MASTER THESIS

DEVELOPING A SCREENING METHODOLOGY FOR THE CORPORATE CREDIT UNIVERSE

NN investment partners

Robert van Wieren

Master specialization Financial Engineering & Management

Company supervisor:
R.E.H. Sanders

University supervisors:
B. Roorda
R.A.M.G. Joosten

October 22, 2019

UNIVERSITY OF TWENTE.

Management summary

The number of bonds issued by companies in emerging markets has grown strongly in the past decade. Between 2007 and 2018, the number of issuers has grown from 200 to over 600. This growth poses a challenge to fund managers. To keep the same level of intensity of coverage, fund managers need to have tripled analysts staffing over the decade, which would increase the costs of asset management and ultimately be to the detriment to the fund managers clients. An effective screening method can assist in allocating resources effectively, keep costs under control and aid focusing on securities that provide an attractive return potential. The output of the screening is a negative recommendation for a part of the universe, indicating that these securities are likely unattractive and that further in-depth analysis is unnecessary.

Research objective: Develop a data-driven, systematic screening model to reduce the corporate credit universe without losing sight of attractive opportunities.

To achieve this objective, we first have to define what attractive opportunities are. Traditional quantitative strategies have focused on screening for "undervalued" bonds which are likely to outperform a benchmark. However, a fund has the ability to have an underweight position in one bond. This allows the fund to increase excess return using underperforming bonds. For this reason, attractive opportunities are not only bonds which are likely to overperform, but also bonds likely to underperform the benchmark. Allocation of research to these two types of bonds makes research analysts able to focus on securities which influence fund performance the most.

In order to measure performance of screening we have to determine which corporate bonds in emerging markets have been attractive at which points in time. For this, we use panel data from 2013 to 2019 of bonds in the "J.P. Morgan CEMBI Broad" benchmark. Using price return data, we calculate for each bond the maximum risk adjusted trading opportunity in a timeframe of six months. While moving this timeframe from 2013 to 2019, we indicate which bonds have been attractive. Each day, the 70% bonds with the highest risk adjusted trading opportunities are indicated as attractive and the other 30% are indicated as unattractive. Because our objective is to not lose sight of attractive opportunities, we should minimize the probability of occurrence of a bond screened negative while it is attractive (Type II error).

To see which screening methods are effective, a Multi Criteria Decision Analysis (MCDA) is executed. Screening based on credit spread, volatility, liquidity and value factor score high on a combination of effectiveness criteria and simplicity criteria. These screening methods are implemented together with combinations of these methods. It turns out that screening based on a combination of volatility and spread performs best on minimizing Type II error. A reduction of 21% could be achieved by a daily negative screening of the lowest 40th quantile of spread, and the lowest 25th quantile of volatility. Executing this screening from 2013 to 2019 resulted in a True Positive Rate of 95%, which equals a Type II error rate of 5%. One side effect of this method is a higher beta than the former universe. However, a lower beta could be achieved by additional investing in undervalued bonds in the negative screened universe.

Contents

Chapter 1

Introduction

The Specialized Fixed-Income business unit of the asset management firm NN Investment Partners hosts several teams that invest in bonds issued by companies, also called corporate credits. Identifying attractive securities through bottom-up analysis is a key component of their investment approach. The fundamentals of a company are considered from numerous perspectives including interpreting its financial statements, assessing quality of the management, understanding its product and services and having insight into competition and suppliers. There are thousands of companies that issue bonds so in practice it is not possible to follow every company in detail. An effective screening method can assist in allocating resources effectively and aid focusing on securities that provide an attractive return potential.

The aim of the research thesis is to investigate data-driven, systematic screening of the credit universe. The output of the screening is a negative recommendation for a part of the universe, indicating that these securities are likely unattractive and that further in-depth analysis is not necessary. As a starting point, the methodology will be tested and developed for emerging market debt corporate credits. However, the screening method should be general enough to be applicable to all types of corporate credits.

1.1   Problem description

The number of bonds issued by companies in emerging markets has grown strongly in the past decade. Figure 1.1 shows this for the total number of issuers and the amount of debt issued by region. Between 2007 and 2018, the number of issuers has grown from 200 to over 600. This growth poses a challenge to fund managers. Matching the growth with analyst capacity would skyrocket costs, which ultimately comes at the expense of return for their clients.

We formulate a problem statement to explain this further. It addresses the issue, and describes the condition to be improved upon. It breaks down in four parts: ideal, reality, consequences and proposal.
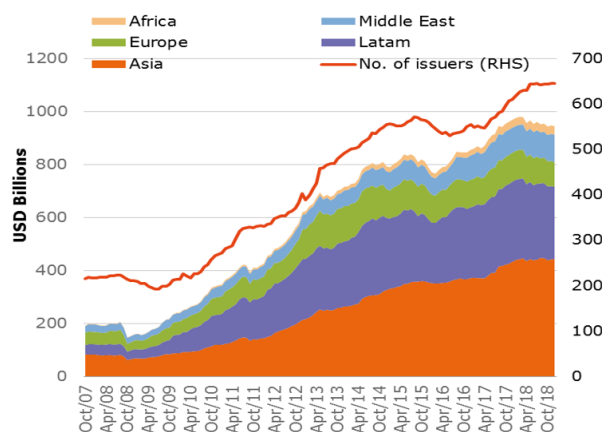
Figure 1.1: Market growth in Emerging Market Debt.

1. **Ideal:** All credit issues are researched extensively while the cost of resources of this research is minimized. Moreover, the value of both humans and computers are optimally deployed in the investment process.

2. **Reality:** There are not a sufficient number of analysts available to cover all corporate bond issues at the highest level of intensity. There is scope to employ computer resources, in particular to create a systematic decision process regarding which bonds issues will not be covered.

3. **Consequences:** There are several consequences. More analysis hours are necessary to cover the additional issuers, this will increase fund costs. When credit research is done with the same number of analysis hours, less time will be spent on each issuer which leads to a lower quality of research. Another option is to cover only one part of the universe. Here, one part of the credit universe is researched extensively while the other part is not taken into account. This leads to an error of omission by missing out attractive opportunities.

4. **Proposal:** Data-driven (computer) resources are going to be used as an addition to the current investment process. The application software will cover all credit issues and makes a recommendation for the analysts regarding which issues are worth to research extensively.

1.2   Research objective and questions

In Section 1.1 the main problem is defined and the proposal describes a way of finding a solution. But what should the solution of the research look like? The research objective is formulated as follows:

Research objective: Develop a data-driven, systematic screening model to reduce the corporate credit universe without losing sight of attractive opportunities.

To reach this objective we first got to find the answer to several questions. First, the word "attractive" needs more explanation. With the use of portfolio performance metrics, we can analyse which bonds deserve more research allocation than others. When we arrive to the second question, we already know roughly what are attractive securities to cover. We research what screening methodologies are possible and how potentially effective they could be. From the selected methodologies we will try to learn from in the literature available similar ones. At last, the methods are applied, the data-driven screening model is programmed and the screening is back tested to measure performance.

Research questions:

1. What are attractive securities to cover in fixed income portfolio management?

2. How can we measure screening performance?

3. What are effective methods for screening?

1.3   Thesis outline

The previous sections have explained *what* to research. This section will describe *how* the research will be conducted. We do this by giving an overview of the chapters of this thesis.

In Chapter 2 the major concepts of this thesis will be given. First, the area of fixed income portfolio management will be explained. The focus of this section is on performance management in fixed income. Secondly, a description is given on how to estimate a yield curve, because it plays a role in the value factor model later in the thesis. Lastly, some elements of an investment process of a manager that invests in EM Corporate debt are highlighted to provide context to the research.

In Chapter 3 the methodology of our research is explained. We first discuss the type of research that is conducted. Section 3.2 describes what data are collected and how they are processed. Section 3.3 and 3.4 discusses the used research methods.

In Chapter 4 we define what attractive bonds are in the context of this research. It is necessary to define this at an early stage, so we can continue with looking for effective screening methods. Section 4.2 and Section 4.3 explains two models on performance measurement. The first one describes how a screening is able to identify the attractive opportunities, as defined in Section 4,1. The second model shows how investing according to this screening affects total return. We do this to gain more knowledge about attribute bias in the positive screened universe.

In Chapter 5 we first determine what criteria a screening method should satisfy. Section 5.2 shows a list for possible screening methods. Each method gets a score for every criterion. Adding up these scores makes it possible to rank each method on both simplicity and effectiveness (Section 5.3). By doing this, we gain knowledge on what screening methods are best to implement. In Section 5.4, the implementation of each screening method is shortly explained, except the value factor model, which needs some more explanation.

In Chapter 6, the effectiveness of each screening method is determined by our performance measurement model. Based on these results, we conclude which screening methods perform best. In the end, the research is finished with a proposal for NNIP on how to shrink the EMD corporate debt universe.

In Chapter 7 a discussion of the research is written. Finally, Chapter 8 concludes this thesis. Conclusions will be made and recommendations for further research will be given.

Chapter 2

Background

In this chapter, some background information will be given on the concepts and methods used in this thesis. Section 2.1 explains the basics of fixed income and how bond performance can be measured. Fixed income is also a topic in Section 2.2, where a model on the term structure of bonds is explained. Section 2.3 concludes this chapter, with a description of some elements of an investment process of a manager that invests in EM Corporate debt.

2.1 Fixed income portfolio management

In fixed income portfolio management, one entity invests in one or more bond securities. A bond is an instrument of indebtedness of the bond issuer to the holder. The most common types of bonds include government bonds and corporate bonds. Although there a quite some differences between these two types of bonds, the method of valuation follows the same procedure.

If you own a bond, you are entitled to a fixed set of cash payoffs. Every year until the bond matures, you collect regular interest payments. At maturity, when you get the final interest payment, you also get back the face value of the bond, which is called the bond's principal (Brealey et al., 2014). The value of a bond is determined by the cash flows following from all the coupons and the principal. The present value is the current value of a future stream of cash flows given a specified rate of return. Future cashflows are discounted at the discount rate, and the higher the discount rate, the lower the present value of future cash flows. Equation 2.1 shows the present value calculation for a bond security.

$$P_0 = \sum_{t=1}^{T} \frac{coupon}{(1+r)^t} + \frac{principal}{(1+r)^T} \qquad (2.1)$$

The discount rate r is also called yield to maturity. Because at one time, all coupon and principal payments are known, a trade off occurs between yield to maturity and the value of a bond: the lower the yield to maturity, the higher the value of a bond. A bond issuer can default and thereby, fail to make an interest or principal payment within the specified period. In general, government bonds have a lower yield to maturity than corporate bonds because they are less likely to default. Rating agencies are specialized in predicting defaults for countries and corporations. Figure 2.1 shows the credit rating of Standard & Poor (S&P) for countries in Europe as at 6 January 2016. The most famous government bond is issued by the U.S. Treasury. Some of these issues do not mature for 20 or 30 years; others, known as notes, mature in 10 years or less.
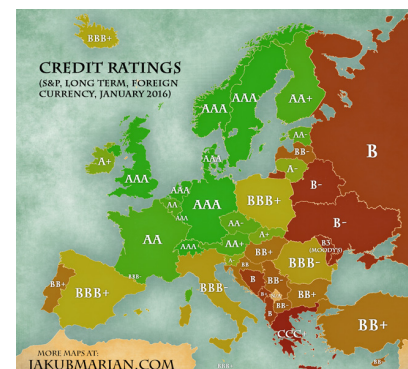
Figure 2.1: Credit country rating in Europe. Retrieved from Marian (2016).

In fixed-income security analysis, it is important to understand why bond prices and yield to maturity change. To do this, it is useful to separate a yield to maturity into two components: the benchmark and the spread. The reason for this separation is to distinguish between macroeconomic and micro economic factors that affect the bond price and, therefore, its yield to maturity. The benchmark captures the macroeconomic factors: the expected rate of inflation in the currency in which the bond is denominated, general economic growth and the business cycle, foreign exchange rates, and the impact of monetary and fiscal policy (Petitt et al., 2015). One approach to determine spread is by calculating the constant yield spread over a government spot curve. This spread is known as the zero-volatility spread

(Z-spread). At a given time t, Z-spread over the benchmark spot curve can be calculated with Equation 2.2

$$P_t = \sum_{n=1}^{N} \frac{coupon}{(1 + z_{[t,n]} + Z_t)^n} + \frac{principal}{(1 + z_{[t,N]} + Z_t)^N} \tag{2.2}$$

The benchmark spot rates ($z_{[t,1]}$, $z_{[t,2]}$, ..., $z_{[t,N]}$) are derived from the government yield curve at time t. Z-spread can be calculated if the coupon, principal and market price P_t are known. The Z-spread is also used to calculate the option-adjusted spread (OAS) on bonds with embedded options. The OAS, like the option-adjusted yield, is based on an option-pricing model and an assumption about future interest rate volatility (Petitt et al., 2015). Then, the value of the embedded option, which is stated in basis points per year, is subtracted from the yield spread. In particular, it is subtracted from the Z-spread (Petitt et al., 2015):

$$OAS = \textit{Z-spread - Option value} \tag{2.3}$$

Separating yield to maturity in a benchmark and a spread component is a very important step for further analysis on corporate bond valuation. For now, we will focus on credit spread. Every corporate bond has its own characteristics, and investors take each of them into consideration to determine a bonds fair value. Therefore, it is interesting to look how different components affect credit spread (Petitt et al., 2015):

- **Default risk component.** Probability of default is an important driver of spread size. The higher this probability, the more uncertain future cash flows from coupons and principal are. There a lot of risk factors contributing to spread. For corporate bonds, two sources of risk can be identified: business risk and industry risk. Examples of business risk are country risk, industry risk and competitive position. Financial risk refers to the quality of fundamentals of a company, like cash flow and leverage position. Business risk and financial risk are both taken into consideration by credit agencies to determine a credit rating.

- **Liquidity component.** Liquidity risk is the risk that the investor will have to sell a bond below its true value where the true value is indicated by a recent transaction. The primary measure of liquidity is the size of the spread between the bid price and the ask price quoted by a dealer (Fabozzi, 2005). For investors who plan to hold an issue to maturity, liquidity risk is not a major concern.

- **Tax component.** Bond spreads are not only explained by default risk and liquidity risk, but also by taxes (Johnson and Qi, 2015). It has been found that personal taxes explain a significant portion of corporate bond spreads. Coupons on corporate bonds are subject to state taxes, while government bond coupons are not (Elton et al., 1999). Another founding is that for higher rating categories, the tax and liquidity factors become more important since they (hardly) vary across rating categories (Driessen, 2005).

Till now we focused on determining the value of a bond at one specific time. It could be more useful to determine bond return on an interval of time. Bond return can be composed into multiple factors:

- **Coupon return.** The coupon return is often defined as 'carry'. It is the income of a portfolio as result of the passage of time. The definition of the coupon return is given by the following equation:

$$CouponReturn_{[t_1,t_2]} = \frac{(AI_{t_2} - AI_{t_1}) + \sum\limits_{t \in [t_1,t_2]} C_t}{P_{t_1}} \tag{2.4}$$

  Here, AI_t is the accrued interest of the bond at time t. C_t is the coupon received at time t, at P_t is the value of the bond at time t. It represents the return due to coupon payments and changes in accrued interest between time t_1 and t_2.

- **Price return.** In the context of price return, clean price is meant. The clean price is the price which does not take into account the coupon return (accrued interest). The price return is the return which results from variations in the clean price of the portfolio. For a standard bond (i.e. without optionality and/or early redemption), the price return can be split into 3 main components: the spread effect, the roll down and the curve effect. The spread effect is described above, the roll down and curve effect will be a topic in Section 2.2. The definition of price return is given by the following formula:

$$PriceReturn = \frac{P_{t_2} - P_{t_1}}{P_{t_1}} \tag{2.5}$$

- **Currency return.** All securities in a portfolio has to be translated to the reference currency of the one who is holding the portfolio. Changes in the exchange rate of the currency of bonds held in the portfolio impact return. The definition of the currency return is given by the following formula:

$$CurrencyReturn = \frac{M_{t_2} S_{t_2}}{M_{t_1} S_{t_1}} - 1 \tag{2.6}$$

Here, M_t represents the market value of the bond at time t. S_{t_1} and S_{t_2} are the spot exchange rates at time t_1 and t_2.

The total return can be obtained from the above equations. Besides absolute total return, a return that is linked to a benchmark, called excess return, is important in fixed income portfolio management. Such benchmarked processes, aim to outperform certain predefined indices.

$$
\begin{aligned}
TotalReturn_{fund} &= CouponReturn_{fund} + PriceReturn_{fund} + CurrencyReturn_{fund} \\
TotalReturn_{index} &= CouponReturn_{index} + PriceReturn_{index} + CurrencyReturn_{index} \\
ExcessReturn &= TotalReturn_{fund} - TotalReturn_{index}
\end{aligned}
\tag{2.7}
$$

The goal for funds that are linked to an index is to achieve an as high as possible excess return. However, this is preferably done by investments that are low risk. Because of the existence of a risk premium in the market, it is easier to achieve a high excess return by taking more risk. To adjust excess return for risk the Information Ratio (IR) is calculated:

$$IR = \frac{ExcessReturn}{\sigma_{R_f - R_i}} \tag{2.8}$$

The denominator of Equation 2.8 is also known as Tracking Error: the standard deviation of the difference between the total returns of the fund and the index. A high IR indicates a consistent outperformance of the benchmark by the fund. A performance measure which is not directly linked to return is the Beta of the fund. The beta of a portfolio is a measurement of its volatility of returns relative to the entire index.

$$\beta = \rho_{fund,index} \frac{\sigma_{fund}}{\sigma_{index}} \tag{2.9}$$

Here, ρ is the correlation between the fund and the index and σ is the volatility of the total return.

2.2 Modeling the term structure

The term structure of interest rates is the relationship between interest rates or bond yields and different terms of maturities for similar debt contracts. When graphed, the term structure of interest rates is known as a yield curve. The yield curves corresponding to the bonds issued by governments in their own currency are called the government bond yield curve. For example, the U.S. dollar interest rates paid on U.S. Treasury securities are closely watched by many traders and the media, because it is an important economic indicator. The shape of the yield curve changes over time. Normally the curve is up sloping, which means that a higher maturity results in a higher yield, but this is not always the case. Figure 2.2 shows how the U.S. treasury yield curve have changed over time.

We are not going in detail about what is affecting the shape of the yield curve. For now, we will focus on modeling the shape of the yield curve. Treasury rates are given for discrete values (3 months, 6 months, 3 years) and it is useful to estimate a continuous curve. There are multiple ways to this. The one described here is the Nelson-Siegel model.

Nelson and Siegel (1987) introduced a yield curve model with the purpose to be simple, parsimonious and flexible enough to represent the range of shapes generally associated with yield curves. This yield structure could be used to determine the values of bonds. The Nelson Siegel model is widely used by central banks (Grkaynak et al., 2006). The yield curve is modelled using three components. The first one remains constant when the term to maturity (τ) varies. The second factor has more impact on short
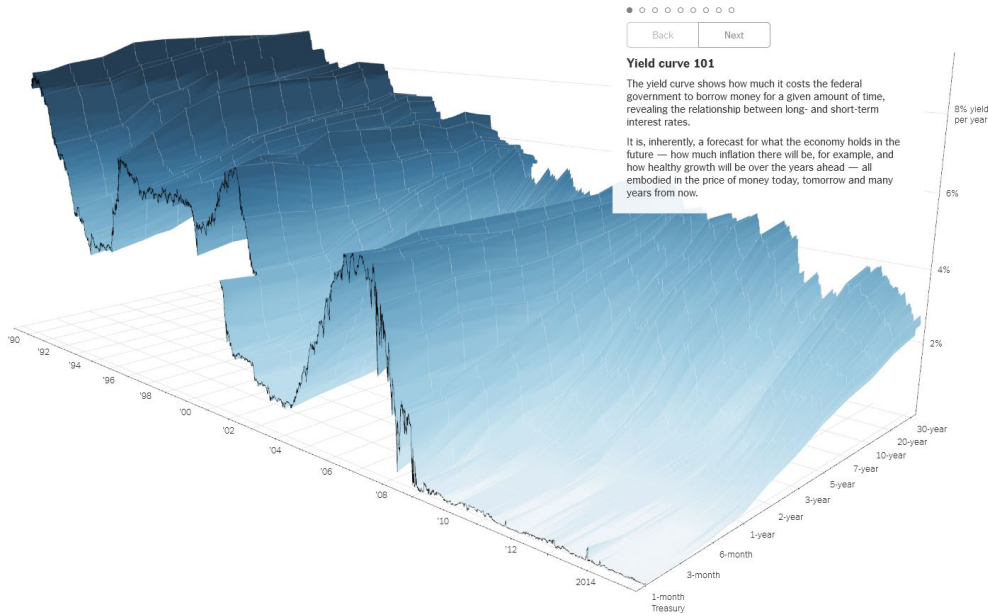
Figure 2.2: On March 18, 2015 the New York Times published an article called "A 3-D view of a chart that predicts the economic future". Gregor Aisch and Amanda Cox made an info graphic about the development of the Treasury yield curve since 1994.

maturities. The impact of the third factor increases with maturity, reaches a peak and then decays to zero (Ibanez, 2015). In their model Nelson and Siegel (1987) specify the forward rate curve $y(\tau)$ as follows:

$$y(\tau) = \beta_0 + \beta_1 \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_2 \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \tag{2.10}$$

where τ is time to maturity, β_0, β_1, β_2 and λ are coefficients, with $\lambda > 0$. β_0 is interpreted as the long run level of interest rates; β_1 is the short term component; β_2 is the medium term component and λ is the decay factor. The effect of both the short and medium term component β_1 respectively β_2 converge to zero over time, which leaves β_0 the singular component to determine the yield on the long run.

Diebold and Li (2003) addresses a key-practical problem with studies performed so far. They make a novel twist of interpretation of the Nelson Siegel model and furthermore go in or out-of-sample of forecasting of yield curves with their model. Their research shows that these parameters can be interpreted as factors that may vary over time and further more shows that the models are consistent with a variety of stylized facts regarding the yield curve.

We will use the Nelson-Siegel model in the value factor model. The position in the term structure influences the value of a bond. For corporate bonds this value is mainly described by the spread. To compare bonds with different term values we have to adjust the spread for this term. To do this, we have to estimate a spread curve based on the data of multiple issues. The Nelson-Siegel model is an appropriate and accepted model for fitting this kind of spread curves.

2.3   Description of common investment process elements

An Emerging Market Debt corporate debt manager invests in bonds issued by companies in countries labelled as emerging markets. These bonds could be housed in a separate EM Corporate Debt fund (EMCD) or in several sub-funds, also called sleeves, of other funds. Every fund is linked to a benchmark; a related index which measures the value of a section of the bond market. These benchmarks are replicated by purchasing a subset of the issues available within it. Thereafter, they are used as a measure of the market portfolio's return to compare the fund's performance with.

Within the EMD Corporate debt universe, two main benchmarks can be distinguished. "J.P. Morgan CEMBI Broad" and "J.P. Morgan CEMBI Diversified". The main difference is the size. The broad index includes all EM corporate bond issuances bigger than 300 million dollars and with 1 year or more to maturity. The diversified index includes issuances larger than 500 million dollars, takes only two issues per issuer and caps country exposure to increase diversification. Because the diversified index does not take all issues (and thereby the whole yield curve), a fund could also invest in bonds that are not part of the index. Besides, issuers can exist in an EMCD Fund that are not in the related benchmark.

9
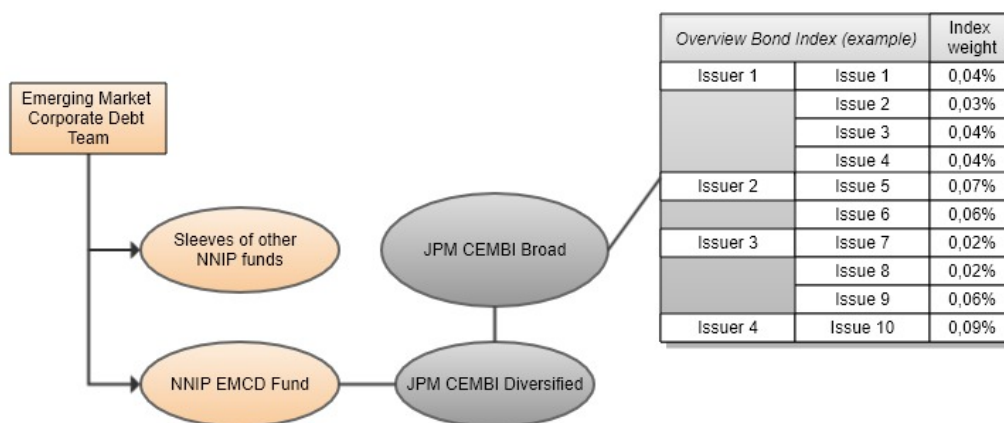
Figure 2.3: Example overview of EMD strategies (orange ellipses), with accompanying benchmarks (grey ellipses).

Every fund is managed by one or more Portfolio Managers (PMs). Position sizing, implementation and risk management are core tasks of a PM. They work together with analysts who asses credit worthiness of the issuers by bottom-up and top-down analysis. At a certain point a decision must be made how one issuer position in the benchmark has to be positioned in an EMCD Fund. There are many ways to determine the position size. One way is to split the process into determining the direction of the trade and the size of the trade. The first one is the recommendation given: underweight, neutral or overweight in comparison to the benchmark. The second one is the conviction level about this weight given. In the introduction we explained that it is not practical nor cost-efficient to have intensive research on all issuers. In practice, there are several ways PMs can deal with this. For instance, some of these issuers could be left as a neutral weight in the fund due to the big market weight in the benchmark. Also, a PM can decide not to invest in the issuer. This results in a underweight position in comparison to the benchmark.

Chapter 3

Methodology

In this chapter, we first discuss the type of research that is conducted in Section 3.1. Section 3.2 describes what data is collected and how it is processed. Finally, in Section 3.3 and 3.4, the methods that are used in this researched are discussed.

3.1 Type of research

We aim to develop a data-driven, systematic screening model to reduce the corporate credit universe, without losing sight of attractive opportunities. This could be labeled as applied research. Applied research is a methodology used to solve a specific, practical problem of an individual or group. In this case, we are solving a problem for the EMD section within NN Investment Partners.

The three research questions necessary to answer before we are able to reach the objective are knowledge questions. The first one we are trying to solve analytically by using performance metrics in fixed income. For the second and third research questions, data analysis is used.

3.2 Data collection

From Section 2.3 we learned that the "J.P. Morgan CEMBI Broad" index covers more issues than "J.P. Morgan CEMBI Diversified" index. Therefore we do data analysis on the broad index. The data are provided by JP Morgan and available via the database of NN Investment Partners. It consists of longitudinal data, which reach from January 2, 2013 to March 6, 2019, except for quarter 4 of 2017 which is missing. It shows multiple characteristics of bonds that are in the index on each date. Normally, the data are updated on each trading day. When an effective screening method is implemented the screening could be used on a daily basis.

3.3 Classification

The aim of this research is to develop a quantitative screening. With screening we actually mean classifying bonds over time. At one time, we can see what securities are in each class and decide where research should be allocated. Section 2.3 describes the research as a combination of a bottom-up and top-down analysis. This is a very time consuming process. It is not possible to do only half of it because the quality of this credit research would be very low. That is why we define two classes. One class, with securities that need to be analyzed further (going through the investment research process), and another class, indicating that these securities that are likely not attractive and further in-depth analysis is not necessary. A classification problem with two classes is also known as binary classification. The following predictions are distinguished:

Prediction 1 (positive): we predict that this security is likely to be attractive and further in-depth analysis is necessary
Prediction 0 (negative): we predict that this security is not attractive enough for further research

Predicting is done based on characteristics of the bond. The screening methods used for this are explained in Chapter 5. One important thing to notice is that this is done after performance measurement which is treated in Chapter 4. In Chapter 4 we try to define which bonds actually has been attractive:

Actual 1 (positive): this security has been attractive and further in-depth analysis was necessary
Actual 0 (negative): this security has not been attractive enough for further research

Model 1 (hypothesis testing) in Chapter 4 aims to define which bonds have actually been attractive and which has been not. The actual values are defined first before we explore methods in Chapter 5. This is preferable, because then we know better what screening methods are likely to have a good performance. In the tables below, an example screening is shown, next to the actual values.

Prediction (Chapter 5)				Actual values (Chapter 4)		
	Bond A	Bond B	Bond C		Bond A	Bond B	Bond C
day 1	1	0	1	day 1	1	0	1
day 2	1	0	1	day 2	1	0	1
day 3	1	1	1	day 3	1	0	0
day 4	0	1	1	day 4	1	0	0

Performance measurement is done by combining the two binary matrices. For example: if our screening methods predicts positive, and the actual value is also positive, a true positive is measured. When this is done for the whole dataset the screening method can be evaluated.

3.4   Multi Criteria Decision Analysis (MCDA)

In Chapter 5 we aim to classify attractive opportunities based on characteristics of the security. There are multiple ways to do this. Multiple methods for screening are known and can be created. However, implementing a screening method takes time and effort. Because of the limited time span of our research (6 months) it is not possible to implement all screening methods. Decisions on which screening method to develop have to be made on multiple criteria. To make the decision easier we make use of a Multi Criteria Decision Analysis. MCDA is a decision-making analysis that evaluates multiple (conflicting) criteria as part of the decision-making process. There are multiple MCDA models to choose from.

For this decision problem, the weighted sum model (WSM) is chosen. In this model, each criterion is given a weight, which is multiplied with the score of the screening method on that criterion. The best alternative is the one that scores high on the weighted sum of all criteria. Weighted summation can only be applied if the attributes are additive. This assumption means that there should be no interaction between the attributes, meaning that the attributes should be independent of each other, which is in many cases an unrealistic assumption. One weakness of this method is the loss of information due to the aggregated value. In this final score, it not possible to trace back a very high score of one alternative on one criterion. Another weakness is the difficult task of assigning weights to each criterion, especially when the number of criteria is large and the criteria are very different in character. However, the aim of this analysis is to quickly and roughly show what screenings could be effective. Therefore, the strengths of WSM in terms of simplicity and interpretation outweigh this weaknesses.

Chapter 4

Performance measurement

In this chapter, we give answer to the first two research questions. In Section 4.1 the research question "what are attractive securities to cover in fixed income portfolio management?" is answered by an analytic approach. In Section 4.2 a model is implemented which shows how screening performance can be measured. In Section 4.3 explains another model, which aims to show whether the remaining universe is biased to other performance measurements in fixed income. With the development of these models, we answer Research Question 2: "How can we measure screening performance?".

4.1 Review of attractiveness criteria

The composition of a fund is, to a great extent, determined by the benchmark (Section 2.3). At the same time, the funds' objective is to outperform this benchmark. This performance is measured by Excess Return (Section 2.1). Equation 4.1 shows how Excess Return can be decomposed in a weight component and a return component.

$$
\begin{aligned}
ExcessReturn = TotalReturn_{fund} &- TotalReturn_{index} \\
&= \sum_{i=1}^{n} w_i r_i - \sum_{i=1}^{n} v_i r_i \\
&= \sum_{i=1}^{n} (w_i - v_i) r_i
\end{aligned}
\tag{4.1}
$$

Here, w_i is the fund weight of bond i, v_i is the index weight of bond i and r_i is the absolute return of bond i within a non specified time interval. The sum of both weights w_i and v_i sum up to 1. It is assumed that there are no bonds invested in outside the benchmark index. Combining the weight component and the return component for one single bond, nine scenarios are distinguished influencing Excess Return (Table 4.1).

	$r_i < 0$	$r_i = 0$	$r_i > 0$
$w_i < v_i$	+	o	-
$w_i = v_i$	o	o	o
$w_i > v_i$	-	o	+

Table 4.1: Funds excess return performance for different weight and return scenario's of an individual bond: "+" stating a positive excess return, "o" no excess return influence and "-" a negative excess return.

Table 4.1 shows that it is possible to increase performance by giving more weight to securities that will have a positive return and less weight to securities that will have a negative return. Bonds with a return $r_i = 0$ do not impact Excess Return and are thereby not attractive to allocate research to. In this scenario sketch, bonds with a positive or negative absolute return are considered attractive. However, there is one problem with this approach. Fixed income securities have an asymmetric return profile. Due to the coupon return, there is a constant drift to positive absolute returns. For classification purposes it is better to translate this absolute return to a relative return (Equation 4.2).

$$ExcessReturn = \sum_{i=1}^{n}(w_i - v_i)r_i$$

$$= \sum_{i=1}^{n}[(w_i - v_i)r_i - w_i r_m + v_i r_m] \qquad (4.2)$$

$$= \sum_{i=1}^{n}(w_i - v_i)(r_i - r_m)$$

Here, r_m is the total return of the benchmark index. The absolute return component is changed to relative return and Table 4.2 shows the nine scenarios that influence performance.

	$r_i < r_m$	$r_i = r_m$	$r_i > r_m$
$w_i < v_i$	+	o	-
$w_i = v_i$	o	o	o
$w_i > v_i$	-	o	+

Table 4.2: Funds excess return performance for different weight and return scenario's of an individual bond: "+" stating a positive excess return, "o" no excess return influence and "-" a negative excess return. Here, r_m refers to the total return of the benchmark index.

We can state that fund performance is not only determined by the long (overweight) positions but also by the short (underweight) positions. To optimally deploy the value of analysts in the investment process it would be best for them to not focus on the neutral return positions. Therefore, the output of the negative screening should exclude as much as possible future under and outperforming positions. This can also be formulated as minimizing error of omission, also called Type II error: "failing to assert what is present, a miss". Table 2.2 introduces the hypothesis and shows how a Type II error occurs when testing it. Figure 4.1 shows which bonds would have been good candidates for a negative screen from 2013 to 2019.

H_1 : *Bond is an attractive opportunity at a given time.*

	H_1 True	H_1 False
Positive screen	True positive	Type I error
Negative screen	**Type II error**	True negative

Table 4.3: Hypothesis testing: its two correct inferences (diagonal cells) and its two errors.

The first step of performance measuring is done by testing the hypothesis for all bonds at every timestep. The output of this test is a True or False. After this step, we know for each screening datapoint if it is a correct inference, Type I error or Type II error. Accumulating these labels for the bonds at all timesteps makes it possible to calculate the error of omission rate (Equation 4.3).

$$\epsilon = \frac{a_{12}}{a_{11} + a_{12}} \qquad (4.3)$$

where: $\epsilon =$ Error of omission rate,

$a_{11} =$ Amount of issues labelled "True positive",

$a_{12} =$ Amount of issues labelled "Type II error".

The objective is to minimize this error of omission rate. However, we have to make one remark. If the whole universe is screened positive then $\epsilon = 0$, and there is no error of omission. A similar situation occurs when all datapoints reject the hypothesis. These situations are not desirable and there should be two additional constrains preventing them (Equation 4.4). Our aim is to screen out 15 to 25%, so the amount of bonds that are negative screened should be at least 15% (Constraint 1). Because we want to be able to minimize Type II error, there has to be more data points where H_1 is true than data points labeled as negative screened. For this reason, we decide that at least 30% have to be labelled as true

according to our hypothesis (Constraint 2).

$$\underset{\epsilon}{\text{minimize}} \quad \epsilon = \frac{a_{12}}{a_{11} + a_{12}}$$

$$\text{subject to} \quad \frac{a_{11} + a_{12}}{a_{21} + a_{22}} \geq 15\%, \tag{4.4}$$

$$\frac{a_{11} + a_{21}}{a_{12} + a_{22}} \geq 30\%$$

4.2 Model 1: hypothesis testing by time frame

In the introduction it stated that bonds are going to be screened on a daily basis. Therefore, it is necessary to measure performance on a daily basis. So the hypothesis "Bond is an attractive opportunity at a given time" will be tested every day for each issue within the benchmark. This approach assumes that we are able to determine a bonds attractiveness on only one day of data. However, when looking at historical performance it is not common to look at individual days. The performance of a security is analyzed within a time frame. For this reason, we separate the time series into multiple time frames. This is done as follows:

First, the length of the time frame has to be chosen. We use 6 months, because this is (approximate) the investment horizon of the EMD team. However, a sensitivity analysis with parameters 3, 6 and 12 months learns us that the time frame length does not influence results. The time frame is shifted from 2013 to the end of 2019. This goes in steps of 2 months. For every time frame the following is calculated for every bond i:

$$\gamma_i = \frac{max(r_i) - min(r_i)}{\beta(rating_i)} \tag{4.5}$$

Where $max(r_i)$ and $min(r_i)$ refers to the maximum and minimum cumulative return within the timeframe. The higher gamma is, the more likely it is to be screened positive. The upper part of the fraction is simple: within a time frame we are looking for the maximum possible trading opportunity. There is no difference here between overperforming bonds or underperforming bonds. In the lower part of the fraction, we are scaling this maximum trading opportunity with the beta of the bonds rating. For example $\beta(rating_i = BB)$ is the beta that corresponds with the combined returns of all BB rated bonds in the benchmark, calculated for the full timespan of the dataset. This is done because we want to adjust for risk. For example, a bond with an A rating is more attractive to research than a BB rated bond, given an evenly large trading opportunity.

Figure 4.1: Visualization of the 6 month timeframe, max(r) and min(r) in one bonds price graph

For every time frame we are able to calculate a gamma for each bond. In the next step we define if a bond is attractive or not. For every time frame the bonds are ranked based on their gamma. In Section 3.1 it is said that at least 30% of the bonds has to be screened negative. For this reason, we define the 30% bonds with the lowest gamma as not attractive. So now we know for every time step which bonds have a true hypothesis and which bonds have a false hypothesis. In figure 4.2 this is visualized for one example security.

4.3 Model 2: subportfolio performance

By applying a screening method, the credit universe is separated into two parts: the negative screened universe and positive screened universe. We expect from a well performing screening method that the positive universe consists of bonds that are likely to overperform or underperform the benchmark. In a negative screened universe we expect bonds that will perform equal to the benchmark. For the purpose of good portfolio management, it is necessary to see if dividing the universe in two parts has

Figure 4.2: Hypothesis testing visualized in a cumulative return graph for one bond. In the green marked area, the hypothesis is true, so the bond is flagged as attractive. In the white parts, the bond is flagged as unattractive.

side effects. Examples of possible consequences are a differences in total return, beta or duration in comparison to the benchmark. For this reason, we will develop two subportfolios: one for the negative screened bonds and one for the positive screened bonds.

Three steps are necessary to calculate return of each subportfolio:

1. At the end of each trading day, our screening method provides us with ah updated list of bonds screened positive and a list of bonds screened negative. Each subportfolio is then rebalanced in comparison to yesterdays subportfolio, with the bonds screened positive and the bonds screened negative correctly allocated.

2. Every bond in a benchmark has a weight which is equal to the bond's market cap divided by the total market cap of the benchmark. The weights of all bonds combined sums up to 1. When making two subportfolios, the weights of bonds in each individual subportfolio do not add up to 1 anymore. To correctly calculate performance, this has to be rebalanced for each subportfolio for every bond on a daily basis.

3. Daily returns of every bond in one subportfolio are multiplied by the weight. The sum of all individual (weighted) returns is the total return of the subportfolio. These daily returns of each subportfolio are accumulated for the full timespan of the dataset.

Chapter 5

Screening methods

In this chapter, we first determine what criteria a screening method should satisfy. In Section 5.2 a list of possible methods for screening is given. In Section 5.3, each method is given a score on each criterion and thereafter we rank each method on both simplicity and effectiveness. By doing this, we gain knowledge on what screening methods are best to implement. In Section 5.4, the implementation of each screening method is shortly explained, except the value factor model, which needs some more explanation.

5.1   Method criteria

The desired output of the screening is now defined and that makes it possible to determine what criteria a screening method should satisfy. The criteria are separated into two categories: effectiveness and simplicity. Effectiveness refers to the ability of a screening method to have a good performance now and in the future. Simplicity tells us about the complexity of one method and its ability to reach the research objective. Criteria 1, 2 and 3 are linked to effectiveness and Criteria 4, 5 and 6 are linked to simplicity:

1. Expected success rate.

   What is the expected ability of the method to minimize Type II error? The method should be able to make a clear distinction between unattractive (neutral) issues and under or over performing issues.

2. Applicability on universe.

   Are we able to screen the all credit issues or only one part of them? For example: suppose we developed a screening that does not work for callable bonds. If the universe consists of 40% callable bonds the screening is only applicable to 60% of the credit universe.

3. Future proof.

   For a screening method to be future proof, the input data should be available at all time and the output of the screening should be relevant in all future market circumstances. For example, screening based on investors behaviour could work when the markets are low volatile. However, in times of recession the same screening could perform a lot worse.

4. Easy to interpret.

   Confidence of analysts and portfolio managers in the screening model is gained when the results are easy to interpret.

5. Easy to implement.

   Easy implementation results in a less time consuming process. Because results can be obtained in short time, an early assessment of the screening model can be done. Another advantage occurs in later stages, where a simpler screening method is easier to adjust.

6. Data availability & quality.

   Data are crucial for the screening method. When data are not easy to obtain it is more time consumable, more expensive or simply not possible to carry out the screening. If the quality of the data is low, the outcome of the screening is not reliable.

5.2   Possible methods

1. Value factor analysis.

   Factor investing is an investment strategy in which securities are chosen based on certain characteristics with the goal of achieving a given investment outcome. The value factor, or value premium, refers to the tendency for mean-reversion in valuations, with 'cheap' stocks outperforming 'expensive' stocks in the long run. The premium has been observed across many different markets, regions and sample periods In corporate bonds, there has not been nearly such a widely accepted definition of value as in equities. By defining a value factor, we can positive screen the biggest outliers from fair value because we expect these bonds to over or under perform.

2. Multi-Factor analysis.

   Multi factor analysis is an extension of value factor analysis. Where the value factor takes into account company specific factors, like fundamental data, multi factor takes into account other factors like ESG (Environmental, Sustainable and Governance) scores, liquidity and momentum. These factors do not relate to the value of a company, but can add value to the security itself.

3. Supply-chain analysis.

   Macroeconomic developments are an important factor for bond prices. Changes in this macroeconomic environment could influence whole sectors or individual companies. For example, in the automotive industry aluminium is one of the main input materials. When prices of aluminium increase or decrease, the income statement of automotive companies change. This causes the probability of default to change and this impacts the bonds value. This is not only applicable to commodity price fluctuations, but also for other scenarios like elections and changes in regulations. During this research, a fin-tech company which is specialized in machine learning applications was interested to implement this. The output of their model could be used as an input for this research by showing whether it has been able to identify under and overvalued bonds.

4. Liquidity.

   An asset's market liquidity describes an asset's ability to sell quickly without having to reduce its price to a significant degree. Because illiquid bonds are less researched than liquid bonds, we except them to be less efficiently priced. Therefore illiquid bonds are a good candidate for a positive screen.

5. Sentiment.

   Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material (towards Data science). For corporate bonds, every type of information provided by the media could be an indicator of future performance. A model could be set up to process this information for making a screening.

6. Spread.

   As we have seen in Section 2.1, a higher Z-spread corresponds with a higher default risk. This higher risk is compensated with a higher expected (coupon) return. Bonds with a relatively high spread that don't default are probably going to over perform the benchmark. However, if these bonds default, they are strongly under performing the benchmark. Therefore, high spread bonds are a good candidate for a positive screen.

7. Volatility.

   When the price return of a bond is relatively volatile, the price of the bond has more upside and downside potential than a non volatile bond. Simply said, high volatility bonds are more likely to out or under perform on the short run. For this reason we except higher volatile bonds as a good candidate for a positive screen.

5.3   Method assessment

It is possible to make an assessment by combining the criteria and the alternative solutions. Every screening method described in Section 5.2 receives a score for each criterion in Section 5.1. The weight of each criterion and the arguments behind every score are provided in Appendix A. The result can be seen in Table 5.1. For every method, we define an effectiveness score and a simplicity score based on the individual criteria scores. In Figure 5.1 this is visualized for every screening method.

	Effectiveness			Simplicity		
	Expected success rate	Applicability on universe	future proof	easy to interpret	easy to implement	Data availability & quality
Value Factor	+	+	+	-	+	++
Multi-Factor	++	+	+	- -	-	+
Supply-chain	+	-	+	-	-	o
Liquidity	o	++	+	+	+	+
Sentiment	o	-	+	-	- -	-
Spread	+	++	++	++	++	+
Volatility	+	++	++	++	++	++

Table 5.1: Criteria scores for each screening method. Explanation is given in Appendix A.
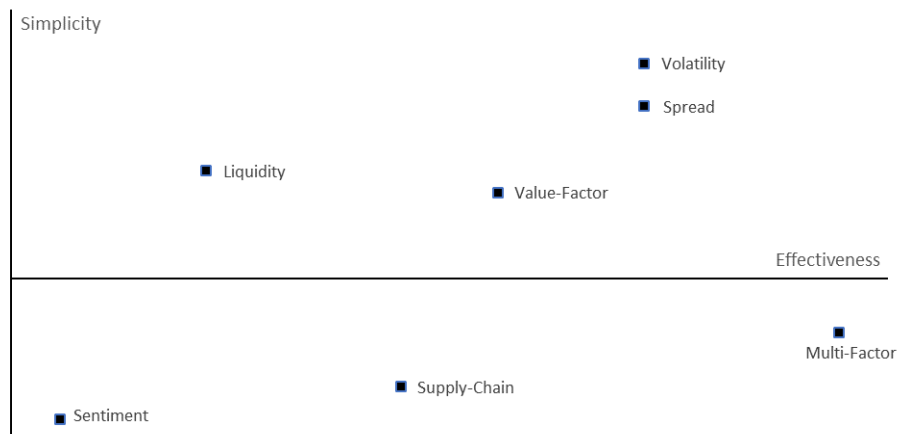
Figure 5.1: Visualization of the scoring of each screening method on both simplicity and effectiveness criteria.

A screening method scores best if it scores high on simplicity and effectiveness (the upper-right part of Figure 5.1). Screening on volatility, spread and value-factor score best on both simplicity and effectiveness and therefore they are chosen to be implemented. Also liquidity is implemented, because this was easy achievable. A Multi-Factor method scores high on effectiveness but low on simplicity. Due to time constrains this method has not been implemented.

5.4 Implementation of screening methods

The chosen screening methods from Section 5.3 are now further explained. For all methods, we will use screening based on each day's quantiles. Quantiles are used, because it makes the screening consistent and future proof. For example, if we decide to negative screen the lowest 30% quantile of spread, it is sure that every day exactly 30% is screened negative.

5.4.1 Spread

When screening on spread, the lowest values are classified as negative. Our hypothesis is that a lower spread corresponds with low under or over performance of the benchmark. Bonds with embedded options are not included, because there is no reliably OAS data available for the CEMBI benchmark.

5.4.2 Volatility of daily returns

When screening on volatility, the lowest values are classified as negative. Our hypothesis is that lower volatility corresponds with low under or over performance of the benchmark.
Daily bond returns are calculated as described in Chapter 2.1. Currency return is not taken into account, because the used currency is dollars for all bonds in CEMBI. Volatility is measures over the past 21 trading days. This is approximately one month of data.

5.4.3 Liquidity

Highly liquid bonds are classified as negative. Our hypothesis is that higher liquidity values correspond with low under or over performance of the benchmark. Liquidity is proxied by the issue size of the bond. Other liquidity indicator data was unfortunately not available. However, issue size is an import liquidity factor. It is important to notice that issue size does not change over time. Therefore, liquidity is the only screening which is static.

5.4.4 Value factor

When screening on value factor, the values that are close to fair value are classified as negative. Our hypothesis is that bonds which are valued far from fair value will under or over perform the benchmark. The most important step in defining fair value is adjusting Option Adjusted Spread (OAS) for spread duration.

The term structure, as decribed in Section 2.2, is also visible in corporate bonds spreads. For investment grade bonds, a higher spread duration corresponds with a higher spread level. In contrast to the spread screen, we will use bonds with embedded options here. If we would not do this, there will not be enough data to determine a reliable term structure. As we will see in the methodology, OAS datapoints are averaged out and are therefore less harm full.

Within each higher level rating category (A, BBB and BB) we are going to make an OAS curve based on spread duration. This is necessary, because it is believed that the spread duration effect is different within each risk category. The OAS curve is estimated with the use of the Nelson-Siegel model, which is described in Section 2.2. However, a simpler version is used here where the curvature effect β_2 is not taken in to account. The curvature effect explains only 3.6% of a US Treasury bonds portfolio (Ibanez, 2015). For this application this small effect is negligible.

Figure 5.2: Points are not evenly distributed, and least squares method does not fit a fair curve.

Figure 5.3: Very rare, but sometimes at lower rated bonds, dispersion is so big that it does not result in a concave curve.

$$y(\tau) = \beta_0 + \beta_1(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau})$$

$$where:$$
$$y = OAS$$
$$\tau = spread\ duration$$

(5.1)

$y(\tau)$ is estimated every day for each main rating category (Equation 5.1). Normally, the input for this equation would be all bond data within each rating category on a given day. However, this has some negative effects. The points are not evenly distributed (Figure 5.2) and the curve is not estimated right across the whole range of spread duration values. Sometimes at lower rated bonds, dispersion is so big (Figure 5.3) that it does not result in a concave curve. To make calibration easier we aim for spread duration buckets. The boundaries of the buckets are chosen, so that each bucket has approximately the same amount of data points. The averages of all data points within the buckets are used as new input values for the Nelson-Siegel model. By doing this, the two negative side effects are partly tackled. Besides, computation time and calibration of parameters in the algorithm used will be easier.

Figure 5.4: Each blue dot corresponds with an A-rated bond on July 29th, 2014. Spread duration is separated into buckets, where the averages of the data points within the buckets are used as new variables.

Equation 5.1 is a nonlinear equation because of the Lambda variable. To solve for β_0, β_1 and λ we will use a solver for non-linear optimization problems. In python, there are a few solver algorithms available. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is suited for this problem. With this algorithm, we can use boundaries for our parameter. Equation 5.2 shows the minimization function and its boundaries.

$$Min.\quad fit\ error(\beta_0, \beta_1, \lambda, \tau) = \sqrt{\sum_n (y(\tau) - y(\beta_0.\beta_1, \lambda, \tau))^2}$$

(5.2)

$$where:$$

$$0 < \beta_0 < 1000$$
$$-1500 < \beta_1 < 0$$
$$0 < \lambda < 1$$

The BFGS algorithm minimizes the sum of the least squares by adjusting the three parameters. Now we are able to fit a daily curve, we are able to adjust OAS for all bonds within a rating category (A, BBB, BB). Therefore, we have to take a base spread duration. We will choose the base as 5, because this is approximately the average of all spread duration data, but it should not matter which value is chosen. Next, all bonds oas are adjusted as if the bonds were having a spread duration of 5 (Equation 5.3).

$$OAS_{adjusted} = \frac{y(\beta_0, \beta_1, \lambda, \tau_{base} = 5)}{y(\beta_0, \beta_1, \lambda, \tau)} * OAS \tag{5.3}$$

Now all OAS are adjusted we can define the fair value. Within the detailed SNP ratings (A-,BBB+, etc.) the average is taken daily. We define this average as fair value.

5.4.5 Combinations

Besides screening on an individual characteristic, we could also screen on a combination of characteristics. This is useful if the combination of screenings works better than the individual screenings. There are two ways this can be achieved:

- Using an OR statement
  The rule at an OR statement is, that when a bond is screened positive on either one screening **or** another screening, the bond is screened positive.

- Using an AND statement
  The rule at and AND statement is, that when a bond both is screened positive on one screening **and** another screening, the bond is screened positive.

Chapter 6

Results

in this chapter, the chosen screening methods from Chapter 5 are combined with the performance measurements models from Chapter 4. Section 6.1 shows the results of Model 1 and Section 6.2 shows the results for Model 2. Finally, in Section 6.3 a proposal is given on what screening method to use. The results are generated by programming in Python. Figure 6.1 shows how the software architecture looks like, going from raw data to actual results.
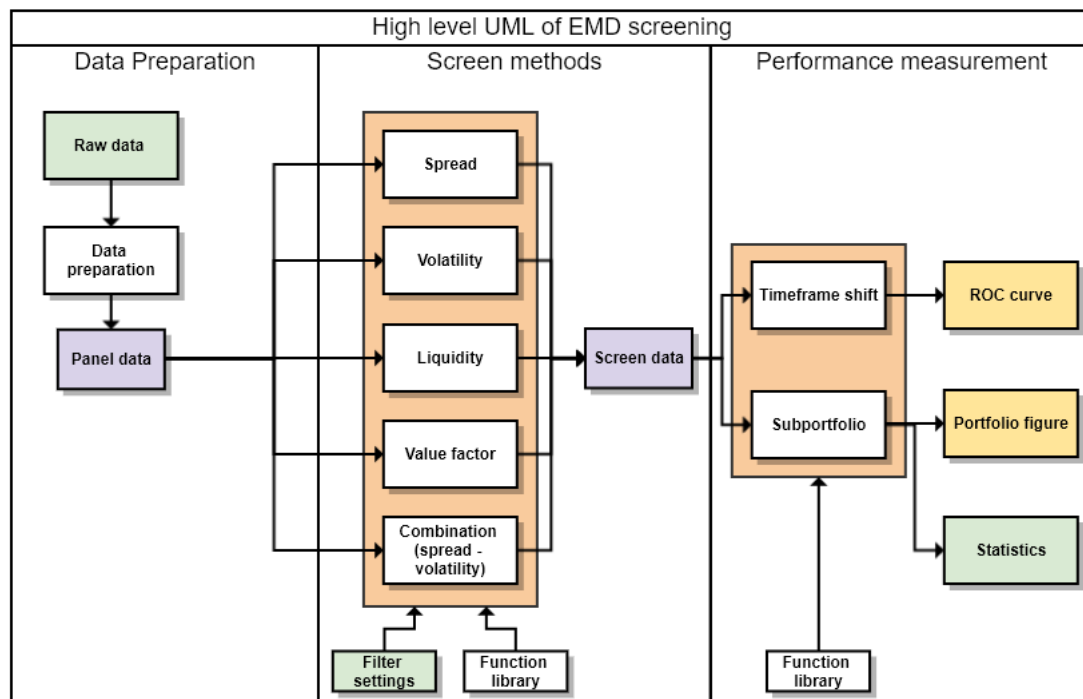
Figure 6.1: UML software architecture of EMD screening. The green rectangles represent excel/csv files, the white rectangles python scripts, the purple rectangles python data pickle files and the yellow rectangles are figures.

6.1   Model 1: Hypothesis testing

In Model 1 we test if our chosen screening method are effectively classifying under and over performing bonds in time. The screening method labels every bond and every date as either positive or negative. For every data point it is checked if the hypothesis "bond is an attractive opportunity" is true. Using this information, the True Positive Rate (TPR) and False Positive Rate (FPR) are calculated. The threshold is given by the quantile that is screened negative. These threshold (quantiles) are lowered in steps of 10% and the TPR and FPR values of the remaining thresholds are interpolated. By doing this, a trade off occurs between the TPR and FPR. This is visualized by the Receiver Operating Characteristic (ROC) curve. The better a screening method is able to minimize FPR and maximize TPR, the better the screening method is performing. Figure 6.2 visualizes this for every screening method.
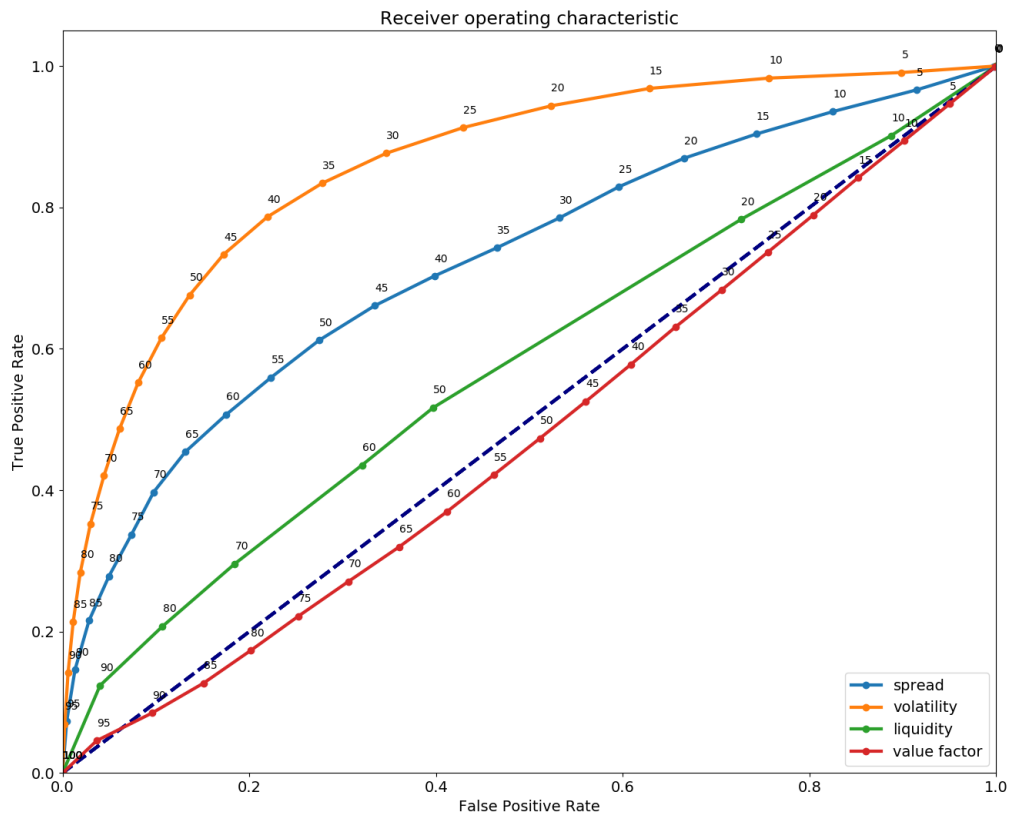
Figure 6.2: In this figure an ROC curve is showed for each screening method. The labels indicate what quantile is screened negative. For example, if a blue dot is labeled as 25, the lowest 25% spreads on each day are screened negative.

The blue dotted line in Figure 6.2 shows how a random screen would perform. If we would randomly assign negative or positive labels to our data points the screening would approximately follow the blue dotted line. An important result of this performance analysis is that volatility is the best classifier, then volatility, then liquidity and as last value factor.

The screening methods that are not performing so well are liquidity and value factor. Liquidity is barely performing better than a random screen would. The hypothesis that low liquidity is able to classify attractive opportunities is therefore false. The value factor screening is performing slightly worse than a random screen would. Our hypothesis that bonds that are far from fair value will over perform or under perform is also false. There could be two explanations for this. It could be that when a bond is over or undervalued according to our model. it will remain in that state over the lifetime of the bond. Another explanation lies in the ability of our value factor model to estimate fair value.

The screening methods that are performing well are volatility and spread. The best classifier is volatility. To see whether we can optimize this screening method we will combine volatility and spread screening by using an OR and an AND statement (Section 5.4.5). The statistics of all combinations are in Appendix B.

The ROC curve is good in showing what screening methods classify best. However, our objective is not to lose sight of attractive opportunities. Therefore, minimizing error of omission (Type II) is more important than minimizing error of commission (Type I). Besides, we aim to reduce the universe by 15 to 25%. To take these two priorities into account a better figure is created (Figure 6.3). The x-axis shows the negative screen ratio and the y-axis shows the TPR. A high TPR, close to 1, corresponds with a low error of omission. The figure includes the best performing screening combinations.
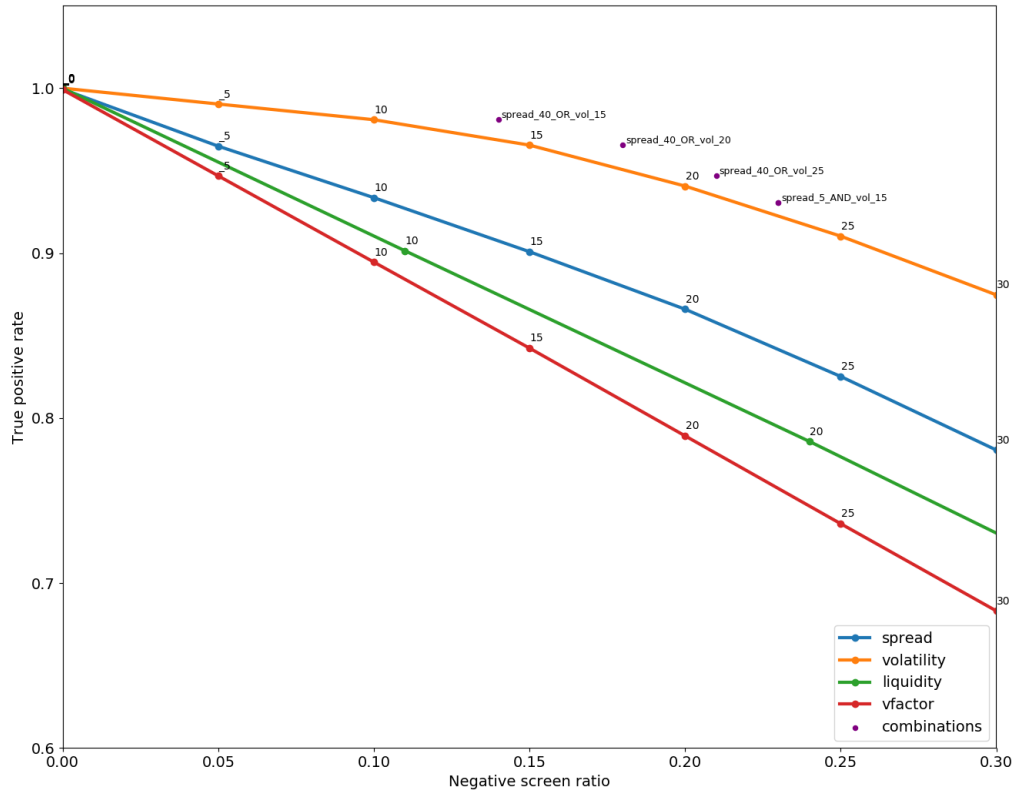
Figure 6.3: Another representation of Figure 6.2, with the best performing screening combinations included. A higher y-axis position of a screen corresponds with a lower Type II error.

It is shown that a combination screening performs better on the TPR ratio than the individual screening methods. The combinations that perform best are described below.

- Reduce the universe with 18%, obtaining a Type II error of 3% by:
  positive screening if spread is higher than the 40th quantile or if volatility is higher than the 20th quantile on a day.

- Reduce the universe with 21%, obtaining a Type II error of 5% by:
  positive screening if spread is higher than the 40th quantile or if volatility is higher than the 25th quantile on a day.

- Reduce the universe with 23%, obtaining a Type II error of 6% by:
  positive screening if spread is higher than the 5th quantile and if volatility is higher than the 15th quantile

	neg. rate	TP	FP	TN	FN	TPR	FPR	accuracy	precision
1. spread	0.20	0.54	0.25	0.13	0.08	0.87	0.67	0.67	0.68
2. volatility	0.20	0.66	0.16	0.14	0.04	0.94	0.52	0.80	0.81
3. spread_5_AND_vol_15	0.23	0.58	0.21	0.17	0.04	0.94	0.55	0.75	0.74
4. spread_40_OR_vol_20	0.18	0.60	0.23	0.15	0.02	0.97	0.61	0.75	0.72
5. spread_40_OR_vol_25	0.21	0.59	0.21	0.17	0.03	0.95	0.55	0.76	0.74

Table 6.1: Statistics for screening on spread, volatility and its best performing combinations.

Table 6.1 shows that Methods 4 and 5 score best on True Positive Rate. Method 4 has a higher TPR (97%) than Method 5 (95%), but this is offset by a lower negative screen rate. A higher negative screen rate, in combination with the higher accuracy and precision score, makes Method 5 slightly more preferable than Method 4. This combination is considered as best screening method and is further analysed in Section 6.2.

6.2 Model 2: Subportfolio returns

Model 2 shows how the two new universes, screened negative and screened positive, perform on total return in comparison to the benchmark. Total return is not of interest in this research. However, it is important to know whether the remaining universe is biased to one or more performance metrices. In Section 6.1 we concluded that liquidity and value factor are not able to classify the attractive opportunities. Subportfolios of both are located in Appendix B. A noticeable result for liquidity, is that bonds with a small issue size outperform the bonds with a higher issue size. This corresponds with Chapter 2.1, which stated that a liquidity premium exists in fixed income.

Another result that is noticeable is an altered version of the value factor screening. The former value factor screen classified bonds on how far they are valued to fair value. A different approach is to positive screen bonds which are undervalued and negative screen bonds that are overvalued. Model 2 shows that for each rating category this method of screening performs very well on Information Ratio (IR). These figures are shown in Appendix B.

In this section subportfolio figures are shown for spread, volatility and the combination of Section 6.1 (Figures 6.4, 6.5 and 6.6). The statistics of each subportfolio are shown in Table 6.1. Noticeable here is the higher beta of the remaining positive universe. This is a negative effect, because higher beta corresponds with taking more risk. However, the geometric average return of the positive screened portfolio is higher than the negative screened portfolio. This means that reducing the universe by volatility and spread, results in a portfolio of bonds that have made higher total returns from 2013 tot 2019.

| | **Positive screen** | | | | | **Negative screen** | | | | |
	geom_avg	stdev	beta	IR	duration	geom_avg	stdev	beta	IR	duration
spread	0.05	0.03	1.14	1.62	4.77	0.01	0.02	0.38	-0.96	3.34
volatility	0.04	0.03	1.17	0.04	5.75	0.03	0.01	0.19	-0.19	3.00
combination	0.05	0.03	1.06	0.69	4.94	0.02	0.01	0.18	-0.44	2.43
benchmark	0.04	0.03	1	0	4.66					

Table 6.2: Statistics for the subportfolios spread, volatility and its combination. As a reference, the benchmark statistics are included in the last row. "geom-avg" refers to the geometric annual average return over the whole lifetime of the portfolio.
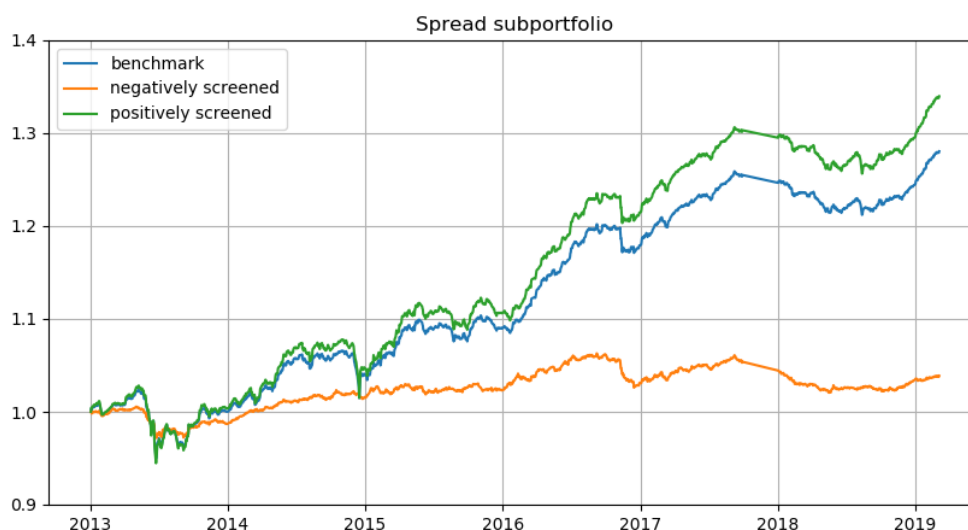
Figure 6.4: Subportfolio for spread, where on a daily basis bonds are screened negative which are located in the lowest 20th quantile of spread.
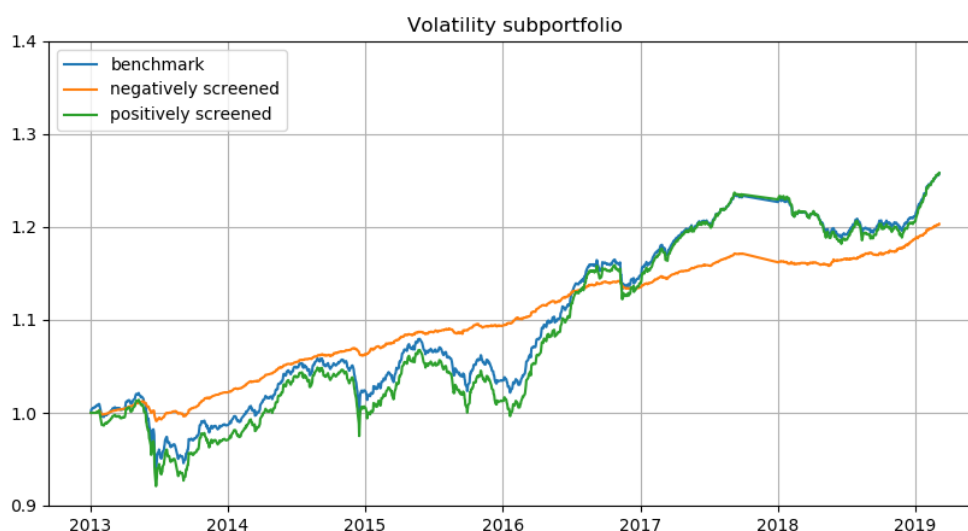
Figure 6.5: Subportfolio for volatility, where on a daily basis bonds are screened negative which are located in the lowest 20th quantile of volatility.
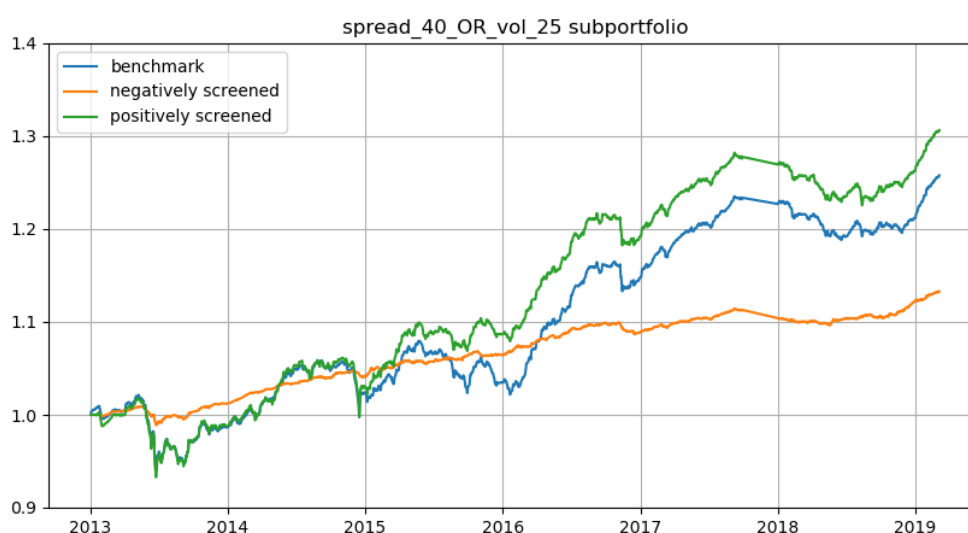
Figure 6.6: Subportfolio for a combination screening. Here, a bond is screened positive if spread is higher than the 40th quantile or if volatility is higher than the 25th quantile on a day.

6.3 Proposal

The EM corporate debt universe could be reduced by screening out bonds that have a low spread and have low volatility. A reduction of 21% could be achieved by a daily negative screening of the lowest 40th quantile of spread, and the lowest 25th quantile of volatility. Executing this screening from 2013 to 2019 resulted in a True Positive Rate of 95%, which equals a Type II error rate of 5%.

The remaining universe has a higher beta than the benchmark. Unfortunately, this is a negative effect because the sum of the remaining bonds are more risky than the overall benchmark. However, the value factor model performs good at obtaining positive returns (Appendix B). According to the value factor model, bonds under fair value are overperforming bonds above fair value. Figure 6.7 shows how a subportfolio of undervalued negative screened bonds perform in comparison with overvalued bonds. A lower beta could be achieved by additional investing in undervalued bonds in the universe screened negative.
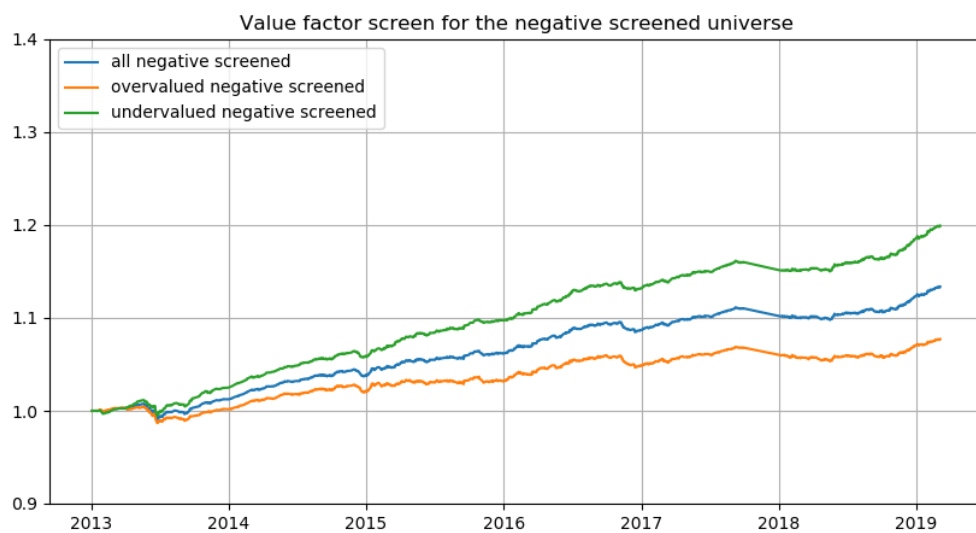
Figure 6.7: When we apply the altered value factor screening on the negative screened universe two subportfolios are distinguished.

Chapter 7

Discussion

In our review of attractiveness we have analysed that research should be allocated to securities that are likely to overperform or underperform the benchmark because they influence fund performance the most. However, it is important that the research analysts are able to overweight and underweight the right securities. If they are not able to, performance will be lower due to this screening method.

Performance Model 1 is the most important model of our research, because this makes us able to show which methods are minimizing Type II error the most. For being a crucial model it has too many parameter values which are arbitrary. The timeframe of 6 months is shifted in steps of 2 months. It would be better to perform a sensitivity analysis based on various time steps. However, a sensitivity analysis is very hard to execute here. Because we define what is attractive or not, a time step which minimizes Type II error the most is not necessarily the best. The best we could do is show that the calculated Type II error does not change much by adjusting this time step. Moreover, in Model 1 gamma is calculated by the maximum trading opportunity which is adjusted for risk. As a proxy for this risk we use the beta of the rating of the bond. We are not sure if the amount of correction here is right. To develop a better proxy, our research should have included an analysis on the trade off between risk and return in emerging market debt.

Implementing a value factor model was a lot more time consuming and difficult than estimated in the MCDA analysis. Defining value by credit rating and spread duration is the best that could be achieved in this research. It would be interesting to take more important factors into account, like sector, liquidity and sovereign information. On the other hand, it seems that this simple model is able to identify undervalued and overvalued bonds in the investment grade range. For further research, the value factor model could be improved by adding more factors that influence the value of a security. This would result in a more accurate definition of fair value.

To see which screening method performs best on Model 1, we used a trial and error approach. Each screening method functions as input for Model 1 and Model 1 outputs a result. This is a simple method, but not the most effective. Because we are working with data, a traditional machine learning approach could be used. By doing this, a test set and a training set could be created where the goal is to minimize Type II error. Techniques like random forest or decision trees could be used here. The effectiveness of this approach would be a lot higher, but it was too complex to implement for the time of this research.

The introduction states that the screening methodology should be general enough to be applicable to all types of corporate credits. However, based on this research we do not know for sure whether this is true. Screening based on spread and volatility is more likely to work in the high yield universe than in the investment grade universe due to the bigger range of spread and volatility in high yield. However, this first has to be tested by using the data of different high yield benchmarks.

Another limit of this research is the time interval of the data. Only data from 2013 till now could be obtained. Unfortunately for this research, 2013 to 2019 were times of economic growth. Adding data from the economic crisis in 2008 would be useful. This way we could test whether our screening methods perform good in another economic cycle. However, you could question whether any systematic screening method is reliable during an economic crisis.

Chapter 8

Conclusion & Recommendations

The objective of this research was to develop a data-driven, systematic screening model to reduce the corporate credit universe, without losing sight of attractive opportunities. To achieve this objective, three research questions have been answered.

Research question 1. By decomposing Excess Return (Equation 4.2) we showed that a fund can increase its excess return using underperforming bonds. For this reason, attractive opportunities are not only bonds likely to overperform, but also bonds likely to underperform the benchmark. Allocation of research to these two types of bonds makes research analysts able to focus on securities which influence fund performance the most.

Research question 2. In order to measure performance of screening we had to determine which corporate bonds in emerging markets were attractive at which points in time (Model 1). For this, we used panel data from 2013 to 2019 of bonds in the "J.P. Morgan CEMBI Broad" benchmark. Using price return data, we calculated for each bond the maximum risk adjusted trading opportunity in a timeframe of six months. While moving this timeframe from 2013 to 2019, we indicate which bonds has been attractive. Each day, the 70% bonds with the highest risk adjusted trading opportunities are indicated as attractive and the other 30% are indicated as unattractive. Model 2 shows how the screening performs in total return context. This increases knowledge about attribute bias in the positive screened universe.

Research question 3. To see which screening method works best, a Multi Criteria Decision Analysis (MCDA) has been executed. Screening based on spread, volatility, liquidity and value factor scores best according to this analysis. These methods of screening are implemented together with combinations of the methods. It turns out that screening based on a combination of volatility and spread performs best. A reduction of 21% could be achieved by a daily negative screening of the lowest 40th quantile of spread, and the lowest 25th quantile of volatility. Executing this screening from 2013 to 2019 resulted in a True Positive Rate of 95%, which equals a Type II error rate of 5%. One side effect of this screening method is a higher beta than the former universe. However, a lower beta could be achieved by additional investing in undervalued bonds in the negative screened universe.

For further research, the value factor model could be improved by adding more factors that influence the value of a security. This would result in a more accurate definition of fair value. Furthermore, higher accuracy can be obtained by using a larger data set which reaches before 2013. By doing this, times of recession are taken into account which are not present now. Last, a machine learning approach could be implemented where we dynamically adjust our screening method algorithm based on new market information. This would help to make the systematic screening model more future proof.

References

Brealey, Myers, and Allen (2014). *Principles of Corporate Finance*. McGraw-Hill Education, New York.

Diebold, F. X. and Li, C. (2003). Forecasting the term structure of government bond yields. CFS Working Paper Series 2004/09, Center for Financial Studies (CFS).

Driessen, J. (2005). Is Default Event Risk Priced in Corporate Bonds? *Review of Financial Studies*, 18(1):165–195.

Elton, E. J., Gruber, M. J., Agrawal, D., and Mann, C. (1999). Explaining the Rate Spread on Corporate Bonds. Finance Department Working Paper Series 99-082, New York University, Leonard N. Stern School of Business.

Fabozzi, F. J. (2005). *The Handbook of Fixed Income Securities*. McGraw-Hill Education, New York.

Grkaynak, R. S., Sack, B. P., and Wright, J. H. (2006). The U.S. Treasury yield curve: 1961 to the present. Finance and Economics Discussion Series 2006-28, Board of Governors of the Federal Reserve System (US).

Ibanez, F. (2015). Calibrating the Dynamic Nelson-Siegel Model: A Practitioner Approach. MPRA Paper 68377, University Library of Munich, Germany.

Johnson, D. and Qi, H. (2015). Explaining bond spreads via default risk, taxes, rating transition, and liquidity. *Financial Decisions*, 27(1).

Marian, J. (2016). Credit ratings by country in europe. `http://web.archive.org/web/20180308043053/https://jakubmarian.com/credit-ratings-by-country-in-europe/`. Accessed: 2019-06-30.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curve. *Journal of Business*, 6(4):473–489.

Petitt, Pinto, and Pirie (2015). *Fixed Income Analysis*. John Wiley & Sons Inc., New Jersey.

Appendices

Appendix A

MCDA Analysis

A.1 Criteria weights

Effectiveness criteria	Weights
Expected succes rate	0,6
Applicability on universe	0,2
Future proof	0,2

The objective of this research is to reduce the universe without losing sight of attractive opportunities. Reducing error of omission (Type II error) is therefore key priority. For this reason, expected success rate is the most important criterion. Applicability on universe and future proof are both equally important.

Simplicity criteria	Weights
Easy interpretable	0,3
Easy implementable	0,3
Data availability & quality	0,4

All simplicity criteria are roughly of similar importance. However, data availability & quality is more important in the short term. Data has to be available and reliable before we are able to implement it. Because a screening entirely depends on its date, this criterion is given more weight.

A.2 Explanation of scoring method

Symbol (score)	Numerical translation
++	2
+	1
o	0
-	-1
- -	-2

A scoring method by symbol is chosen because it quickly visualizes what a screening methods strong and weak points are. The numerical translation is done by punishing a screening method if it scores negative on one criterion and rewarding it if it scores positive. In the scoring of alternatives (Section A.3) it is important to take into account that a '++' score is rewarded twice as good as a '+' score. The same applies for the relation between '- -' and '-' scores.

A.3 Alternative scores

Value factor	Score	Explanation
Expected success rate	+	If we are able to make a method that flags undervaluation or overvaluation of bonds attractive opportunities can be correctly identified. One weakness is that the value factor does not take all factors into account that describes the price of a bond.
Applicability on universe	+	It is expected that it is applicable to the whole universe. However, because we are not sure what (fundamental) factors are taken into account, we are not fully sure.
future proof	+	If a value factor method has worked good from 2013, there are reasons to assume this will work in the future. However, changing economic environments could be a risk here.
Easy to interpret	-	Because different factors are taken into account and regression models are used here, results are not easy to interpret.
Easy to implement	+	Using prediction statistics, it is good implementable. However, fixed effects and collinearity of bond data are difficult to tackle.
Data availability & quality	++	Fundamental data of companies is expected to be easily available and are expected of good quality.

Multi-factor	Score	Explanation
Expected success rate	++	If we are able to make a method that flags undervaluation or overvaluation of bonds attractive opportunities can be correctly identified. The expected success rate would be very high, because most factors describing the value of a bond are taken into account.
Applicability on universe	+	It is expected that it is applicable to the whole universe. However, because we are not sure what factors are taken into account, we are not fully sure.
Future proof	+	If a multi-factor method has worked good from 2013, there are reasons to assume this will work in the future. However, changing economic environments could be a risk here.
Easy to interpret	- -	A lot of different factors are taken into account here so the model could act like a black box.
Easy to implement	-	Prediction statistics is used here and this is common knowledge. However, a lot of factors are taken into account and fixed effects and collinearity of bond data are difficult to tackle.
Data availability & quality	+	Fundamental data is available. ESG scores and liquidity data are more difficult to obtain.

Supply-Chain	Score	Explanation
Expected success rate	+	A supply chain analysis method using machine learning is expected to flag attractive opportunities.
Applicability on universe	- -	This method does not cover the whole credit universe. When it is implemented, it will first only cover the issues with a large market size.
Future proof	+	It is expected that a good supply chain analysis method will work in future economic cycles. One risk occurs with fixed income markets becoming more liquid. Events are more quickly represented in the credit spreads.
Easy to interpret	-	Results are not easy to interpret because the model acts like a black box. However, an analyst could analyse itself how one event could influence the supply chain and check if this corresponds with the model.
Easy to implement	-	An external company would implement this model. Although it is a very complex problem, they already have experience with similar problems.
Data availability & quality	-	Supply chain data is available. However, different sources has to be used which increases complexity here.

Liquidity	Score	Explanation
Expected success rate	o	Illiquid bonds are less researched and less efficiently priced. Therefore, we expect less movements in the price of a bond. However, when an event occurs we expected bigger movements in the price of these bonds. We are not sure if this method is able to identify attractive opportunities
Applicability on universe	+ +	Every bond has a liquidity measure so it is applicable to the whole universe.
Future proof	+	It is future proof except for an economic crisis. In that situation a lot of bonds can suddenly become a lot more illiquid.
Easy to interpret	+	Liquidity is easily understood, but there are multiple measures for it.
Easy to implement	+	It is easy to implement. Only when multiple liquidity measures are combined, it could get more difficult.
Data availability & quality	+	Liquidity data is expected to be available. We are not sure of its quality.

Sentiment	Score	Explanation
Expected success rate	o	It is proven that in equity markets, sentiment plays a role. It is not certain that sentiment has an impact in fixed income markets.
Applicability on universe	-	We expect that this will not work for all issues. Only the bigger ones that are researched a lot have sentiment data.
Future proof	+	Sentiment is an universal expressions, which will not change over time. Also here, economic crisis is a risk for the validation of the model.
Easy to interpret	-	Analysts understand what sentiment means. However, how sentiment is measured is not always easy to describe and understood.
Easy to implement	- -	Although there have been a lot of development in the past years on word-processing models, it is still very difficult to implement.
Data availability & quality	-	Data should come from news sources. The data are difficult to obtain but it is not impossible. Also, there are new start-ups who collect data of all news sources. However, subscription to these services is costly.

Spread	Score	Explanation
Expected success rate	+	Bonds with a high spreads are expected to have a lot of coupon return, but carry more credit risk. Because they are more likely to overperform or under perform the benchmark the expected success rate is high.
Applicability on universe	++	All issues have a credit spread, so it is applicably to the whole universe.
Future proof	++	Although spread levels change a lot over time, the concept of spread will not change.
Easy to interpret	++	Spread is one of the fundamental metrics in fixed income.
Easy to implement	++	In general only one data field is taken into account, so it is very easy to implement.
Data availability & quality	+	Spread data is available. However, we are not sure how reliable the spread data is for bonds with an embedded option.

Volatility	Score	Explanation
Expected success rate	+	Bonds with a high volatility are expected to underperform or overperform the benchmark on the short term. Therefore, the expected success rate is high.
Applicability on universe	++	Volatility can be measured for every issue.
Future proof	++	Volatility moves along with economic cycles, which makes it future proof.
Easy to interpret	++	Volatility is a simple measure.
Easy to implement	++	By calculating daily returns, we can easily implement a volatility screen.
Data availability & quality	++	High quality data is available to calculate daily returns.

Appendix B

Model 1 statistics

B.1 Spread

	neg rate	TP	FP	TN	FN	TPR	FPR	accuracy	precision	f1 score
spread 0	0.00	0.62	0.38	0.00	0.00	1.00	1.00	0.62	0.62	0.11
spread 5	0.05	0.60	0.35	0.03	0.02	0.97	0.91	0.63	0.63	0.11
spread 10	0.10	0.58	0.31	0.07	0.04	0.94	0.83	0.65	0.65	0.11
spread 15	0.15	0.56	0.28	0.10	0.06	0.90	0.74	0.66	0.67	0.11
spread 20	0.20	0.54	0.25	0.13	0.08	0.87	0.67	0.67	0.68	0.11
spread 25	0.25	0.51	0.23	0.15	0.11	0.83	0.60	0.67	0.70	0.11
spread 30	0.30	0.49	0.20	0.18	0.13	0.79	0.53	0.66	0.71	0.11
spread 35	0.35	0.46	0.18	0.20	0.16	0.74	0.47	0.66	0.72	0.11
spread 40	0.40	0.44	0.15	0.23	0.18	0.70	0.40	0.66	0.74	0.11
spread 45	0.45	0.41	0.13	0.25	0.21	0.66	0.33	0.66	0.76	0.11
spread 50	0.50	0.38	0.10	0.27	0.24	0.61	0.27	0.65	0.79	0.11
spread 55	0.55	0.35	0.08	0.29	0.27	0.56	0.22	0.64	0.80	0.11
spread 60	0.60	0.31	0.07	0.31	0.31	0.51	0.18	0.63	0.83	0.11
spread 65	0.65	0.28	0.05	0.33	0.34	0.45	0.13	0.61	0.85	0.11
spread 70	0.70	0.25	0.04	0.34	0.37	0.40	0.10	0.59	0.87	0.10
spread 75	0.75	0.21	0.03	0.35	0.41	0.34	0.07	0.56	0.88	0.10
spread 80	0.80	0.17	0.02	0.36	0.45	0.28	0.05	0.53	0.90	0.10
spread 85	0.85	0.13	0.01	0.37	0.49	0.22	0.03	0.50	0.93	0.09
spread 90	0.90	0.09	0.01	0.37	0.53	0.15	0.01	0.46	0.95	0.08
spread 95	0.95	0.05	0.00	0.38	0.58	0.07	0.00	0.42	0.96	0.07
spread 100	1.00	0.00	0.00	0.38	0.62	0.00	0.00	0.38	0.00	0.00

B.2 Volatility

	neg rate	TP	FP	TN	FN	TPR	FPR	accuracy	precision	f1 score
vol 0	0.00	0.70	0.30	0.00	0.00	1.00	1.00	0.70	0.70	0.11
vol 5	0.05	0.69	0.27	0.03	0.01	0.99	0.90	0.72	0.72	0.11
vol 10	0.10	0.69	0.23	0.07	0.01	0.98	0.76	0.76	0.75	0.11
vol 15	0.15	0.68	0.19	0.11	0.02	0.97	0.63	0.79	0.78	0.11
vol 20	0.20	0.66	0.16	0.14	0.04	0.94	0.52	0.80	0.81	0.11
vol 25	0.25	0.64	0.13	0.17	0.06	0.91	0.43	0.81	0.83	0.11
vol 30	0.30	0.61	0.10	0.20	0.09	0.88	0.35	0.81	0.85	0.11
vol 35	0.35	0.58	0.08	0.22	0.12	0.83	0.28	0.80	0.87	0.11
vol 40	0.40	0.55	0.07	0.23	0.15	0.79	0.22	0.78	0.89	0.11
vol 45	0.45	0.51	0.05	0.25	0.19	0.73	0.17	0.76	0.91	0.11
vol 50	0.50	0.47	0.04	0.26	0.23	0.68	0.14	0.73	0.92	0.11
vol 55	0.55	0.43	0.03	0.27	0.27	0.62	0.11	0.70	0.93	0.11
vol 60	0.60	0.39	0.02	0.28	0.31	0.55	0.08	0.66	0.94	0.11
vol 65	0.65	0.34	0.02	0.28	0.36	0.49	0.06	0.62	0.95	0.11
vol 70	0.70	0.29	0.01	0.29	0.41	0.42	0.04	0.58	0.96	0.10
vol 75	0.75	0.25	0.01	0.29	0.45	0.35	0.03	0.54	0.96	0.10
vol 80	0.80	0.20	0.01	0.30	0.50	0.28	0.02	0.49	0.97	0.10
vol 85	0.85	0.15	0.00	0.30	0.55	0.21	0.01	0.45	0.98	0.09
vol 90	0.90	0.10	0.00	0.30	0.60	0.14	0.01	0.40	0.98	0.08
vol 95	0.95	0.05	0.00	0.30	0.65	0.07	0.00	0.35	0.99	0.06
vol 100	1.00	0.00	0.00	0.30	0.70	0.00	0.00	0.30	0.00	0.00

B.3 Liquidity

	neg rate	TP	FP	TN	FN	TPR	FPR	accuracy	precision	f1 score
liq 0	0.00	0.70	0.30	0.00	0.00	1.00	1.00	0.70	0.70	0.11
liq 10	0.11	0.63	0.27	0.03	0.07	0.90	0.89	0.66	0.70	0.11
liq 20	0.24	0.55	0.22	0.08	0.15	0.78	0.73	0.63	0.72	0.11
liq 50	0.53	0.36	0.12	0.18	0.34	0.52	0.40	0.54	0.75	0.11
liq 60	0.61	0.30	0.10	0.20	0.40	0.44	0.32	0.51	0.76	0.11
liq 70	0.74	0.21	0.06	0.24	0.49	0.30	0.18	0.45	0.79	0.10
liq 80	0.83	0.15	0.03	0.27	0.55	0.21	0.11	0.41	0.82	0.09
liq 90	0.90	0.09	0.01	0.29	0.61	0.12	0.04	0.37	0.88	0.08
liq 100	1.00	0.00	0.00	0.30	0.70	0.00	0.00	0.30	0.00	0.00

B.4 Value factor

	neg rate	TP	FP	TN	FN	TPR	FPR	accuracy	precision	f1 score
value factor 0	0.00	0.70	0.30	0.00	0.00	1.00	1.00	0.70	0.70	0.11
value factor 5	0.05	0.66	0.29	0.01	0.04	0.95	0.95	0.68	0.70	0.11
value factor 10	0.10	0.63	0.27	0.03	0.07	0.89	0.90	0.66	0.70	0.11
value factor 15	0.15	0.59	0.26	0.04	0.11	0.84	0.85	0.63	0.70	0.11
value factor 20	0.20	0.55	0.24	0.06	0.15	0.79	0.80	0.61	0.70	0.11
value factor 25	0.25	0.52	0.23	0.07	0.18	0.74	0.76	0.59	0.69	0.11
value factor 30	0.30	0.48	0.21	0.09	0.22	0.68	0.71	0.57	0.69	0.11
value factor 35	0.35	0.44	0.20	0.10	0.26	0.63	0.66	0.54	0.69	0.11
value factor 40	0.40	0.40	0.18	0.12	0.30	0.58	0.61	0.52	0.69	0.11
value factor 45	0.45	0.37	0.17	0.13	0.33	0.53	0.56	0.50	0.69	0.11
value factor 50	0.50	0.33	0.15	0.15	0.37	0.47	0.51	0.48	0.68	0.11
value factor 55	0.55	0.30	0.14	0.16	0.40	0.42	0.46	0.46	0.68	0.10
value factor 60	0.60	0.26	0.12	0.18	0.44	0.37	0.41	0.44	0.68	0.10
value factor 65	0.65	0.22	0.11	0.19	0.48	0.32	0.36	0.42	0.67	0.10
value factor 70	0.70	0.19	0.09	0.21	0.51	0.27	0.31	0.40	0.67	0.10
value factor 75	0.75	0.16	0.08	0.22	0.54	0.22	0.25	0.38	0.67	0.09
value factor 80	0.80	0.12	0.06	0.24	0.58	0.17	0.20	0.36	0.67	0.09
value factor 85	0.85	0.09	0.05	0.25	0.61	0.13	0.15	0.34	0.66	0.08
value factor 90	0.90	0.06	0.03	0.27	0.64	0.08	0.10	0.33	0.67	0.07
value factor 95	0.95	0.03	0.01	0.29	0.67	0.05	0.04	0.32	0.74	0.05
value factor 100	1.00	0.00	0.00	0.30	0.70	0.00	0.00	0.30	0.00	0.00

B.5 Combinations

	rate	TP	FP	TN	FN	TPR	FPR	accuracy	precision	f1_score
spread_5_AND_vol_10	0.17	0.59	0.26	0.12	0.03	0.95	0.68	0.71	0.70	0.11
spread_5_AND_vol_15	0.23	0.58	0.21	0.17	0.04	0.94	0.55	0.75	0.74	0.11
spread_10_AND_vol_5	0.15	0.58	0.28	0.10	0.04	0.93	0.75	0.67	0.67	0.11
spread_10_AND_vol_10	0.20	0.57	0.24	0.14	0.05	0.92	0.62	0.71	0.71	0.11
spread_10_AND_vol_15	0.25	0.56	0.20	0.18	0.06	0.91	0.52	0.75	0.74	0.11
spread_15_AND_vol_5	0.19	0.56	0.26	0.12	0.06	0.90	0.68	0.68	0.68	0.11
spread_15_AND_vol_10	0.24	0.55	0.22	0.16	0.07	0.89	0.58	0.71	0.72	0.11
spread_20_AND_vol_0	0.20	0.54	0.25	0.13	0.08	0.87	0.67	0.67	0.68	0.11
spread_20_AND_vol_5	0.24	0.54	0.23	0.15	0.08	0.86	0.61	0.68	0.70	0.11
spread_20_OR_vol_35	0.15	0.59	0.26	0.12	0.03	0.94	0.69	0.70	0.69	0.11
spread_20_OR_vol_40	0.17	0.58	0.26	0.12	0.04	0.93	0.68	0.70	0.69	0.11
spread_25_OR_vol_30	0.17	0.59	0.25	0.13	0.03	0.95	0.65	0.72	0.70	0.11
spread_25_OR_vol_35	0.18	0.58	0.24	0.14	0.04	0.93	0.63	0.72	0.71	0.11
spread_25_OR_vol_40	0.20	0.57	0.23	0.15	0.05	0.92	0.62	0.71	0.71	0.11
spread_30_OR_vol_25	0.17	0.60	0.24	0.14	0.02	0.96	0.64	0.73	0.71	0.11
spread_30_OR_vol_30	0.19	0.59	0.23	0.15	0.04	0.94	0.60	0.74	0.72	0.11
spread_30_OR_vol_35	0.21	0.57	0.22	0.16	0.05	0.92	0.58	0.73	0.72	0.11
spread_30_OR_vol_40	0.23	0.56	0.21	0.17	0.06	0.90	0.56	0.73	0.72	0.11
spread_35_OR_vol_25	0.19	0.59	0.23	0.15	0.03	0.96	0.59	0.75	0.72	0.11
spread_35_OR_vol_30	0.22	0.58	0.21	0.17	0.04	0.94	0.55	0.75	0.73	0.11
spread_35_OR_vol_35	0.24	0.57	0.20	0.18	0.05	0.91	0.52	0.75	0.74	0.11
spread_40_OR_vol_20	0.18	0.60	0.23	0.15	0.02	0.97	0.61	0.75	0.72	0.11
spread_40_OR_vol_25	0.21	0.59	0.21	0.17	0.03	0.95	0.55	0.76	0.74	0.11
spread_40_OR_vol_30	0.24	0.58	0.19	0.19	0.04	0.93	0.51	0.76	0.75	0.11

Appendix C

Model 2 figures and statistics

C.1 Subportfolio figures

Liquidity subportfolio

Figure C.1
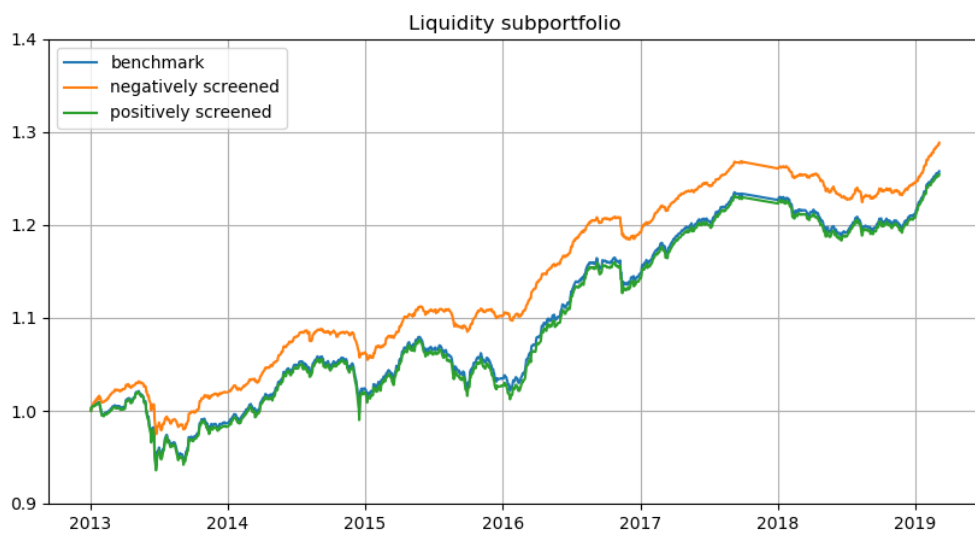
Value factor subportfolio
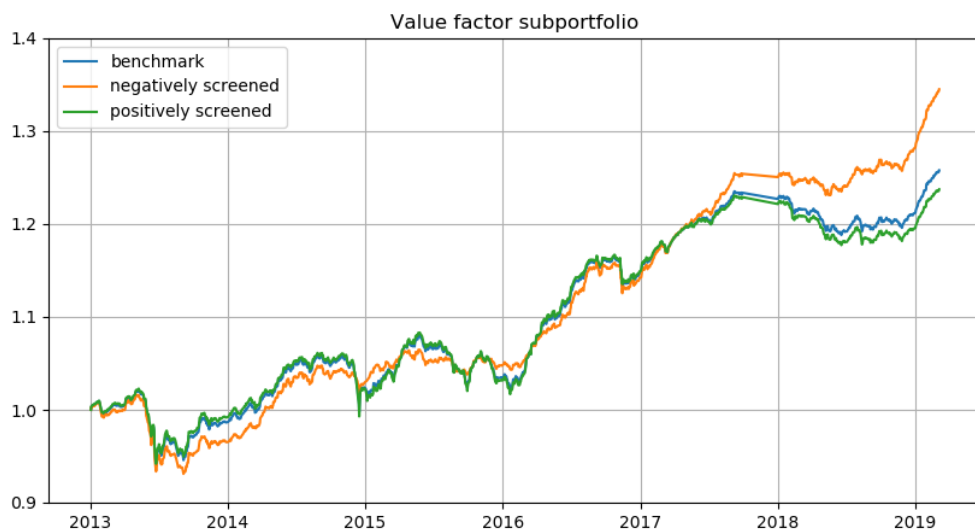
Figure C.2

C.2   Value factor alpha screening figures
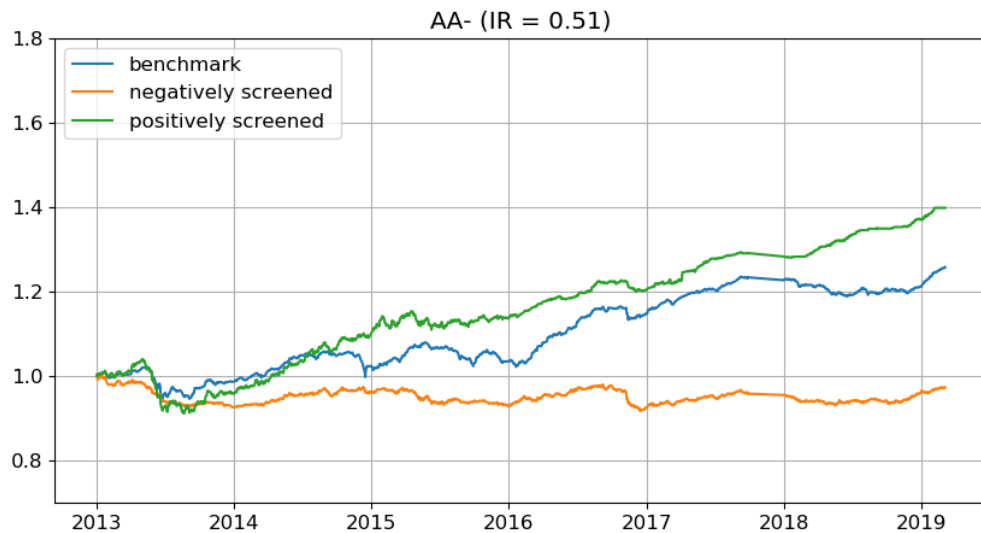
Figure C.3
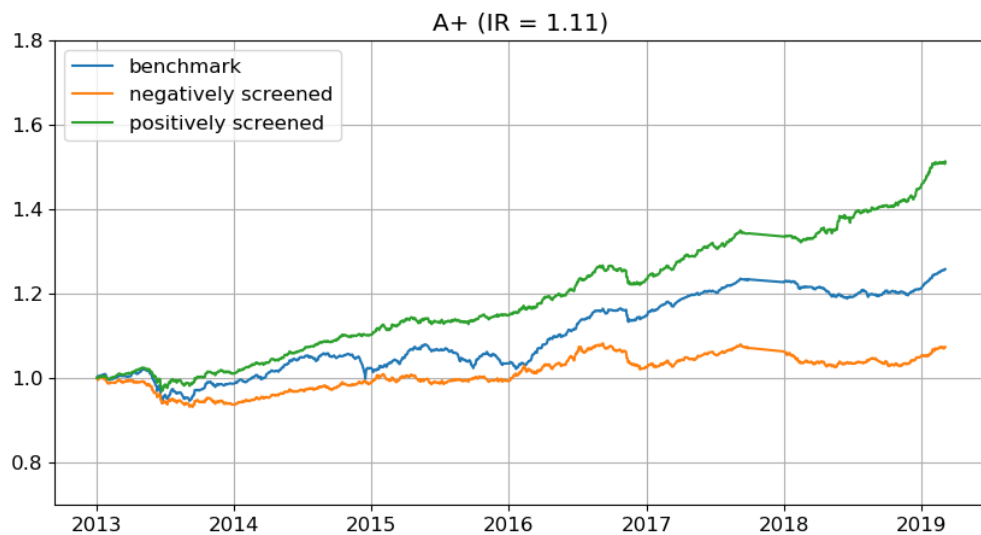
Figure C.4
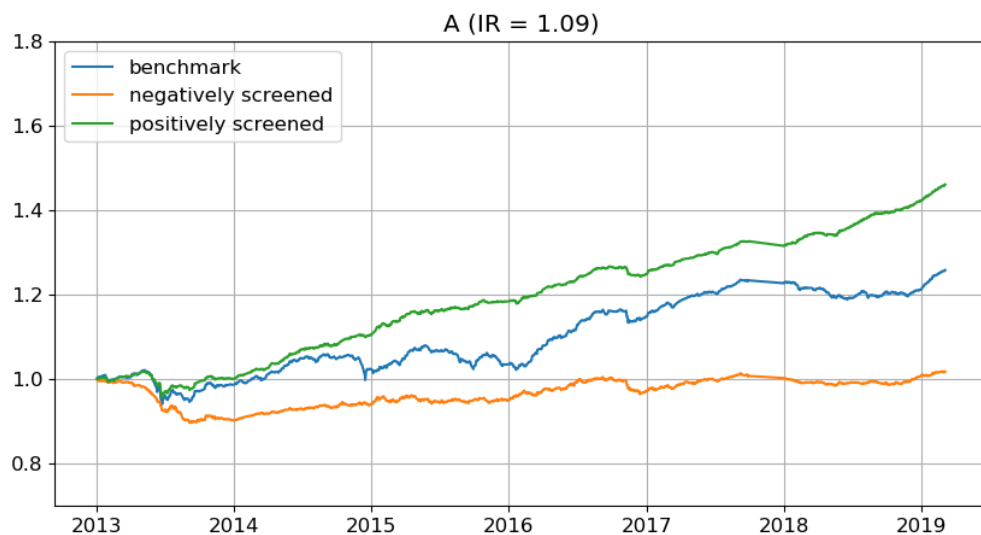
Figure C.5
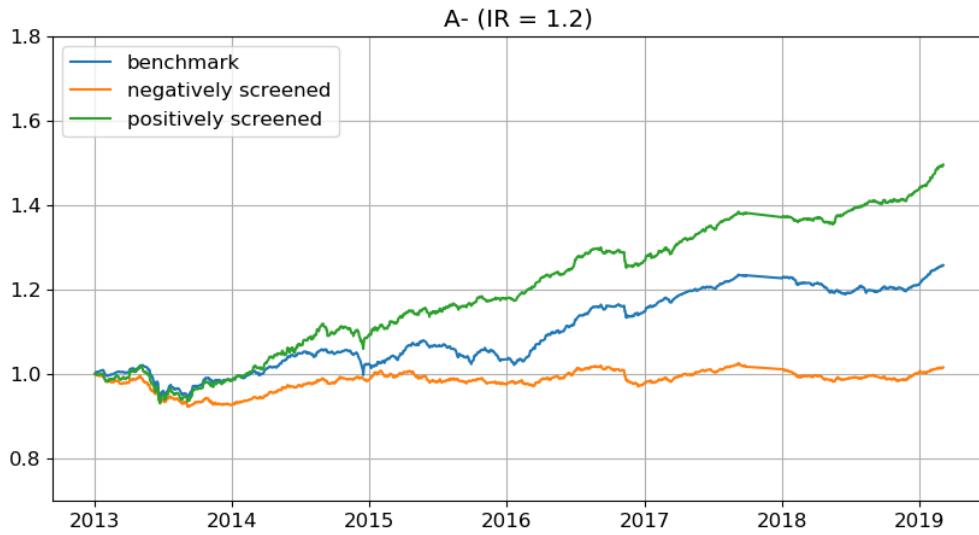
Figure C.6

C.3 Subportfolio statistics

C.3.1 Spread

	geom_avg	stdev	beta	IR	duration	geom_avg	stdev	beta	IR	duration
spread_0	0.04	0.03	1.00	1.62	4.51	-0.04	0.03	0.24	-1.58	2.95
spread_5	0.05	0.03	1.04	2.35	4.58	-0.02	0.02	0.35	-1.51	3.04
spread_10	0.05	0.03	1.07	2.03	4.66	-0.01	0.02	0.34	-1.26	3.05
spread_15	0.05	0.03	1.11	1.75	4.72	0.00	0.02	0.35	-1.06	3.18
spread_20	0.05	0.03	1.14	1.62	4.77	0.01	0.02	0.38	-0.96	3.34
spread_25	0.05	0.03	1.17	1.52	4.79	0.01	0.02	0.43	-0.88	3.53
spread_30	0.05	0.03	1.21	1.38	4.80	0.01	0.02	0.47	-0.80	3.71
spread_35	0.06	0.04	1.25	1.27	4.82	0.02	0.02	0.50	-0.73	3.84
spread_40	0.06	0.04	1.29	1.18	4.84	0.02	0.02	0.53	-0.68	3.93
spread_45	0.06	0.04	1.34	1.10	4.86	0.02	0.02	0.55	-0.62	4.02
spread_50	0.06	0.04	1.40	1.03	4.86	0.02	0.02	0.57	-0.57	4.11
spread_55	0.06	0.04	1.46	0.96	4.83	0.02	0.02	0.59	-0.52	4.21
spread_60	0.07	0.05	1.53	0.93	4.78	0.03	0.02	0.62	-0.49	4.30
spread_65	0.07	0.05	1.60	0.92	4.74	0.03	0.02	0.64	-0.47	4.36
spread_70	0.08	0.05	1.69	0.96	4.68	0.03	0.02	0.67	-0.46	4.42
spread_75	0.08	0.06	1.82	0.98	4.59	0.03	0.02	0.70	-0.46	4.47
spread_80	0.09	0.07	1.98	0.96	4.50	0.03	0.02	0.74	-0.43	4.50
spread_85	0.11	0.08	2.17	1.02	4.38	0.03	0.02	0.79	-0.41	4.52
spread_90	0.11	0.09	2.29	0.99	4.17	0.04	0.02	0.87	-0.32	4.53
spread_95	0.13	0.11	2.50	0.85	3.83	0.04	0.03	0.94	-0.19	4.53
spread_100	0.00	0.00	0.00	-1.53	0.00	0.04	0.03	1.00	0.00	4.50

C.3.2 Volatility

	geom_avg	stdev	beta	IR	duration	geom_avg	stdev	beta	IR	duration
vol_0	0.04	0.03	1.00	NaN	5.31	0.00	0.00	0.00	-0.86	0.00
vol_5	0.04	0.03	1.03	-0.47	5.38	0.03	0.01	0.09	-0.14	2.55
vol_10	0.04	0.03	1.08	-0.15	5.51	0.03	0.01	0.12	-0.21	2.59
vol_15	0.04	0.03	1.12	-0.02	5.64	0.03	0.01	0.15	-0.21	2.81
vol_20	0.04	0.03	1.17	0.04	5.75	0.03	0.01	0.19	-0.19	3.00
vol_25	0.04	0.04	1.22	0.05	5.87	0.03	0.01	0.22	-0.16	3.16
vol_30	0.04	0.04	1.27	0.07	5.99	0.03	0.01	0.26	-0.15	3.30
vol_35	0.04	0.04	1.32	0.04	6.11	0.04	0.01	0.29	-0.12	3.44
vol_40	0.04	0.04	1.38	0.06	6.23	0.04	0.01	0.32	-0.13	3.57
vol_45	0.04	0.04	1.44	0.04	6.36	0.04	0.01	0.36	-0.11	3.70
vol_50	0.04	0.05	1.51	0.07	6.48	0.04	0.01	0.39	-0.11	3.83
vol_55	0.04	0.05	1.59	0.10	6.61	0.04	0.02	0.43	-0.12	3.97
vol_60	0.04	0.05	1.68	0.12	6.75	0.04	0.02	0.47	-0.13	4.10
vol_65	0.04	0.05	1.78	0.16	6.92	0.04	0.02	0.51	-0.14	4.22
vol_70	0.04	0.06	1.90	0.17	7.09	0.03	0.02	0.55	-0.15	4.34
vol_75	0.04	0.06	2.04	0.17	6.89	0.03	0.02	0.59	-0.15	4.63
vol_80	0.04	0.07	2.23	0.15	6.87	0.04	0.02	0.64	-0.15	4.80
vol_85	0.04	0.08	2.47	0.12	6.99	0.04	0.02	0.70	-0.13	4.91
vol_90	0.04	0.10	2.86	0.05	7.03	0.04	0.02	0.77	-0.10	5.04
vol_95	0.04	0.13	3.39	0.11	6.76	0.04	0.03	0.88	-0.11	5.18
vol_100	0.00	0.00	0.00	-1.34	0.00	0.04	0.03	1.00	-0.07	5.27

C.3.3 Liquidity

	geom_avg	stdev	beta	IR	duration	geom_avg	stdev	beta	IR	duration
liq_0	0.04	0.03	1.00	NaN	5.31	0.00	0.00	0.00	-0.86	0.00
liq_10	0.04	0.03	1.02	-0.20	5.35	0.04	0.02	0.53	0.10	4.63
liq_20	0.04	0.03	1.06	-0.17	5.42	0.04	0.02	0.56	0.12	4.55
liq_50	0.04	0.03	1.17	-0.10	5.35	0.04	0.02	0.65	0.06	5.23
liq_60	0.04	0.04	1.21	-0.04	5.45	0.04	0.02	0.69	0.03	5.09
liq_70	0.04	0.04	1.32	-0.02	5.63	0.04	0.02	0.73	0.02	5.04
liq_80	0.04	0.05	1.46	-0.13	5.74	0.04	0.02	0.76	0.02	5.10
liq_90	0.04	0.05	1.60	-0.04	5.91	0.04	0.02	0.81	0.01	5.12
liq_100	0.00	0.00	0.00	-1.34	0.00	0.04	0.03	1.00	0.00	5.31

C.3.4 Value factor

	geom_avg	stdev	beta	IR	duration	geom_avg	stdev	beta	IR	duration
value_factor_0	0.04	0.03	1.00	-0.13	5.31	0.08	0.09	0.67	0.39	4.93
value_factor_5	0.04	0.03	1.01	-0.83	5.31	0.05	0.03	0.80	0.43	5.37
value_factor_10	0.04	0.03	1.02	-0.92	5.30	0.05	0.03	0.82	0.44	5.37
value_factor_15	0.04	0.03	1.03	-1.00	5.30	0.05	0.03	0.83	0.45	5.37
value_factor_20	0.04	0.03	1.04	-0.89	5.29	0.05	0.03	0.83	0.39	5.39
value_factor_25	0.04	0.03	1.06	-0.87	5.28	0.05	0.03	0.82	0.38	5.40
value_factor_30	0.03	0.03	1.07	-0.90	5.27	0.05	0.03	0.82	0.39	5.40
value_factor_35	0.03	0.03	1.09	-0.95	5.27	0.05	0.03	0.82	0.40	5.40
value_factor_40	0.03	0.03	1.11	-0.96	5.26	0.05	0.03	0.83	0.40	5.40
value_factor_45	0.03	0.03	1.13	-0.89	5.25	0.05	0.03	0.83	0.36	5.39
value_factor_50	0.03	0.04	1.16	-0.81	5.24	0.05	0.03	0.83	0.32	5.38
value_factor_55	0.03	0.04	1.20	-0.80	5.23	0.05	0.03	0.82	0.31	5.38
value_factor_60	0.03	0.04	1.25	-0.73	5.23	0.05	0.03	0.82	0.28	5.37
value_factor_65	0.03	0.04	1.30	-0.62	5.25	0.05	0.03	0.82	0.23	5.34
value_factor_70	0.03	0.05	1.37	-0.59	5.28	0.05	0.03	0.83	0.21	5.32
value_factor_75	0.02	0.05	1.46	-0.48	5.34	0.04	0.03	0.83	0.17	5.30
value_factor_80	0.02	0.06	1.59	-0.39	5.43	0.04	0.03	0.84	0.14	5.28
value_factor_85	0.03	0.07	1.78	-0.26	5.68	0.04	0.03	0.86	0.10	5.25
value_factor_90	0.03	0.08	2.01	-0.14	6.36	0.04	0.03	0.90	0.07	5.22
value_factor_95	0.01	0.10	2.32	-0.26	8.75	0.04	0.03	0.95	0.07	5.18
value_factor_100	0.00	0.00	0.00	-1.34	0.00	0.04	0.03	1.00	0.00	5.31