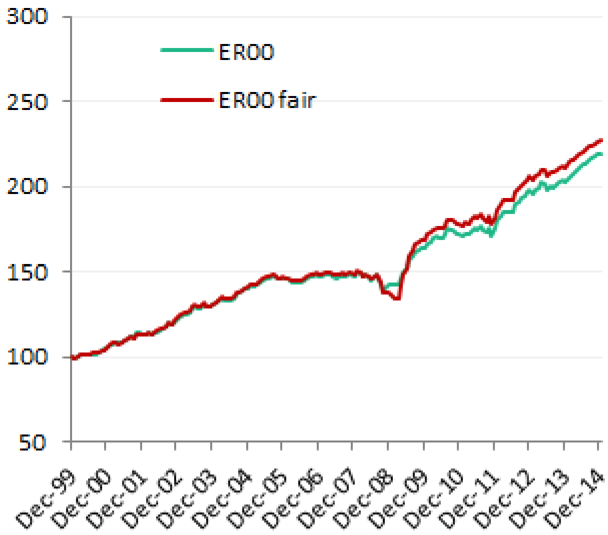
Table 11: Bid/ask spread summary for the ER00 and C0A0 indices

Instead of assuming a fixed bid/ask spread in our backtest results below, we use a break-even analysis. For each investment strategy we calculate the bid/ask spread for which the strategy performance equals the benchmark performance. The bid/ask spread is positive hence the break-even analysis makes only sense for investment strategies that outperform the benchmark. The calculation is done numerically with a grid-search with 0.1% steps.

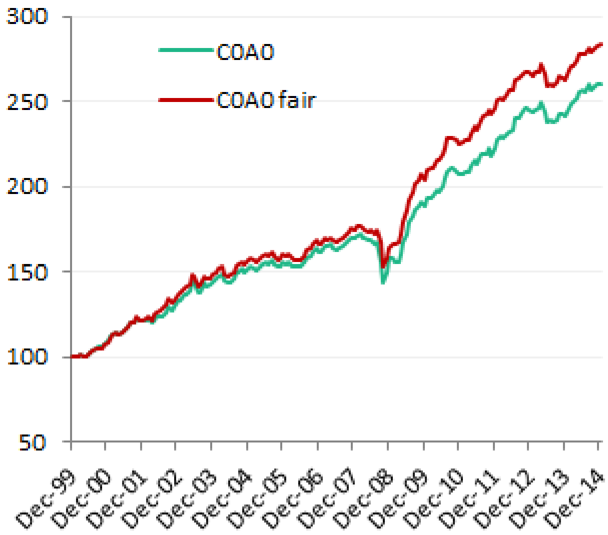
To evaluate the economic significance of an investment strategy one can compare the break-even transaction costs with the actual bid/ask spreads quoted on the market.

3.2 Fair benchmarks

We evaluate the results of our backtests by comparing the backtest performance to a benchmark (in addition to evaluating the break-even analysis as mentioned above). The natural benchmarks of our strategies are the two corporate bond indices ER00 and C0A0. These benchmarks would also be used in any practice for an invested portfolio. However, since we restricted out investment universe to the firms that are in the compositions of the equity indices STOXX 600 and S&P 500, these benchmarks would not be fair from a purely economic point-of-view. Therefore we calculate a fair version of the benchmarks were we allow only bonds of firms that are in the composition of the respective indices to be included. We include these bonds with original benchmark weight rescaled to 100 percent. In the Figure 18 below, we compare the total return indices of the original benchmarks ER00 and C0A0 with their fair subsets:



(a) European dataset



(b) US dataset

Figure 18: Comparison of the original ER00 and C0A0 benchmarks with their fair subsets

As we can see from the graphs, the fair benchmarks outperform the original benchmarks restricted most of the historical timespan we consider (the return of the ER00 benchmark is 5.38%, the C0A0 6.58% and their fair counterparts have 5.63% and 7.19% respectively, see Table 12 below). During the turmoil of the financial crisis the drawdowns of the fair benchmarks are higher what may stem from a higher concentration of the fair benchmark.

	ER00	ER00 fair	C0A0	C0A0 fair
2000	5.90%	5.72%	9.13%	8.46%
2001	6.85%	6.59%	10.70%	11.69%
2002	8.45%	9.34%	10.17%	13.20%
2003	6.47%	6.36%	8.31%	8.01%
2004	7.58%	7.76%	5.41%	5.43%
2005	4.04%	4.03%	1.97%	2.24%
2006	0.59%	0.91%	4.38%	4.00%
2007	0.22%	0.31%	4.64%	5.02%
2008	-3.28%	-7.72%	-6.82%	-6.45%
2009	14.90%	22.82%	19.76%	25.19%
2010	4.82%	5.56%	9.52%	10.02%
2011	1.99%	2.11%	7.51%	9.70%
2012	13.03%	13.15%	10.37%	8.31%
2013	2.39%	2.64%	-1.46%	-1.72%
2014	8.25%	7.86%	7.51%	7.95%
ret.pa	5.38%	5.63%	6.58%	7.19%

Table 12: Comparison of the original ER00 and C0A0 benchmarks with their fair subsets

In all analysis below, we work with the fair benchmarks, as these are the most objective indicators from an economic point of view. Since now, by "benchmark" we mean the fair version of the benchmark.

3.3 Backtest results

The results from our panel regression analysis in the previous chapter imply that:

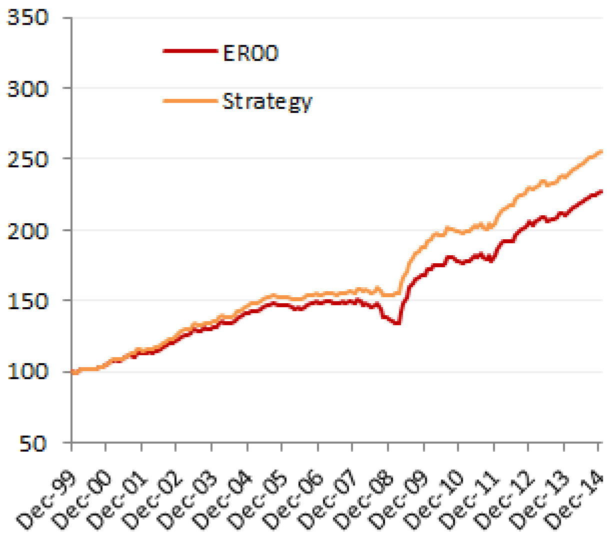
1.) the equity momentum factor (*Eq.mom*) is significant in the regression using the whole timespan 1999-2014 in both the European and US dataset,

2.) in the post-crisis period there are three significant factors *Beta*, *Long Term Leverage* (*LTLev*) and *Market Value* (*MV*) common to both datasets.

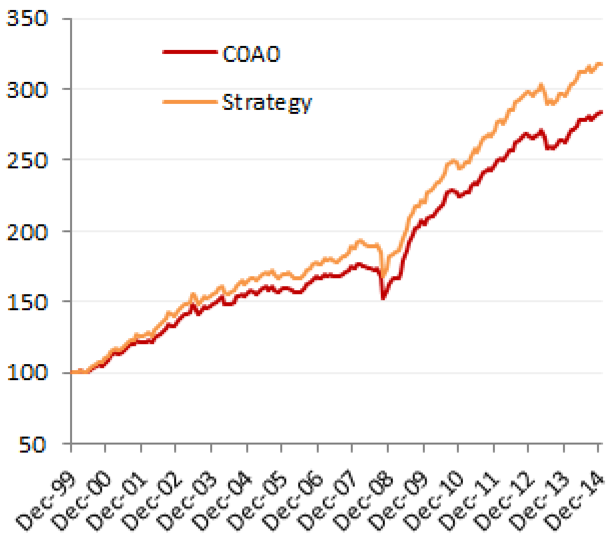
Based on these findings, we run the backtest with two different scoring strategies – in the first one (the Momentum Strategy) we score only on *Eq.mom* (positive direction) and in the second one (the Post-crisis Strategy) we score on *Beta* (positive direction), *LTLev* (positive direction) and *MV* (negative direction). For both strategies we provide a graphical performance overview, numerical performance figures, the break-even analysis and a robustness check based on the performance of quartile portfolios.

3.3.1 The Momentum strategy

First we present the results of our backtest with the investment strategy based solely on the *Eq.mom* factor. In the Figure 19 below we compare the total returns of the benchmarks and the strategy total returns for both the European and the US datasets. The numerical results are attached in Appendix B.



(a) European dataset



(b) US dataset

Figure 19: The ER00/C0A0 benchmarks vs. the Momentum Strategy

Our strategy is more successful than benchmark during the whole period from year 1999 until year 2014. Table 13 below gives numerical overview of the results of the strategy. The detailed description of the statistical information is provided in Appendix C.

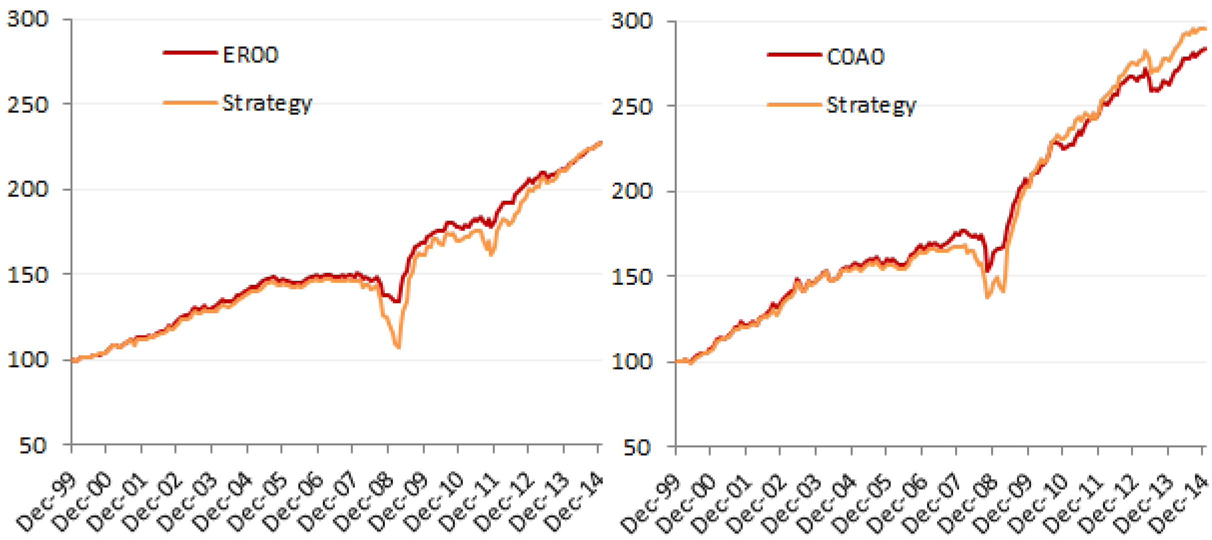
Information	EU 1999-2014	US 1999-2014
pf.ret	6.45%	8.02%
bm.ret	5.63%	7.19%
rf.ret	2.18%	2.18%
pf.exretbm	0.82%	0.83%
pf.exretrf	4.27%	5.83%
bm.exretrf	3.45%	5.00%
pf.vola	3.21%	5.37%
bm.vola	3.92%	5.29%
pf.te	0.1	0.02
pf.mdd	3.58%	12.82%
bm.mdd	10.67%	13.51%
pf.turnover	239%	324%
pf.tickets	246.52	328.78
pf.sharpe	1.33	1.09
bm.sharpe	0.88	0.95
pf.ir	3.55	2.78

Table 13: Statistical information for the Momentum Strategy

The break-even analysis shows low break-even costs of only 0.5% for the European and 0.6% for the US datasets. Comparing this with the actual mean bid/ask spreads of 0.52% and 0.96% for Europe and US respectively, one can conclude that this strategy is not very economically significant.

3.3.2 The Post-crisis strategy

Now we turn to the investment strategy based on three factors – *Beta*, *LTLev* and *MV*. In the Figure 20 below we compare the the total returns of the benchmarks and the strategy total returns for both the European and the US datasets. The numerical results are provided in Appendix B.



(a) European dataset (b) US dataset

Figure 20: The ER00/C0A0 benchmarks vs. the Post-crisis Strategy

As expected, this strategy is becoming successful only for the period where the factors were significant in regression - after the financial crisis. Before the crisis it consistently underperforms the benchmarks in both dataset.

In the Table 14 below we provide numerical overview of the performance of the strategy for the post-crisis period. The description of the statistical information is provided in Appendix C.

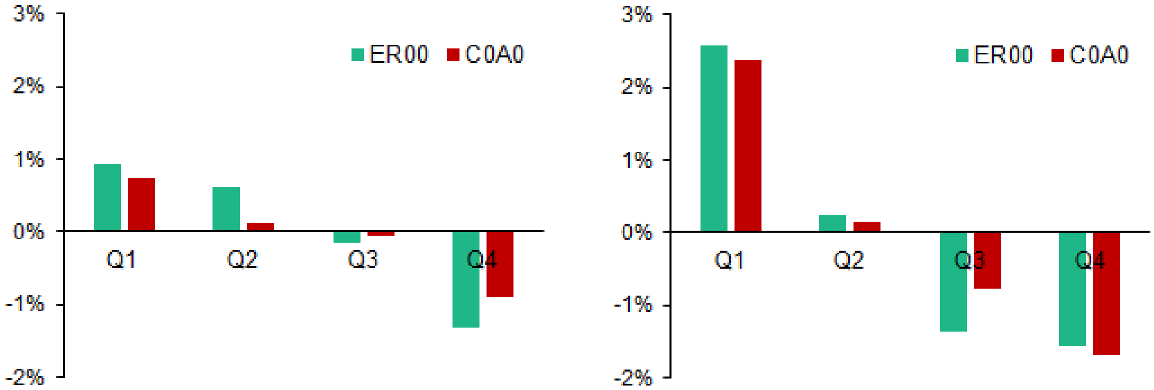
Information	EU Apr2009-2014	US Apr2009-2014
pf.ret	11.69%	12.58%
bm.ret	8.61%	9.27%
rf.ret	0.47%	0.47%
pf.exretbm	3.08%	3.31%
pf.exretrf	11.22%	12.11%
bm.exretrf	8.14%	8.80%
pf.vola	6.42%	6.31%
bm.vola	4.02%	4.48%
pf.te	0.04	0.03
pf.mdd	8.14%	4.42%
bm.mdd	2.91%	4.64%
pf.turnover	113%	150%
pf.tickets	138.06	152.73
pf.sharpe	1.56	1.92
bm.sharpe	2.02	1.96
pf.ir	2.52	3.53

Table 14: Statistical information for the Post-crisis Strategy

The turnover of 113% and 150% in the European and US dataset respectively is much lower than in the Momentum strategy. This is reflected in much higher break-even costs of 3.7% for the European dataset and 3.1% for the US dataset. Comparing these numbers with the actual mean bid/ask spreads of 0.52% for the ER00 and 0.96% for the C0A0 index one can conclude that the outperformance of this strategy is economically significant.

3.3.3 Quartile portfolios

To evaluate the robustness of the scoring factor(s) we look at the performance of equally weighted quartile portfolios. These portfolios are built by investing into firms in the respective scoring-quartile, e.g. each month we build four non-overlapping portfolios. If the outperformance generated by scoring on the chosen factor(s) is not purely random we expect that the performance of the quartile portfolios falls with the quartile. In Figure 21 we can see that this is indeed the case for the *Eq.mom* factor as well as for the Post-crisis investment strategy in both datasets. In Figure 21b the results are presented for the relevant post-crisis period.



(a) Momentum Strategy (b) Post-crisis Strategy

Figure 21: Quartile portfolios excess returns for the Momentum Strategy and the Post-crisis Strategy

The Q1 and Q2 portfolios composed from companies with the highest score have positive excess returns, while Q3 and Q4 which are composed of the second half of the benchmark have negative excess returns over benchmarks. This analysis suggests that both of our investment strategies are robust.

Conclusion

The focus of this thesis was the preparation of a single issues corporate bond/firm fundamentals database, an analysis of the predictability of corporate bond returns and the construction of an investment strategy for European and US corporate bond portfolio.

In the first chapter we described the data collection and provided some descriptive analysis of our database. This part of thesis was the most time consuming. The two databases we used didn't have a common key so we had to devise a complex process of merging the two datasets partly relying on manual internet search. The European dataset contains 337 firms and 86 333 single bond issues whereas the US contains 636 firms and 421 286 single bond issues. Both dataset have a monthly frequency in the timespan between December 1999 and December 2014.

In the second chapter we used panel data regression analysis to identify which of our selected firm fundamentals significantly predict corporate bond returns. To select which fundamental data we include in our regression analysis we focused on economically meaningful factors with low correlation and enough data. Along the lines of Goyal *et al.*, two bond specific factors were added. The significance levels were calculated using a robust covariance matrix estimation which is appropriate for fixed effects panel models where heteroscedasticity is present. The fluctuations of the financial crisis period seemed to have an adverse impact on predictability so we analyzed the pre-crisis, crisis and post-crisis periods separately.

Based on results from the second chapter we analyzed two different investment strategies. The first strategy is based solely on the momentum factor, which we found to be significant on the whole time span and consistently for both the European and US dataset. The second strategy is based on our findings from the post-crisis period, where the *Beta*, *Long Term Leverage* and *Size* factors were significant in both datasets.

Both strategies outperform the respective benchmarks. However after accounting for trading costs in form of bid-ask spreads, our break-even analysis suggests that due to the high turnover the momentum strategy is economically not significant. In the post-crisis strategy the turnover is much lower and the break even-analysis suggests that this strategy would have been economically significant.

Possible improvements to this work include e.g. enhancement of the algorithm for merging the bond database with firm data. A larger database could include also non-investment grade firms, where also the results of Goyal *et al.* were stronger. In the regression study more or maybe other predictors could be used. To bring the investment strategy nearer to real-world portfolio managers requirements one could introduce different constraint e.g. on sector weights, duration, rating etc. Further improvements could include robustness checks, e.g. not to take the mean of all the firms bonds which are available in the index at a given point in time but choose single issues based on some criteria (or randomly).

We conclude that we successfully fulfilled the aims we set out to in the introduction. Our suggested investment strategy is to invest into the corporate issued by firms with high *Beta* and *Long Term Leverage* scores and low *Market Value* score. This strategy was both statistically and economically significant in recent years in both regions we studied. This investment strategy could prove to be of value to a portfolio manager assuming that no disruptive crises hit the market in coming years.

References

[1] Baltagi, B. H.: *Econometric Analysis of Panel Data*, John Wiley & Sons Ltd, 2005.

[2] BofA Merrill Lynch Bond Indices: *Bank of America Bond Index System User Guide May 2013*, `http://www.mlindex.ml.com/`.

[3] Castilla, C.: *Heteroskedasticity in Fixed-Effects One-way Error Components Models: Evaluating the Performance of Standard Tests*, The Ohio State University, 2008, `https://www.researchgate.net/publication/228275611_Heteroskedasticity_in_Fixed-Effects_One-Way_Error_Component_Models_Evaluating_the_Performance_of_Standard_Tests`.

[4] Croissant, Y., Millo G.: *Panel Data Econometrics in R: The plm package*: `https://cran.r-project.org/web/packages/plm/plm.pdf`.

[5] Fama E.F., MacBeth J.D.: *Risk, Return, and Equilibrium: Empirical Tests*, Journal of Political Economy Vol. 81, No. 3, pp. 607-636, 1973.

[6] Goyal, A. *et al.*: *Are Capital Market Anomalies Common to Equity and Corporate Bond Markets?*, 2015, `http://papers.ssrn.com/sol3/papers`.

[7] Petersen, M. A.: *Estimating Standard Errors in Finance Panel*, Northwestern University, 2016, `http://rfs.oxfordjournals.org/`.

[8] STOXX Ltd., `https://www.stoxx.com/index-details?symbol=SXXGR`.

[9] Wooldridge, J.M.: *Correcting estimated standard errors in the presence of heteroskedaticity*, The American Economic Review, Vol. 93, No. 2, pp. 133-138, 2003, `http://people.stern.nyu.edu/wgreene/Lugano2013/Wooldridge-Cluster.pdf`.

[10] Thomson Financial: *Worldscope Database: Datatype Definitions Guide*, Issue 6, 2007.

A

Manually matched companies

Ticker	DSID	Name
SELIM	729439	Montedison Spa
SELIM	772670	Edison Spa
TELEFO	257686	Telefonica Europe
TELEFO	929534	Telefonica S.A.
BPLO	905478	Banca Popolare
BPIIM	905478	Banca Popolare
DANBNK	772753	Sampo Bank
DANBNK	929846	Danske Bank
HEIANA	905001	Heineken n.v.
VOD	953133	Vodafone
NGGLN	870181	National Grid Gas plc
FRTEL	885569	Orange S.A.
TIIM	923374	Telecom Italia Spa

Table 15: List of manually matched companies for the European dataset

Ticker	DSID	Name
T	945388	AT&T Corp.
TWX	325364	Time Warner Company
USB	951046	U.S. Bancorp
INC	32480Q	CBS Corp.
FON	689416	Centel Capital
FON	904864	Sprint Corp.
HON	906191	Honeywell,Inc.
NOC	905809	Northrop-Grumman
WB	923253	Wachovia Bank
WB	951048	First Union
AGN	541863	Allergan Plc.
MON	268193	Monsanto company
MON	902221	Pharmacia Corporation
MO	51605D	Philip Morris
MO	904853	Altria Group Inc.
CAH	944704	Cardinal Health
CAH	67751X	CareFusion
KO	741521	Coca-Cola Enterprises Inc.
COP	901666	ConocoPhillips
COP	87035H	Phillips
ABT	916328	Abbott Laboratories
ABT	87851X	AbbVie Inc.

Table 16: List of manually matched companies for the US dataset

B

Numerical results from backtesting

	ER00	Momentum Strategy for ER00	C0A0	Momentum Strategy for C0A0
2000	5.72%	6.01%	8.46%	12.25%
2001	6.59%	8.13%	11.69%	11.65%
2002	9.34%	10.45%	13.20%	15.48%
2003	6.36%	6.93%	8.01%	7.49%
2004	7.76%	8.19%	5.43%	6.19%
2005	4.03%	4.08%	2.24%	2.94%
2006	0.91%	1.18%	4.00%	3.86%
2007	0.31%	0.95%	5.02%	6.60%
2008	-7.72%	-1.31%	-6.45%	-2.95%
2009	22.82%	22.53%	25.19%	20.67%
2010	5.56%	5.86%	10.02%	11.07%
2011	2.11%	2.93%	9.70%	10.85%
2012	13.15%	12.34%	8.31%	9.56%
2013	2.64%	2.95%	-1.72%	-0.78%
YTD	7.86%	7.70%	7.95%	7.76%
ret.pa	5.63%	6.45%	7.19%	8.02%

Table 17: The ER00/C0A0 benchmarks vs. the Momentum Strategy

	ER00	Post-crisis Strategy for ER00	C0A0	Post-crisis Strategy for C0A0
2000	5.72%	5.62%	8.46%	7.49%
2001	6.59%	5.53%	11.69%	11.32%
2002	9.34%	9.24%	13.20%	12.25%
2003	6.36%	5.47%	8.01%	10.27%
2004	7.76%	8.02%	5.43%	4.43%
2005	4.03%	4.21%	2.24%	1.33%
2006	0.91%	0.99%	4.00%	4.19%
2007	0.31%	-0.23%	5.02%	2.19%
2008	-7.72%	-16.50%	-6.45%	-12.39%
2009	22.82%	32.48%	25.19%	38.82%
2010	5.56%	5.56%	10.02%	13.42%
2011	2.11%	-2.11%	9.70%	7.49%
2012	13.15%	20.23%	8.31%	11.27%
2013	2.64%	5.40%	-1.72%	0.40%
YTD	7.86%	7.59%	7.95%	6.69%
ret.pa	5.63%	5.62%	7.19%	7.48%

Table 18: The ER00/C0A0 benchmarks vs. the Post-crisis Strategy

	ER00	Post-crisis Strategy for ER00	C0A0	Post-crisis Strategy for C0A0
2009	18.32%	32.49%	19.18%	34.81%
2010	5.56%	5.74%	10.02%	13.33%
2011	2.11%	-1.98%	9.70%	7.78%
2012	13.15%	20.19%	8.31%	11.04%
2013	2.64%	5.40%	-1.72%	0.29%
2014	7.86%	7.58%	7.95%	6.76%
ret.pa	8.61%	11.69%	9.27%	12.58%

Table 19: The ER00/C0A0 benchmarks vs. the Post-crisis Strategy restricted to the post-crisis period

C

Description of statistical information for investment strategies

- Total return p.a.*(pf.ret, bm.ret, rf.ret)*: describes total return of portfolio, benchmark or risk-free return.
- Return over benchmark p.a. *(pf.exretbm)*: calculates the difference between our portfolio and the benchmark return.
- Return over risk free p.a. *(pf.exretrf)*: calculates the difference between our portfolio and the risk-free return.
- Benchmark over risk free p.a. *(bm.exretrf)*: calculates the excess returns of the benchmark.
- Volatility *(pf.vola,bm.vola)*: represents the dispersion of returns of our portfolio or benchmark.
- Tracking error *(pf.te)*: is the divergence between the price behavior of the portfolio and the price behavior of the benchmark.
- Maximum Drawdown *(pf.mdd,bm.mdd)*: is the maximum loss from a peak to a trough of a portfolio (or benchmark), before a new peak is reached and is calculated as (Trough Value – Peak Value) / Peak Value.
- Turnover *(pf.turnover)*: measures how frequently are the assets in portfolio traded. It is annualized rate found by dividing the purchases or sales (which is less) by the average of portfolio value.
- Trading tickets *(pf.tickets)*: represents number of trades in portfolio.
- Sharpe ratio*(pf.sharpe, bm.sharpe)*: is the average return earned in excess of the risk-free rate per unit of volatility. This is an information about risk-adjusted return and is calculated for our portfolio as well as for the benchmark.
- Information ratio *(pf.ir)*: is the ratio of portfolio returns above the returns of a benchmark *(pf.exretbm)* to the volatility of these returns.