Invest In Your Peers

A Fundamental Approach to Peer Selection and Investment Strategy

Master Thesis

M.Sc. Finance and Investments

Copenhagen Business School 2019

Samuel Klein (92859)

Edvinas Belousovas (94016)

September 16th, 2019

Supervisor: Michael Ahm

Number of pages (characters): 120 (223,932)

Abstract

In this paper, we analyze a proprietary model that selects peer groups for individual firms based on fundamental valuation principles. Using the sum of absolute values approach, we identify peers based on single and multiple factors representing profitability, risk and growth characteristics. To determine whether peer selection is accurate, we compare the relative value of the base company to the peer group through enterprise value and price multiples. Furthermore, we tested the robustness of the peer selection by using different peer group sizes, 6 or 12 peers, and firm fundamentals based on single year or 3-year average values. This model was tested in developed and emerging markets, like indices in the U.S., Europe, and MSCI Emerging Markets. Results differed between markets; in Europe and emerging markets we reject the idea of enhanced robustness, however in the U.S. the accuracy tends to improve. Overall between markets, the analysis illustrates that peer selection accuracy typically improves when multiple fundamentals are adopted.

The investment strategy component of the paper is based on the theory of mean reversion. We select the 20 most undervalued firms relative to their respective peer group to form a portfolio. The argument for this approach is based on the theory that our peer group should represent what the base company is worth, so buying the undervalued firm would result in convergence towards the mean relative value and therefore price appreciation all else equal. We measure the return and risk performance from 2004 to 2018 through a historical simulation. The results show portfolios in the emerging markets have the highest total return and most significant alpha, suggesting that markets are the least efficient. Risk-adjusted return for the U.S. is on par with emerging markets, whereas Europe has the lowest. The portfolio risk as measured by maximum drawdown was lowest in emerging markets despite highest volatility of the indices. Overall, the various portfolio strategies outperformed their respective benchmarks on risk-adjusted return and presented positive alpha, though some strategies are riskier.

Table of Contents

1. Introduction

The availability and aggregation of financial data has led to a significant transformation of investment strategies and portfolio management. Specifically, the growth of quantitative strategies offers investors the opportunity to take advantages of inefficiencies, mispricing's and anomalies in the market based on data rather than qualitative assessments. Quantitative strategies are highly customizable, and investors can take advantage of various tactics and objectives to get exposure in strategies like momentum, value or arbitrage.

The research in this paper outlines a proprietary model for selecting investments based on identification of peer groups for individual firms within developed and emerging market indices. Therefore, we have divided the paper into two main sections of analysis; peer selection and portfolio performance. Peer selection builds on the notion of comparing firms based on fundamental characteristics. Comparing and contrasting firms based on attributes like profitability, growth and risk are not new concepts, but we build on this idea in a systematic manner to select a peer group for every company. To expand on this model, we adopt the method of equity valuation based on market multiples to identify mispriced securities from a fundamental perspective. Several relative valuation multiples, enterprise value and price multiples, are applied to identify any discrepancies that appear in various strategies and/or markets.

Besides processing and responding to large amounts of data, building a quantitative model offers a key advantage to asset managers and investors. In particular, the ability to backtest a strategy provides invaluable information regarding performance and behavior of a quantitative model. Backtesting our strategies within the different markets illustrates a picture of the historical performance, however it does not guarantee future results. Running many simulations for each strategy within the different markets enables us to analyze peer selection as well as the risk and returns of the portfolios.

The motivation for this paper stems from the idea of creating a strategy that is transparent with relatively low implementation costs, making the portfolio strategy attractive to both retail and institutional investors. The focus on intrinsic value is a natural and clear concept to grasp, so presenting multiple strategies in various markets that can capitalize on this is an attractive proposition. Though quantitative strategies are not a perfect solution for investors and managers, they have established their place as a useful investment tool and are becoming an essential part of investors' portfolios. Continuously innovating and developing new strategies is the key for investment managers to stay ahead of the competition. This is a central motivation for our research: to construct, test and analyze a unique quantitative strategy.

2. Hypothesis and Research Questions

Portfolio managers and investors are constantly looking to outperform the market. The growth of quantitative strategies has led to many new ideas and models that try to beat a benchmark. This will be the main ambition of this research paper. As we will strive to outperform the markets, we have developed the following hypothesis:

The investment strategies will beat their corresponding benchmarks in terms of both risk and return.

In order to accept or reject the main hypothesis of this paper, a thorough analysis of various parameters is necessary. Therefore, we have formulated several questions that we kept in mind during the whole research and analysis processes:

- Do single or multi-factor strategies lead to the most accurate peer selection?
  - How does it relate to the portfolio performance?
- How do larger peer groups and 3-year average fundamentals affect the robustness of peer selection accuracy and portfolio performance?
- What markets yield the most significant alphas and the highest risk-adjusted returns?
  - How does it relate to the efficient market hypothesis?

2.1. Delimitations

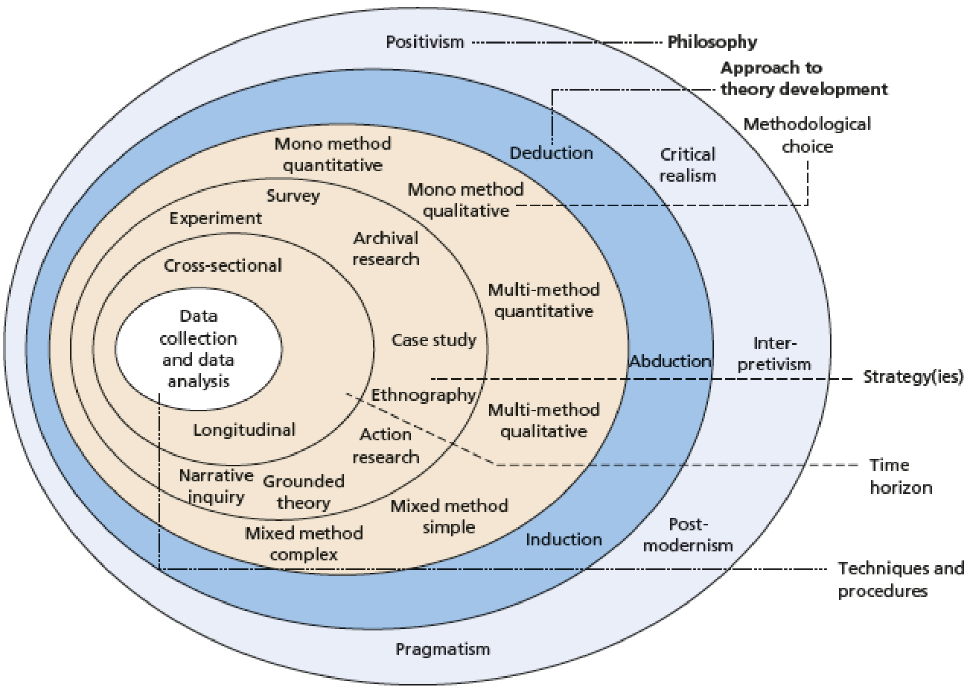
This section describes the delimitations needed to limit the scope of the research. It is necessary to maintain focus on collecting relevant data as well as developing supporting analysis to answer the above questions and conclude on the hypothesis.

As we have developed an investment model based on valuation and company metrics, the choice of fundamentals had to be limited to financial measures that have been extensively covered in literature. Furthermore, the requirement for firms to be included in the analysis was them being a member of the chosen indices and have data available across all factors during the relevant year. Selecting listed firms that are constituents of major indices ensure that there is sufficient liquidity to support assumptions in the model. Firms with negative market multiples were deliberately excluded because the model would primarily invest in those companies. The remaining firms must have had calendar year or quarterly financial results ending December 31st available, regardless of their fiscal year, in order to be included and to justify that firms are compared on equal terms. For example, Apple Inc's fiscal year ends on September 30th, meaning that data for four calendar quarters can be obtained, however Target Corporation was excluded because their accounting year ends on January 31st. Data on the STOXX Europe 600 index was available from 2001, which limits the backtest to starting in 2004 to account for having data on 3-year average fundamentals. The positions taken in the model are exclusively long, as shorting requires a plethora of assumptions like availability and borrowing costs. Lastly, the term performance is used throughout and should be interpreted as measuring the execution and outcome of the portfolio with regard to risk and return.

3. Methodology

Much of this section is concerned with how the data was collected and what techniques we adhered to on our way to finding arguments for or against our mentioned postulate. A research onion is a great visual summary of what methodological considerations require researchers' utmost attention (see figure 3.1). We will therefore structure this chapter so that we carefully peel it off step by step.

Figure 3.1 – Research Onion



Source: Saunders, Lewis & Thornhill (2016)

3.1. Philosophy

The very top layer suggests starting from the research philosophy. The question we should be asking ourselves at this point is how we are going to approach the development of new knowledge?[1] Some of the most respected academics within methodology field like Saunders, Lewis, Thornhill or Veal distinguish between two main philosophical assumptions: epistemological and ontological. Let us discuss those two in detail as they are going to underpin our methodological choices along the way.

3.1.1. Epistemology

Epistemology relates to assumptions about human knowledge. More specifically, it concerns researchers' choice of what is going to make up legitimate knowledge. Even fictitious data can be considered acceptable as per researchers' decision. Therefore, different projects can adopt different epistemologies.[2] The thing to consider under this philosophical stance is whether the researchers prefer acting as positivists or perhaps taking interpretivists' perspective feels more natural.[3]

As positivists, epistemologically researchers strive to rely on observable and measurable facts and deliver credible and meaningful results.[4] Law-like generalizations are often reached through discovering causal relationships within the studied subjects. Unlike interpretivists, positivists' viewpoint is not driven by human perceptions or interpretations, instead the human biased data are largely avoided and more emphasis is put on numerical evidence.[5] On the contrary, interpretivists' perspective often allows for reaching deep understanding of surrounding realities or developing a radically new theory based on complex individual experiences.[6] Here, human beings and their social worlds are awarded utmost attention as they create meanings.[7] In other words, interpretivist is explicitly subjectivist and can be viewed as an antagonist of positivist.[8]

As it has been explained in the introductory chapter, this report aims to test whether or not our investment model can outperform the benchmark. Not only the outperformance measures are entirely numerical, but the whole investment selection process is quantitative and structured so that

[1] Saunders, Lewis & Thornhill (2016), p. 124
[2] Saunders, Lewis & Thornhill (2016), p. 127
[3] Veal (2011), p. 30
[4] Crotty (1998) in Saunders, Lewis & Thornhill (2016), p. 136
[5] Saunders, Lewis & Thornhill (2016), p. 136
[6] Saunders, Lewis & Thornhill (2016), p. 127
[7] Saunders, Lewis & Thornhill (2016), p. 140
[8] Saunders, Lewis & Thornhill (2016), pp. 140-141

human interference is very limited. In fact, the only stage where we as researchers were involved in decision making was the creation of the investment model itself. However, each choice in favor or against the usage of certain measures or techniques was either heavily supported by the literature or logically argued for, all signaling our adherence to a positivist paradigm.

3.1.2. Ontology

Ontology relates to realities researchers encounter throughout the research process.[9] More specifically, ontological assumptions influence the way we see the world and what we want to study about it.[10] For instance, researchers can choose to see reality as either external to social actors or built up of perceptions of those social characters, and it is this approach to studied subjects that lies in core of this philosophical stance. When it comes to ontological considerations, Bryman (2012) most sharply distinguishes between positions of objectivism and constructionism.[11]

From objectivist standpoint, the reality is the same for all social actors regardless our interpretations or perceptions of the social world.[12] By taking this position, researchers assume that the reality is beyond our reach or influence [13] and that *"physical phenomena exists independently, being universal and enduring in character"*.[14] An alternative ontological paradigm, constructionism, considers reality in a continuous state, where people create it through ongoing interactions.[15] This position is much more of a subjective type, since researchers are interested in different opinions and narratives which are subjects to constant revision by social actors.[16]

Clearly, for our quantitative study objectivists' viewpoint seems to be a natural choice. In our research, we are not taking a perspective of somebody else, but rather construct an objective investment model which, once built, cannot be influenced by subjective judgements or opinions of users. Moreover, this ontological choice nicely matches positivists' perspective opted to in the section about epistemology. These two philosophical views already hint at the usage of deductive approach and quantitative data collection methods, which will be discussed subsequently.

---

9 Veal (2011), p. 30
10 Saunders, Lewis & Thornhill (2016), p. 127
11 Bryman (2012), p. 32
12 Saunders, Lewis & Thornhill (2016), p. 128
13 Bryman (2012), p. 32
14 Saunders, Lewis & Thornhill (2016), p. 128
15 Bryman (2012), p. 34
16 Saunders, Lewis & Thornhill (2016), p. 130

3.2. Approach to Theory Development

We now continue with the second layer of the research onion referring to approach to theory development. There are two main alternatives to choose from, namely, induction and deduction. The mixture of the two is also possible and it goes under the name abduction.

To begin with, let us look at a deductive approach. Deduction normally starts with a hypothesis which is usually based on the theory developed in prior studies.[17] The process is then followed by subjecting it to empirical scrutiny in a certain application or context.[18] Deductive approach often goes hand in hand with quantitative data in order for findings to be objectively measurable.[19] This is very important for fair conclusions to be drawn resulting in either acceptance or rejection of original hypothesis. Generalization is another feature of deduction. However, in order to arrive at a generalizable inferences, it is essential to select a representative sample and for it to be of a reasonable size. Conversely, induction does not seek to generalize results and often works with smaller samples to unveil different views of phenomena of studied subjects.[20] Furthermore, this research approach usually starts with a research question and aims to create new knowledge rather than test already available evidence.[21] Therefore, induction is more common for social studies, whereas deduction finds its root in natural sciences.[22] Finally, a mixture of the two can be applied as well. This kind of approach to theory development is called abduction and it moves back and forth from theory to data or data to theory at different stages of the research process.[23]

In our case, as we are testing the hypothesis of our model outperforming the benchmarks, deductive approach obviously prevails. We have even stressed it in the preceding section as both, our philosophical considerations and the fact the hypothesis testing was adhered to over the answering of the research question, signaled the use of this approach.

3.3. Research Design

The next three layers of the research onion consolidate under one topic, namely, research design. Therefore, this section is going to uncover our methodological choice, research strategy and time

[17] Saunders, Lewis & Thornhill (2016), p. 145
[18] Bryman (2012), p. 24
[19] Saunders, Lewis & Thornhill (2016), p. 146
[20] Saunders, Lewis & Thornhill (2016), p. 147
[21] Veal (2011), p. 39
[22] Saunders, Lewis & Thornhill (2016), pp. 146-147
[23] Saunders, Lewis & Thornhill (2016), p. 148

horizon for this scrutiny, all of which will be influenced by our research philosophy and corresponding approach to theory development discussed above.[24]

3.3.1. Methodological Choice

The decision to make here is whether to follow a quantitative or qualitative research design. The two can also be mixed together in different ways.[25]

To put it simply, quantitative inquiry deals with numeric data while qualitative research collects and analyzes non-numeric information such as words, images, etc..[26] However, there is much more to that. In a quantitative research, data are often collected though questionnaires or accessed from various administrative sources or authorities. Such methods are especially useful when generalizability is sought for as they ensure enough data through relatively large sample sizes.[27] Therefore, quantitative research often utilizes deductive approach and is associated with positivists' perspective.[28] In reference to qualitative research type, it usually goes together with an interpretive philosophy and any kind of approach to theory development, although induction and abduction are more commonly used in practice.[29] The data here are normally gathered through semi-structured interviews or observations, what naturally constraints the scale of the research, but allows for in-depth understanding of the studied sample.[30] These are some distinctive characteristics of quantitative or qualitative research designs. However, in many cases the difference between the two becomes uncertain and narrow. This happens when a combination of the two methods is used at data collection and/or analysis stage(s). For instance, researchers might choose to conduct interviews but ask respondents some statistical questions. The opposite might also be true when a research design requires data to be collected through surveys containing some open-ended questions instead of it being structured as a regular multiple-choice questionnaire. Moreover, these qualitative and quantitative methods can also go one after another and serve as complementary tools to each other.[31]

---

[24] Saunders, Lewis & Thornhill (2016), p. 162
[25] Saunders, Lewis & Thornhill (2016), p. 165
[26] Veal (2011), pp. 34-35
[27] Veal (2011), p. 34
[28] Saunders, Lewis & Thornhill (2016), p. 166
[29] Saunders, Lewis & Thornhill (2016), p. 168
[30] Veal (2011), p. 35
[31] Saunders, Lewis & Thornhill (2016), p. 165

In this master's thesis, we believe that objective facts act as the best scientific evidence of our investment model outperforming corresponding indices. We have therefore decided to take a positivist standpoint combined with a deductive approach. All of the above decisions directed us to the use quantitative tools such as Bloomberg, which was the main data source for our tests.

3.3.2. Strategy

A choice of research strategy is a very important element in designing research as it is "*the methodological link between your philosophy and subsequent choice of methods to collect and analyze data*".[32] It is sort of a coherent plan of how to answer a research question or accept or reject a proposed hypothesis.[33]

Saunders et al. (2016) emphasizes eight research strategies to choose from. The span is then further limited to four which can be applied with quantitative methods. As we have neither investigated a specific company, event, problem or phenomenon nor have we used survey as one of our data collection tactics, case study and survey strategies were both casted away. In fact, we believe our research strategy lies somewhere between the remaining two. The first one is defined as archival and documentary research.[34] This strategy is very much linked with the use of online archives or databases. Due to digitalization of data, wide range of information has become accessible to potential researchers from any place and at any time,[35] which is exactly what we made use of. As it was stated previously, we took advantage of our university's access to Bloomberg data bank and extracted all needed financial records from there before putting them to use in our model. Our investment idea itself, has not been applied or presented in the literature earlier, at least not to our knowledge. Therefore, using SARD approach for index constrained peer selection and combining it with mean reversion theory for investment decision making can also be seen as experimental. Namely, experiment is another quantitative strategy suggested by Saunders et al. (2016). The purpose of this strategy is usually to test researchers' predictions, often expressed in form of hypotheses, about the relationship between some dependent and independent variables.[36] In our case, we have varied financial fundamentals for peer selection, experimented with valuation multiples for investment decisions and observed our model performing in different economies. As a result, it can be said that in

[32] Denzin & Lincoln (2011) in Saunders, Lewis & Thornhill (2016), p. 177
[33] Saunders, Lewis & Thornhill (2016), p. 177
[34] Saunders, Lewis & Thornhill (2016), p. 178
[35] Saunders, Lewis & Thornhill (2016), p. 183
[36] Saunders, Lewis & Thornhill (2016), p. 178

this master's thesis we have carried out a documentary research with elements of experiment, both of which are totally in line with our previous considerations.

3.3.3. Time Horizon

The option variety is less of an issue when it comes to time horizon. Researchers can either choose to illustrate a 'snapshot' at a certain point in time or make their study longitudinal.[37]

The first option refers to cross-sectional research, where a particular phenomenon is investigated within a short time frame. Such a time horizon can either be inferred from the research topic or come from externally imposed time constraints. In relation longitudinal studies, they normally require many years of data.[38] If deadlines allow, these data can be collected by researchers themselves. In the opposite scenario, information can simply be accessed from public archives or certain online sources.[39] Due to the convenience and availability of access to Bloomberg database, we have managed to carry out a longitudinal research with the financial records starting from December 31, 2001 and ending April 30, 2018. More than 17 years of data were extracted for three indices, then analyzed and converted into meaningful conclusions. Such a scale nicely coincides with the previously mentioned methodological instruments.

3.4. Collection of Data

The core and the final layer of the research onion concerns data collection. Therefore, this section is going to be attributed to the description and evaluation of tactics used to gather the necessary information.

3.4.1. Primary & Secondary Data

Generally, there are two types of data, primary and secondary. The former is regarded to newly collected records and facts with the researcher being their primary user. This is definitely an advantage of this data type, however, costs and time needed to accumulate enough data pull in the

[37] Saunders, Lewis & Thornhill (2016), p. 200
[38] Saunders, Lewis & Thornhill (2016), p. 200
[39] Saunders, Lewis & Thornhill (2016), p. 201

opposite direction.[40] Instead of embarking on time-consuming and expensive data collection procedures, researchers can choose to re-use reports, statistics or literature published some time earlier.[41] In this case, not only that researchers do not have to spend time collecting the data, they can evaluate the existing data prior to use.[42] The quality of secondary data is as it is, researchers are free to refuse using data which do not meet their standards and resort to another source at no cost. However, this kind of 'trick' is not that easy perform with primary data. We have therefore largely relied on secondary data sources in our report. As it was noted several times previously, Bloomberg database was made use of. The Bloomberg Excel API provided access to all Bloomberg data which is a powerful data analysis tool when paired with Excel. Moreover, Damodaran's country default spreads in combination with sovereign credit ratings were applied to appropriately account for the extra risk in emerging markets. Finally, some statistical data was obtained from Yahoo Finance for descriptive purposes of presented indices.

3.4.2. Quality Assessment

It is also important to evaluate the quality of the data used. An assessment criteria of quantitative data is divided into the following: internal and external validity, reliability and objectivity.[43]

The validity aspect refers to *"the extent to which the information presented in the research truly reflects the phenomenon which the researcher claims it reflects"*. To be precise, external validity concerns the matter of representativeness and generalizability, i.e. whether the same findings apply to other samples. Internal validity, in its turn, evaluates how accurately all relevant variables are identified.[44] As researchers, we did our best to score as high as possible in both validity aspects. We have carried out a number of tests in different markets in order to make sure our results were backed up by a lot of evidence within and across the studied economies, thus, ensuring high external validity. To secure high rank in terms of internal validity as well, we have heavily relied on academic peer-reviewed literature when describing the elements of our model. Each company fundamental or multiple was presented and explained with the great precision, including their underlying financial drivers. Additionally, we have resorted to Yahoo Finance, a respectable source within finance, for a comprehensive picture of the examined indices.

[40] Veal (2011), p. 186
[41] Saunders, Lewis & Thornhill (2016), p. 319
[42] Saunders, Lewis & Thornhill (2016), p. 335
[43] Veal (2011), p. 47
[44] Veal (2011), p. 46

Reliability within quantitative research considers the replicability of the research. The question here is whether the results would have been the same if different time frame or other samples were used.[45] The time horizon is not an issue in our case as the study is of a longitudinal type. We have accumulated almost two decades of data to capture the performance of our investment model and its corresponding benchmark in different market circumstances. It should be noted, however, that little attention was given to the specific periods within the studied time horizon, but rather the overall picture was in focus. Finally, we believe that altering samples in our test would likely result in similar conclusions as made in the end of this report, though with the condition of staying within the confines of developed and emerging markets.

Objectivity concept was somewhat touched upon in this chapter earlier. The term basically means that researchers should report honestly on the results and remain as objective as possible throughout the enquiry.[46] This concept was already laid down in our chosen philosophy. Moreover, dealing with quantitative data, our sources themselves were of an objective numeric kind. Finally, the code for the model as well as all the complementary calculations and/or raw data can be found in the appendices to this thesis, making it possible for the reader to verify reported findings and conclusions made.

3.5. Summary of Methodological Choices

To prove our hypothesis right or wrong, a number of methodological considerations were thoroughly discussed and the following choices were made:

- Philosophy – positivists-objectivists' perspective;
- Approach to theory development – deductive;
- Methodological choice – quantitative;
- Research strategy – experiment combined with archival and documentary research;
- Time horizon – longitudinal;
- Data collection tactics – secondary.

[45] Veal (2011), p. 46
[46] Veal (2011), p. 44

4. Literature Review

As finance students, we have read a lot investment related material over the course of a two-year Master programme at CBS. We have explored even more academic books and articles and spent lots of time brainstorming prior to embarking on this study. As a result of this process, we have identified several themes that interested us the most. Namely, valuation, “quant” equity and factor investing were the topics that made it to the shortlist and, soon after, laid the foundation for the structure of our investment model. Due to all of them being so crucial to this report, we are going to devote a separate section to each and present what other academics have to say about them.

4.1. Valuation

The topic of valuation has been researched a lot in the past. Many scholars have come up with and tested various techniques for valuing companies. However, almost all of them fall under one of the two main categories: discounted cash flow or relative valuation. The former approach values companies’ assets given their cash flow, growth and risk characteristics. In fact, to estimate those parameters, a number of assumptions that require researchers’ subjective judgement need to be made. Conversely, relative valuation approach is more intuitive to use as far fewer assumptions are needed. The objective here is to identify similar firms in the market and compare their market prices. While this technique is simple, it can also be easily misused because no companies are identical and it is difficult to find similar ones even within the same industry.[47] To some extent we have attempted to combine both of these approaches in our model. Specifically, we start out by determining peers within indices based on their growth, risk and profitability characteristics and then select which companies to invest in based on their valuation relative to the previously identified peer group. However, more on that later (in chapter 6). Now let us introduce some of the articles that helped us arrive at this idea.

Alford (1992), in his research on the accuracy of the price-earnings valuation method, identifies comparable firms based on the industry, risk an earnings characteristics to then determine how close are their price-earnings multiples. Particularly, companies are grouped on the basis of similarity of their Standard Industrial classification (SIC) codes, total assets, return on equity, both individually as well as in combinations. He applies a number of strict selection criteria, which are then loosened until the minimum of 6 comparable firms is found. According to Alford (1992) it is important to extend the

[47] Damodaran (2001), p. 237

peer group to more than one equally comparable company as the price estimate can be predicted with relatively high standard error otherwise. In his study he tests whether the industry classification alone accurately captures firm's risk and growth aspects when selecting peers or the use of total assets coupled with return on equity can add to the valuation accuracy. The results of experiment show that these two peer selection strategies account for much of the same information, but all three criteria put together provide the most effective way of selecting comparable companies. Additionally, Alford (1992) finds that the peer selection accuracy improves with the number of SIC digits used, but only up to the third digit, as otherwise selection becomes too narrow and the needed minimum of 6 comparables cannot be returned. Moreover, the accuracy of price-earnings valuation as well as peer selection goes up with size (proxied by total assets) in many cases. Alford also studies the influence that adjusting the estimated price-earnings multiples for leverage has on the valuation accuracy. Interestingly, his sample produces results that are contrary to a *"recommendation to control for differences in leverage across comparable firms"*.

Nel, Bruwer & Roux (2014) examine how a careful selection of valuation fundamentals impacted the valuation accuracy of multiples. While developed market literature suggests that peer group selection based on valuation measures can improve valuation accuracy, the same perspective for emerging markets has not been offered. Net et al. (2014) aims to provide insight into emerging market peer group selection within South African companies listed on the Johannesburg Securities Exchange. Specifically, they measure magnitude of potential accuracy improvement using different selection fundamentals, like return on equity, total assets and revenue growth. Moreover, they test the explanatory power of the valuation fundamentals and how valuation accuracy differs with multiple selection criteria. Nel et al. (2014) select peer groups based on firms that fall within a 30% deviation margin of fundamentals relative to the target company. Their results exhibit mixed conclusions between the single and multifactor tests and the 16 relative valuation multiples. However, they conclude that valuation accuracy is generally improved vis-à-vis the single factor tests. They find that the combination of fundamentals which presents the most potential improvement of valuation accuracy is return on equity and revenue growth for all 16 multiples, however the improvement is insubstantial. Their evidence also suggests that selected multiples using multiple fundamentals can provide significant improvement over single factors like with price-to-book value of equity multiple which improved between 29-71%. They conclude that a combination of valuation fundamentals offers more accurate valuations because multiple fundamentals help define peer groups that resemble more closely the characteristics of the target company. Lastly, they wanted to argue that return on equity, total assets and growth presented the highest degree of valuation accuracy, but the lack of size of the South African market limited their ability to form enough peer groups.

The article by Knudsen, Kold & Plenborg (2017) starts from an overview of two schools of thought in reference to ways of selecting peer companies. The first one argues for peer selection based on industry classification, whereas the second one states that only valuation fundamentals are key, regardless of companies' industry affiliation. Both methods have been closely studied by practitioners/academics who express interesting views in relation to those two. According to Lee et al. (2014), the industry classification method rests too heavily on the harsh assumption that companies operating in the same industry should be comparable on a number of economic factors. Dittmann and Weiner (2005) find this method being less accurate in identifying peers compared to the one based on different financial proxies. Knudsen et al. (2017) themselves write that *"prior studies have demonstrated that identifying comparable companies on the basis of different proxies for profitability, growth, and risk may be a useful alternative to the industry classification method"*. However, those same studies were largely limited on data with peer selection criteria becoming more and more rigid. This is due to lack of firms that would be highly comparable on a number of variables. As a result, a working solution to this shortcoming is developed. In their article, Knudsen et al. (2017) offer an alternative approach to identifying comparable companies, namely, they introduce the sum of absolute rank differences (SARD) concept. This is a flexible model that can accommodate any number of selection variables, including industry classification, what makes it a highly convenient tool in a financial analyst's toolbox. It ranks firms on a set of measures used to identify peers relative to the rest of the sample. SARD score is simply the sum of those ranks, and the smaller it is, the closer the company is considered to be to the target firm. Apart from the description of the SARD approach, Knudsen et al. (2017) also put it to use and run a number of tests to conclude on its quality and applicability. In their article, Kundsen et al. (2017) use S&P Composite 1500 as a sample consisting of large-, mid- and small-cap companies. They estimate all relevant variables using the latest available 12-month financial information and rank them from smallest to largest within the index. The mentioned SARD approach is then used to identify peers for the target company. Namely, companies with the lowest SARD are assumed to share similar characteristics with respect to chosen variables, which include return on equity, growth, market capitalization, EBIT margin and a leverage ratio. The authors then examine the accuracy of the selected peers using a measure of absolute percentage error between the multiple predicted by the peer group and that of the target company. Similar method is also applied by Alford (1992), Dittmann and Weiner (2005) and Nel et al. (2014). Knudsen et al. (2017) outline enterprise value to earnings before interest and taxes (EV/EBIT), enterprise value to sales (EV/Sales), price-to-book (P/B) and price-earnings (P/E) multiples paired with Wilcoxon signed-rank tests for the assessment of the peer selection accuracy. In line with Alford (1992) they limit their choice to 6 comparable companies and re-run the test once a year, as companies' fundamentals may change and they might not be comparable for extended periods of time. Knudsen et al. (2017) find that SARD framework produces valuation estimates that are more accurate than the common industry classification approach. However, the combination of these two techniques yields even better results, suggesting that the two account for slightly different aspects of peers' fundamentals. They claim these findings are robust across time, company size, and varying numbers of peers. Moreover, the authors confirm the match between EBIT margin and EV/sales multiple. The results suggest that the EBIT margin is the best measure of a company's operational performance (and better than industry) and may be an important channel for capturing companies' industry-specific, operational characteristics. This reconfirms the flexibility and relevance of the SARD approach.

4.2. "Quant" Equity

The development of quantitative equity strategies is one of the motivators for the premise of this thesis. Therefore, a fundamental understanding of what quantitative equity is, and why it is important had to be reviewed prior to delving into the details of the specific strategies.

According to AQR Principal and Professor Lasse Heje Pedersen, *"quants define their trading rules explicitly and build systems that implement them systematically".*[48] Quantitative strategies require models and other sophisticated tools to process and interpret data. They incorporate ideas from finance, statistics, mathematics, economics and computer science that in combination generate trading signals. Quants seek to take advantage of discrepancies in markets that ordinary participants may not have incorporated into prices.[49]

Quantitative strategies differ from other equity strategies like discretionary long-short equity and dedicated short bias. The latter two strategies put the investment decision making at the discretion of the investment manager. *"Discretionary long-short equity managers typically go long or short stocks based on a fundamental analysis of each company, comparing its profitability to its valuation and studying the growth prospects".*[50] In addition, such a fundamental analysis is often accompanied by

[48] Pedersen (2015), p. 10
[49] Pedersen (2015), p. 10
[50] Pedersen (2015), p. 9

meetings with management and seeing the businesses. Discretionary investment strategies benefit by thorough analysis of every trade and incorporate qualitative data like conversations and other personal information. However, there are drawbacks especially considering the reliance on the trader to trade without any psychological biases, whereas computer-based models are objective and are not swayed by outside analysts or by corporate investor relations.[51] This method of investing is also very time consuming, and a confined amount of companies can be analyzed.

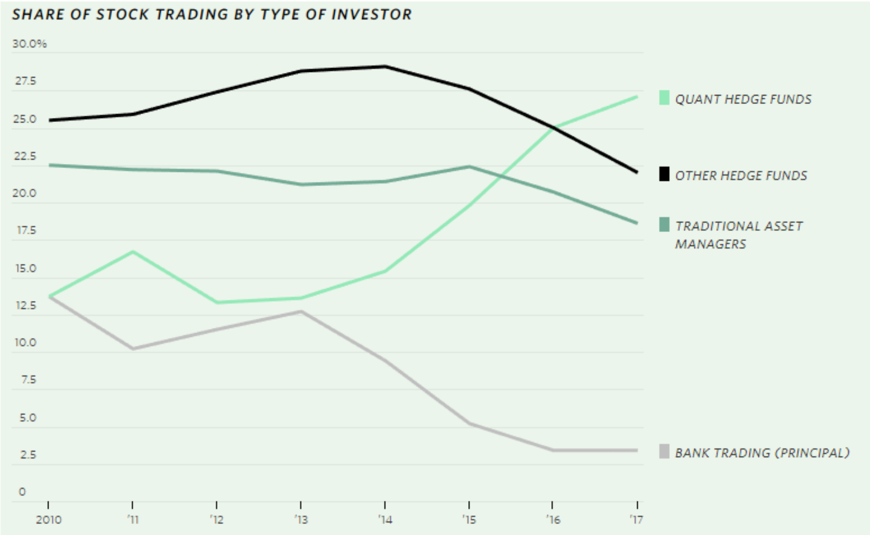
Like discretionary investment managers approach to security selection varies significantly, so do quantitative equity strategies. For example, some focus on high frequency trading, where traders exit a trade in mere milliseconds or minutes after it was opened. Other strategies involve statistical arbitrage, which involves daily trading frequency on numerous statistical patterns. Also, some take a lower trading frequency approach called fundamental quant, where computer systems use a systematic approach to identify cheap stocks to buy and expensive stocks to sell using similar factors as discretionary managers.[52] While these strategies and approaches to trading vary drastically, the root of a disciplined investment process is the same and two key concepts lay the foundation: risk and return.[53] An advantage of quantitative investing is the ability to "backtest" it or simulate its performance using historical data. Furthermore, a computer system is disciplined and follows rules, which enables managers to quantify risk exposure.

All the aforementioned advantages of quant-based investing have led to significant growth of assets under management in such strategies. As of 2018, quant hedge funds tracked by HFR, a data provider, manage almost $1 trillion or double the level in 2010.[54] In correlation with the money flow, the share of equity trading volume for quant hedge funds has about doubled in the same period, and their trading accounts for almost 30% of market trading[55].

---

[51] Ahmed & Nanda (2005)
[52] Pedersen (2015), p. 10
[53] Becker & Reinganum (2018)
[54] Wigglesworth (2018), https://www.ft.com/content/ff7528bc-ec16-11e7-8713-513b1d7ca85a
[55] Zuckerman & Hope (2017), https://www.wsj.com/articles/the-quants-run-wall-street-now-1495389108

Figure 2.1.



Source: Wall Street Journal

4.3. Factor Investing

In 1992, Eugene Fama and Kenneth French changed the way investors approach investing. They revolutionized portfolio management through the creation and application of 'factor models'. The first factors tested were SMB (small minus big) and HML (high minus low). SMB is the difference between the returns of small and big stock portfolios with the same book-to-market equity, which mimics a risk factor in returns related to size.[56] The HML factor is the difference in returns of the high value portfolios less low value, where the value is calculated as book-to-market equity.[57] Fama and French's results show that the factors are good at explaining returns, which is applicable to portfolio selection, evaluation of portfolio performance and measuring abnormal returns.[58]

Following Fama and French's discoveries, investors are searching for ways to take advantage of alternative strategies to generate returns. Investors are increasingly looking at "style premia", or systematic sources of returns typically uncorrelated with traditional assets, which led to an increase in

[56] Fama & French (1993)
[57] Fama & French (1993)
[58] Fama & French (1993)

investment options.[59] A style can be anything, so long as it is based on a variable of some sort. This is a good thing for investors, as they are presented with limitless possibilities for portfolio construction. Being able to tailor investment strategies for investors by mixing styles provides significant value, and helps manage portfolio construction, risk and trading costs.[60]

Unfortunately, the grass is not always greener on the other side, even when considering quantitative investment strategies. Practitioners like Osman Ali of Goldman Sachs Asset Management's quantitative investment strategies group says crowding, or over-population in a strategy, is something they think about a lot.[61] Crowding comes from other quantitative investors, as well as other market participants and results in diminishing returns and reduced effectiveness. While backtesting does provide valuable information as to whether a strategy would have succeeded historically, it is not be able to decipher whether a trade is crowded. This is especially an issue since quantitative investing is still relatively new, so a lot of the strategies are quite young and may not have had an impact yet. Though there likely is no remedy for crowded trades, continuously developing new strategies could likely be the solution for investment managers to stay ahead of over saturated trades.

5. Indices

The investment strategy presented in this research will be tested and analyzed on several reference indices. The index will act as a benchmark to the strategy performance, and the index constituents provide a starting point for companies to be included in the test. The indices used are Standard & Poor's 500, STOXX Europe 600 and MSCI Emerging Markets Index. Further details on index attributes are discussed in their respective sections.

Large index providers like S&P Global, STOXX and MSCI have significant influence over liquidity and market accessibility of companies added to their indices. When one of these providers adds a company, country or region to an index, then investment flows rapidly to the addition. Passive index managers rush to begin trading the new components because they have to match the holdings of the benchmark. Inclusion to an index is positive for a company and investors due to the liquidity and accessibility, however being dropped or excluded can have the opposite effect. Testing an investment

[59] Israel, Jiang & Ross (2017)
[60] Israel, Jiang & Ross (2017)
[61] Oyedele (2016)

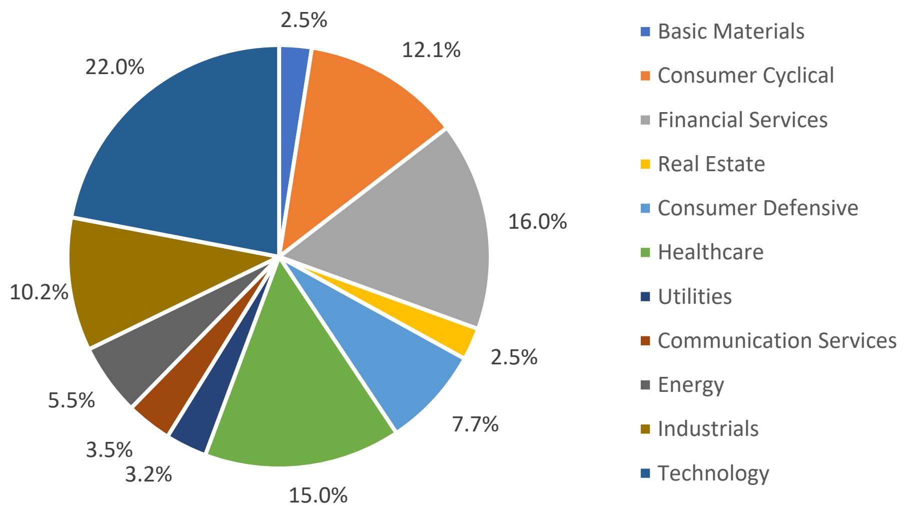
strategy on companies that are components of popular indices ensures that liquidity and market access are not a concern.

5.1. S&P 500 (SPX Index)

The S&P 500 was launched in 1957 and is considered the best benchmark of large-capitalization U.S. equities. S&P Global estimates that over $9.9 trillion is indexed or benchmarked to the index as of January 31st, 2019.[62] The S&P 500 index is float-adjusted market-cap-weighted, meaning that the market cap is calculated based on the total shares available for public trading, or shares ”floated”.[63] The index captures around 80% of available free-float market capitalization.

Constituents in the index are exclusively U.S. companies, and the index does not include companies listed on U.S. exchanges that are not common shares or American Depositary Receipts (ADRs). Currently, for companies to be eligible to be included in the index, they must have a market cap of at least $6.1 billion. Constituents have at least 50% of outstanding shares available to trade, and the companies' common shares are highly traded with active and deep markets.[64]

Figure 5.1



Source: Yahoo Finance, February 22, 2019

[62] S&P Dow Jones Indices, https://us.spindices.com/indices/equity/sp-500
[63] S&P Dow Jones Indices, https://us.spindices.com/indices/equity/sp-500
[64] S&P Dow Jones Indices, https://us.spindices.com/indices/equity/sp-500

The S&P 500 index is diversified in the distribution of industries in which the constituents operate, as pictured in figure 5.1. above. The selection committee selects companies to be included in the index, such that the index represents the overall performance and strength of the U.S. economy. These attributes make the index attractive to test on because of the large index size and the amount of analyst coverage that the index companies receive.

Table 5.1

Index Descriptive Characteristics (as of January 31st, 2019 unless otherwise specified)

Number of Constituents	505	10 Year Annualized Returns	14.86%
Total Index Market Cap*	23,830	10 Year Annualized Standard Deviation	13.43%
Mean Market Cap*	47.19	Price / Book**	2.85
Median Market Cap*	20.27	Price / Earnings (Forward)**	16.08
Top 10 Holdings Weight**	20.94%		

*In USD billions

**As of February 22, 2019

Source: Yahoo Finance, S&P Dow Jones

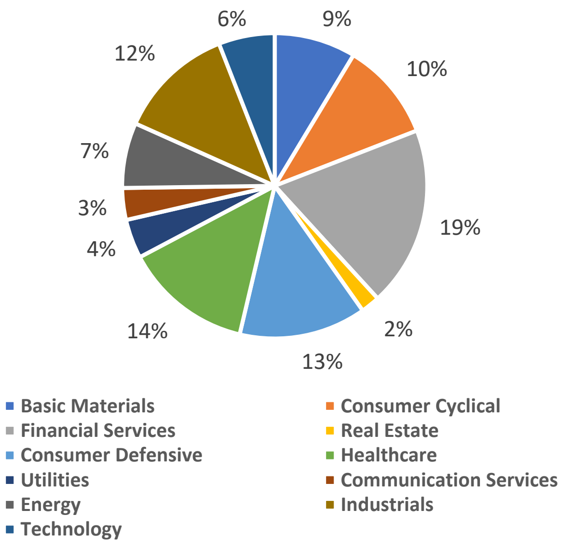
Following the financial crisis, the index has performed well over the past 10 years, having an annualized net total return of 14.86%. The price-to-book ratio shows the market value per share is roughly 2.85 times the book value, or net assets, per share on average. An investor is currently willing to pay $16.08 for $1 in current earnings.

5.2. STOXX Europe 600 (SXXP Index)

The STOXX Europe 600 Index was introduced in 1998, and represents large, mid and small capitalization companies across 17 countries in the European region: Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland and the United Kingdom.[65] The index consists of the largest developed market stocks in Europe in terms of free-float market capitalization, and it covers about 90% of the whole market in this regard. The constituent weights are also calculated on a free-float market capitalization basis.[66]

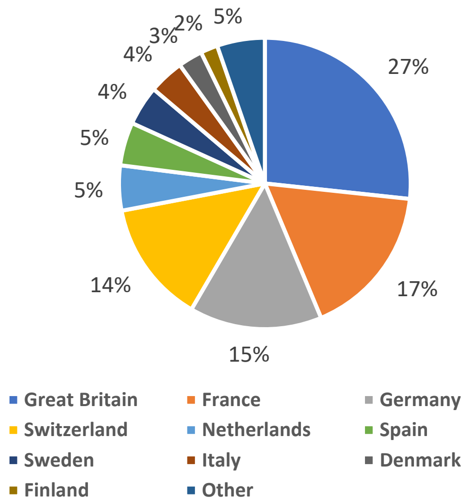
[65] IISTOXX® Europe 600, https://www.stoxx.com/index-details?symbol=SXXP
[66] STOXX® Europe 600, https://www.stoxx.com/index-details?symbol=SXXP

Figure 5.2



Source: Yahoo Finance, February 22, 2019

Figure 5.3



Source: STOXX, February 22, 2019

The Europe 600 index has over a quarter of its holdings in Great Britain, with France and Germany bringing the total to 59% (see figure 5.3). The index is attractive to test on because of the varying constituent sizes through small, mid and large cap companies, and the country diversification. While all the included countries are in developed Europe, their respective economies can differ greatly.

Table 5.2

Index Descriptive Characteristics (as of January 31st, 2019 unless otherwise specified)

Number of Constituents	600	10 Year Annualized Returns**	9.81%
Total Index Market Cap*	9,697	10 Year Annualized Standard Deviation**	13.34%
Mean Market Cap*	12.90	Price / Book**	1.52
Median Market Cap*	5.60	Price / Earnings (Forward)**	12.95
Top 10 Holdings Weight	17.77%		

*In EUR billions

**As of February 22, 2019

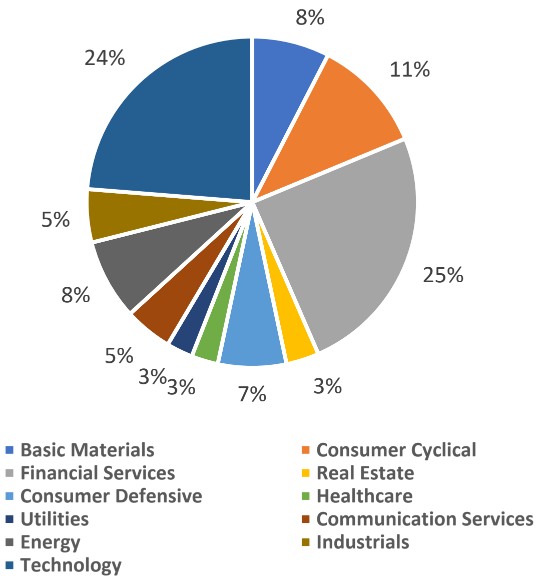
Source: Yahoo Finance, STOXX

The total index market cap is roughly €9.7 trillion or about half the S&P 500.

5.3. MSCI Emerging Markets (MXEF Index)

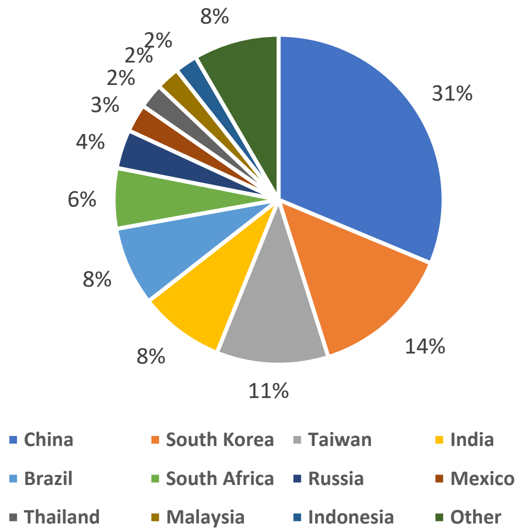
The MSCI Emerging Markets Index was launched in 1988, and initially consisted of 10 countries that accounted for less than 1% of the world's market capitalization.[67] Many economic developments have occurred since the initial launch and regions previously classified as frontier markets have shifted to emerging regions. The index has expanded significantly, and currently consists of 24 emerging countries from the Americas, EMEA and Asia: Brazil, Chile, China, Colombia, Czech Republic, Egypt, Greece, Hungary, India, Indonesia, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, Qatar, Russia, South Africa, Taiwan, Thailand, Turkey and United Arab Emirates.[68] The index now accounts for roughly 10% of the world's market capitalization and it covers about 85% of the available free-float market capitalization in these 24 countries.

Figure 5.4



Source: Yahoo Finance, February 22, 2019

Figure 5.5



Source: Bloomberg, February 22, 2019

The Emerging Markets index is an interesting market because of the continuous economic and social developments in the associated regions. Adjustments to the index are attributed to the growth and relative wealth level in countries. While developed markets have high levels of per capita income, emerging areas have low, middle and upper middle levels of income and rapidly growing economies.[69] This leads to constant assessment and updates to the constituents of the index over time. For

---

[67] MSCI Emerging Markets Index, https://www.msci.com/emerging-markets
[68] MSCI Emerging Markets Index, https://www.msci.com/emerging-markets
[69] Wall (2018), https://www.morningstar.in/posts/47141/difference-emerging-frontier-market.aspx

example, in 2013, MSCI upgraded UAE and Qatar to emerging markets status and downgraded Greece from developed to emerging. Furthermore, in June 2018 MSCI added China A-shares to the Emerging Markets index.[70] Prior to this change, Chinese companies were only included if they were listed in Hong Kong, whereas A-shares are firms listed in mainland China.

Table 5.3

Index Descriptive Characteristics (as of January 31st, 2019 unless otherwise specified)

Number of Constituents	1,124	10 Year Annualized Returns**	8.75%
Total Index Market Cap*	5,223	10 Year Annualized Standard Deviation**	19.40%
Mean Market Cap*	4.65	Price / Book**	1.4
Median Market Cap*	1.92	Price / Earnings (Forward)**	10.92
Top 10 Holdings Weight**	23.99%		

*In USD billions

**As of February 22, 2019

Source: Yahoo Finance, MSCI

5.4. Index Comparisons

The S&P 500 index is the largest of the 3 indices in terms of total market cap, and is purely focused on large cap stocks. MSCI EM has almost a quarter of the index in 10 holdings, even though the index has the highest number of holdings. The Europe 600 index is currently the most balanced in terms of country and sector distribution, as well as the smallest weights for the top holdings. The S&P 500 is overweight technology, financials and healthcare, while MSCI EM is also overweight technology and financial services.

The Emerging Markets index is the riskiest in terms of 10 year annualized standard deviation, which makes sense given that emerging regions have high growth prospects, which can lead to volatility if the economic forecasts are not realized. MSCI EM index has the lowest price-to-book ratio, suggesting that the balance sheet value for money is highest in the emerging market areas. In addition, the price-earnings ratio is the lowest in MSCI EM index and investors are currently not willing to pay as much for forward earnings as in the U.S. or Europe. Finally, returns of the S&P 500 index have been the highest over the past 10 years, showing investors may have been willing to pay more for large cap firms, possibly due to their lower level of risk.

[70] Wall (2018), https://www.morningstar.in/posts/47141/difference-emerging-frontier-market.aspx

The indices described in this section have diverse firm specific attributes related to size, earnings and risk, and broader socio-economic distinction across countries and regions. These varying and unique attributes provide an interesting foundation to test peer selection and the investment strategy.

5.5. Efficient Market Hypothesis

A main reason for selecting several indices was to spur discussion regarding performance of the various investment strategies and how it differs between markets. Each market is unique because they have differing characteristics like availability of information. The concept of financial markets processing and reflecting all information in stock prices is known as the efficient market hypothesis. Theoretically, risk-adjusted returns greater than the market should not be possible to achieve because the market reflects all available information and therefore there should be no mispricing's.[71] Empirical studies have shown that markets are efficient to a large extent, and we want to analyze whether the investment strategy performance supports this, or if certain markets are mispriced. Market efficiency varies and is to an extent based on the amount of coverage companies receive, where smaller stocks which receive less attention or stocks in countries with less rigorous accounting standards are expected to be less efficient.[72] Lastly, the idea of markets being efficiently inefficient suggests that markets are efficient as a result of competition between active managers, but inefficient enough that professionals receive profits to compensate for their additional risks and costs.[73]

6. Investment Model

An investment model lays the framework for how an investment manager selects securities and the mandate with which they manage a portfolio. For an investor it is crucial to know how the model works in order to fully understand the process, risks and how the model correlates with other investments. The following section outlines how our model is constructed and the components that drive investment decision making.

[71] Munk (2017), p. 345
[72] Munk (2017), p. 346
[73] Pedersen (2015), p. 4

6.1. Structure

The model used in this research to find comparable firms and trade on relative mispricing's is partially adapted from Knudsen, Kold and Plenborg (2017). The model is based on a ranking system called *Sum of Absolute Ranked Differences* or SARD. A ranking system enables us to generalize and arrange a set of information such that each observed data point is given a natural number that makes it possible to evaluate complex data. Ranking is a common method used in science to measure differences, particularly Bhojraj & Lee (2002) rank multiples to determine comparable firms and identify this as a more objective method for investors and analysts to use. Ranking is also quite prominent in quantitative equity investing, as presented by Pedersen, where ranking is applied to factor models.[74]

SARD is given by a matrix of sum of ranked differences between companies:

$$SARD_{i,j} = |r_{x,i} - r_{x,j}| + |r_{y,i} - r_{y,j}| + \cdots + |r_{z,i} - r_{z,j}|$$

Where SARD is the sum of absolute ranked differences between company i and company j, and where $r_{x,i}$ is the rank of company i in terms of factor x, $r_{x,j}$ is the rank of company j in terms of factor x and so on.[75] The model measures comparable companies by summing the absolute differences in ranks of different factors, where the score closest to zero represents the firm that is most comparable with the base company, and vice-versa.

6.1.1. Steps for Peer Selection

Initially we start out with an index, and the underlying components of that index, or constituents. As outlined in the Indices section, the peer selections process and investment strategy will be tested on three indices representing three different markets: S&P 500, STOXX Europe 600 and MSCI Emerging Markets Index. Of the 3 indices, STOXX Europe 600 had the least historical data and started in December 2001. We will invest on the first available trading day of May every year, using the last four quarterly reports published up to 31st December for every company. The date of investment is chosen because publicly traded companies will likely have reported their 4th calendar quarter earnings over the first four months of the new year.

Let's consider an example of the data processing and peer selection steps for the S&P 500 index in 2017/2018. As we will be investing on the first day of May, the initial company scope will come from

[74] Frazzini & Pedersen (2014), p. 9
[75] Knudsen, Kold and Plenborg (2017)

the list of constituents on the last trading day of March 2018 to make sure the companies are relevant and actively traded at that point in time.

First things first, the list of firms was checked for duplicates because companies with more than one share class can appear multiple times. For example, Alphabet Inc. has a GOOG stock that represents class C shares with no voting rights and GOOGL stock that represents class A shares with voting rights. After removing duplicates from the components list, we retrieved all the financial data over the past year from Bloomberg. Again we filtered out and removed companies from the list if some of the data was unavailable, or not applicable like a negative earnings multiple. In this case the same data set was used for every simulation in a given year, even if there was more data available for specific simulations that would not require the whole data set. The reason for keeping the sample consistent was to provide an equal measure of performance throughout all the strategy simulations.

After the data set was cleaned, each valuation fundamental was ranked as in the example below.

Table 6.1

	Fundamental Values			Ranks		
Company	**ROE (%)**	**Total Assets ($ millions)**	**Growth (%)**	**ROE**	**Total Assets**	**Growth**
American Express	13.66	181,159	15.28	4	3	6
Broadcom	8.64	54,418	33.20	6	9	1
Caterpillar	5.62	76,962	17.97	9	8	5
Chevron	6.26	253,806	23.40	7	2	3
Coca-Cola	6.22	87,896	-15.41	8	7	10
Walt Disney	21.23	95,789	-0.89	3	6	8
Exxon Mobil	11.10	348,691	18.21	5	1	4
Phillips 66	21.51	54,371	25.96	2	10	2
IBM	32.11	125,356	-0.98	1	5	9
Johnson & Johnson	1.99	157,303	6.34	10	4	7

Source: Authors' findings

The table above consists of 10 companies that were part of the S&P 500 on March 31st, 2018. The fundamental values are from calendar year 2017 and each fundamental is ranked against the other companies in the sample, where 1 is the highest value and 10 the lowest. Now that each company is ranked based on ROE, Total Assets and Growth, the SARD can be computed. Using the formula given previously, 6 peer companies are found for each base company. The number in parentheses indicates the SARD score.

Table 6.2

Base Company	Peer 1	Peer 2	Peer 3	Peer 4	Peer 5	Peer 6
American Express	Exxon Mobil (5)	Walt Disney (6)	Chevron (7)	IBM (8)	Johnson & Johnson (8)	Caterpillar (11)
Broadcom	Phillips 66 (6)	Caterpillar (8)	Chevron (10)	Exxon Mobil (12)	American Express (13)	Coca-Cola (13)
Caterpillar	Coca-Cola (7)	Johnson & Johnson (7)	Broadcom (8)	Chevron (10)	American Express (11)	Walt Disney (11)
Chevron	Exxon Mobil (4)	American Express (7)	Johnson & Johnson (9)	Broadcom (10)	Caterpillar (10)	Coca-Cola (13)
Coca-Cola	Caterpillar (7)	Walt Disney (8)	Johnson & Johnson (8)	IBM (10)	American Express (12)	Broadcom (13)
Walt Disney	IBM (4)	American Express (6)	Coca-Cola (8)	Johnson & Johnson (10)	Caterpillar (11)	Exxon Mobil (11)
Exxon Mobil	Chevron (4)	American Express (5)	Walt Disney (11)	Johnson & Johnson (11)	Broadcom (12)	Caterpillar (12)
Phillips 66	Broadcom (6)	Walt Disney (11)	Caterpillar (12)	American Express (13)	IBM (13)	Chevron (14)
IBM	Walt Disney (4)	American Express (8)	Coca-Cola (10)	Johnson & Johnson (12)	Exxon Mobil (13)	Phillips 66 (13)
Johnson & Johnson	Caterpillar (7)	American Express (8)	Coca-Cola (8)	Chevron (9)	Walt Disney (10)	Exxon Mobil (11)

Source: Authors' findings

After identifying the group of peer companies for each base company, the next step is to decide which firms to invest in. The process for selecting investment opportunities is given by:

$$V_i = \frac{\overline{m}_{x,p} - m_{x,i}}{m_{x,i}}$$

Where V_i is the value of the base company compared to its group of peers. In addition, $\overline{m}_{x,p}$ is the average of multiple *x* for peers *p*, and $m_{x,i}$ is the multiple *x* for company *i*. The base company is undervalued compared to its peers when $V_i > 0$ and overvalued when $V_i < 0$. An example using the 10 companies can be seen in table 6.3.

Table 6.3

	Company Multiples				Peer Group Average Multiples				% Undervalued (+) or Overvalued (-)			
Company	**EV/Sales**	**EV/EBITDA**	**P/CF**	**P/B**	**EV/Sales**	**EV/EBITDA**	**P/CF**	**P/B**	**EV/Sales**	**EV/EBITDA**	**P/CF**	**P/B**
American Express	2.86	15.01	6.48	4.68	2.72	35.51	13.12	4.70	-5%	137%	103%	1%
Broadcom	6.39	47.29	15.63	5.10	2.63	40.64	14.73	4.75	-59%	-14%	-6%	-7%

Caterpillar	2.09	21.26	16.34	6.88	4.33	39.32	15.32	5.44	108%	85%	-6%	-21%
Chevron	2.15	110.32	11.59	1.61	4.08	27.90	16.00	6.04	90%	-75%	38%	275%
Coca-Cola	6.33	29.52	27.84	11.45	3.62	21.97	12.90	5.75	-43%	-26%	-54%	-50%
Walt Disney	3.16	12.63	12.52	3.58	3.37	22.44	14.82	6.53	7%	78%	18%	82%
Exxon Mobil	1.69	33.17	11.84	1.89	3.63	37.94	13.41	4.68	115%	14%	13%	148%
Phillips 66	0.67	34.56	14.28	2.03	3.13	36.83	11.85	4.98	365%	7%	-17%	146%
IBM	2.14	14.51	8.56	8.04	3.30	24.34	15.14	4.97	54%	68%	77%	-38%
Johnson & Johnson	5.12	21.15	17.86	6.23	3.04	36.98	14.43	5.01	-40%	75%	-19%	-20%

Source: Authors' findings

The left part of the table shows the multiples of the base companies. The middle columns are the averages of those same multiples for the peer group. For example, the peer group average EV/Sales for American Express consists of multiples of Exxon Mobil, Walt Disney, Chevron, IBM, Johnson & Johnson and Caterpillar. The right part of the table shows how the relative value of the base company compares with its peer group. The red cells represent firms with overvalued multiples, whereas green show undervalued. Comparing American Express on EV/Sales again, shows that it is 5% overvalued compared to its peer group. Interestingly, American Express is undervalued when compared on EV/EBIT and P/CF. This illustrative example has a small sample size and is not necessarily the most representative of the results with the full data set, however it is important to note that firms' relative value dynamics can change significantly depending on what underlying multiples are considered.

Through this example, the steps of data selection, processing and analysis were presented. The simulations in the following chapters will include the years from 2004 to 2018 in each of the 3 indices. On top of this, the tests will include various fundamentals for selecting peers and relative valuation multiples for determining investments. The fundamentals used in combination are given below:

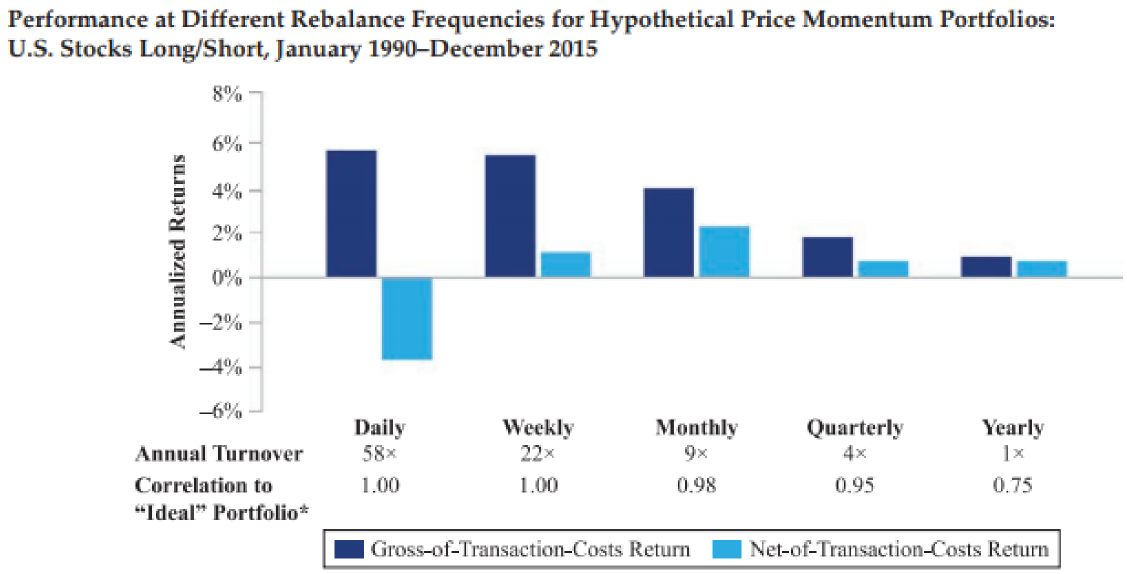
Single Factor Tests	**Multi Factor Tests**
Return on Equity (ROE)	ROE + TA
Total Assets (TA)	ROE + TA + G
Growth (G)	ROIC + TA + G + WACC
Return on Invested Capital (ROIC)	ROE + TA + G + WACC + ROIC
Weighted Average Cost of Capital (WACC)	

Furthermore, duplicate tests using the 3-year average for the fundamentals ROE, ROIC, TA, G and WACC were simulated. A 3-year average could provide stability to the fundamentals compared to a single year, since a valuation metric will be less skewed in the instance of a one-off charge, capital structure shift, or arbitrary business development etc. However, the drawback could be that historical accounting results and performance would not necessarily provide an accurate representation of the current state of the firm.

The number of peers that was used for the tests were 6 and 12 companies, where the former is in line with previous studies conducted by Alford (1992) and Knudsen et al. (2017), and the latter is just a double of that. As more peers are selected, the deviation of the SARD score to the base company will likely increase, but the average of peer group multiples will possibly gain in stability. The two different size peer groups will provide insights as to how the selection model performs.

Another important factor in the implementation of the investment strategy was rebalancing of the portfolio. Rebalancing of a strategic portfolio was required to effectively execute a theoretical portfolio strategy. Market movements caused deviation in the portfolio from the optimal strategic allocation that we had defined. Therefore, it was imperative to consider rebalancing frequency and the implications of trading frequency. We are admirers of AQR as researchers and practitioners due to their application and implementation of quantitative portfolio management in reality, and rely on their methods to execute our own strategy. More specifically, Israel et al. (2017) from AQR discuss the application of style investing where they consider the ideal rebalancing frequency (see figure 6.1).

Figure 6.1

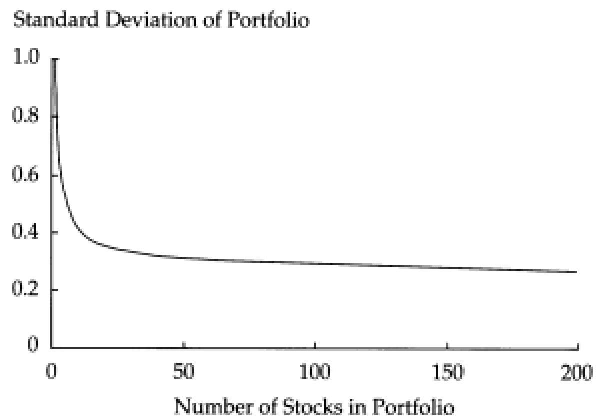


Source: Israel, Jiang & Ross (2017)

Israel et al. explains that a portfolio which is rebalanced daily can result in the "freshest" portfolio, meaning that the correlation is closest to the ideal portfolio and as a result has the highest gross return. On the other hand, a portfolio that is rebalanced annually will typically be "stale", meaning that its performance is not as highly correlated with the ideal portfolio but it also does not have high transaction costs.[76] Portfolio rebalancing requires a fine balance between transaction cost savings at lower frequencies and performance degradation because of stale prices. The optimal rebalancing frequency based on the test of momentum portfolios was once a month, because it resulted in the highest return net of transaction costs. Furthermore, Frazzini & Pedersen (2014) also applied a monthly rebalancing rule to their Betting Against Beta style portfolio.[77] In accordance with prior research, we applied a monthly rebalancing approach to our portfolio to ensure that the portfolio holdings did not deviate from the strategic weights for too long.

As explained, rebalancing is used to achieve the strategic weights for all securities in a portfolio. Therefore, it is important to define the amount of securities selected and the allocation of weights within the portfolio. Portfolio diversification is dependent on the amount of stocks in a portfolio, and expected standard deviation declines as a portfolio becomes more diverse.[78]

Figure 6.2



Note: The correlation between the returns of any two stocks is 0.08, and the standard deviation of any stock is 1.0.

Source: Statman (2004)

[76] Israel, Jiang & Ross (2017), p. 12
[77] Frazzini & Pedersen (2014), p. 9
[78] Statman (2004), p. 47

Diversification also depends heavily on the correlation between assets, where figure 6.2 from Statman (2004) shows how the standard deviation of a 20-stock portfolio is only 35% of a 1 stock portfolio. Standard deviation of a portfolio with n positions is given by:

$$\sigma_n = \sqrt{\sum_{j=1}^{n} \sum_{i=1}^{n} w_i w_j cov(r_i, r_j)}$$

Where w_i and w_j are the portfolio weights of stocks i and j and $cov(r_i, r_j)$ is the covariance of the returns between i and j. While a 20-stock portfolio is not necessarily the optimal, it retains much of the benefits of diversification. Increasing the number of stocks in a portfolio typically lowers the standard deviation, but other factors like cost of buying and holding stocks and the expected equity premium must also be considered.[79] Furthermore, buying more stocks means that the stocks which are selected last should be less undervalued based on the SARD model. To avoid this but also ensure that the portfolio is diversified, 20 stocks were chosen for the long portfolios. Having determined the amount of securities to buy, we had to decide how to weight each investment. AQR research implemented their style portfolios by allocating equal weight to each stock, with monthly rebalancing to maintain equal weights.[80] This method of portfolio management ensures that the portfolio is not overexposed to a specific company. Using non-equal weights would introduce a form of bias to the implementation process, since favoring one stock over another could require a subjective view of the security selection process. Therefore, we have decided to use the same method as AQR in our test.

6.1.2. Model – Data Processing and Analysis

The choice of software platform to use for backtesting is an important consideration because it can increase productivity and enable the broadest possible spectrum of strategies to be simulated.[81] To process and analyze the collected data, STATA and Excel were used. STATA is a data science program that enables us to simulate the portfolio selections. We have written the code that is dynamic and can be applied across all years and indices. Furthermore, it allows for changes in inputs that the peer selection process should be executed on, i.e. different fundamentals or number of peers can be used. The code is provided in appendix 1 for reference. The overall data set was imported to STATA via a

[79] Statman (2004), p. 47
[80] Frazzini & Pedersen (2014) p. 13; Israel, Jiang & Ross (2017), p. 10
[81] Chan (2013), p. 1

readable datafile, where the code was run individually per year and with the specific selection criteria. Running the code applies the steps for processing the data in accordance with the SARD process that was outlined in the previous section. The output results in the 20 most undervalued companies.

After determining the companies that we wanted to trade, the company selections were transferred to Excel. Through Bloomberg Excel API we were therefore able to retrieve monthly returns on each holding and in-depth statistics that provided analytical insight to the investment strategy.

6.1.3. Backtesting

Now that we have established a trading idea, backtesting will provide valuable insights to the performance of the strategy. To execute a backtest means to simulate how a trading strategy would have performed historically. That being said, historical performance does not necessarily forecast how a strategy will perform in the future, but it is very useful nevertheless. Probably the most useful information that a backtest provides, is an indication of whether a trading strategy has a possibility of good performance. If a backtest shows that the strategy was bad, then investors can save time and money not pursuing such an investment idea. A backtest can also provide valuable information that can lead to ideas for improvements.[82]

There are several important aspects to consider when running a backtest: the universe of securities, signals, trading rules and time lags. The universe of securities that are to be traded is a fundamental part of backtesting, because it is only possible to trade securities that were listed historically. In our case, the reference point for the equity's universe was the index components in a given year, so that we knew what securities had existed at that point in time. The signals refer to the data used in the backtest, which in this report was primarily from Bloomberg due to its quality and availability. The trading rule comprises of how we trade on the data collected, including how often to make trades, when to rebalance positions and what position sizes to take. In our case, we were selecting the securities once a year using the previous years' earnings data, and sizing all the positions equally with monthly rebalancing. The time lags required to implement a strategy are crucial because it is not possible to trade on information that was not available at the time of the backtest. This is called look-ahead bias, in other words, using tomorrow's price to determine today's trading signal.[83] Our strategy

[82] Pedersen (2015), p. 47
[83] Chan (2013), p. 4

has a 3-month lag, because a firm's full year earnings report summarizes its annual performance at the year end, but that report is first presented some time later.

Data mining biases are very difficult to avoid in backtests, as Pedersen (2015) explains: "*when you are analyzing a trading idea, you end up looking at a number of different implementations and gravitate toward one that has worked well in the past. Hence, you (consciously or subconsciously) pick this implementation of your trading idea because it has worked well in the past, but you could not have known this back then. Furthermore, some version will have worked the best in the past, perhaps just by chance, but, if this is by chance, it probably will not work well in the future, when you are actually trading on it".*[84] This concept explained by Pedersen, can be considered data-snooping bias. This bias is a result of having too many free parameters that are fitted to market patterns to make historical performance look good.[85] A strategy can be adapted to random market movements to improve investment results, however it does not guarantee that future fluctuations will be happening in a similar manner. The easiest way to avoid data-snooping bias is to keep the strategy as simple as possible, because fewer conditions are less susceptible to bias. Our philosophy was to keep the strategy transparent, objective and systematic such that we avoided manipulating data and the results in our favor. Essentially, look-ahead bias was accounted for and the peer selection process was purely quantitative to avert any prejudice in the strategy.

Survivorship bias is another thing to consider when dealing with historical data. Including delisted stocks in a backtest is very important, because by examining currently listed stocks only, implicitly means that only companies that have survived are researched. Survivorship bias skews returns positively, because a strategy that does not account for these issues will never invest in a company that will later be delisted because of bankruptcy or for any reason. We took this into consideration since we updated our investment universe every year based on the constituents of prominent indices.

Another very important aspect of backtesting is the implementation, because it can have a large impact on profitability. For example, when trading should we consider the bid-ask price or the last price? What about implementation shortfall when trading large positions or less liquid stocks? To begin with, the universe of stocks we considered were all components of large indices, which meant that funds which used the indices as benchmarks were trading the underlying securities continuously. For example, passive index funds that track an index will buy and sell the index constituents constantly

[84] Pedersen (2015), p. 49
[85] Chan (2013), p. 4

in order to replicate the composition of underlying indices. While market liquidity may not be an issue, deciding to use close prices can to an extent be negligence. Some exchanges use closing auctions, like New York Stock Exchange or Hong Kong Stock Exchange, which is an auction to determine the final closing price for investors to trade in for a short period of time after the market closes. So, in theory it is possible to trade at nearly closing prices after the close at the exchanges that offer such auctions.

Trading costs like commission are a very important to consider when investing because they have the potential to ruin a strategy if trading is too expensive. For example, the concept of high frequency trading explained in the "quant"equity would not be possible with high trading costs simply because they would eat large part of the return. However, the advancement of electronic trading has improved liquidity and brought trading prices down. A good example of this is online broker Robinhood, which offers its clients commission free trading. Instead of making money off charging clients for executing a trade, Robinhood makes money from order routing.[86] This concept known as 'Payment for Order Flow', is where brokers are paid fractions of a cent per trade for routing their orders to market makers, such as Citadel Securities.[87] Market makers profit off the spread between buying and selling securities, and are willing to pay for more business. All this means that trading is democratized as never before, and in some cases even costless. Considering our investment strategy has relatively low turnover and does not require constant trading, it seemed reasonable to assume no trading costs in the backtest.

There can be seemingly limitless amount of implementation considerations when executing a backtest, which is why many published articles gloss over these issues. Understandably, focusing on strategy implementation can distract and detract from the main idea of simulating investment strategy performance.[88]

6.2. Choice of Fundamentals

A central reason for using a quantitative approach to investing is the practice of objective decision making. The process of selecting investments in this research is purely systematic and without bias, due to the use of our computer model. That being said, the inputs which the model acts on are based on valuation fundamentals that we choose, and which should best represent characteristics used to compare firms. The peer accuracy tests were meant to determine which fundamentals are useful in

---

[86] Sraders (2019), https://www.thestreet.com/investing/how-does-robinhood-make-money-14856528
[87] Massa (2017), https://www.bloomberg.com/quicktake/payment-for-order-flow
[88] Chan (2013), p. 2

determining peers, however a shortlist needed to be formed to determine which factors could be most accurate for peer selection, as it is be out of the scope of the project to test every single fundamental.

What Makes a Firm Comparable, or Peer?

Aswath Damodaran states that a comparable firm has similar cash flows, growth potential and risk to the firm being valued.[89] Traditionally, an analyst can use an industry or sector to compare firms within that group. However, operating in the same industry does not guarantee similar firm characteristics in terms of growth, risk and cash flows. Selecting peers that operate in the same business naturally limits the amount of available comparables, especially if firm size is considered. Market size also curbs the amount of comparable companies, and if the sector is defined too broadly then differences across firms might also be large.[90]

All fundamentals, with the exception of WACC, used in the peer selection process were derived from companies' financial statements, i.e. no market values/estimates were used. This is due to the subjectivity and uncertainty that comes in when dealing with market expectations about companies' future performance. Moreover, it was decided to average beginning and end-of-year balance sheet numbers to capture company's performance over the year and not just at the year end.

We are reliant on the reported numbers being both true and correct from the firms themselves, as well as Bloomberg reporting accurate data.

6.2.1 Growth

Growth is important to all stakeholders of a firm, and therefore a viable metric to compare firms on. For instance, shareholders view growth as attractive because it allegedly creates value, while lenders and suppliers see growth as a business opportunity.[91]

Revenue (or sales) growth is an important factor for determining peer firms, because fundamentally it shows how a business is growing and whether consumers want to buy the company's goods and services. Nel et al. (2014) implements revenue growth as one of the valuation fundamentals used to

[89] Damodaran (2006), p. 472
[90] Damodaran (2006), p. 472
[91] Petersen, Plenborg & Kinserdal (2017), p. 183

determine peers in emerging markets. They find that revenue growth, especially when combined with other factors, improved valuation accuracy and indicated lower median relative valuation errors. The revenue growth metric is straightforward, and is given by:

$$Revenue\ growth_i = \frac{Revenue_{i,t}}{Revenue_{i,t-1}} - 1$$

where the growth for company i is given by revenue at time t divided by revenue at time $t-1$, i.e. the previous period. The revenue growth used will be from the previous end of year to the end of year. In this case, the growth will be the actual trailing growth and not estimated one. Estimates are largely based on individual company and analyst predictions, which can be subject to bias. Damodaran (2011) identifies that analysts' subjective judgements are often based on guesswork, and that incentives and compensation can influence their ratings and predictions.[92] Therefore, using the most recent growth figures will provide a less subjective approach to comparing firms.

Revenue growth uses the Bloomberg field "TRAIL_12M_NET_SALES" from the most recent year end and the previous year end.

6.2.2. Risk

One of the proxies used for risk, and adopted from Nel et al. (2014) is firm size. In this case, size is given by total assets as reported on the company's financial statements. A company's total assets during the year are given by:

$$\overline{Total\ assets_{i,t}} = \frac{Total\ assets_{i,t} + Total\ assets_{i,t-1}}{2}$$

where total assets for company i are the sum of current and non-current assets indicated on the balance sheet as of time t:

$$Total\ assets_i = Current\ assets_i + Non\ current\ assets_i$$

The reasoning for using total assets as a measure of risk is based on the value that assets have. A firm with a lot of assets, will in most cases be able to sell them if they need cash. Financial distress can be avoided through sale of tangible and intangible assets or restructuring the organization as Jensen (1989) explains in his "privatization of bankruptcy" argument. First, having substantial assets can make

[92] Damodaran (2011), pp. 8-10

it easier for companies to obtain financing like loans or debt sales because the liabilities can be backed by more collateral. Second, even if levered companies are more likely to distress, they are actually not more likely to enter into formal bankruptcy. This is due to the large going concern value which can be proxied by balance sheet value of each asset. In an adverse scenario, debt holders would not let a company go bankrupt, but rather be motivated to restructure and get it out of distress, as otherwise a huge fraction of their value would be lost.[93] Similarly, we have seen a number of bank bailout cases, where governments could not afford large banks going out of business as it would put global economies at risk. Therefore, in our model, we have assumed that the greater the going concern value or value of total assets of a company, the safer it is. Safer in the sense that shareholders can be more certain their money will not be lost completely even in the worst case scenario, as the company would then either be restructured or bailed out by a government.

Total assets are given by the Bloomberg field "BS_TOT_ASSET", and for the non-average tests we used the balance sheet value of total assets reported in the final quarter of the calendar year, because it is a more relevant measure of firm size at the time of selecting investments. In addition, for Europe all TA are calculated in EUR and USD for emerging markets. This ensured that the value could be compared, and different currencies were not causing the model to misinterpret size.

6.2.3. Profitability

Profitability analysis is a huge part of the financial analysis. Companies always strive to generate satisfactory returns for their investors and keep positive relationships with their customers and suppliers through sound economic performance.[94] Achieving positive profit is often viewed as satisfactory, but in reality it is not that simple and company's performance should be evaluated against a number of indicators:[95]

- Required rate of return – positive profit (in terms of positive return on equity) cannot be considered satisfactory unless it exceeds investors' required rate of return;
- Recurring/non-recurring items – profit reached through e.g. the sale of some tangible assets, is not as value creating as profit which comes from company's core business and can be sustained over a long period of time;

[93] Jensen (1989)
[94] Petersen, Plenborg & Kinserdal (2017), p. 139
[95] Petersen, Plenborg & Kinserdal (2017), p. 140

- Performance of peers – high profit in isolation might not look as good when compared to peers who achieved much more during the same period and/or under similar conditions;
- Own prior performance – even negative profit might be viewed positively if it is significantly better from prior years.

The following profitability ratios were chosen to measure the operating profitability of studied companies and account for the above points.

Return on Equity (ROE)

By far the most popular profitability ratio is ROE. A number of academics and analysts such as Alford (1992), Cheng and McNamara (2000), Nel et al. (2014) and Knudsen et al. (2017) to name a few, have used this measure in their articles on valuation. ROE examines profitability from the common equity holders' perspective as it shows returns available to equity-holders:[96]

$$ROE_{i,t} = \frac{Net\ Earnings_{i,t}}{\left(\frac{BVE_{i,t} + BVE_{i,t-1}}{2}\right)}$$

where BVE is company i's book value of common equity and denominator measures its average worth over the year t. The company's earnings are estimated net of preferred dividends and tax.[97] ROE is given by the Bloomberg field "RETURN_COM_EQY".

Although ROE is frequently used in the literature, the fact that the ratio is driven by the effect of financial leverage and return on net operating assets cannot be overlooked.[98] The concern here is that a firm with high financial gearing can declare high ROE despite its actual business performance being rather moderate. It is therefore often very useful to capture that operational business element in isolation. This is where return on invested capital (ROIC) comes into play. The dependence between ROIC and ROE can be shown through introduction of two new variables, namely net borrowing cost (NBC) and net interest-bearing liabilities (NIBL):[99]

$$ROE = ROIC + (ROIC - NBC) * \frac{NIBL}{BVE}$$

where

[96] Petersen, Plenborg & Kinserdal (2017), p. 169
[97] Damodaran (2011), pp. 25-26
[98] Viebig, Poddig & Varmaz (2008), p. 74
[99] Petersen, Plenborg & Kinserdal (2017), p. 168

$$ROIC = operational\ performance \quad \text{and} \quad (ROIC - NBC) * \frac{NIBL}{BVE} = effect\ of\ gearing$$

Therefore, we have chosen ROIC as our next profitability indicator.

Return on Invested Capital (ROIC)

According to Petersen et al. (2017), return on invested capital (ROIC) is a more reasonable measure of overall profitability.[100] This is because ROIC not only shows the returns the company generates from operations, but also the returns available to capital providers:[101]

$$ROIC_{i,t} = \frac{NOPAT_{i,t}}{\left(\frac{IC_{i,t} + IC_{i,t-1}}{2}\right)}$$

where NOPAT is company i's net operating profit after tax in year t and IC stands for invested capital, which is calculated as:

$$IC_i = operating\ assets_i - operating\ liabilities_i = NIBL_i + BVE_i$$

As can be seen from the formulas, ROIC ratio considers investments made by both debt and equity holders and is therefore not affected by changes in capital structure. However, it is important to point out that ROIC might not be comparable among companies due to differences in tax rates or accounting policies.[102] This, together with other limitations of this report, will be touched upon in the discussion chapter.

ROIC is given by the Bloomberg field "RETURN_ON_INV_CAPITAL".

Weighted Average Cost of Capital (WACC)

To make sure that companies that are comparable in terms of the above measures are good peers, their ROE and ROIC could also be evaluated against their corresponding required rates of return. The owners' required rate of return is cost of equity,[103] whereas the most adequate measure reflecting the

[100] Petersen, Plenborg & Kinserdal (2017), p. 169
[101] Petersen, Plenborg & Kinserdal (2017), p. 147
[102] Petersen, Plenborg & Kinserdal (2017), p. 147
[103] Petersen, Plenborg & Kinserdal (2017), p. 171

operating risk of the whole firm is weighted average cost of capital (WACC).[104] WACC can be considered a measure of risk, since firm risk corresponds to its required rate of return. However, the fundamental is also closely related to profitability analysis and was therefore presented in this section. Moreover, WACC was selected for further use as it accounts for the required rate of return to both lenders and shareholders, which is important when dealing with companies that have mixed capital structures. This ratio is calculated as:

$$WACC_{i,t} = \frac{MVE_i}{MVE_i + NIBL_i} * r_{e,i} + \frac{NIBL_i}{MVE_i + NIBL_i} * r_{d,i} * (1 - \tau)$$

Where $r_{e,i}$ and $r_{d,i}$ are company i's required returns on equity and debt, respectively, while MVE and NIBL are their corresponding market values of equity (i.e. market cap) and debt (i.e. net interest bearing liabilities).[105] The calculation can also be extended to account for the cost of preferred equity, when needed.

The formula weights proportional costs of debt and equity employed to finance the company's assets and accounts for the tax deductibility of interest payments by multiplying the second part of the formula by one minus tax rate, τ.[106] The use of market values for this element is necessary as else it would simply not make economic sense. In order for investors to be willing to invest in a company, it should offer more attractive returns compared to its alternatives in the market. As a result, the required rate of return should be estimated using market rates and market value weights.[107] However, the book value of NIBL is often used instead due to many firms making use of bank financing which is only specified in book and not market value terms. While this estimate might work well for mature companies in developed markets, in less stable environments with volatile interest rates and default spreads, market and book values might differ significantly.[108] Nevertheless, we decided to include WACC in our model due to its importance for the profitability analysis and lack of alternatives using book values.

The value of WACC used is given by the Bloomberg field "WACC".

[104] Damodaran (2006), p. 384.
[105] Petersen, Plenborg & Kinserdal (2017), p. 143
[106] Viebig, Poddig & Varmaz (2008), p. 36
[107] Damodaran (2006), p. 143
[108] Damodaran (2006), pp. 147-148

6.3. Choice of Multiples

In relative valuation, assets are assessed based on how similarly they are valued in the market. Contrary to discounted cash flow models, relative valuation has a different philosophy. DCF models try to estimate the intrinsic value of an asset through forecasts of future cash flows, which are largely based on estimates and projections. Intrinsic value is based on uncertain projects and arguable discount rates, and as a hedge fund manager states "you have to figure out where you are relative to everybody else. It's an investment decision overlaid by game theory".[109] Relative valuation provides an assessment of what the asset is worth based on what the market is currently paying for similar assets.[110] However, relative valuation is not a perfect technique because the market can systematically undervalue or overvalue groups of assets. This issue was addressed earlier, where it was discussed that we selected portfolio holdings once a year in order to account for changes in market conditions.

6.3.1. Equity Multiples

Price-to-Book

Bhojraj & Lee (2002), Pedersen (2015) and Knudsen et al. (2017) use price-to-book (hereafter P/B) as a valuation multiple for comparing firms. The P/B ratio relates the market value of equity to the firm's net assets that are available to shareholders. The ratio is a somewhat simplistic measure of the fundamental value, especially considering the issues related to accounting principles, however it still provides useful scaling for market values.[111] Price-to-book is given by:

$$\frac{P}{B} = \frac{ROE \ \times PO \times (1+g)}{r_e - g}$$

where ROE is return on equity, PO is payout ratio, r_e is required return on equity and g is growth. The ratio can be controlled for by using discount rate, growth and ROE. However, as stated with equity measures, P/B can be misleading due to inconsistent capital structures when comparing firms. Nonetheless, P/B is popular among analysts and widely used because it is a good and rather simple valuation parameter.

Essentially, a low P/B ratio (like 1 or under) is insurance for an investor because in the event of business failing, assets would be sold, and investor would get his/her money back. The P/B ratio's

[109] Hooke (2010), p. 217
[110] Damodaran (2006), p. 447
[111] Pedersen (2015), p. 136

proxy to value is widely researched and used by practitioners. Fama & French (1993) and Pedersen (2015) illustrate that high value stocks have tended to outperform low value over the long term.

The price per share is given by Bloomberg field "PX_LAST" and book value "BOOK_VAL_PER_SH".

Price-to-Cash Flow

The second equity ratio that was applied to the model was price-to-cash flow (hereafter P/CF). P/CF is given by:

$$\frac{P}{CF} = \frac{Price\ per\ share}{Operating\ Cash\ Flow\ per\ share}$$

where operating cash flow is calculated as $Net\ Income + Depreciation + Amortization + Changes\ in\ Non\ cash\ Working\ Capital$, and is calculated on a trailing 12-month basis when available. Cash flow per share is a measure of a firm's financial strength which represents the net cash a firm produces, on a per share basis.[112]

Using a cash flow multiple can be preferable because cash flows are generally linked to value (e.g. discounted cash flow valuations). Compared to earnings-based multiples, cash flow is not noisy measure of value change and will not be skewed by potentially irrelevant historical costs.[113] Furthermore, cash flows are more difficult to manipulate than earnings, e.g. through depreciations and other non-cash items, and can therefore provide a more realistic representation of a firm's value generation.

The P/CF multiple can also be useful for valuing companies that have a positive operating cash flow, but negative earnings that may be a result of incurred large non-cash expenses.

The price is given by Bloomberg field "PX_LAST" and cash flow "TRAIL_12M_CASH_FLOW_PER_SH".

[112] Bloomberg (2019)
[113] Pandey (2012), p. 442

6.3.2. Firm Multiples

The firm multiples considered in this section are income statement heavy, meaning that they are more focused on how effective a company is at generating income and the value of the revenues, and not influenced by the capital structure.[114]

To calculate enterprise value on the 30th April, the date of selecting portfolios, we used the enterprise value at the end of the previous year and adjusted for the change in market capitalization over the four months. The difference between the products of the previous price per share with previous shares outstanding and the current price and current shares outstanding resulted in the change in market cap.

Enterprise Value to Sales

Contrary to equity multiples, enterprise or firm multiples value the whole firm or its operating assets.[115] Firm multiples are more flexible than equity multiples because they deal with leverage and thus take into account a firm's debt ratio.

Bhojraj & Lee (2002) analyze the use of the enterprise value-to-sales (hereafter EV/Sales) in selection of comparable firms, and they find sharp improvements to selection compared to other techniques, such as industry classification.

$$\frac{EV}{Sales} = \frac{ROIC - g}{WACC - g} \times \frac{1}{ROIC} \times (1 - t) \times (1 - depreciation\ rate) \times EBITDA\ margin$$

The EV/Sales multiple has several distinct advantages, particularly when comparing many different companies with various characteristics. For example, EV/Sales is strictly positive as sales cannot be negative and they represent a gross figure. With this ratio all companies can be compared, even if they have a negative bottom line. The obvious drawback to this multiple is the breadth and lack of more precise accounting figures that detail the position of firm’s earnings. This becomes especially important when comparing companies that have positive and negative net earnings.

The sales figure is given by the Bloomberg field “TRAIL_12M_NET_SALES”.

[114] Hooke (2010), p. 223
[115] Damodaran (2006), p. 553

Enterprise Value to EBITDA

The enterprise value-to-earnings before interest, tax, depreciation and amortization (hereafter EV/EBITDA) is a popular multiple and widely used in corporate finance. EBITDA and EBIT are quite similar and both are used in relative valuation. However, EBIT adjusts for non-cash charges from the depreciation and amortization, or write-downs of balance sheet assets. Naturally, EBITDA is more likely to be positive because EBIT has a larger charge to the revenue. This also means that our dataset is less limited, since we decided to exclude negative multiples. Another reason for selecting EV/EBITDA relates to Lie & Lie (2002) finding that the EBITDA multiple generally yields better results in assessing company values compared to EV/EBIT.[116] The multiple can be seen as:

$$\frac{EV}{EBITDA} = \frac{ROIC - g}{WACC - g} \times \frac{1}{ROIC} \times (1 - \tau) \times (1 - depreciation\ rate)$$

where growth is given by g, and the tax rate by τ. The multiple is a function of value creation or efficiency at appropriating capital to profitable investments measured by ROIC. Also, the multiple is a function of such discount rate as WACC, and the depreciation rate. Hooke (2010) states that the multiple incorporates growth and risk aspects of a stock, which is depicted in the above equation. EV/EBITDA also considers the cost of revenue, unlike EV/Sales.

EBITDA can be useful to analyze and compare profitability between companies and industries because it eliminates the effects of financing and accounting decisions.[117] Hooke claims that Wall Street defines operating cash flow as EBITDA, which means EV/EBITDA and P/CF have similar denominators.[118] The EV multiple is clearly a firm multiple and considers capital structure, while the equity multiple does not take leverage into account.

The value of EBITDA is given by the Bloomberg field "EBITDA".

Multiples – Wrap up

The main issue with using multiples is that they do not tell investors whether a sector or group of companies is cheap or expensive at a specific point in time.[119] Relating this problem to our investment strategy means that our model may choose a company to invest in because it is cheap compared to its

[116] Lie & Lie (2002), p. 53
[117] Bloomberg (2019)
[118] Hooke (2010), p. 229
[119] Hooke (2010), p. 224

particular group of peers, but those peers may all be expensive. The model does not use any historical multiples to determine justifiable relative value levels because that can require speculating in macroeconomic outlooks. Businesses constantly change and geopolitics can affect industries and countries continuously, making historical multiples potentially irrelevant.

The usage of relative valuation multiples in our model enables us to compare firms with their peers in the current business landscape with the belief that company selection will outperform in all market conditions. For example, consider company XYZ that has price-to-book ratio of 3x whereas the average P/B for XYZ's peer group is 4x (Market values of equity are 4 times the book values). In normal market conditions where we assume that the peer group average is priced fairly, we would expect a 33% upside in XYZ. Now consider that there are many geopolitical risks such as the peer group's P/B should actually be 2x, which means that the peer group market values will collectively decline by 50%. Expecting XYZ to revert to the mean P/B would imply a decline of 33%, however that is still an outperformance of 17% compared to the peer group.

The multiples described in this section are all applicable to money-losing firms, such as companies with negative net earnings. Generally speaking it may make more sense to compare firms on multiples with fewer variables that have been adjusted for costs like taxes, depreciation, etc., but there are a few good reasons to use the chosen multiples: (1) casting a wider net thus having more observations, (2) our model will not choose young high growth companies that may be losing money as a comparable to an established companies because their valuation fundamentals will differ significantly. Negative net earnings do not necessarily make a company an unattractive investment opportunity, so we would not want to exclude them. Furthermore, companies in the S&P 500 and STOXX Europe 600 are typically established companies that have proven their operational abilities to become part of the index, so negative earnings produced by one of these companies can occasionally be due to one-off charges, in which case we would still want to have the company in the dataset.

Relative valuation is only effective when the firms have true comparables, because varying characteristics will misrepresent the relative value. The problem is that no company has a perfect peer on every accounting level. The model used in this research will test which of the chosen multiples are the most accurate in selecting cheap firms relative to their peer groups, and whether that translates to a superior portfolio performance.

6.4. Descriptive Statistics

6.4.1. SPX Index

Table 6.4

		Average								
Year	**Available Companies**	**ROE (%)**	**ROIC (%)**	**WACC (%)**	**TA ($ millions)**	**G (%)**	**EV/Sales**	**EV/EBITDA**	**P/CF**	**P/B**
2003	152	24.98	11.43	8.13	31,231	12.30	2.79	44.75	13.78	4.24
2004	165	20.74	12.18	7.66	34,383	13.73	2.76	40.50	9.72	4.55
2005	183	22.08	13.13	8.50	26,519	13.38	2.73	42.64	16.79	3.89
2006	196	22.81	13.56	9.36	28,953	12.20	2.82	13.98	12.95	4.93
2007	209	21.61	13.33	9.34	29,829	12.79	2.53	39.72	10.25	3.66
2008	222	17.36	11.24	9.40	27,947	10.76	1.79	34.51	7.04	2.76
2009	242	14.57	9.83	9.27	35,135	-9.26	2.69	40.09	13.45	3.04
2010	260	20.54	11.39	9.72	35,914	12.93	3.12	35.76	4.53	3.95
2011	289	19.11	11.24	9.59	36,635	12.22	2.92	48.63	12.17	5.74
2012	294	17.58	4.03	8.79	44,691	4.90	3.30	30.97	15.18	6.39
2013	303	17.58	10.08	8.65	46,021	5.90	3.67	41.09	15.07	3.91
2014	301	20.81	10.38	8.30	47,253	7.04	3.98	39.14	20.57	5.75
2015	298	19.50	8.43	7.80	48,240	-0.73	4.20	90.55	18.37	5.80
2016	304	21.85	9.33	7.46	49,287	3.22	4.52	61.54	9.27	6.51

Source: Bloomberg

The table above illustrates the historical development of the key factors used for the S&P 500 index. The number of available companies is based on the companies in the index with all data available across factors. The beginning years have the fewest available observations, but the dataset is growing throughout the years on average. We can identify times of crisis and market decline by identifying the negative growth years which occurred in 2009 and 2015. Throughout the 17 years of data, WACC had remained relatively constant within the range of 7 - 9. Total assets are generally growing except around the financial crisis. While table 6.4 describes the data set that is used in this research, it may not be an accurate representation of the total S&P 500 index because of the exclusion of companies that did not meet the criteria, as explained previously.

6.4.2. SXXP Index

Table 6.5

		Average								
Year	**Available Companies**	**ROE (%)**	**ROIC (%)**	**WACC (%)**	**TA (€ millions)**	**G (%)**	**EV/Sales**	**EV/EBITDA**	**P/CF**	**P/B**
2003	89	16.50	11.22	7.74	11,327	8.17	1.80	24.01	10.76	2.99
2004	122	22.85	12.20	7.19	15,435	6.93	2.05	23.39	9.51	3.12
2005	148	20.83	12.72	7.42	17,214	12.29	2.49	30.56	10.53	3.46
2006	189	23.56	14.44	8.23	20,286	13.01	2.55	32.61	16.78	4.12
2007	224	23.58	13.39	8.95	21,003	10.88	2.56	19.09	11.70	4.90
2008	230	19.58	11.29	8.71	22,458	12.40	1.76	7.19	9.88	3.05
2009	202	13.60	7.68	7.86	24,652	-7.58	1.95	3.47	10.35	3.97
2010	237	19.43	10.77	9.20	25,695	9.33	2.08	29.60	15.34	3.52
2011	344	16.80	11.40	10.06	24,961	10.96	2.34	25.35	13.35	2.97
2012	311	16.49	14.51	8.61	24,955	6.59	2.89	29.28	13.48	5.08
2013	309	18.78	38.29	8.37	24,370	11.37	3.14	32.24	68.83	4.88
2014	312	22.07	43.22	8.05	29,300	4.76	3.42	20.01	18.85	9.38
2015	314	14.97	9.90	7.64	31,770	9.66	3.20	16.87	11.90	4.31
2016	354	15.20	9.70	7.38	32,221	4.06	3.39	43.99	17.40	4.31

Source: Bloomberg

The table above illustrates the historical development of the key factors used for the STOXX Europe 600 index. The available company financials are increasing over time, where a little over half the index constituents are included in our dataset towards the last years. ROE increased until the financial crisis, where growth in 2009 was about -7.6%. ROIC mirrors the trend of ROE, and in the latest couple of years returns on equity and invested capital have been lower amid tougher euro conditions. The cost of capital for the firms in the index are between 7 - 8 in most years. Total assets are increasing and are lower compared to constituents of S&P 500 even after accounting for EUR/USD rates. The index members for SXXP vary between small, mid and large cap and are more diverse than S&P 500 which explains the lower average of total assets. EV/Sales has generally been increasing overtime, suggesting that valuations have increased faster as a function of revenues.

6.4.3. MXEF Index

Table 6.6

		Average								
Year	**Available Companies**	**ROE (%)**	**ROIC (%)**	**WACC (%)**	**TA ($ millions)**	**G (%)**	**EV/Sales**	**EV/EBITDA**	**P/CF**	**P/B**
2003	69	16.99	12.50	6.60	2,990	10.80	2.61	34.39	13.92	2.45
2004	91	21.09	14.00	6.39	3,028	18.96	2.31	35.07	14.92	2.76
2005	113	21.68	14.76	7.52	4,155	16.46	2.57	35.27	14.72	3.43
2006	132	21.43	13.22	8.56	7,055	19.94	3.09	39.11	14.69	3.51
2007	152	23.23	14.39	9.13	8,515	19.94	3.12	36.97	13.54	3.26
2008	118	20.98	15.37	10.57	12,812	18.33	2.10	26.30	8.92	2.39
2009	168	19.24	12.90	10.85	14,153	5.46	3.87	55.38	14.45	3.57
2010	181	23.11	14.91	11.34	13,260	24.07	3.73	46.62	19.76	4.20
2011	301	18.42	11.65	11.28	12,344	17.75	2.41	43.32	17.12	3.30
2012	371	15.94	10.25	10.00	13,923	12.85	2.71	34.57	15.23	3.30
2013	433	14.26	8.38	9.53	17,306	9.05	2.46	38.68	16.35	2.95
2014	425	13.47	9.09	9.52	17,852	8.28	2.65	45.23	27.29	2.99
2015	418	13.59	8.23	9.40	17,539	9.63	2.62	44.77	17.46	2.96
2016	423	13.54	9.15	9.26	15,710	6.11	2.79	27.94	20.22	3.08

Source: Bloomberg

The table above illustrates the historical development of the key factors used for the MSCI Emerging Markets index. The index is characterized by the development of emerging market countries and the continuous amendments to the index constituents. As previously explained, countries included in the index are reviewed frequently, so the composition is constantly changing, and therefore historic vs current comparisons are not as relevant compared to SPX & SXXP indices. That being said, amount of companies in the data set increases significantly across the period. Interestingly, ROE and ROIC are broadly decreasing over time suggesting firms are less effective at deploying capital. Growth fluctuates significantly over the period, and despite the financial crisis growth was still positive. The financial crisis is reflected in the multiple valuations where the 4 multiples are lower in 2008 than in the pre and post crisis years.

6.4.4. Descriptive Statistics – Conclusion

The fundamentals and multiples within each index vary, as it was shown in the previous sections. The S&P 500 is characterized by large companies where the emerging markets index is the most diverse in terms of size. This is also illustrated in total assets, where SPX is roughly 3.5 times larger than MXEF in 2016. Over the period of the dataset MXEF assets have grown 425% compared to 58% for SPX and 184% for SXXP. ROE is highest on average and most steady in the United States. Growth rates of SPX and SXXP firms are closely related, though there is a notable exception in 2015 where growth in the US went negative while Europe gave roughly 9.5%. On the other hand, the emerging markets firms are growing much quicker as their economies are developing rapidly.

In terms of enterprise relative valuations, U.S. companies have historically been more expensive than European and emerging market companies. SXXP constituents generally have the lowest EV/Sales and EV/EBITDA multiples suggesting they are cheap. However, the price an investor has to pay for equivalent cash flows is lowest in the U.S. at 12.80 and highest in Europe at 17.05. When looking at P/B, U.S. companies appear more expensive with P/B ratios over 4 where EM have historically been at 3.15 on average. Granted, it is difficult to state that one region or group of companies is cheaper than another, because many factors like interest rates, politics and economic health play into the valuations.

7. Peer Analysis

In the previous sections, we outlined the valuation fundamentals that were the factors used to determine peer groups for each company based on profitability, risk and growth. We have run the backtest simulations 6,048 times in order to determine the portfolio of chosen stocks for every year, strategy, sample and market. Before delving into the performance in terms of return and risk compared to the benchmark, we had to analyze how effective the SARD model was in determining comparable peer groups.

7.1. Evaluating Peer Selection

The first part of evaluating peer selection was measuring the absolute percentage error of the base company's multiple relative to the average across the peer group. The absolute percentage error is used in Knudsen et al. (2017), Alford (1992) and Nel et al. (2014) because it is simple to understand and has a useful economic interpretation.

$$Absolute\ percentage\ error_i = \left|\frac{\overline{m}_{x,i} - m_{x,i}}{\overline{m}_{x,i}}\right|$$

Where $\overline{m}_{x,i}$ is the average peer group multiple x of company i, and $m_{x,i}$ is the base company multiple x for company i. Essentially, the absolute percentage error shows how different the base company is from the mean. The strategy was designed to select the most undervalued companies relative to the average, with the expectation of mean reversion. Theoretically, in a perfect world, the average multiple would be the true multiple with which the base company should trade at, so if the company was underpriced then we would expect the base company to revert towards the mean multiple. This means that the absolute percentage error is expected to decrease one year after the initial investment, assuming that the selected peer group and the base company will remain comparable over the investment period. To measure the selection accuracy, we are going to compare the absolute percentage error at the time of investment and one year later, where the base company multiple in 1 year and the average multiple in one year will be used. Markets tend to go up on average, so if we only compared the average multiple at time t with the base company at time t and t+1, the results would likely be skewed and they would not provide the real picture. Comparing our study to Alford's, he does not take multiple valuation method for granted but rather aims to test it assuming that the observable market stock prices of the studied companies fully reflect their fundamentals. In our model, we are saying that the value should be determined by multiples. And if relative values of comparable companies are not the same, then this is due to mispricing and not because of inaccurate selection criteria. However, we of course cast doubt on our assumptions and resort to significance tests for statistical evidence.

As it was just highlighted, the second part of evaluating requires testing the significance of the absolute percentage errors. The expectation of converging towards the mean suggests that the absolute percentage error at the time of investment less the error 1 year later should be positive:

$$APE\ difference = APE_t - APE_{t+1}$$

The positive difference would mean the absolute percentage error between the base company and its peers decreased over the year and vice versa. To determine whether these differences are significant, Wilcoxon's signed rank test will be used.

The Wilcoxon Signed Rank Test is a nonparametric test for paired data. It analyzes the signs of the differences, but also accounts for their magnitude.[120] The paired data to analyze will be the absolute percentage error at time t with t+1. The null and alternative hypotheses of the test are given below:

H_0 = the difference between absolute ranked differences is 0

H_A = the difference between absolute ranked differences is positive

The test statistic W is given by:

$$W = \sum_{i=1}^{N_r} [sign(\varepsilon_{t,i} - \varepsilon_{t+1,i}) \times R_i]$$

Where N_r is the sample size, $\varepsilon_{t,i}$ is the absolute percentage error at time t, $\varepsilon_{t+1,i}$ is the absolute percentage error at time t+1, and R_i is the rank of the absolute difference in the errors of company i in the sample. Wilcoxon's test can be calculated as one or two sided, but as we are only interested in identifying whether the differences were positive, one sided test was deemed sufficient. Therefore, we need to calculate W-, i.e. sum the R_is (ranks) of negative differences in APEs. Let us use SPX strategy 1 for 2004 (see table 7.1) as an example:

$$W\text{-} = 1 + 6 + 5 + 2 + 19 = 33$$

Finally, the z-score for the sample is calculated by:

[120] LaMorte (2017), http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Nonparametric/BS704_Nonparametric6.html

$$z\text{-}score = \frac{W\text{-} - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Where n is the number of observations. Using the data from table 7.1 the z-score is -2.69. A negative z-score means that the sum of signed ranks is primarily positive, and the z-score is deemed significant at 1% level if it is below -2.33, meaning that the null hypothesis can be rejected.

Table 7.1 – Wilcoxon Signed Rank Test Example (2004)

Company	APE (t)	APE (t+1)	APE difference	Rank
MCKESSON CORP	0.956	0.953	0.003	3
AON PLC	0.989	0.519	0.470	20
BEMIS COMPANY	0.675	0.651	0.024	8
GOODRICH CORP	0.674	0.586	0.087	17
WW GRAINGER INC	0.760	0.673	0.086	16
HUMANA INC	0.920	0.879	0.041	10
JOHNSON CONTROLS INC	0.803	0.750	0.053	13
LEGGETT & PLATT INC	0.754	0.697	0.057	14
NORTHROP GRUMMAN CORP	0.748	0.749	-0.001	1
RAYTHEON COMPANY	0.688	0.706	-0.018	6
RYDER SYSTEM INC	0.767	0.722	0.045	11
MARATHON OIL CORP	0.838	0.813	0.025	9
WHIRLPOOL CORP	0.770	0.748	0.022	7
CARDINAL HEALTH INC	0.886	0.897	-0.011	5
AMERISOURCEBERGEN CORP	0.959	0.961	-0.002	2
LEXMARK INTERNATIONAL INC-A	0.672	0.625	0.047	12
VALERO ENERGY CORP	0.824	0.765	0.059	15
AETNA INC	0.810	0.801	0.009	4
RS LEGACY CORP	0.719	0.826	-0.107	19
EXPRESS SCRIPTS HOLDING CO	0.881	0.793	0.088	18

Source: Authors' findings

In the following sections, the Wilcoxon signed ranked test results will denote the sign of the strategy on average, where a plus sign will illustrate the strategy followed the alternative and a minus sign will mean we cannot accept the alternative hypothesis. In addition, the significance of the signs across 1%, 5% and 10% levels will be given.

Finally, it is important to note that the Wilcoxon Signed Rank Test is calculated based on the 20 investments selected yearly as a sample. We want to identify if the significance of peer selection also translates into better portfolio returns, which will be discussed in the next chapter. However, the mean and median absolute percentage errors are based on the whole sample for the given year. This will indicate how the peer selection accuracy was on average including both under and overvalued companies.

7.2. SPX Strategies

7.2.1. Non-average & 6 Peer Sample

Table 7.2.1 below illustrates the peer selection results for the S&P 500 using the 1-year (i.e. non-average) fundamentals and 6 peers for the investment selection process.

Table 7.2.1 – Peer Selection Accuracy

(SPX, non-average, 6 peers)

	ROE	**TA**	**G**	**ROIC**	**WACC**	**ROE TA**	**ROE TA G**	**ROIC TA G WACC**	**ROE TA G WACC ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.341 (5)	1.371 (9)	1.345 (6)	1.346 (7)	1.370 (8)	1.271 (4)	1.252 (3)	1.206 (2)	1.155 (1)
Median	0.528 (9)	0.519 (8)	0.485 (5)	0.501 (6)	0.502 (7)	0.478 (4)	0.472 (3)	0.439 (1)	0.464 (2)
EV/EBITDA									
Mean	0.585 (5)	0.671 (9)	0.628 (8)	0.613 (7)	0.604 (6)	0.551 (4)	0.546 (3)	0.522 (1)	0.528 (2)
Median	0.326 (8)	0.330 (9)	0.304 (5)	0.310 (6)	0.312 (7)	0.302 (4)	0.286 (3)	0.276 (2)	0.275 (1)
P/CF									
Mean	0.769 (7)	0.735 (6)	0.784 (9)	0.773 (8)	0.702 (5)	0.692 (4)	0.666 (3)	0.590 (1)	0.601 (2)
Median	0.397 (9)	0.394 (8)	0.378 (5)	0.392 (7)	0.366 (4)	0.379 (6)	0.350 (3)	0.315 (2)	0.309 (1)
P/B									
Mean	0.564 (4)	1.018 (8)	1.029 (9)	0.669 (6)	0.938 (7)	0.530 (2)	0.533 (3)	0.619 (5)	0.513 (1)

Median	0.328 (4)	0.478 (8)	0.516 (9)	0.371 (6)	0.476 (7)	0.311 (3)	0.304 (2)	0.344 (5)	0.291 (1)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+*	+*	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+*	+	+*	+	+	+*	+
3. Average SARD score of the 20 company sample									
EV/Sales	2.020	2.002	2.021	2.043	2.003	19.091	49.982	82.618	112.119
EV/EBITDA	2.030	2.016	2.037	2.057	2.026	19.837	52.169	86.950	117.060
P/CF	2.026	2.016	2.052	2.046	2.039	19.286	50.786	84.888	117.290
P/B	2.029	2.011	2.026	2.043	2.017	20.033	51.239	85.196	118.790

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

EV/Sales has a much higher variation between mean and median compared to the other 3 multiples, however this is also to be expected because the multiple is using numbers highest up on the income statement. Peer selection using more than 1 fundamental (strategies 6 – 9) mostly see improvements in the absolute percentage error, whereas strategies 8 and 9 typically perform best on this metric across the multiples. EV/EBITDA has the lowest absolute percentage error relative to the other multiples. P/CF is similar to EV/EBITDA except that P/CF is an equity multiple and does not account for capital structure, which may be a reason for EV/EBITDA performing best in this case. Notably, all tests have a positive trend on average, suggesting that the model works in all cases, however the significance levels vary. P/CF and EV/EBITDA are significant at the 5% level for each strategy, which may be related to the fact that they include the more detail from financial statements compared to EV/Sales and P/B. Strategy 5 using WACC is significant at 10% or better across all strategies which indicates that cost of capital is a useful indicator for determining peers because WACC accounts for capital structure and implicit risk.

The average SARD scores for strategies 1 – 5 will always be close to 2 when using 6 peers, so we are more interested in the last 4 strategies with multiple factors. A lower SARD score is considered better because it means that the ranks across fundamentals are closer to the base company. Since the model is selecting the 20 most undervalued companies, a low score means the base company is close to the peers in terms of fundamentals and we would not expect a skewed average multiple. A high SARD means that the base company is not similar to its peers, so we should not rely on the accuracy of the base multiple compared to the average one. In general, the SARD scores for each strategy are similar for each multiple, but EV/Sales has the lowest score for each test.

7.2.2. Non-average & 12 Peer Sample

Selecting 12 peers for each company provides more comparables to evaluate the base company's value against, however with the drawback that each additional peer selected is farther away from the base as measured by the SARD. The underlying reasoning for backtesting with more peers relates to the stability of selection criteria. With few peers, a single outlier, i.e. skewed data point, can misrepresent how a company is valued compared to the peer group.

Table 7.2.2 below illustrates the peer selection results for the S&P 500 using the 1-year (i.e. non-average) fundamentals and 12 peers for the investment selection process.

Table 7.2.2 – Peer Selection Accuracy

(SPX, non-average, 12 peers)									**ROE**
								ROIC	**TA**
							ROE	**TA**	**G**
						ROE	**TA**	**G**	**WACC**
	ROE	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.319 (6)	1.365 (9)	1.334 (7)	1.308 (4)	1.344 (8)	1.319 (5)	1.252 (3)	1.183 (2)	1.159 (1)
Median	0.505 (9)	0.501 (8)	0.474 (4)	0.493 (7)	0.488 (5)	0.491 (6)	0.452 (2)	0.443 (1)	0.455 (3)
EV/EBITDA									
Mean	0.586 (7)	0.582 (5)	0.601 (8)	0.624 (9)	0.582 (6)	0.540 (4)	0.525 (3)	0.508 (2)	0.506 (1)
Median	0.297 (5)	0.314 (9)	0.312 (8)	0.306 (7)	0.304 (6)	0.289 (4)	0.281 (3)	0.275 (1)	0.281 (2)
P/CF									
Mean	0.770 (9)	0.722 (6)	0.744 (8)	0.735 (7)	0.673 (4)	0.684 (5)	0.636 (3)	0.580 (2)	0.580 (1)
Median	0.389 (9)	0.382 (7)	0.369 (6)	0.386 (8)	0.362 (4)	0.368 (5)	0.340 (3)	0.310 (1)	0.317 (2)
P/B									
Mean	0.552 (4)	0.962 (8)	1.001 (9)	0.632 (6)	0.903 (7)	0.519 (2)	0.530 (3)	0.611 (5)	0.508 (1)
Median	0.317 (4)	0.500 (8)	0.520 (9)	0.362 (6)	0.490 (7)	0.298 (3)	0.297 (2)	0.353 (5)	0.289 (1)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+*	+*	+	+	+*	+	+	+	+

EV/EBITDA	+*	+**	+**	+**	+**	+**	+**	+**	+**
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+*	+	+*	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	3.570	3.521	3.569	3.543	3.521	26.238	62.767	100.885	134.274
EV/EBITDA	3.638	3.568	3.649	3.674	3.564	27.322	65.390	105.940	140.014
P/CF	3.609	3.624	3.685	3.612	3.578	27.126	64.795	103.937	137.633
P/B	3.618	3.557	3.585	3.626	3.558	27.630	65.343	104.907	139.935

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The absolute percentage errors are quite similar to the tests with 6 peers, where EV/Sales has the highest mean because it is also the least detailed multiple. On the other hand, this 12-peer test is notably more different on the single factor tests compared to 6 peers, where strategy 1 is now significant for EV/Sales but less significant for EV/EBITDA. While absolute percentage error in EV/Sales is still higher, the added number of peers may have added stability to the selection thus rendering the multiple more useful. WACC is again significant across all multiples which signals that the factor can find undervalued companies, but the absolute percentage error is relatively high, so it may also be more probable to have improvements. Strategies 6 through 9 have lower absolute percentage errors compared to the single factor tests on average for all multiples. This signals that peer selection is more accurate using more than one fundamental to determine peers, but more factors do not select closer peers in every instance. Strategy 9 that includes all five fundamentals, has the lowest absolute percentage error in 5 out of 8 cases in line with the general trend of improvement with the addition of more fundamentals. Despite seeing a positive progression in absolute percentage errors, EV/EBITDA and P/CF were the only multiples to display significant improvement of the base company with its respective peer group.

The average SARD scores for single fundamentals are roughly 3.5, and higher than the 6-peer average due to more companies that are farther away based on the ranks. Like in the test with 6 peers, this experiment has the lowest SARD score for EV/Sales.

7.2.3. Average & 6 Peer Sample

This test uses 3-year average values for each of the fundamentals, and the peer group consists of 6 companies. The 3-year average tests are designed to adhere a sense of stability across each firms' financial characteristics. This balanced approach tests whether peer selection is more accurate when a longer operating period is considered. Firms may report substantial performance changes due to non-recurring events when analyzing a short horizon. By using 3 years approach, we seek to minimize the impact of an extraordinary earnings release that deteriorates the actual picture of financial strength. However, a company's performance 3 years ago may not be an accurate representation of the current situation, thus voiding the selection method.

Table 7.2.3 – Peer Selection Accuracy

(SPX, average, 6 peers)									**ROE**
								ROIC	**TA**
							ROE	**TA**	**G**
						ROE	**TA**	**G**	**WACC**
	ROE	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.321 (5)	1.354 (7)	1.388 (9)	1.335 (6)	1.359 (8)	1.288 (4)	1.274 (3)	1.177 (2)	1.139 (1)
Median	0.521 (9)	0.503 (6)	0.478 (4)	0.506 (7)	0.500 (5)	0.510 (8)	0.468 (3)	0.438 (1)	0.445 (2)
EV/EBITDA									
Mean	0.625 (7)	0.617 (5)	0.620 (6)	0.635 (8)	0.643 (9)	0.565 (3)	0.538 (1)	0.575 (4)	0.554 (2)
Median	0.318 (9)	0.309 (7)	0.306 (4)	0.309 (6)	0.316 (8)	0.307 (5)	0.293 (3)	0.278 (2)	0.271 (1)
P/CF									
Mean	0.754 (7)	0.714 (6)	0.793 (9)	0.754 (8)	0.712 (5)	0.673 (4)	0.664 (3)	0.591 (1)	0.595 (2)
Median	0.392 (8)	0.384 (7)	0.382 (6)	0.393 (9)	0.363 (5)	0.358 (4)	0.346 (3)	0.325 (2)	0.318 (1)
P/B									
Mean	0.583 (4)	1.004 (7)	1.031 (8)	0.666 (6)	1.058 (9)	0.547 (1)	0.566 (3)	0.638 (5)	0.559 (2)
Median	0.343 (5)	0.482 (8)	0.516 (9)	0.380 (6)	0.472 (7)	0.322 (3)	0.310 (2)	0.340 (4)	0.294 (1)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+	+*	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**

P/B	+	+	+	+	+*	+	+	+*	+
3. Average SARD score of the 20 company sample									
EV/Sales	2.026	2.005	2.019	2.034	2.001	18.384	49.128	82.243	113.343
EV/EBITDA	2.021	2.025	2.061	2.042	2.006	19.546	51.664	87.911	115.831
P/CF	2.028	2.030	2.046	2.020	2.026	19.134	49.728	83.458	113.853
P/B	2.033	2.011	2.032	2.029	2.021	19.427	50.996	84.080	114.563

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

Comparing the results of average and non-average tests shows relatively similar results for the absolute percentage errors. While errors are not exactly the same, mean and median values follow the non-average test quite closely across strategies. Reasons for this phenomenon may be a result of companies in the sample providing stable performance, so similar peer groups are being selected.

EV/EBITDA and P/CF are again significant for all strategies at the 5% confidence level. WACC is also significant for all the multiples and compared to the mean values from the non-average tests, the maximum difference is 0.004 in terms of absolute percentage error. WACC appears to be an inherently stable fundamental, since a large capital structure shift or individual cost of capital would be needed to change WACC considerably.

The SARD scores of this test reflect the outcomes of the previous tests, where EV/Sales has the lowest score across the multi factor tests.

7.2.4. Average & 12 Peer Sample

This test uses 3-year average values for each of the fundamentals, and the peer group consists of 12 companies. Of the four peer selection tests, it would be expected that this test yields the most stable accuracy results since averages of fundamentals and larger peer groups should help reduce the impact of outliers.

Table 7.2.4 – Peer Selection Accuracy

(SPX, average, 12 peers)

									ROE
								ROIC	**TA**
							ROE	**TA**	**G**

(Strategy)	ROE (1)	TA (2)	G (3)	ROIC (4)	WACC (5)	ROE TA (6)	TA G (7)	G WACC (8)	WACC ROIC (9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.283 (4)	1.344 (8)	1.350 (9)	1.298 (6)	1.339 (7)	1.287 (5)	1.238 (3)	1.158 (2)	1.141 (1)
Median	0.507 (8)	0.508 (9)	0.475 (4)	0.487 (6)	0.485 (5)	0.489 (7)	0.450 (3)	0.431 (1)	0.433 (2)
EV/EBITDA									
Mean	0.624 (9)	0.586 (5)	0.594 (6)	0.613 (8)	0.607 (7)	0.528 (2)	0.530 (4)	0.529 (3)	0.526 (1)
Median	0.307 (7)	0.301 (5)	0.309 (8)	0.307 (6)	0.314 (9)	0.294 (4)	0.281 (3)	0.268 (1)	0.271 (2)
P/CF									
Mean	0.767 (9)	0.703 (6)	0.765 (8)	0.728 (7)	0.676 (5)	0.656 (4)	0.650 (3)	0.584 (2)	0.576 (1)
Median	0.391 (9)	0.372 (7)	0.370 (6)	0.386 (8)	0.354 (4)	0.363 (5)	0.341 (3)	0.309 (1)	0.316 (2)
P/B									
Mean	0.561 (4)	0.965 (7)	1.009 (9)	0.659 (6)	0.981 (8)	0.532 (1)	0.539 (2)	0.632 (5)	0.549 (3)
Median	0.327 (4)	0.502 (8)	0.531 (9)	0.368 (6)	0.476 (7)	0.306 (2)	0.312 (3)	0.344 (5)	0.302 (1)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+*	+*	+	+*	+*	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+*
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+*	+*	+*	+	+	+*	+
3. Average SARD score of the 20 company sample									
EV/Sales	3.571	3.520	3.566	3.580	3.514	26.084	63.233	99.314	133.678
EV/EBITDA	3.587	3.555	3.621	3.605	3.524	27.658	64.763	104.351	137.151
P/CF	3.624	3.604	3.644	3.559	3.557	26.828	63.148	101.569	134.047
P/B	3.614	3.556	3.570	3.633	3.568	27.390	64.502	100.517	133.096

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The absolute percentage errors generally follow the trend of improvement as more fundamentals are combined to select peers. Though ROE + TA shows a lot of variation in terms of the rankings across multiples. This test appears to have quite similar results as the average test with 6 peers, which suggests that peer selection using averages does not require more peers to provide stable results. In this case, using both averages and 12 peers seems to be redundant.

A general theme of P/B is that absolute percentage errors are highest for TA, G and WACC when looking at means in this sample and previous samples. These factors for P/B also appear to have the highest variation, because the median absolute percentage error shows a significant improvement to the accuracy, illustrating that tails are fat.

The Wilcoxon signed ranked test shows a few interesting points for this peer analysis test. Firstly, the return fundamentals (ROE & ROIC) have a significant improvement of errors using 12 peers compared to 6 peers and averages. While it was previously stated that comparing absolute percentage errors for the two samples of averages (6 & 12 peers) showed similar results, this outcome illustrates that having more peers adds a degree of stability such that the conclusion of accuracy is not just due to luck. Secondly, adopting more peers presents more error improvements particularly on single factor tests. Strategy 1, 2 and 4 became significant in relation to EV/Sales, and strategies 3 and 4 became significant for P/B. EV/Sales is the least detailed in terms of income statement and balance sheet depth and therefore the most likely to have greater variation. Thus, using more peers provides more stable results for single factor strategies.

7.2.5. SPX Conclusion

In this section, we have analyzed the peer selection of firms in the S&P 500 index using combinations of peer groups and average and non-average fundamentals. The general reoccurring conclusion is that peer selection accuracy improves with the use of more factors to find peers. This is to be expected because the more similar characteristics firms have with each other, the more comparable they should be. Another persistent theme is the significance in absolute percentage error improvements for EV/EBTIDA and P/CF multiples. These multiples generally have the lowest errors and least fluctuation across strategies which suggests that the more in-depth accounting figures are better at selecting firms. EV/Sales is overall the worst multiple to select peers, because revenue figures are too wide-ranging when used as a standalone measure. P/B appears to work well in certain instances, like in strategies 1 and 4 based on returns, but not in the other single factor tests. In addition, P/B sees improvement in multifactor strategies except for strategy 8 where ROE is excluded. We know from the previous chapter, where the multiples were described, that P/B can be rewritten using ROE, which explains why P/B is most useful when incorporating ROE into the peer selection process.

7.3. SXXP Strategies

7.3.1. Non-average & 6 Peer Sample

The table below illustrates the peer selection statistics for the test with 6 peers and non-average fundamentals in STOXX Europe 600 index.

Table 7.3.1 – Peer Selection Accuracy

(SXXP, non-average, 6 peers)

(Strategy)	ROE (1)	TA (2)	G (3)	ROIC (4)	WACC (5)	ROE TA (6)	ROE TA G (7)	ROIC TA G WACC (8)	ROE TA G WACC ROIC (9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.351 (7)	1.347 (6)	1.332 (5)	1.363 (8)	1.410 (9)	1.313 (4)	1.228 (3)	1.180 (2)	1.151 (1)
Median	0.611 (4)	0.639 (8)	0.617 (5)	0.633 (7)	0.639 (9)	0.630 (6)	0.606 (3)	0.576 (2)	0.563 (1)
EV/EBITDA									
Mean	0.821 (8)	0.806 (5)	0.815 (6)	0.831 (9)	0.818 (7)	0.780 (4)	0.760 (3)	0.701 (1)	0.730 (2)
Median	0.448 (6)	0.444 (4)	0.447 (5)	0.458 (9)	0.455 (8)	0.454 (7)	0.432 (3)	0.421 (1)	0.425 (2)
P/CF									
Mean	1.335 (6)	1.258 (3)	1.367 (7)	1.390 (8)	1.288 (4)	1.541 (9)	1.335 (5)	1.060 (1)	1.106 (2)
Median	0.448 (6)	0.444 (5)	0.474 (9)	0.442 (4)	0.468 (8)	0.448 (7)	0.435 (3)	0.415 (2)	0.412 (1)
P/B									
Mean	0.691 (4)	0.867 (7)	1.231 (9)	0.707 (5)	1.206 (8)	0.609 (1)	0.640 (2)	0.724 (6)	0.646 (3)
Median	0.355 (2)	0.503 (7)	0.518 (8)	0.390 (6)	0.539 (9)	0.356 (3)	0.360 (4)	0.381 (5)	0.343 (1)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+	+	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+*
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+*	+	+	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	2.018	2.009	2.021	2.026	2.036	18.347	48.366	80.891	112.130
EV/EBITDA	2.034	2.014	2.024	2.032	2.027	18.791	49.294	84.663	114.734
P/CF	2.058	2.033	2.059	2.020	2.040	18.745	49.979	84.689	114.409

P/B	2.059	2.028	2.038	2.033	2.038	18.838	49.048	84.557	114.557

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

In the first test of SXXP, EV/Sales and P/CF show the most variation between mean and median absolute percentage errors. EV/Sales and P/CF have similar average errors at around 130over the 9 strategies, however the median for P/CF appears more like that of EV/EBITDA at approximately 44%. Price-to-book peer selection is the most accurate for the strategies that include ROE as a selection factor, where the mean and median errors with ROE are around 65% and 35%, and without ROE are 95% and 47%, respectively. The two enterprise value multiples and price-to-cashflow generally see improvement as factors are combined to result in more accurate peer selection.

In Wilcoxon's signed rank test, none of the EV/Sales strategies had any significant improvement overall throughout the 20 company samples, though the statistic was positive on average. EV/EBITDA and P/CF both display significance at the 5% level, except for the former in strategy 9. This test is an exception and looking back at the errors, strategy 9 is less accurate than 8. Remember that EV/EBITDA can be rewritten using ROIC, thus adding ROE does not necessarily positively impact the selection. Interestingly, P/B only sees significant improvement in strategy 3 with growth despite having the highest and 2nd highest errors in terms of mean and median. This result contradicts the expectation that inaccurate (i.e. with larger errors) selection would see significance, however growth is an important variable in P/B which may explain this outcome.

Lastly, SARD scores are lowest for EV/Sales showing that its peers are closer in terms of fundamental ranks. The other 3 multiples have nearly identical scores throughout.

7.3.2. Non-average & 12 Peer Sample

The table below illustrates the peer selection statistics for the test with 12 peers and non-average fundamentals in STOXX Europe 600 index.

Table 7.3.2 – Peer Selection Accuracy

(SXXP, non-average, 12 peers)

									ROE
								ROIC	**TA**
							ROE	**TA**	**G**
						ROE	**TA**	**G**	**WACC**

	ROE	TA	G	ROIC	WACC	TA	G	WACC	ROIC
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.302 (5)	1.317 (6)	1.319 (8)	1.319 (7)	1.370 (9)	1.259 (4)	1.195 (3)	1.162 (2)	1.159 (1)
Median	0.634 (5)	0.647 (8)	0.636 (6)	0.641 (7)	0.653 (9)	0.606 (3)	0.612 (4)	0.580 (2)	0.578 (1)
EV/EBITDA									
Mean	0.782 (5)	0.792 (6)	0.793 (8)	0.793 (7)	0.799 (9)	0.750 (4)	0.733 (3)	0.690 (1)	0.717 (2)
Median	0.452 (9)	0.445 (7)	0.442 (4)	0.443 (5)	0.447 (8)	0.445 (6)	0.429 (3)	0.427 (1)	0.427 (2)
P/CF									
Mean	1.268 (6)	1.229 (3)	1.290 (7)	1.355 (8)	1.241 (4)	1.243 (5)	1.458 (9)	1.089 (2)	1.066 (1)
Median	0.459 (7)	0.447 (6)	0.473 (9)	0.441 (4)	0.471 (8)	0.436 (3)	0.441 (5)	0.425 (1)	0.431 (2)
P/B									
Mean	0.662 (4)	0.842 (7)	1.203 (9)	0.673 (5)	1.175 (8)	0.591 (1)	0.634 (3)	0.679 (6)	0.622 (2)
Median	0.346 (2)	0.501 (7)	0.553 (9)	0.393 (6)	0.545 (8)	0.344 (1)	0.354 (4)	0.389 (5)	0.347 (3)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+	+	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+*	+**	+*
P/CF	+**	+*	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+*	+	+	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	3.563	3.532	3.572	3.602	3.617	24.936	61.181	97.247	131.347
EV/EBITDA	3.605	3.562	3.618	3.633	3.599	26.081	61.777	100.468	136.017
P/CF	3.663	3.599	3.659	3.639	3.620	26.110	62.960	102.734	137.013
P/B	3.658	3.629	3.662	3.629	3.629	25.772	62.190	103.492	136.246

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

Initially the results show the least amount of variation of absolute percentage errors using EV/EBTIDA and P/B in most cases. Overall, the sample errors are on par with the previous test using 6 peers, suggesting that using more peers in this non-average dataset does not yield peer selection improvements. For EV/Sales, strategy 9 performed best for the whole sample which is as expected because EV/Sales is the least detailed measure and more factors should improve peer selection.

EV/EBITDA has the lowest absolute percentage error using strategy 8, where the factors WACC and ROIC can be derived from the multiple formula. Like in the previous test, P/CF has similar mean errors to EV/Sales at 125% and 127%, respectively, however its median errors of about 45% are similar to those of EV/EBITDA. Lastly, P/B still shows the most accurate peer selection in tests using ROE as a selection parameter.

The Wilcoxon signed rank test for the 20-company sample also shows similar results to the previous test with 6 peers. However, selection accuracy appears to be slightly worse using more comparables. While EV/Sales and P/B results are unchanged between the two tests, EV/EBITDA strategy 7 is now only significant at the 10% level as is P/CF strategy 2. Single-factor tests are not expected to perform as well as multifactor strategies, which is why it is interesting to see strategy 7 becoming less significant with more peers. The problem may be that the extra 6 peers are significantly less comparable than the original 6 companies chosen. This defeats the purpose of trying to create a more stable peer group, if the added companies are much less comparable. However, it is difficult to prove whether this is the case because only 1 of the multifactor strategies was affected.

EV/Sales has the lowest SARD scores on average, and strategy 8 has the most variation compared to the other strategies. The most undervalued firms using P/CF and P/B are the farthest away from the base company in terms of the sum of absolute ranked differences system. This could mean that the base company is more undervalued because the peers are fundamentally more different and should not necessarily be compared in the first place. However, the difference is still relatively small, and the APE/Wilcoxon tests do not support this theory.

7.3.3. Average & 6 Peer Sample

The table below illustrates the peer selection statistics for the test with 6 peers and average fundamentals in STOXX Europe 600 index.

Table 7.3.3 – Peer Selection Accuracy

(SXXP, average, 6 peers)									
									ROE
								ROIC	**TA**
							ROE	**TA**	**G**
						ROE	**TA**	**G**	**WACC**
	ROE	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

1. Absolute percentage errors of the entire sample

EV/Sales									
Mean	1.303 (5)	1.327 (6)	1.328 (7)	1.334 (8)	1.413 (9)	1.280 (4)	1.199 (3)	1.191 (2)	1.169 (1)
Median	0.617 (5)	0.636 (8)	0.621 (7)	0.620 (6)	0.638 (9)	0.614 (4)	0.590 (3)	0.569 (1)	0.576 (2)
EV/EBITDA									
Mean	0.762 (3)	0.825 (7)	0.840 (8)	0.821 (6)	0.841 (9)	0.764 (4)	0.749 (1)	0.769 (5)	0.761 (2)
Median	0.457 (9)	0.444 (2)	0.445 (3)	0.457 (8)	0.456 (6)	0.453 (5)	0.446 (4)	0.444 (1)	0.457 (7)
P/CF									
Mean	1.322 (6)	1.348 (7)	1.377 (8)	1.398 (9)	1.292 (4)	1.243 (2)	1.140 (1)	1.313 (5)	1.273 (3)
Median	0.467 (9)	0.443 (5)	0.458 (8)	0.434 (3)	0.451 (7)	0.443 (6)	0.434 (4)	0.423 (2)	0.409 (1)
P/B									
Mean	0.672 (4)	0.850 (7)	1.140 (8)	0.734 (5)	1.282 (9)	0.611 (1)	0.653 (2)	0.748 (6)	0.660 (3)
Median	0.374 (4)	0.487 (7)	0.507 (8)	0.384 (6)	0.533 (9)	0.357 (2)	0.360 (3)	0.379 (5)	0.346 (1)

2. Wilcoxon Signed Rank Test of the 20 company sample

EV/Sales	+	+	+	+	+	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+*	+**	+**
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+	+	+*	+	+	+	+

3. Average SARD score of the 20 company sample

EV/Sales	2.027	2.001	2.028	2.023	2.032	17.671	47.532	79.442	110.846
EV/EBITDA	2.033	2.027	2.036	2.033	2.024	18.508	48.138	83.664	112.910
P/CF	2.049	2.047	2.051	2.044	2.032	18.082	49.020	85.181	115.290
P/B	2.052	2.030	2.014	2.037	2.023	18.197	48.862	82.494	112.551

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

In this section are analyzing the peer selection results of using 3-year average fundamentals. Starting with EV/Sales, the absolute percentage errors mostly improve with the addition of more selection factors, and the results are very similar to the non-average 6 peer test. On the other hand, EV/EBITDA's results seem to be more sporadic in terms of ranks because several single factor tests have lower errors than the multifactor strategies. Because the overall errors for EV/EBTIDA are on par with the non-average test, the result of single factors performing better can be attributed to the 3-year averages. The averages can implicitly act as a multifactor strategy themselves as they provide a more realistic representation of a firm's qualities, such that comparing average fundamentals acts as an accurate peer selector. P/CF has a similar trend, however the non-average test is now more similar to this one in terms of errors, so the theory appears less conclusive in this instance.

The signed rank tests also illustrate similar results to the non-average test, where EV/EBITDA and P/CF are significant for all strategies. This shows that using average values does not necessarily provide better peer selection, when looking at selection accuracy of the 20 most undervalued firms.

The SARD scored have a similar tendency to the previous test with non-average 12 peers, where there is greater variation in the scores of the multifactor tests. While EV/Sales has the lowest score in both the average and non-average test, the other 3 multiples have varying scores. The scores for strategies 8 and 9 in the non-average 6 peer test were about 84.5 and 114.5 for EV/EBITDA, P/CF and P/B. This test results in EV/EBITDA and P/B several points under the non-average test, while P/CF has a higher score. Based on SARD, the undervalued base companies based on EV/EBITDA and P/B are more comparable to their peer groups than on P/CF.

7.3.4. Average & 12 Peer Sample

The table below illustrates the peer selection statistics for the test with 12 peers and average fundamentals in STOXX Europe 600 index.

Table 7.3.4 – Peer Selection Accuracy

(SXXP, average, 12 peers)

	ROE	**TA**	**G**	**ROIC**	**WACC**	**ROE TA**	**ROE TA G**	**ROIC TA G WACC**	**ROE TA G WACC ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.286 (5)	1.322 (7)	1.314 (6)	1.323 (8)	1.360 (9)	1.255 (4)	1.185 (3)	1.169 (2)	1.158 (1)
Median	0.626 (6)	0.644 (9)	0.613 (4)	0.632 (7)	0.638 (8)	0.616 (5)	0.592 (2)	0.594 (3)	0.581 (1)
EV/EBITDA									
Mean	0.772 (5)	0.807 (9)	0.798 (7)	0.796 (6)	0.802 (8)	0.728 (2)	0.728 (3)	0.726 (1)	0.730 (4)
Median	0.458 (8)	0.441 (2)	0.457 (7)	0.461 (9)	0.454 (6)	0.442 (4)	0.445 (5)	0.440 (1)	0.442 (3)
P/CF									

Mean	1.243 (4)	1.279 (5)	1.436 (9)	1.305 (6)	1.237 (3)	1.307 (7)	1.314 (8)	1.175 (1)	1.176 (2)
Median	0.461 (8)	0.459 (7)	0.467 (9)	0.457 (6)	0.454 (5)	0.434 (3)	0.435 (4)	0.415 (1)	0.418 (2)
P/B									
Mean	0.653 (4)	0.852 (7)	1.191 (8)	0.694 (5)	1.225 (9)	0.604 (1)	0.635 (2)	0.699 (6)	0.647 (3)
Median	0.379 (4)	0.500 (7)	0.530 (8)	0.392 (5)	0.547 (9)	0.356 (1)	0.370 (3)	0.397 (6)	0.363 (2)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+	+	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+*
P/CF	+**	+**	+**	+**	+**	+**	+*	+*	+**
P/B	+	+	+*	+	+	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	3.610	3.525	3.571	3.605	3.612	24.456	59.184	96.418	129.031
EV/EBITDA	3.599	3.548	3.604	3.589	3.557	26.380	60.853	97.712	128.814
P/CF	3.654	3.609	3.693	3.607	3.600	26.420	63.192	101.569	136.240
P/B	3.668	3.639	3.631	3.623	3.627	25.378	62.001	99.676	132.129

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The final peer selection test in Europe examines whether 3-year average fundamentals are able to complement the usage of 12 peers and improve selection accuracy. Initially, EV/Sales sees a slight improvement in general across the 9 strategies. Since this multiple is the shallowest in terms of financial details, additional peers reduce the errors. EV/EBTIDA and P/CF show small improvements in some of the strategies with fewer factors, but there are slight increases in the errors for strategies 8 and 9 (with more factors). Interestingly, both EV/EBITDA and P/CF have the lowest errors in strategy 8 where ROIC and WACC are used to derive the multiple. P/B also has very similar absolute percentage errors as in the non-average test, and strategy 6 still performs the best.

The signed rank test provides similar results to the non-average 12 peer test, where EV/EBITDA and P/CF are significant for all strategies. In addition, strategy 3 using growth is the only significant one for P/B.

SARD scores have increased variation in strategies 7 – 9 compared to the previous 12 peer test. In strategy 7, only P/CF has a higher SARD. In strategies 8 and 9 all the scores are lower, particularly

EV/Sales, EV/EVITDA and P/B. So, using average values for 12 peer tests provide similar results regarding absolute percentage errors, while also selecting closer theoretical peers.

7.3.5. SXXP Conclusion

In this section, we analyzed the peer selection of firms in the STOXX Europe 600 index using combinations of peer groups and average and non-average fundamentals.

Through a holistic perspective, EV/EBITDA and P/CF were the most significant regarding peer selection accuracy in the 20 company samples. Every strategy using these 2 multiples was significant at either the 10% or 5% level. EV/Sales is not significant at any level across any of the 4 tests, whereas P/B is significant in strategy 3 (growth) in 3 of the 4 tests. Despite many strategies not being significant, the results show a positive development in the absolute percentage errors in each strategy.

The mean APE's of the 12 sample tests are slightly lower than the 6 sample tests for both non-average and average at approximately 126% vs 128% for EV/Sales. The median values are nearly identical at 62%, suggesting that using more peers minimizes the variation, however EV/Sales is still the most inaccurate of the 4 multiples. EV/EBITDA has a similar trend, where the mean samples with 12 peers have slightly lower APE's than 6 peers, at 79% vs 76%. Again, the median values are also nearly identical at 45% for APE overall. Mean P/CF errors for 12 peer tests are approximately 125% compared to 130% for 6 peer samples. The median for all the P/CF samples are 44% on average. P/CF and EV/EBITDA have similar median values suggesting that they select peers in a similar fashion, however the mean value for P/CF is much higher and therefore the variation is too. This may be due to P/CF being an equity multiple and not considering the whole picture of the firm, like the amount of debt. Lastly, P/B has a mean APE of approximately 81% compared to a median of 42% for all 4 tests. P/B has the lowest errors on average when compared to the other 3 multiples. At first, this appears to contradict the significance of the Wilcoxon test, however this test is based on improvement of the base company towards the peer group. P/B has consistently a higher absolute percentage error for strategies 3 and 5, which are the only strategies to have resulted in significance. This might suggest that strategies based on P/B select peers that are close to the base company in terms of fundamentals, as well as multiple valuations. Thus, reversion towards the mean is less likely because the base firms are already closely comparable with their peers.

SARD scores in the non-average tests with 6 and 12 peers are fairly consistent, though they are generally lower for EV/Sales. However, the average tests have lower scores compared to the non-average tests with equivalent peer sizes. As we previously stated, absolute percentage errors are quite similar between all the tests, however, the SARD scores illustrate that the average test select peers that are fundamentally closer to the base company while maintaining similar accuracies.

The general trend for EV/Sales and EV/EBITDA is a reduction in selection errors when combining more fundamentals to select peers. The two equity multiples also see improvements, however more fundamentals used in the selection do not necessarily result in higher accuracy. Variations for P/CF and P/B are lowest in strategies 7 and 6, respectively. In conclusion, peer selection in terms of lowest errors is typically best with P/B, while EV/EBITDA and P/CF have the most significant improvement of the base company towards the peer group.

7.4. MXEF Strategies

This index consists of companies in less developed markets compared to SPX and SXXP. The countries included in the index are quite diverse as illustrated in the index description, however the countries are less politically tied together, compared to the other indices. When many firms in STOXX Europe 600 are part of the Eurozone and share a common currency and monetary policy to a large extent, countries in MXEF are more independent of each another. This means qualitative factors like geopolitics can play into firm valuation. Nonetheless, peer selection is going to be analyzed in the same manner.

7.4.1. Non-average & 6 Peer Sample

The table below illustrates the peer selection statistics for the test with 6 peers and non-average fundamentals in MSCI Emerging Markets index.

Table 7.4.1 – Peer Selection Accuracy

(MXEF, non-average, 6 peers)									
									ROE
								ROIC	**TA**
							ROE	**TA**	**G**
						ROE	**TA**	**G**	**WACC**
	ROE	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.234 (5)	1.267 (6)	1.277 (7)	1.283 (8)	1.297 (9)	1.204 (3)	1.211 (4)	1.180 (2)	1.160 (1)
Median	0.529 (4)	0.539 (6)	0.543 (8)	0.540 (7)	0.568 (9)	0.533 (5)	0.510 (2)	0.505 (1)	0.512 (3)
EV/EBITDA									
Mean	0.958 (8)	0.917 (5)	0.932 (6)	0.938 (7)	0.968 (9)	0.898 (4)	0.858 (3)	0.802 (2)	0.770 (1)
Median	0.485 (9)	0.459 (7)	0.456 (6)	0.468 (8)	0.452 (4)	0.451 (3)	0.452 (5)	0.428 (1)	0.431 (2)
P/CF									
Mean	1.252 (7)	1.080 (5)	1.272 (8)	1.293 (9)	1.238 (6)	1.038 (3)	1.077 (4)	1.023 (2)	1.000 (1)
Median	0.554 (8)	0.514 (4)	0.539 (6)	0.555 (9)	0.550 (7)	0.534 (5)	0.513 (3)	0.512 (2)	0.499 (1)
P/B									
Mean	0.564 (4)	0.825 (7)	0.987 (9)	0.599 (6)	0.963 (8)	0.512 (1)	0.528 (3)	0.591 (5)	0.516 (2)
Median	0.342 (4)	0.512 (7)	0.534 (8)	0.372 (6)	0.548 (9)	0.332 (2)	0.336 (3)	0.367 (5)	0.320 (1)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+*	+	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/CF	+***	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+	+	+	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	2.014	2.008	2.026	2.029	2.039	17.979	46.971	81.632	109.924
EV/EBITDA	2.054	2.045	2.040	2.051	2.025	18.742	48.397	84.122	112.082
P/CF	2.029	2.011	2.024	2.063	2.038	18.596	49.265	82.204	112.909
P/B	2.027	2.026	2.021	2.044	2.018	18.335	48.765	83.787	116.105

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

Initially, the absolute percentage errors show that EV/Sales is the least accurate. This is reflected in both the mean and median values and suggests that EV/Sales has the most variation in peer selection errors. Accuracy for this multiple is improved when more valuation fundamentals are used to select peers. EV/EBITDA has an error of approximately 89% on average, and a median error of 45% across the 9 strategies. Errors are lowest for strategies 8 and 9, illustrating that peer selection improves by having more selection criteria when using EV/EBITDA as a valuation variable. P/CF has similar error values as EV/Sales, meaning variation is high based on the big difference between median and mean errors. On the other hand, P/B is the most accurate with the mean error of 68% and median 40%. P/B results again show that strategies involving ROE (strategy 1, 6, 7, 9) have the lowest absolute percentage errors. This reflects that ROE can be derived as a variable from P/B.

Wilcoxon's signed rank test shows that EV/Sales and P/B are not significant in the vast majority of strategies regarding the decrease of absolute percentage errors for the 20-company sample. P/B is the most accurate on an overall level, so there is less room for improvement to begin with, which also means that the base company is likely more correctly priced against the selected peer group. However, EV/Sales has much more room for improvement, but since the multiple is the shallowest in terms of detail, then the improvements could be more random and therefore not very significant. EV/EBTIDA and P/CF both show significant improvement of accuracy at a 5% level or higher. This means that the 20 undervalued companies selected every year were actually undervalued compared to our peer group, and that the company's relative valuations became closer to their peers over time. In this test, strategy 1 was very significant at a 1% level suggesting that P/CF has been a good metric to determine undervalued companies that are peers based on their ROE's. ROE is fundamentally driven by net income and therefore also operating cash flow, where an increase in cash generated from operations typically leads to more cash generated for equity holders.

Lastly, SARD scores are lowest for EV/Sales and P/B saw a large jump in its score from strategy 8 to 9. While strategy 8 was fundamentally closer to its peers, the absolute percentage errors show that the test was not the most accurate in reference to relative valuation. Even though the SARD score of strategy 9 was the highest, signaling the base company being farthest away from its peers, APE-wise it was the most accurate at selecting peers.

7.4.2. Non-average & 12 Peer Sample

The table below illustrates the peer selection statistics for the test with 12 peers and non-average fundamentals in MSCI Emerging Markets index.

Table 7.4.2 – Peer Selection Accuracy

(MXEF, non-average, 12 peers)									**ROE**
								ROIC	**TA**
							ROE	**TA**	**G**
						ROE	**TA**	**G**	**WACC**
	ROE	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**

(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.220 (5)	1.237 (6)	1.271 (8)	1.243 (7)	1.273 (9)	1.211 (4)	1.186 (3)	1.153 (2)	1.143 (1)
Median	0.527 (7)	0.521 (5)	0.540 (8)	0.522 (6)	0.559 (9)	0.519 (4)	0.502 (3)	0.491 (2)	0.483 (1)
EV/EBITDA									
Mean	0.917 (7)	0.897 (5)	0.923 (9)	0.896 (4)	0.920 (8)	0.898 (6)	0.835 (3)	0.809 (2)	0.792 (1)
Median	0.454 (6)	0.461 (8)	0.476 (9)	0.451 (5)	0.459 (7)	0.440 (3)	0.446 (4)	0.429 (1)	0.434 (2)
P/CF									
Mean	1.212 (7)	1.063 (3)	1.248 (9)	1.209 (6)	1.215 (8)	1.122 (5)	1.120 (4)	1.043 (2)	1.036 (1)
Median	0.574 (9)	0.535 (4)	0.549 (6)	0.554 (7)	0.569 (8)	0.543 (5)	0.509 (1)	0.516 (2)	0.529 (3)
P/B									
Mean	0.549 (4)	0.812 (7)	0.946 (9)	0.575 (5)	0.939 (8)	0.515 (1)	0.521 (3)	0.580 (6)	0.517 (2)
Median	0.332 (2)	0.504 (7)	0.536 (8)	0.367 (5)	0.566 (9)	0.327 (1)	0.333 (3)	0.369 (6)	0.336 (4)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+	+	+	+	+	+
EV/EBITDA	+**	+**	+*	+**	+**	+**	+**	+**	+**
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+	+	+	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	3.564	3.542	3.587	3.602	3.636	24.732	59.047	97.809	127.917
EV/EBITDA	3.653	3.612	3.671	3.684	3.569	25.607	61.968	100.699	130.519
P/CF	3.606	3.571	3.585	3.644	3.645	25.053	61.107	100.702	135.156
P/B	3.601	3.568	3.589	3.599	3.606	25.583	61.754	99.770	135.044

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The results of this test reflect similar statistics as the preceding one, however there are a few key findings to note. Particularly, EV/Sales sees improvement in the absolute percentage errors for all strategies. For example, the mean error of the 12-peer test is 121% compared to 124% for 6 peers, and median values are 51% and 53%, respectively. Since the variation of errors is greatest for EV/Sales across the strategies in the two tests, a decrease in errors with more peers suggests that a larger peer group reduces the impact of outliers that may be falsely selected. As it was previously mentioned, EV/Sales is the least detailed in terms of income statement and is therefore not the best multiple to select undervalued firms.

EV/EBITDA has the smallest errors for strategies with 3 or more selection fundamentals, which illustrates that peer selection accuracy is improved. EV/EBITDA performs best when ROIC and WACC are combined, supporting the fact that they can be derived from the multiple. P/CF has slightly lower variation than EV/Sales since the mean error is 114% and median 54%. However, the selection errors are still quite high meaning that P/CF might not be a useful measure to determine undervalued firms. Lastly, P/B is the most accurate regarding the SARD selection model. Of the single factor tests, ROE and ROIC show the lowest absolute percentage errors which is as expected considering ROE is a main factor in calculating P/B. Interestingly, P/B performs best in strategy 6 where the size-risk factor complements the profitability parameter. Adding growth does not lead to improvement, and neither does ROIC. Growth may not be as important or distinct in emerging markets, because emerging countries are already characterized by rapid growth, so profitability and risk could be the most important factors in relative valuation.

Wilcoxon's signed rank test shows a positive improvement of absolute percentage errors for the 20-company sample for all tests. However, the significance of this improvement is limited to EV/EBITDA and P/CF where the majority is significant at a 5% level. Despite P/CF having larger variation, the selection of the most undervalued firms appears to be better since the base companies converge towards their peers valuation. As previously suggested for P/B, tests may not be significant because the strategies are already relatively accurate.

SARD scores are lowest for EV/Sales, whereas the other 3 multiples have similar scores for strategies 6 – 8. In strategy 9, the two equity multiples select peers that are farther away from the base companies' fundamentals compared to the enterprise value multiples.

7.4.3. Average & 6 Peer Sample

The table below illustrates the peer selection statistics for the test with 6 peers and average fundamentals in MSCI Emerging Markets index.

Table 7.4.3 – Peer Selection Accuracy

(MXEF, average, 6 peers)			**ROE**
		ROIC	**TA**
	ROE	**TA**	**G**

						ROE	TA	G	WACC
	ROE	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.218 (5)	1.260 (7)	1.327 (9)	1.244 (6)	1.319 (8)	1.191 (4)	1.182 (3)	1.178 (2)	1.162 (1)
Median	0.524 (4)	0.543 (8)	0.560 (9)	0.534 (6)	0.538 (7)	0.526 (5)	0.520 (3)	0.511 (2)	0.493 (1)
EV/EBITDA									
Mean	0.945 (6)	0.915 (5)	0.971 (9)	0.949 (7)	0.970 (8)	0.895 (3)	0.900 (4)	0.853 (2)	0.800 (1)
Median	0.486 (9)	0.469 (8)	0.455 (5)	0.463 (7)	0.462 (6)	0.448 (4)	0.420 (2)	0.425 (3)	0.414 (1)
P/CF									
Mean	1.214 (8)	1.136 (5)	1.196 (7)	1.292 (9)	1.190 (6)	1.033 (1)	1.080 (4)	1.073 (3)	1.059 (2)
Median	0.563 (9)	0.516 (3)	0.530 (6)	0.545 (7)	0.545 (8)	0.526 (5)	0.518 (4)	0.507 (1)	0.510 (2)
P/B									
Mean	0.589 (5)	0.826 (7)	1.007 (9)	0.604 (6)	1.005 (8)	0.540 (3)	0.521 (2)	0.586 (4)	0.521 (1)
Median	0.372 (5)	0.508 (7)	0.537 (8)	0.402 (6)	0.570 (9)	0.361 (3)	0.337 (2)	0.371 (4)	0.334 (1)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+	+*	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+	+	+	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	2.014	2.005	2.031	2.018	2.033	17.099	45.524	83.132	109.940
EV/EBITDA	2.037	2.055	2.026	2.056	2.020	18.219	48.413	84.130	111.267
P/CF	2.050	2.017	2.024	2.033	2.034	18.414	47.481	84.291	113.689
P/B	2.045	2.020	2.030	2.020	2.027	18.795	49.193	83.066	113.418

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

EV/Sales has a mean absolute percentage error of 123% and median of 53% on average across all strategies. This is nearly identical to the first test in emerging markets, with 6 peers and non-average values. Average values are used to provide a picture of a firm's characteristics over a 3 year period because current fundamentals can be skewed by non-recurring performance. This is expected to be useful for EV/Sales since the multiple is the least intricate, however average fundamentals do not make a difference. This may also be due to the traits that characterize firms in emerging markets, particularly risk and growth. These factors can swing here more than in developed markets because of political and economic factors, so historical factors do not necessarily provide an accurate perception of the current situation or the future.

For EV/EBITDA, the mean values show absolute percentage errors a few percent higher than for the non-average tests. This test has slightly worse peer selection especially for the multifactor strategies. This supports the earlier statement of historic fundamentals not adding value to the selection process. P/CF still has a high amount of variation with average values illustrating that using cash flows to determine relative value has a lot of outliers even when average fundamentals are considered. P/B appears to have the most accurate peer selection again in terms of the lowest absolute percentage errors. Strategies involving either ROE or ROIC have the best performances selection-wise, like in the non-average test.

The signed rank test shows similar results to the non-average test, where all strategies for EV/EBITDA and P/CF are significant at the 5% level. This shows that of the 20 firms chosen as peers for each underlying test, the base company reverts towards the mean multiple of the given peer sample. Interestingly, strategy 5 involving WACC is significant for EV/Sales. This strategy has some of the largest errors, so it is possible that the 20 most undervalued companies chosen are converging because they are generally undervalued and not specifically undervalued compared to the peer group. However, it is important to note that WACC combined with other fundamentals also does not provide the same result of significance.

7.4.4. Average & 12 Peer Sample

The table below illustrates the peer selection statistics for the test with 12 peers and average fundamentals in MSCI Emerging Markets index.

Table 7.4.4 – Peer Selection Accuracy

(MXEF, average, 12 peers)									
									ROE
								ROIC	**TA**
							ROE	**TA**	**G**
						ROE	**TA**	**G**	**WACC**
	ROE	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**
(Strategy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

1. Absolute percentage errors of the entire sample									
EV/Sales									
Mean	1.204 (5)	1.229 (7)	1.307 (9)	1.222 (6)	1.293 (8)	1.175 (4)	1.172 (3)	1.162 (2)	1.135 (1)
Median	0.514 (6)	0.520 (7)	0.549 (9)	0.502 (3)	0.542 (8)	0.513 (5)	0.504 (4)	0.487 (1)	0.489 (2)
EV/EBITDA									
Mean	0.933 (6)	0.894 (4)	0.951 (9)	0.934 (7)	0.935 (8)	0.879 (3)	0.913 (5)	0.832 (2)	0.784 (1)
Median	0.471 (9)	0.456 (6)	0.459 (7)	0.456 (5)	0.461 (8)	0.449 (4)	0.427 (3)	0.427 (2)	0.422 (1)
P/CF									
Mean	1.178 (8)	1.071 (2)	1.170 (6)	1.260 (9)	1.175 (7)	1.048 (1)	1.072 (3)	1.163 (4)	1.167 (5)
Median	0.571 (9)	0.517 (1)	0.550 (7)	0.566 (8)	0.534 (5)	0.538 (6)	0.523 (3)	0.523 (2)	0.534 (4)
P/B									
Mean	0.565 (4)	0.805 (7)	0.963 (8)	0.591 (6)	0.963 (9)	0.522 (1)	0.525 (3)	0.581 (5)	0.524 (2)
Median	0.349 (4)	0.510 (7)	0.549 (8)	0.379 (6)	0.566 (9)	0.339 (2)	0.327 (1)	0.378 (5)	0.341 (3)
2. Wilcoxon Signed Rank Test of the 20 company sample									
EV/Sales	+	+	+	+	+	+	+	+	+
EV/EBITDA	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/CF	+**	+**	+**	+**	+**	+**	+**	+**	+**
P/B	+	+	+	+	+	+	+	+	+
3. Average SARD score of the 20 company sample									
EV/Sales	3.580	3.529	3.618	3.584	3.632	24.142	58.867	99.143	129.278
EV/EBITDA	3.618	3.615	3.592	3.687	3.573	25.219	60.858	102.282	132.025
P/CF	3.641	3.566	3.594	3.621	3.605	24.978	59.810	102.194	133.929
P/B	3.661	3.578	3.622	3.591	3.595	25.828	61.450	99.843	133.661

In section 1, ranks of absolute percentage error across the 9 strategies are given in brackets

Significance levels are denoted for Wilcoxon Test: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

Absolute percentage errors of EV/Sales in this test have similar results to the previous one with non-average fundamentals and 6 peers. The strategy does see clear improvement when adding more selection factors, with strategies 8 and 9 being the most accurate. EV/EBITDA also follows a similar trend of reducing selection errors when adding more factors. The mean error for this multiple is 90% on average compared to 87% for the non-average test. Both tests have a median of 45%, but this test has slightly worse results overall. The P/CF test does not appear to have a clear relation to improved selection with more factors since the error ranks are scattered. While strategy 6 has the lowest mean, strategy 2 using only total assets has the lowest median and second lowest mean. Total assets could have a relation to cash flow, as larger more established firms with more assets have cash flows similar to those of other large firms with comparable value of total assets. P/B has similar results to the non-average test, where absolute percentage errors are the lowest for the 4 multiples. In addition, strategies involving ROE and ROIC clearly perform best at peer selection, with the former yielding the most accurate results. Again, this is as expected since ROE can be derived from P/B.

Wilcoxon's signed rank test shows nearly identical results to the non-average test with 12 peers. All strategies for EV/EBITDA and P/CF show significance at a 5% level, which illustrates that the base companies chosen converge towards their respective peer group relative valuations. EV/Sales is the least accurate and is not significant at any level. P/B is the most accurate as shown in part 1 of the table, which also means that it is less likely to show significant improvement of the errors.

The SARD scores show that EV/Sales selects undervalued firms that are closest to the peer group using the ranking system. For strategy 9, P/CF and P/B have lower scores than the non-average test, while EV/Sales and EV/EBITDA have higher. This could show that equity multiples improve in terms of selection scores when averages are used, whereas enterprise value multiples are negatively impacted.

7.4.5. MXEF Conclusion

In this section, we have analyzed the peer selection of firms in the MSCI Emerging Markets index using combinations of peer groups and average and non-average fundamentals.

Initially, we observe the general trend of peer selection accuracy for EV/Sales being quite low. EV/Sales has the largest difference between the mean and median absolute percentage errors across strategies and the 4 different peer selection tests. This is expected to an extent, because EV/Sales is the least detailed in terms of income statement as it does not account for the cost of sales. That being said, the results also show a clear tendency of decreasing errors when using more selection factors.

EV/EBITDA has lower variation than EV/Sales, and lower absolute percentage errors on average. The accuracy also follows the trend of improving as more factors are used to select peers. There are no clear indications that using more peers or 3-year average fundamentals provide more accurate results. However, between the tests using average values, the 12-peer test had slightly less variation compared to the non-average test. All tests show significance, illustrating that undervalued base companies tend to converge towards the peer group average. 35 of the 36 strategies tested in MXEF with this multiple are significant at the 5% level.

P/CF is the least accurate of the four multiples based on median absolute percentage errors. From a relative valuation perspective, P/CF ratios of the base company are the most unlike their respective peer groups. The selection tests using P/CF generally have the lowest errors in the strategies that use multiple factors to select peers. Interestingly, total assets as a single factor selection parameter are consistently more accurate than the other single factor tests. This illustrates that the size of a firm correlates with the value of the firm per unit of cash flow. While the full sample does not show impressive results in terms of absolute percentage errors, the Wilcoxon signed rank test shows that convergence of the 20-company sample towards the peer group average is significant at the 5% level or better for each strategy. Here, the undervalued firms are likely to be more undervalued compared to the firms using the other multiples, however their relative valuation also reverts towards the mean of the peer group.

Lastly, P/B resulted in the lowest absolute percentage errors and least variation for many of the strategies. In particular, the strategies including ROE clearly had the lowest errors which can be attributed to the fact that ROE can be derived from the P/B ratio. ROIC also demonstrates the ability to pick peers better than the other single factor tests, except ROE. None of the strategies are significant based on the 20-company sample, though this is likely due to the fact that the peers selected already have more similar valuations compared to the other multiples, so there is less room for improvement.

EV/Sales had consistently the lowest SARD scores, whereas EV/EBITDA and P/CF had similar scores. Overall, there were not any imminent advantages to choosing 3-year averages or 12 peers as selection variables. The results were relatively similar, illustrating that the model is able to exclude outliers and other potential oddities that could skew the data.

7.5. Peer Selection – Conclusion

This section covered the peer selection accuracy and statistics related to the selection for each of the 9 strategies. Furthermore, each strategy was tested with a combination of the number of peers used in a peer group and firm fundamentals that were over a one-year period and average over 3 years. Lastly, the undervalued companies were selected based on 4 different multiples and across 3 markets. The results show interesting strategy specific results, as well as trends that reoccur on certain strategies, markets and selection criteria as summarized below.

Strategies based on selection using EV/Sales are clearly the least accurate for selection purposes, however the absolute percentage errors tend to decrease as more fundamentals are used to select peers. In addition, the use of 12 peers and average fundamentals resulted in slight error improvements. However, the Wilcoxon signed rank test is rarely significant which also shows that EV/Sales is not good at identifying investments that are going to revert to the mean of the peer group, likely due to being the least detailed in terms of results on the income statement. Tests using the STOXX Europe 600 index have the highest variation for EV/Sales compared to other indices, while the S&P 500 has the lowest.

The 2nd multiple used, EV/EBITDA, is characterized by less variation between mean and median absolute percentage errors and a clear tendency of improvement with multiple fundamentals compared to the use of single selection factors. Since EV/EBITDA can be derived from ROIC and WACC, it was expected that strategy 8 will have the lowest error compared to the other strategies. Strategy 8 does well but is not consistently the best strategy when using EV/EBITDA. Its mean and median are typically in the top 33%, however if we account for strategy 9 which also includes all factors then it can clearly be seen that accuracy is improved. Wilcoxon's test illustrates that the convergence of the base company multiple towards the peer group is almost always significant at the 10% level at least. This supports the reduction in errors as the base companies are moving closer to their peers in terms of valuation.

P/CF has more mixed results between different indices, but it clearly provides the best results in the S&P 500 index. For STOXX Europe 600, absolute percentage errors are only slightly lower than those of EV/Sales and for MSCI Emerging Markets, those errors are modestly higher. Therefore, variation of the results between mean and median is quite a bit higher for the two indices compared to S&P 500, where the results are similar to EV/EBITDA. This could mean that cash flows are not as appropriate for relative valuation in Europe and emerging markets, since other factors may influence what determines a peer. An important factor to note is the geography of the indices. Companies in SXXP and MXEF indices are based in different jurisdictions and are therefore subjects to different regulation. However, most of the tests of the different strategies are significant on the 10% level or better for all 3 markets. Even though selection is not as accurate in Europe or emerging markets, using P/CF shows considerable reversion of the base company towards the mean valuation of the peer group.

Lastly, peer selection tests including P/B illustrate some trends that are persistent between the different markets and selection criteria. Price-to-book can be derived from ROE and Growth, so we expected strategies involving these factors to perform best. In fact, those containing ROE (strategies 1, 6, 7 and 9) clearly have the lowest absolute percentage errors, and adding growth in the multifactor tests tends to further reduce the error. However, growth by itself does not seem to be a good fundamental for selecting undervalued peers. Tests based on P/B are rarely significant based on

Wilcoxon's signed rank test, though variation is low (especially in strategies involving ROE + G) and the improvement is less likely because selected peers are already more accurate compared to those of other multiples.

Having used 6 and 12 peers in addition to average and non-average fundamentals, it is difficult to clearly conclude whether some samples are better suited for selecting peers than others. While using 12 peers may reduce the variation between the mean and median values, the extra peers are also fundamentally more different from the base company and are presumed to have a less comparable valuation. The same goes for the type of fundamental, because a current financial measure shows how a firm is performing in the present, but a 3-year average provides a part of historical picture as well. Since there is little variation between tests with different peer group sizes and fundamental types, it can be concluded that there might be few outliers if any to skew the data, otherwise we would expect 12 peers and 3-year averages to provide the most stable result of peer selection accuracy.

8. Portfolio Performance

As previously mentioned, backtesting an investment strategy is highly useful as it shows how well it would have performed historically and helps avoid using models that have never worked in the past.[121] Investors are especially interested in strategies that not only generate high returns, but also manage to reduce the amount of risk taken. In the following section we describe the statistics and measurements used to accurately understand the performance of the different portfolios with respect to return and risk.

Return

Return of a portfolio is a fairly straightforward measurement as it illustrates the amount a portfolio appreciates or depreciates in value over time. In this chapter, the return is calculated as the geometric average annualized return to include compounding of returns over the 14 years. In addition, the returns include dividends and 100% dividend reinvestment in the underlying security or benchmark is assumed. Lastly, returns for SPX are calculated in USD and SXXP are given in EUR. MXEF is also

[121] Pedersen (2015), p. 11

calculated in USD because it is the currency of the benchmark, and emerging market debt is often dollar denominated because it is more attractive to foreign investors and can be included in bond indices like JP Morgan's emerging market bond index (EMBI). More on this in the risk-free rate section.

Risk-free rate

The risk-free rate is used in risk and return models and the rate is determined by using an asset that is defined as risk free. The risk-free rate is required to measure the excess return from risky assets. Damodaran states that for an asset to be risk free there can be no default risk, which is typically a security issued by a government, and the term structure of the security should be long-term rate such as a 10-year government bond rate.[122] For the U.S. market we will use the 10-year United States Government bond and for Europe the 10-year German Bund because Germany is the largest member state of the Eurozone with a stable outlook and high credit quality justifying the notion of no default risk. The monthly yield for the two government securities is retrieved from Bloomberg for the same period as the portfolio returns, 01-05-2004 to 30-04-2018. The annualized yield for the U.S. risk free bond is 3.11% and 2.32% for Germany.

Determining the risk-free rate for emerging markets is more complicated, because the countries included in the index are not as mature and are typically characterized by higher default risk. To account for the higher default risk, we have calculated the weighted average default spread based on the countries in the MXEF index. The individual country default spreads (i.e. how much it costs to insure against default) are adapted from Damodaran where a combination of CDS spreads and sovereign credit ratings is used to extrapolate what rate over the U.S. market would compensate for the default risk.[123] We evaluate the overall default spread to be 1.37% (see appendix 2), and add the spread to the U.S. rate such that the estimated annualized risk-free rate in emerging markets is 4.53%. Emerging market debt is often dollar denominated because foreign investors want the opportunity to invest in higher yielding securities but want to avoid an additional layer of risk in form of currency risk. Therefore, emerging markets rates are very correlated with the U.S. treasury rates and it's reasonable to assume that the default spread accounts for the return an investor would require for an asset which is risk free.

---

[122] Damodaran (2006), p. 81-82

[123] Damodaran (2019), http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/ctryprem.html

Alpha & Beta

Alpha and Beta are two statistical measures used to analyze a portfolio and are related to one another as they are calculated by running a regression of a portfolio's excess returns on the excess return of the market. This calculation is given by:

$$(R_t - R_t^r) = \alpha + \beta(R_t^M - R_t^r) + \varepsilon_t$$

Where $(R_t - R_t^r)$ is the asset return in excess of the risk-free rate and $(R_t^M - R_t^r)$ is the market return in excess of the risk-free rate. Beta β measures the market exposure of a given strategy and its tendency to follow market movements. Market has a beta of 1, so if a portfolio has a beta of 2 then a market return of 5% would result in 10% increase for the portfolio, all else equal. Beta represents a systematic risk, or a risk that cannot be diversified away. Firm specific risk, or idiosyncratic risk, given by ε_t is independent of market fluctuations and can be diversified away. In the classic capital asset pricing model (CAPM), the expected return is based on the beta or market risk taken. However, we strive to achieve alpha α because it defies CAPM and represents return in excess of compensation for higher risk.[124] In other words, alpha is the excess or abnormal return after accounting for beta or market movements. Alpha is not always meaningful because it can be a result of luck, so we will determine whether alpha is significant to reliably claim that abnormal return is an outcome of the strategy. We will use the t-statistic (alpha divided by standard error) to determine the statistical significance of alpha, where a t-stat greater than 1.645 means the measure is significant at 10% confidence level, greater than 1.96 for 5% significance, and greater than 2.576 represents statistical significance at 1% confidence interval.

Sharpe Ratio

Return is a key factor in determining the attractiveness of an investment, but for risky investments the return is often evaluated in excess of the risk-free rate. Also known as risk premium, excess return illustrates the additional return for taking the risk. Risk in finance is commonly defined as a standard deviation or volatility of the returns because it shows the distribution of the rate of return. Investors obviously seek higher returns, but the trade-off between additional risk is an important consideration as well. Therefore, we often resort to Sharpe ratio, which is given by

---

[124] Pedersen (2015), p. 28

$$Sharpe\ ratio = \frac{(R_t - R_t^r)}{\sigma_t}$$

It is used to measure the risk-return trade-off because it describes the relationship between risk premium and standard deviation.[125] In other words, the ratio shows the excess return per unit of risk taken.

Maximum Drawdown

The final measure used to evaluate portfolio performance analyzes the historical risk of an investment strategy. Drawdown is the cumulative loss since the first loss started, and is the percentage loss from the peak, otherwise known as a high water mark (HMW).[126]

$$DD_t = \frac{HWM_t - P_t}{HWM_t} \qquad\qquad MDD_t = max_{t \le T} DD_t$$

Here drawdown (DD) shows the loss from the HWM relative to P_t, the cumulative return at time t. The maximum drawdown (MDD) takes the largest drawdown over the period t to T, where T is the time at present. Maximum drawdown is a useful measure of risk because it quantifies how much an investor would risk losing if he/she have invested in a strategy. Since the timeline used for backtesting includes the financial crisis, the maximum drawdown essentially shows how much was lost during that period. Historical losses are not necessarily representative of future risks, but the drawdown provides investors with a sense of possible loss in the future.

8.1. SPX Strategies

The following sections are dedicated to presenting the results of the various portfolios backtested from beginning of May 2004 to end of April 2018. The results shown in the upcoming sections illustrate the return and risk of the portfolio based on the performance statistics outlined above.

[125] Munk (2017), p. 51
[126] Pedersen (2015), p. 35

8.1.1. Non-average & 6 Peer Sample

This backtest outlines the results of the portfolios from the S&P 500 strategy using non-average selection fundamentals and 6 peers.

Table 8.1.1 – Returns and Performance

(SPX, non-average, 6 peers)

(Strategy)	SPX	ROE (1)	TA (2)	G (3)	ROIC (4)	WACC (5)	ROE TA (6)	ROE TA G (7)	ROIC TA G WACC (8)	ROE TA G WACC ROIC (9)
1. Annualized Return										
EV/Sales	8.7%	10.7% (6)	11.4% (3)	11% (5)	10.3% (9)	11.9% (2)	13% (1)	10.6% (8)	11.3% (4)	10.6% (7)
EV/EBITDA	8.7%	11.7% (5)	12.4% (2)	12% (4)	10.8% (8)	12.5% (1)	12.1% (3)	9.3% (9)	11% (7)	11.2% (6)
P/CF	8.7%	11.3% (3)	11.6% (2)	10.3% (7)	10.8% (4)	12.8% (1)	9.4% (9)	10.2% (8)	10.8% (5)	10.7% (6)
P/B	8.7%	14.5% (1)	10% (9)	14.1% (2)	12.3% (6)	11% (8)	13.4% (3)	12.9% (4)	11% (7)	12.7% (5)
2. Alpha										
EV/Sales		0.15%	0.21%	0.19%	0.12%	0.25%	0.34%	0.15%	0.21%	0.16%
EV/EBITDA		0.25%	0.33%	0.28%	0.2%	0.32%	0.31%	0.08%	0.19%	0.24%
P/CF		0.22%	0.26%	0.15%	0.19%	0.32%	0.1%	0.14%	0.17%	0.2%
P/B		0.45% *	0.12%	0.44% *	0.3%	0.17%	0.41%	0.34%	0.2%	0.35%
3. Beta										
EV/Sales	1.00	1.24	1.21	1.13	1.21	1.18	1.13	1.18	1.16	1.16
EV/EBITDA	1.00	1.08	1.03	1.08	1.05	1.05	1.02	1.10	1.12	1.02
P/CF	1.00	1.08	1.05	1.09	1.08	1.12	1.02	1.10	1.15	1.04
P/B	1.00	1.12	1.12	1.04	1.08	1.14	1.02	1.11	1.12	1.04
4. Sharpe Ratio										
EV/Sales	0.41	0.35 (8)	0.40 (5)	0.40 (4)	0.35 (9)	0.43 (2)	0.51 (1)	0.37 (7)	0.41 (3)	0.37 (6)
EV/EBITDA	0.41	0.46 (5)	0.51 (3)	0.48 (4)	0.42 (7)	0.53 (1)	0.51 (2)	0.32 (9)	0.41 (8)	0.45 (6)
P/CF	0.41	0.45 (3)	0.45 (2)	0.39 (6)	0.42 (5)	0.52 (1)	0.35 (9)	0.37 (8)	0.39 (7)	0.42 (4)
P/B	0.41	0.61 (2)	0.36 (9)	0.63 (1)	0.48 (6)	0.41 (7)	0.57 (3)	0.51 (5)	0.40 (8)	0.52 (4)
5. Maximum Drawdown										
EV/Sales	-50.9%	-66.7%	-73.1%	-70.7%	-70.7%	-72.7%	-67.6%	-72.5%	-71.4%	-71.2%
EV/EBITDA	-50.9%	-55.0%	-65.1%	-63.9%	-62.3%	-62.2%	-61.7%	-69.4%	-66.7%	-61.5%
P/CF	-50.9%	-54.4%	-61.4%	-61.2%	-62.1%	-64.1%	-60.2%	-63.7%	-61.7%	-55.5%
P/B	-50.9%	-52.5%	-68.0%	-61.4%	-58.1%	-66.8%	-53.2%	-58.9%	-58.4%	-54.4%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

Initially, the annualized returns show that they outperform the index with a return of 8.7%. Alpha, or abnormal excess return, for this sample is positive for all strategies, illustrating that the model is able to consistently provide return greater than the index. However, most of the alpha values are not significant, so it is not possible to determine whether this alpha arose due to the model or just luck.

Betas of the strategies are all higher than the market beta of 1 with the average for all portfolios being 1.10 or 10% extra exposure to market movements. A beta above 1 shows that the model is selecting firms that tend to be slightly riskier than the market, whereas defensive stocks would have betas under 1.

The Sharpe ratio shows the risk adjusted return and compared to the SPX index about half of the portfolios have a higher risk adjusted return. Strategies 1 and 3 of the test with P/B where the only 2 portfolios to have a significantly positive alpha, and these two portfolios also have the highest Sharpe ratios at 0.61 and 0.63 respectively compared to the benchmark of 0.41. The maximum drawdown can be tied into the Sharpe ratio results, as higher volatility negatively influences the Sharpe ratio and might lead to a higher drawdown. The average of all the portfolio maximum drawdowns is -63.6% or about 0.25 times higher than that of the benchmark. The model portfolios are considerably riskier than the index and an investor would have lost much more on average being invested in these portfolios. P/B strategies 1 and 3 have lower drawdowns than the average, which also helps explain why the Sharpe ratio and alpha look good.

When comparing the market performance to the peer selection, the results are not necessarily correlated. The expectation of more accurate peer selection does not clearly translate into better market performance. EV/Sales had the lowest accuracy, and in this case the lowest Sharpe ratio on average at 0.40. EV/EBITDA and P/CF both had significant improvement of errors and were more accurate in the strategies with multiple selection factors, however the performance here does not illustrate that. Strategy 5 using WACC has the highest SR in both cases and some of the highest alphas for the strategies with these multiples.

8.1.2. Non-average & 12 Peer Sample

This backtest outlines the results of the portfolios from the S&P 500 strategy using non-average selection fundamentals and 12 peers.

Table 8.1.2 – Returns and Performance

(SPX, non-average, 12 peers)

				ROE
			ROIC	**TA**
		ROE	**TA**	**G**
	ROE	**TA**	**G**	**WACC**

	SPX	ROE	TA	G	ROIC	WACC	TA	G	WACC	ROIC
(Strategy)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Annualized Return										
EV/Sales	8.7%	15.2% (2)	16.5% (1)	13.3% (9)	14.4% (7)	14.7% (6)	15.2% (3)	14.7% (5)	14.9% (4)	13.9% (8)
EV/EBITDA	8.7%	12.6% (5)	11.7% (7)	11.6% (8)	13.7% (1)	13.5% (2)	12.9% (4)	11.4% (9)	13.5% (3)	12.5% (6)
P/CF	8.7%	13.8% (1)	13.4% (5)	11.3% (8)	13.4% (4)	11.3% (9)	13.5% (3)	13.7% (2)	13.4% (6)	13.2% (7)
P/B	8.7%	13% (6)	13.2% (3)	13.1% (5)	15.3% (1)	12.7% (7)	13.1% (4)	12.1% (8)	13.9% (2)	11.5% (9)
2. Alpha										
EV/Sales		0.44% **	0.53% **	0.33%	0.41% *	0.4% *	0.45% **	0.43% *	0.41% *	0.33%
EV/EBITDA		0.19%	0.12%	0.14%	0.3%	0.26%	0.24%	0.12%	0.27%	0.17%
P/CF		0.29%	0.27%	0.12%	0.3%	0.07%	0.3%	0.31%	0.26%	0.23%
P/B		0.25%	0.25%	0.27% *	0.42% **	0.21%	0.25%	0.19%	0.32% *	0.12%
3. Beta										
EV/Sales	1.00	1.15	1.15	1.11	1.10	1.16	1.14	1.11	1.18	1.17
EV/EBITDA	1.00	1.23	1.25	1.19	1.19	1.23	1.18	1.20	1.21	1.27
P/CF	1.00	1.23	1.20	1.20	1.16	1.30	1.16	1.19	1.23	1.25
P/B	1.00	1.19	1.20	1.14	1.17	1.20	1.19	1.18	1.17	1.21
4. Sharpe Ratio										
EV/Sales	0.41	0.67 (2)	0.74 (1)	0.57 (9)	0.65 (5)	0.63 (7)	0.66 (3)	0.65 (4)	0.64 (6)	0.58 (8)
EV/EBITDA	0.41	0.51 (5)	0.45 (9)	0.47 (7)	0.58 (1)	0.56 (3)	0.54 (4)	0.46 (8)	0.57 (2)	0.49 (6)
P/CF	0.41	0.58 (3)	0.56 (5)	0.45 (8)	0.58 (2)	0.41 (9)	0.59 (1)	0.57 (4)	0.54 (6)	0.52 (7)
P/B	0.41	0.53 (6)	0.55 (4)	0.59 (3)	0.70 (1)	0.53 (7)	0.55 (5)	0.50 (8)	0.62 (2)	0.46 (9)
5. Maximum Drawdown										
EV/Sales	-50.9%	-48.7%	-47.4%	-55.7%	-54.3%	-56.5%	-52.6%	-53.5%	-56.6%	-58.5%
EV/EBITDA	-50.9%	-50.9%	-53.4%	-51.0%	-49.4%	-53.9%	-52.7%	-56.0%	-54.0%	-57.0%
P/CF	-50.9%	-50.9%	-55.0%	-53.4%	-47.0%	-58.3%	-55.0%	-49.4%	-52.6%	-52.2%
P/B	-50.9%	-55.0%	-61.6%	-60.9%	-56.9%	-63.0%	-52.3%	-60.0%	-65.7%	-62.4%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

In the 12-peer sample, average return across all strategies is 13.4%, which is 54% higher than the return of the benchmark. This sample already shows more significant alphas and an average alpha of 0.28% compared to 0.24% for the previous test with 6 peers. All the alphas are positive, but interestingly EV/Sales now shows that 7 of the 9 strategies are significant. Since EV/Sales is the least detailed multiple, it was not expected to have such alphas. Strategies 1, 2 and 5 were significant in the peer selection analysis, and this has also translated into significant alphas for these strategies. For P/B, growth strategy is significant again in addition to tests 4 and 8 where ROIC is included, and ROE strategies are no longer significant.

Beta is higher in this sample with the average between all strategies being 1.19 compared to 1.10 for the previous sample. Since beta and portfolio market exposure is higher in this sample, it would be expected that volatility is higher as well. However, looking at maximum drawdown risk seems to generally be lower given by the average of -54.8% compared to -63.6% from earlier. The drawdown has now improved by almost 14% but is still higher than the market one.

In this sample, all the Sharpe ratios are equal to or higher than the market, suggesting that using more peers improves risk adjusted return. With this sample we get 0.57 on average compared to 0.45 with 6 peers. Risk adjusted return is higher due to better annualized returns as well as lower drawdown and therefore lower volatility.

A holistic view of the results in this sample shows again that better returns are not necessarily correlated with more accurate peer selection. The majority of significant alphas are linked to the single factor tests, which are typically associated with less accurate selection of comparable companies. EV/EBITDA and P/CF illustrated significant APE improvement within the 20-company sample, but that has not translated into any significant alphas. Lastly, the portfolio performance of the multifactor strategies only outperforms the single factor strategies occasionally, which could mean that the return is more a result of luck than accurate peer selection.

8.1.3. Average & 6 Peer Sample

This backtest outlines the results of the portfolios from the S&P 500 strategy using average selection fundamentals and 6 peers.

Table 8.1.3 – Returns and Performance

(SPX, average, 6 peers)

(Strategy)	SPX	ROE (1)	TA (2)	G (3)	ROIC (4)	WACC (5)	ROE TA (6)	ROE TA G (7)	ROIC TA G WACC (8)	ROE TA G WACC ROIC (9)
1. Annualized Return										
EV/Sales	8.7%	13.6% (5)	14% (3)	15.2% (2)	12.8% (8)	15.6% (1)	13.9% (4)	13.4% (6)	13.2% (7)	12.4% (9)
EV/EBITDA	8.7%	14.5% (1)	12.2% (6)	13.1% (4)	14.2% (2)	13.5% (3)	12.3% (5)	12.1% (8)	11.8% (9)	12.1% (7)
P/CF	8.7%	14.2% (4)	15.3% (1)	13.4% (6)	13.1% (7)	11.8% (9)	13.7% (5)	12.7% (8)	14.6% (2)	14.3% (3)
P/B	8.7%	11.9% (6)	13.4% (2)	12.7% (3)	11.5% (9)	13.6% (1)	12.4% (4)	11.5% (8)	11.6% (7)	12.3% (5)
2. Alpha										
EV/Sales		0.35% *	0.4% *	0.49% **	0.3%	0.49% **	0.39% *	0.35% *	0.33%	0.27%

EV/EBITDA		0.4% **	0.24%	0.31% *	0.38% **	0.33% *	0.22%	0.22%	0.18%	0.21%
P/CF		0.4% **	0.48% ***	0.34% *	0.31%	0.18%	0.35% *	0.29%	0.37% *	0.37% *
P/B		0.22%	0.33% *	0.28% *	0.16%	0.33% *	0.25%	0.19%	0.19%	0.25%
3. Beta										
EV/Sales	1.00	1.17	1.11	1.12	1.15	1.19	1.12	1.14	1.15	1.15
EV/EBITDA	1.00	1.19	1.19	1.17	1.18	1.22	1.24	1.18	1.25	1.20
P/CF	1.00	1.14	1.14	1.14	1.19	1.28	1.17	1.15	1.31	1.25
P/B	1.00	1.16	1.15	1.12	1.24	1.21	1.19	1.14	1.18	1.16
4. Sharpe Ratio										
EV/Sales	0.41	0.57 (5)	0.62 (3)	0.68 (1)	0.54 (8)	0.67 (2)	0.61 (4)	0.57 (6)	0.55 (7)	0.51 (9)
EV/EBITDA	0.41	0.63 (1)	0.50 (7)	0.56 (4)	0.61 (2)	0.56 (3)	0.49 (8)	0.51 (5)	0.46 (9)	0.51 (6)
P/CF	0.41	0.65 (2)	0.70 (1)	0.58 (5)	0.54 (8)	0.44 (9)	0.59 (4)	0.54 (7)	0.58 (6)	0.59 (3)
P/B	0.41	0.50 (6)	0.59 (1)	0.58 (3)	0.45 (9)	0.58 (2)	0.51 (5)	0.49 (7)	0.48 (8)	0.53 (4)
5. Maximum Drawdown										
EV/Sales	-50.9%	-49.0%	-50.5%	-53.3%	-62.8%	-59.4%	-61.5%	-63.4%	-63.1%	-63.9%
EV/EBITDA	-50.9%	-48.5%	-51.2%	-49.6%	-49.3%	-58.3%	-61.1%	-63.1%	-63.9%	-63.3%
P/CF	-50.9%	-48.6%	-47.8%	-49.6%	-49.3%	-65.7%	-59.6%	-61.8%	-58.1%	-56.0%
P/B	-50.9%	-59.3%	-58.8%	-60.7%	-68.4%	-64.2%	-64.6%	-67.5%	-71.8%	-67.9%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

Initially it can be observed that returns are higher in this sample on average than those in the non-average sample with 6 peers, as returns are 13.2% and 11.5% respectively. Compared to the benchmark, returns are about 52% higher in this instance. EV/Sales has the highest average annualized return between the 9 strategies at 13.8%.

The result of the regression illustrates that alpha is positive for each portfolio and the average is about 0.31%. The significance of alpha in the sample is considerably higher than that of the 6-peer sample as roughly half of the portfolios are significant. This result is important evidence that the model can provide abnormal return, rather than outperformance being a matter of luck. EV/Sales and P/CF have the most significant alphas, what makes sense for P/CF because Wilcoxon's test also showed that the selected undervalued companies reverted towards the peer group mean. However, EV/Sales multiple did not reflect the same level of peer selection accuracy, so its returns are more difficult to explain.

Beta for this sample is about 1.18 signaling 18% extra exposure to the market, and is higher than that of the non-average sample which was estimated at 1.10. Since market exposure is higher, we would expect maximum drawdown to be higher as well because the portfolios would typically experience more drastic declines compared to the market. The average drawdown for this sample is about -58.7% or almost 5% less of a decline than for the sample with non-average fundamentals. Interestingly, the average for the first 5 strategies is -55.2% compared to -63.2% for the multi factor tests. We associated the multifactor tests with a better peer selection, however from a risk perspective the portfolios based on the single factor tests are less risky. This could be due to concentration of firms within the same sectors. For example, accurate peer selection could select firms within the same sector thus reducing the diversification and increasing exposure to certain parts of the market.

The Sharpe ratio averages about 0.57 and is therefore about 39% higher than the benchmark's. This outperformance in risk adjusted return can be attributed to much higher annualized returns but it is also negatively impacted by slightly higher volatility. There is also a clear link between risk adjusted return and absolute return, where a higher and more significant alpha typically results in a higher Sharpe ratio.

8.1.4. Average & 12 Peer Sample

This backtest outlines the results of the portfolios from the S&P 500 strategy using average selection fundamentals and 12 peers.

Table 8.1.4 – Returns and Performance

(SPX, average, 12 peers)

	SPX	ROE	TA	G	ROIC	WACC	ROE TA	ROE TA G	ROIC TA G WACC	ROE TA G WACC ROIC
(Strategy)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Annualized Return										
EV/Sales	8.7%	15.7% (2)	14.2% (6)	15.5% (3)	12.7% (9)	15.9% (1)	14.4% (5)	13.4% (8)	13.9% (7)	14.6% (4)
EV/EBITDA	8.7%	13.1% (5)	12% (8)	13.8% (2)	14.8% (1)	13.3% (3)	12.1% (7)	11.9% (9)	13.2% (4)	12.4% (6)
P/CF	8.7%	13.3% (8)	14.2% (3)	14.3% (2)	14% (4)	13.9% (5)	13% (9)	13.8% (6)	14.4% (1)	13.3% (7)
P/B	8.7%	11.4% (8)	9.8% (9)	12.3% (4)	12.1% (7)	13.9% (1)	13.5% (2)	12.3% (5)	13.2% (3)	12.2% (6)
2. Alpha										
EV/Sales		0.51% **	0.4% *	0.51% **	0.31%	0.52% **	0.43% **	0.34%	0.37% *	0.44% **
EV/EBITDA		0.29%	0.21%	0.36% *	0.45% **	0.31%	0.21%	0.19%	0.28%	0.23%
P/CF		0.31% *	0.38% *	0.41% **	0.38% **	0.34% *	0.29%	0.36% *	0.39% *	0.29%
P/B		0.17%	0.04%	0.27% *	0.22%	0.35% **	0.33% *	0.24%	0.3% *	0.23%
3. Beta										
EV/Sales	1.00	1.17	1.16	1.13	1.09	1.18	1.12	1.17	1.18	1.14
EV/EBITDA	1.00	1.23	1.21	1.17	1.13	1.21	1.23	1.26	1.25	1.24
P/CF	1.00	1.19	1.19	1.13	1.14	1.27	1.21	1.19	1.24	1.28
P/B	1.00	1.18	1.20	1.09	1.17	1.18	1.20	1.19	1.19	1.21

4. Sharpe Ratio										
EV/Sales	0.41	0.69 (1)	0.61 (6)	0.68 (3)	0.55 (9)	0.69 (2)	0.63 (5)	0.55 (8)	0.58 (7)	0.64 (4)
EV/EBITDA	0.41	0.54 (4)	0.48 (8)	0.60 (2)	0.67 (1)	0.55 (3)	0.49 (7)	0.46 (9)	0.54 (5)	0.50 (6)
P/CF	0.41	0.56 (6)	0.60 (3)	0.64 (1)	0.62 (2)	0.56 (7)	0.54 (8)	0.59 (5)	0.59 (4)	0.52 (9)
P/B	0.41	0.47 (8)	0.37 (9)	0.56 (2)	0.51 (6)	0.61 (1)	0.56 (3)	0.51 (5)	0.56 (4)	0.50 (7)
5. Maximum Drawdown										
EV/Sales	-50.9%	-48.4%	-52.4%	-54.3%	-61.2%	-55.6%	-58.9%	-62.7%	-62.0%	-57.0%
EV/EBITDA	-50.9%	-50.7%	-51.1%	-48.8%	-50.0%	-54.2%	-62.0%	-61.7%	-59.9%	-60.9%
P/CF	-50.9%	-55.0%	-50.3%	-50.5%	-47.3%	-56.4%	-61.0%	-58.3%	-54.8%	-54.1%
P/B	-50.9%	-59.6%	-64.2%	-63.2%	-64.4%	-63.1%	-63.1%	-65.7%	-66.7%	-66.5%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The final sample for S&P 500 delivers the highest annualized returns compared to the previous 3 samples and averages at 13.4%. This is 54% premium to the annualized return of the benchmark. In addition, this sample has the highest average portfolio alpha of 0.33% and is significant for 20 of the 36 portfolios. This significance shows that using 12 peers and 3-year average fundamentals provides the highest abnormal returns. However, the returns do not necessarily relate to the peer selection accuracy because the multifactor portfolios were generally the most accurate, but in this case they do not always have the highest alpha. Since the multifactor tests were the most accurate in terms of APEs, there is potentially less room to improve towards the mean of peer group's relative value. As a result, the single factor portfolios might have more room for improvement, but the returns are expected to be more random.

This sample has the beta of 1.19, which is the same as the 12-peer sample with non-average fundamentals. Therefore, the risk of these portfolios tends to be greater than the benchmark's. The maximum drawdown is higher on average compared to the 12 peer non-average sample as we get -57.7% and -54.8%, respectively. In particular, the multifactor tests illustrated a higher level of risk because the average for strategies 6 through 9 is roughly 61% compared to 55.7% for the non-average test.

The risk adjusted return given by the Sharpe ratio displays an average of 0.57. This is the same as for the two previous tests, and each portfolio has a higher Sharpe ratio than the benchmark. For EV/EBITDA and P/CF the greatest Sharpe ratios were associated with strategies 3 and 4, which were some of the least accurate strategies in peer selection. However, growth and ROIC can both be derived from the EV/EBITDA multiple which may be an explanation for the high risk adjusted return.

8.1.5. SPX Returns and Performance Conclusion

The results in performance section for the portfolios from the S&P 500 index highlighted the returns had an investor used the model historically. The backtest also outlined the risk associated with investing in the various portfolios compared to the benchmark.

Observations for the different samples illustrate that total return of the portfolio strategies resulted in a higher annualized return than that of the benchmark in each case. Compared to the peer selection accuracy, returns do not follow the same trend of improvement when using multifactor portfolios. The alpha closely mimics the annualized return. The former tends to be highest and most significant for EV/Sales, which may seem contradictory to APE tests. However, since selected base companies are relatively highly undervalued compared to their theoretical peers, there is also more opportunity to grow in value towards the relative valuation of the peer group.

Beta of each of the portfolios in the 4 samples is higher than the benchmark's, which implies that exposure to the market is greater than 1 and it adds to the riskiness. Assuming that markets are efficient, and the peer selection model works perfectly, the firms chosen for the investment portfolio should not be undervalued if they have similar characteristics as the peer group. That being said, there may be an underlying reason for a firm to trade at a discount that is not reflected in its financials, i.e. it could be due to negative news, management, or governance for example. This could mean that some of the investments made are in undervalued companies because they trade at a discount for a reason, which would likely have an impact on performance. However, if the model only chose troubled companies, then high returns and positive alpha would not be expected. It can be observed that larger market exposure in the form of beta translates to higher maximum drawdowns. The drawdown for the various portfolios is typically larger than the market's illustrating a deeper drop in asset value in a worst-case scenario.

Contrary to peer selection where the different samples did not have a significant impact on accuracy, using 12 peers and average fundamentals result in higher risk adjusted and abnormal returns. The last sample with 12 peers and 3-year averages had the highest alpha, Sharpe ratio and relatively lower risk. This implies that using average financial measures provide a better fundamental perspective to assessing peers, while using more peers likely reduces the impact of outliers.

To conclude, an investor would have seen risk adjusted returns higher than those of the benchmark on average over the time period based on the historical backtest of our model.

8.2. SXXP Strategies

8.2.1. Non-average & 6 Peer Sample

This backtest outlines the results of the portfolios from the STOXX Europe 600 index using non-average selection fundamentals and 6 peers.

Table 8.2.1 – Returns and Performance

(SXXP, non-average, 6 peers)

(Strategy)	SXXP	ROE (1)	TA (2)	G (3)	ROIC (4)	WACC (5)	ROE TA (6)	ROE TA G (7)	ROIC TA G WACC (8)	ROE TA G WACC ROIC (9)
1. Annualized Return										
EV/Sales	7.1%	10.7% (6)	11.4% (3)	11% (5)	10.3% (9)	11.9% (2)	13% (1)	10.6% (8)	11.3% (4)	10.6% (7)
EV/EBITDA	7.1%	11.7% (5)	12.4% (2)	12% (4)	10.8% (8)	12.5% (1)	12.1% (3)	9.3% (9)	11% (7)	11.2% (6)
P/CF	7.1%	11.3% (3)	11.6% (2)	10.3% (7)	10.8% (4)	12.8% (1)	9.4% (9)	10.2% (8)	10.8% (5)	10.7% (6)
P/B	7.1%	14.5% (1)	10% (9)	14.1% (2)	12.3% (6)	11% (8)	13.4% (3)	12.9% (4)	11% (7)	12.7% (5)
2. Alpha										
EV/Sales		0.21%	0.27%	0.25%	0.19%	0.31%	0.4% **	0.22%	0.28%	0.22%
EV/EBITDA		0.31% *	0.38% **	0.35% *	0.26%	0.4% **	0.36% **	0.14%	0.25%	0.29%
P/CF		0.29% *	0.31% *	0.22%	0.25%	0.4% **	0.15%	0.2%	0.23%	0.24%
P/B		0.53% ***	0.18%	0.52% ***	0.36% *	0.26%	0.46% **	0.41% **	0.26%	0.4% **
3. Beta										
EV/Sales	1.00	1.39	1.34	1.27	1.32	1.32	1.25	1.29	1.28	1.29
EV/EBITDA	1.00	1.20	1.18	1.18	1.17	1.14	1.16	1.24	1.24	1.16
P/CF	1.00	1.18	1.21	1.20	1.20	1.22	1.17	1.24	1.28	1.20
P/B	1.00	1.22	1.24	1.13	1.22	1.22	1.17	1.22	1.26	1.18
4. Sharpe Ratio										
EV/Sales	0.34	0.39 (9)	0.43 (5)	0.44 (4)	0.39 (8)	0.47 (2)	0.55 (1)	0.41 (7)	0.45 (3)	0.41 (6)
EV/EBITDA	0.34	0.51 (5)	0.55 (3)	0.52 (4)	0.47 (7)	0.57 (1)	0.55 (2)	0.36 (9)	0.45 (8)	0.49 (6)
P/CF	0.34	0.49 (3)	0.50 (2)	0.44 (6)	0.46 (5)	0.56 (1)	0.40 (9)	0.41 (8)	0.43 (7)	0.46 (4)
P/B	0.34	0.65 (2)	0.40 (9)	0.67 (1)	0.52 (6)	0.46 (7)	0.61 (3)	0.55 (5)	0.44 (8)	0.56 (4)
5. Maximum Drawdown										
EV/Sales	-53.9%	-66.7%	-73.1%	-70.7%	-70.7%	-72.7%	-67.6%	-72.5%	-71.4%	-71.2%
EV/EBITDA	-53.9%	-55.0%	-65.1%	-63.9%	-62.3%	-62.2%	-61.7%	-69.4%	-66.7%	-61.5%
P/CF	-53.9%	-54.4%	-61.4%	-61.2%	-62.1%	-64.1%	-60.2%	-63.7%	-61.7%	-55.5%
P/B	-53.9%	-52.5%	-68.0%	-61.4%	-58.1%	-66.8%	-53.2%	-58.9%	-58.4%	-54.4%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The first sample backtested in Europe shows annualized returns in excess of the benchmark. Compared to the benchmark return of 7.1%, the portfolios yielded 11.5% per year on average. Looking back at the US sample, the average return here is identical, however the benchmark return in Europe is 1.6% lower in absolute terms. The price-to-book ratio has the highest average return across the different strategies reaching 12.5%. ROE, growth and multifactor strategies including these fundamentals have the highest annualized return. ROE and G can be derived from P/B, so the outcome of the backtest illustrates that the model is able to provide higher returns by selecting undervalued companies. Furthermore, the alpha supports the aforementioned conclusions. The strategies including ROE and G (1, 3, 6, 7, 9) are all significant at the 5% level or higher for P/B. Portfolios under EV/EBITDA and P/CF presented the most significant improvement in relative valuation, though the result of alpha can only be explained in a handful of the portfolios. WACC performs well for both multiples, suggesting that capital structure and the cost associated with capital provision to select undervalued firms is able to provide abnormal returns.

Beta is 1.23 on average representing 23% overexposure to the market. EV/Sales has the highest beta at 1.31, which is likely a result of the inaccurate peer selection and insignificant relative valuation improvement. The selected companies included in the portfolio would be chosen in a more random fashion since the model is not working well with this multiple, thus a higher volatility and market exposure are not surprising. In this case, the high beta is correlated with the maximum drawdown for several portfolios as the average drawdown is approximately -63.6%.

Lastly, the Sharpe ratio in part 4 of the table is closely related to the return and risk measures discussed. The average of 0.49 is roughly a 44% increase in risk adjusted return versus the Europe 600 index. This is a result of a drastically larger annualized return, and a slightly higher risk profile. The Sharpe ratio's do not explicitly mimic the finding of the peer selection, where the multifactor strategies typically displayed the lowest APEs, because the single factor portfolios tend to have the most attractive risk adjusted return.

8.2.2. Non-average & 12 Peer Sample

This backtest outlines the results of the portfolios from the STOXX Europe 600 strategy using non-average selection fundamentals and 12 peers.

Table 8.2.2 – Returns and Performance

(SXXP, non-average, 12 peers) **ROE**

(Strategy)	SXXP	ROE (1)	TA (2)	G (3)	ROIC (4)	WACC (5)	ROE TA (6)	ROE TA G (7)	ROIC TA G WACC (8)	TA G WACC ROIC (9)
1. Annualized Return										
EV/Sales	7.1%	11.1% (5)	10.9% (6)	10.7% (7)	11.7% (4)	12% (3)	13.1% (1)	12.3% (2)	10.5% (8)	10% (9)
EV/EBITDA	7.1%	10.9% (5)	10.8% (6)	11.7% (3)	10.9% (4)	12.5% (1)	9.5% (9)	10.2% (8)	12.3% (2)	10.7% (7)
P/CF	7.1%	13.1% (2)	11.3% (3)	11.3% (4)	9.8% (8)	13.4% (1)	10.5% (7)	9.3% (9)	11.3% (5)	10.8% (6)
P/B	7.1%	13.4% (3)	9.6% (9)	14.5% (1)	11.8% (6)	10% (8)	14.3% (2)	13.2% (4)	12.5% (5)	11.4% (7)
2. Alpha										
EV/Sales		0.24%	0.24%	0.21%	0.3%	0.33% *	0.42% **	0.34% *	0.21%	0.17%
EV/EBITDA		0.26%	0.26%	0.31% *	0.28%	0.4% **	0.17%	0.2%	0.37% **	0.26%
P/CF		0.43% **	0.29%	0.29% *	0.17%	0.45% ***	0.24%	0.12%	0.27%	0.25%
P/B		0.46% **	0.16%	0.53% ***	0.34% *	0.2%	0.54% ***	0.43% **	0.38% *	0.31%
3. Beta										
EV/Sales	1.00	1.36	1.28	1.39	1.31	1.25	1.26	1.34	1.31	1.35
EV/EBITDA	1.00	1.17	1.16	1.20	1.16	1.11	1.13	1.25	1.21	1.17
P/CF	1.00	1.17	1.22	1.18	1.22	1.19	1.15	1.27	1.28	1.23
P/B	1.00	1.16	1.23	1.22	1.18	1.18	1.12	1.22	1.23	1.18
4. Sharpe Ratio										
EV/Sales	0.34	0.42 (6)	0.43 (5)	0.39 (8)	0.46 (4)	0.50 (2)	0.54 (1)	0.48 (3)	0.40 (7)	0.36 (9)
EV/EBITDA	0.34	0.48 (4)	0.47 (6)	0.51 (3)	0.47 (5)	0.58 (1)	0.41 (8)	0.40 (9)	0.53 (2)	0.46 (7)
P/CF	0.34	0.60 (2)	0.48 (4)	0.49 (3)	0.39 (8)	0.61 (1)	0.46 (5)	0.36 (9)	0.46 (6)	0.45 (7)
P/B	0.34	0.61 (3)	0.37 (9)	0.65 (2)	0.51 (6)	0.42 (8)	0.68 (1)	0.56 (4)	0.52 (5)	0.49 (7)
5. Maximum Drawdown										
EV/Sales	-53.9%	-66.1%	-71.2%	-72.2%	-68.4%	-69.8%	-64.8%	-71.3%	-72.9%	-72.4%
EV/EBITDA	-53.9%	-56.6%	-61.1%	-61.5%	-61.6%	-58.8%	-64.2%	-68.2%	-63.3%	-66.0%
P/CF	-53.9%	-50.5%	-59.0%	-57.9%	-62.8%	-58.9%	-57.1%	-66.0%	-63.0%	-63.9%
P/B	-53.9%	-50.1%	-65.6%	-52.5%	-57.0%	-66.0%	-48.3%	-59.0%	-59.0%	-57.5%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

This sample which uses non-average fundamentals and 12 peers continues the trend of high annualized returns. Like for the previous sample, the return is also 11.5% and P/B has the highest average one at around 12.4%. The portfolio returns display quite similar results as for the previous sample, therefore it is difficult to say whether more peers in the model improve portfolio performance.

The number of significant abnormal portfolio returns is unchanged for this sample and is a result of similar returns. The P/B ratios are no longer significant in all the strategies with ROE, however both portfolios with ROIC are significant at a 10% level. Relating this back to the peer selection, we could see APEs improving for ROIC strategies when 12 peers were used. Looking at EV/Sales and EV/EBITDA where growth/WACC/ROIC can be derived from, it is not possible to present a clear relation between these factors and significant alpha. When considering EV/EBITDA, growth and WACC as single factor tests as well as strategy 8, all illustrate abnormal returns that can be explained. However, there is not a clear trend of increasing significance alphas when using more selection factors.

The average beta of all the portfolios is 1.22, which is just under that of the previous sample. So far, the distinction between 6 and 12 peers is quite minimal for portfolios based on the European index. Since market exposure is nearly identical, the riskiness and risk-adjusted return is expected to show similar results. Comparing the maximum loss from a peak to a trough described by the maximum drawdown, the data results in a slightly lower drawdown of -62.3% versus -63.3% for the previous sample. P/B has the lowest drawdown risk given by the average of -57.2%. It is lowest for the portfolios using ROE, growth and a combination of these two fundamentals. On the other hand, EV/Sales has the highest drawdown of about -70% which clearly shows that using EV/Sales to select investments leads to higher volatility. This correlates with our findings in the peer selection chapter.

Lastly, the average risk-adjusted return as given by the Sharpe ratio is unchanged from the previous sample, i.e. 0.49. This is still an outperformance relative to the benchmark of roughly 44%. As expected, P/B has the highest Sharpe ratio on average, while EV/EBITDA and P/CF both have it at 0.48. EV/EBITDA tends to have more accurate peer selection with lower variation compared to P/CF, so it is interesting to see that the portfolios often have similar results.

8.2.3. Average & 6 Peer Sample

This backtest outlines the results of the portfolios from the STOXX Europe 600 strategy using average selection fundamentals and 6 peers.

Table 8.2.3 – Returns and Performance

(SXXP, average, 6 peers)										ROE
									ROIC	TA
								ROE	TA	G
							ROE	TA	G	WACC
	SXXP	ROE	TA	G	ROIC	WACC	TA	G	WACC	ROIC
(Strategy)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Annualized Return										
EV/Sales	7.1%	11.7% (4)	10% (9)	11.7% (3)	10.9% (7)	13% (1)	11.5% (6)	10.4% (8)	11.7% (2)	11.5% (5)

EV/EBITDA	7.1%	12.4% (3)	13.5% (1)	9.4% (9)	11.3% (5)	12.7% (2)	12.3% (4)	10.8% (6)	10.4% (7)	10.2% (8)
P/CF	7.1%	11.9% (3)	10.5% (7)	11.8% (4)	9.5% (9)	12.3% (2)	12.4% (1)	10.6% (6)	10.2% (8)	10.9% (5)
P/B	7.1%	13.1% (1)	10.7% (7)	12.5% (2)	11.2% (4)	11% (5)	9.6% (9)	10.8% (6)	9.8% (8)	12.1% (3)
2. Alpha										
EV/Sales		0.3%	0.17%	0.3%	0.24%	0.4% **	0.3%	0.21%	0.32% *	0.3% *
EV/EBITDA		0.4% **	0.47% ***	0.14%	0.32% *	0.39% **	0.37% **	0.27%	0.23%	0.21%
P/CF		0.36% **	0.24%	0.31% *	0.16%	0.34% *	0.38% **	0.24%	0.2%	0.27% *
P/B		0.44% **	0.24%	0.37% **	0.29%	0.24%	0.17%	0.27%	0.19%	0.37% *
3. Beta										
EV/Sales	1.00	1.30	1.32	1.32	1.29	1.27	1.22	1.24	1.20	1.21
EV/EBITDA	1.00	1.10	1.12	1.22	1.11	1.22	1.19	1.12	1.18	1.17
P/CF	1.00	1.11	1.15	1.26	1.18	1.30	1.20	1.17	1.20	1.16
P/B	1.00	1.16	1.22	1.23	1.20	1.30	1.21	1.12	1.18	1.14
4. Sharpe Ratio										
EV/Sales	0.34	0.46 (5)	0.37 (9)	0.46 (6)	0.42 (8)	0.55 (1)	0.48 (4)	0.42 (7)	0.51 (2)	0.49 (3)
EV/EBITDA	0.34	0.59 (2)	0.64 (1)	0.37 (9)	0.51 (5)	0.54 (3)	0.54 (4)	0.49 (6)	0.44 (7)	0.43 (8)
P/CF	0.34	0.56 (1)	0.46 (7)	0.50 (3)	0.39 (9)	0.49 (4)	0.55 (2)	0.47 (6)	0.43 (8)	0.49 (5)
P/B	0.34	0.60 (1)	0.44 (6)	0.54 (3)	0.47 (5)	0.43 (7)	0.38 (9)	0.48 (4)	0.40 (8)	0.54 (2)
5. Maximum Drawdown										
EV/Sales	-53.9%	-65.0%	-71.6%	-63.0%	-71.0%	-60.5%	-65.2%	-66.5%	-61.9%	-62.0%
EV/EBITDA	-53.9%	-50.5%	-49.4%	-62.7%	-63.5%	-61.9%	-65.2%	-56.6%	-57.5%	-59.2%
P/CF	-53.9%	-50.6%	-56.7%	-55.5%	-63.9%	-60.4%	-59.8%	-57.7%	-55.7%	-53.8%
P/B	-53.9%	-48.3%	-58.8%	-65.9%	-60.6%	-62.9%	-60.7%	-55.0%	-59.3%	-55.2%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

In section 8.1.3 when analyzing S&P 500 strategies, we saw an improvement in the risk-adjusted returns and now we will understand whether this trend continues. At a first glance, the annualized returns seem 20 basis points lower than those of the non-average sample (11.3%). Compared to the benchmark, this sample's annualized return is about 59% higher and therefore significantly outperforms.

The mean alpha for this sample is 0.29% and each portfolio has a positive alpha illustrating that the portfolios are all able to produce abnormal return. However, the significance of the alphas exhibits that not all portfolio alphas can be explained by the model. In this case, ROE as a single factor strategy presents significant alpha for EV/EBITDA, P/CF and P/B suggesting that ROE is useful in selecting undervalued investments in the model. Looking at strategy 9, which uses all fundamentals, we see the significant alpha for all multiples but EV/EBITDA. Though ROE is significant for EV/EBITDA as are TA, ROIC and WACC, it is surprising that strategy 9 alpha is not statistically better. This demonstrates that combined fundamentals do not necessarily add value. Interestingly, the median peer selection absolute percentage errors for EV/EBITDA were equal for strategies 1, 4 and 9.

The market exposure given by beta indicates a high market exposure over the backtest and returns a value of 1.20. This is slightly lower than for the non-average sample, and therefore the drawdown is also expected to be lower. The maximum loss over the period is 3.5% less than in the other sample. The beta of this sample is 20% higher than the market's, whereas the maximum drawdown is higher by 11%. This illustrates that an investors would have captured more of the upside because they are more exposed to market fluctuations but have proportionally less downside risk.

Once again, the risk-adjusted return presented by the Sharpe ratio results in an excess return per unit of risk of 0.49. Though the riskiness of this sample analyzed is slightly lower than in the non-average sample, returns are also marginally lower resulting in no net Sharpe ratio change on average. EV/EBITDA has the highest mean Sharpe ratio due to portfolio 1 and 2 having drawdowns under the benchmark and thus comparably lower volatilities.

8.2.4. Average & 12 Peer Sample

This backtest outlines the results of the portfolios from the STOXX Europe 600 strategy using average selection fundamentals and 12 peers.

Table 8.2.4 – Returns and Performance

(SXXP, average, 12 peers)

	SXXP	ROE	TA	G	ROIC	WACC	ROE TA	ROE TA G	ROIC TA G WACC	ROE TA G WACC ROIC
(Strategy)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Annualized Return										
EV/Sales	7.1%	10.1% (6)	10.1% (7)	11.5% (2)	10.7% (4)	12.2% (1)	9.2% (9)	10.5% (5)	9.7% (8)	10.9% (3)
EV/EBITDA	7.1%	12% (3)	10.8% (7)	10.7% (9)	11.7% (5)	12.9% (1)	11.3% (6)	12.6% (2)	11.8% (4)	10.8% (8)
P/CF	7.1%	12.2% (4)	11.9% (5)	12.3% (2)	11.3% (6)	14% (1)	12.3% (3)	10.1% (9)	11.2% (8)	11.2% (7)
P/B	7.1%	13.3% (3)	11.7% (8)	13.6% (2)	14.5% (1)	7.9% (9)	12.5% (4)	12.4% (5)	11.8% (7)	12.1% (6)
2. Alpha										
EV/Sales		0.19%	0.18%	0.3%	0.21%	0.34% *	0.12%	0.23%	0.15%	0.25%
EV/EBITDA		0.35% **	0.29%	0.25%	0.31% *	0.42% **	0.3%	0.41% **	0.34% *	0.26%
P/CF		0.37% **	0.35% **	0.35% **	0.29% *	0.48% ***	0.36% **	0.2%	0.27%	0.28%
P/B		0.45% **	0.33% *	0.48% ***	0.53% ***	0.03%	0.39% *	0.4% **	0.33% *	0.37% *
3. Beta										

EV/Sales	1.00	1.27	1.33	1.25	1.39	1.27	1.29	1.21	1.31	1.26
EV/EBITDA	1.00	1.14	1.08	1.18	1.20	1.18	1.17	1.11	1.19	1.18
P/CF	1.00	1.16	1.15	1.26	1.18	1.24	1.23	1.19	1.25	1.20
P/B	1.00	1.18	1.19	1.16	1.22	1.24	1.18	1.13	1.23	1.17
4. Sharpe Ratio										
EV/Sales	0.34	0.39 (5)	0.38 (7)	0.47 (2)	0.39 (6)	0.50 (1)	0.34 (9)	0.43 (4)	0.36 (8)	0.44 (3)
EV/EBITDA	0.34	0.54 (3)	0.50 (6)	0.46 (8)	0.51 (5)	0.57 (2)	0.49 (7)	0.58 (1)	0.51 (4)	0.46 (9)
P/CF	0.34	0.55 (2)	0.54 (3)	0.52 (5)	0.49 (6)	0.61 (1)	0.53 (4)	0.42 (9)	0.47 (8)	0.48 (7)
P/B	0.34	0.59 (3)	0.50 (7)	0.62 (2)	0.65 (1)	0.29 (9)	0.54 (5)	0.56 (4)	0.49 (8)	0.52 (6)
5. Maximum Drawdown										
EV/Sales	-53.9%	-64.3%	-72.0%	-59.9%	-69.7%	-67.0%	-74.0%	-68.7%	-72.9%	-71.3%
EV/EBITDA	-53.9%	-51.0%	-58.0%	-64.2%	-58.2%	-63.3%	-67.9%	-56.3%	-59.1%	-62.7%
P/CF	-53.9%	-53.2%	-57.2%	-59.7%	-54.4%	-61.3%	-62.4%	-61.4%	-62.3%	-59.3%
P/B	-53.9%	-48.1%	-58.3%	-57.8%	-52.5%	-72.0%	-62.0%	-58.9%	-57.9%	-59.4%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The annualized returns present the highest return (11.6%) per year on average across all the portfolios. Portfolios using P/B continue to exhibit the highest return among multiples reaching 12.3%. EV/Sales has the lowest annualized return on average at 10.6%. The WACC portfolios for EV/Sales, EV/EBITDA and P/CF have the highest return for each respective multiple, but comparing it back to the absolute percentage errors it is difficult to conclude that a substantial level of accuracy is the cause of these returns.

Looking into the alphas as a result of the regression shows an overall abnormal return generated from this sample as the average alpha is 0.32%. This is about 30 basis points greater than for the previous sample using just 6 peers. ROE continues to provide significant abnormal returns for EV/EBTIDA, P/CF and P/B suggesting that the undervalued firms selected based on ROE as the main peer determining fundamental provides consistent outperformance relative to the benchmark. P/B is clearly the best multiple to use for achieving high abnormal return because 8 of the 9 portfolios exhibit an alpha that can be explained. Though Wilcoxon's test does not show a large degree of significant relative valuation improvement, the absolute percentage errors are typically lowest for P/B which supports the solid alpha results. Portfolios 7 and 8 for EV/EBITDA can be related back to peer selection because these two strategies are some of the best performing in terms of low absolute percentage errors. The significance levels and alphas also improved for P/CF in this sample which suggests that the previous sample has some variation between the comparables and therefore the use of 12 peers provides a more stable result. In addition, the mean absolute percentage errors for P/CF were slightly lower for the 12-peer sample compared to 6 peers.

The average beta for all the portfolios in the sample is 1.21 or 1% greater than for the previous sample in absolute terms. The model continues to build a portfolio that has market exposure greater than 1. This result is skewed upwards because the average beta for EV/Sales multiple is 1.29. Despite EV/Sales biasing the beta upwards, the risk of the various portfolios is still expected to be greater than that of the market on average. For example, the maximum drawdown is -61.6% on average based on historical data. Few portfolios would have lost less than the benchmark over the period of time, but ROE performed quite well for EV/EBITDA, P/CF and P/B.

Finally, the Sharpe ratio describes the excess return per unit of risk and in this sample the ratio is 0.50 on average. This is very slightly higher than for the previous 3 samples, but it presents the case that using 3-year averages and 12 peers is the best model to use for earning the greatest risk-adjusted returns.

8.2.5. SXXP Returns and Performance Conclusion

The results and performance section for the portfolios based on the STOXX Europe 600 highlighted the returns had an investor used the model historically. The backtest illustrates the performance of the different strategies relative to the benchmark, which highlights key statistics such as risk and return.

The annualized returns of all the portfolios outperformed the benchmark in terms of return since the average return of the four samples is between 11.3% and 11.6%. Using the median average yearly return of 11.5% over the 14-year time period results in a total return of about 359%, compared to the benchmark of 161% resulting in almost 200% outperformance over the period. From a purely return perspective, an investor would clearly have benefitted from using this model to invest. However, relative to the peer selection, the annualized returns do not illustrate clear performance improvements when multiple factors are used to select peers.

The outcome of the alphas exhibits a few clear trends that persist for the different samples. In particular, EV/Sales has the lowest absolute returns of the multiples which ties into the peer selection as the multiple also has the highest absolute percentage errors and rarely shows improvement regarding this matter. On the other hand, P/B presents the highest absolute returns and many of its underlying portfolios are significant. This is clearly correlated with peer selection since P/B has the least variation and lowest absolute percentage errors. ROE also has a tendency to select undervalued firms based on the multiples like EV/EBITDA, P/CF and P/B. Another single factor strategy that results in alpha that is explainable is WACC for EV/EBITDA and P/CF. WACC is interesting because it encompasses firm leverage, i.e. the capital structure, in addition to the riskiness of the capital based on the required rate of return for debt and equity. However, relating this to selection accuracy, WACC is not the best at selecting firms with similar valuations but this also means that undervalued firms are going to be more undervalued from a relative valuation perspective.

The portfolio betas calculated in the regressions illustrate only small differences between samples as the four sample averages are between 1.20 and 1.23. Therefore, the portfolios are generally about 20% overexposed to market movements and should be susceptible to a higher degree of volatility and risk. Since we have defined risk as the maximum loss an investor would have suffered from it historically. The statistics show that a loss of around 60% is not unlikely.

Given that a portfolio would have an exposure 20% greater than the market, but a maximum loss that is about 11% higher than the market, it is possible to assume that the model takes more risk but captures even more return on the upside. The Sharpe ratio also puts this into perspective since all portfolios equal or outperform the benchmark’s Sharpe ratio of 0.34. This means that the return in excess of the risk-free rate is higher than the market’s per unit of risk taken. While this is clearly beneficial for an investor, the relation to peer accuracy is less clear because multifactor strategies only result in greater risk-adjusted returns on occasion.

In conclusion, the sample with 3-year averages and 12 peers presented a small improvement in performance relative to the other samples and P/B is the best multiple for investors to select their portfolio.

8.3. MXEF Strategies

8.3.1. Non-average & 6 Peer Sample

This backtest outlines the results of the portfolios from the MSCI Emerging Markets strategy using non-average selection fundamentals and 6 peers.

Table 8.3.1 – Returns and Performance

(MXEF, non-average, 6 peers)

				ROE
			ROIC	TA
		ROE	TA	G
	ROE	TA	G	WACC

	MXEF	ROE	TA	G	ROIC	WACC	TA	G	WACC	ROIC
(Strategy)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Annualized Return										
EV/Sales	10.1%	16.2% (9)	18.2% (6)	17% (7)	19% (2)	18.9% (3)	19.3% (1)	18.2% (5)	17% (8)	18.8% (4)
EV/EBITDA	10.1%	14.7% (8)	16.6% (4)	16.6% (3)	16% (6)	16.4% (5)	12.4% (9)	14.8% (7)	17.9% (2)	18% (1)
P/CF	10.1%	17.5% (3)	16.8% (7)	17.4% (4)	15.1% (9)	17.4% (5)	18.1% (2)	18.5% (1)	16.8% (6)	15.7% (8)
P/B	10.1%	17.5% (2)	13.4% (9)	15.1% (8)	17% (3)	16.4% (4)	15.9% (6)	15.9% (5)	15.5% (7)	17.5% (1)
2. Alpha										
EV/Sales		0.5% **	0.65% ***	0.55% **	0.71% ***	0.7% ***	0.73% ***	0.65% ***	0.55% **	0.69% ***
EV/EBITDA		0.39% **	0.53% ***	0.52% ***	0.49% **	0.51% ***	0.22%	0.42% **	0.62% ***	0.65% ***
P/CF		0.61% ***	0.58% ***	0.6% ***	0.44% **	0.6% ***	0.66% ***	0.69% ***	0.56% ***	0.48% **
P/B		0.59% ***	0.31%	0.42% *	0.55% **	0.53% ***	0.48% **	0.49% **	0.45% **	0.59% **
3. Beta										
EV/Sales	1.00	0.96	0.94	1.01	0.95	0.93	0.94	0.97	1.04	0.98
EV/EBITDA	1.00	0.90	0.91	0.97	0.93	0.90	0.92	0.87	0.93	0.89
P/CF	1.00	0.86	0.81	0.88	0.85	0.85	0.86	0.81	0.87	0.86
P/B	1.00	0.93	0.88	0.95	0.98	0.88	0.93	0.91	0.96	0.97
4. Sharpe Ratio										
EV/Sales	0.26	0.50 (8)	0.60 (5)	0.51 (7)	0.63 (3)	0.65 (2)	0.65 (1)	0.58 (6)	0.49 (9)	0.60 (4)
EV/EBITDA	0.26	0.48 (8)	0.56 (3)	0.53 (5)	0.52 (6)	0.56 (4)	0.36 (9)	0.49 (7)	0.59 (2)	0.61 (1)
P/CF	0.26	0.62 (4)	0.62 (5)	0.62 (6)	0.51 (9)	0.64 (3)	0.66 (2)	0.72 (1)	0.59 (7)	0.54 (8)
P/B	0.26	0.58 (1)	0.41 (9)	0.46 (8)	0.53 (4)	0.56 (2)	0.50 (6)	0.51 (5)	0.47 (7)	0.55 (3)
5. Maximum Drawdown										
EV/Sales	-61.4%	-50.4%	-57.2%	-60.8%	-58.4%	-64.3%	-62.5%	-67.2%	-73.1%	-67.5%
EV/EBITDA	-61.4%	-50.7%	-50.1%	-52.5%	-52.3%	-58.8%	-69.3%	-58.5%	-59.0%	-58.1%
P/CF	-61.4%	-46.0%	-47.4%	-46.5%	-59.5%	-59.7%	-60.0%	-49.4%	-53.5%	-54.3%
P/B	-61.4%	-48.2%	-60.3%	-57.4%	-52.1%	-59.9%	-60.1%	-54.0%	-55.7%	-46.9%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

First impression of the emerging markets performance is that annualized returns are higher for both the benchmark and portfolios. Specifically, the average annualized return for the different portfolios is 16.8%, which is 66% higher than the benchmark's of 10.1%. Only analyzing return, illustrates more lucrative investment opportunities in emerging markets as the benchmark has a higher annualized return than SPX and SXXP. In addition, the returns are more relatable to the peer selection, because several multifactor portfolios generate high returns compared to the other portfolios based on the same multiples. For example, strategy 9 results in the highest return for EV/EBITDA and P/B which corresponds accurate peer selection. EV/Sales displays an average annualized return of 18.1% whereas EV/EBITDA and P/B both yield about 16%.

Moving on to alpha in part 2, the initial observations is that many of the portfolios have significant alphas, particularly at the 1% significance level. 20 of the 36 portfolios are significant at 1% and the average alpha is 0.55% between the various portfolios. Alpha is greatest and significant for every strategy in EV/Sales and P/CF. This is seemingly contradictory to the absolute percentage errors in the peer selection because these two multiples show the highest degree of variation between mean and median values and the largest absolute percentage errors. This suggests that for inaccurate strategies, the selected investments are more undervalued and as a result have more upside.

The beta for this sample shows interesting statistics compared to the other indices because the average beta is 0.92. Only 2 portfolios, 3 and 8 for EV/Sales, have betas that are slightly higher than the market's which means that the exposure to market movements is typically less than the benchmark's. Considering that annualized returns are higher than the benchmark, it is surprising that the beta is lower because less risk usually means lower return. Since beta is lower than the market, then the volatility and overall riskiness of the strategies are expected to be lower compared to the benchmark. Analyzing the maximum drawdown, the risk clearly appears to be lower than that of the market because the average drawdown is about -56.7%. Compared to the benchmark, investors would have lost 4.7% less based on historically.

Lastly, the average Sharpe ratio of 0.56 is 115% greater than the benchmark's which suggests that the model considerably outperforms in terms of returns while simultaneously mitigating risk. P/CF has the highest Sharpe ratio on average (0.62) which correlates with the significant returns, but unlike EV/Sales that also displays good returns, P/CF has considerably lower risk which is demonstrated by a 10% lower drawdown. This suggests that the per dollar value of cash flow in emerging markets is very applicable to finding undervalued firms.

8.3.2. Non-average & 12 Peer Sample

This backtest outlines the results of the portfolios from the MSCI Emerging Markets strategy using non-average selection fundamentals and 12 peers.

Table 8.3.2 – Returns and Performance

(MXEF, non-average, 12 peers)

										ROE
									ROIC	**TA**
								ROE	**TA**	**G**
							ROE	**TA**	**G**	**WACC**
	MXEF	**ROE**	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**

(Strategy)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Annualized Return										
EV/Sales	10.1%	18.2% (5)	19.3% (2)	16.9% (7)	18.2% (4)	17.9% (6)	18.6% (3)	19.7% (1)	15.2% (9)	16.2% (8)
EV/EBITDA	10.1%	12.8% (9)	15.1% (3)	13.5% (8)	13.6% (7)	15.5% (2)	13.8% (6)	16.7% (1)	14.9% (4)	14.7% (5)
P/CF	10.1%	17.5% (4)	17.6% (3)	17.7% (2)	15% (8)	16.5% (6)	15.2% (7)	18.6% (1)	14.1% (9)	17.5% (5)
P/B	10.1%	15% (7)	15.1% (6)	14.3% (9)	16.4% (3)	15.8% (4)	14.6% (8)	15.4% (5)	17.5% (1)	17.5% (2)
2. Alpha										
EV/Sales		0.64% ***	0.72% ***	0.55% **	0.66% ***	0.63% ***	0.68% ***	0.75% ***	0.43% *	0.5% **
EV/EBITDA		0.26%	0.42% **	0.3%	0.31%	0.45% **	0.34% *	0.53% ***	0.42% *	0.39% **
P/CF		0.62% ***	0.63% ***	0.63% ***	0.45% **	0.53% ***	0.45% **	0.69% ***	0.37% *	0.61% ***
P/B		0.4% *	0.44% **	0.37%	0.51% **	0.47% **	0.39% *	0.43% *	0.59% **	0.59% ***
3. Beta										
EV/Sales	1.00	0.96	0.97	1.00	0.94	0.97	0.94	0.98	0.98	0.96
EV/EBITDA	1.00	0.91	0.92	0.96	0.91	0.93	0.88	0.96	0.92	0.90
P/CF	1.00	0.85	0.83	0.85	0.83	0.87	0.82	0.85	0.85	0.86
P/B	1.00	0.96	0.86	0.95	0.94	0.92	0.95	1.00	0.99	0.93
4. Sharpe Ratio										
EV/Sales	0.26	0.58 (5)	0.63 (2)	0.50 (7)	0.59 (4)	0.57 (6)	0.62 (3)	0.63 (1)	0.44 (9)	0.50 (8)
EV/EBITDA	0.26	0.38 (9)	0.48 (3)	0.39 (8)	0.42 (7)	0.50 (2)	0.44 (6)	0.54 (1)	0.46 (5)	0.48 (4)
P/CF	0.26	0.63 (4)	0.65 (2)	0.65 (3)	0.52 (8)	0.58 (6)	0.54 (7)	0.69 (1)	0.46 (9)	0.62 (5)
P/B	0.26	0.46 (6)	0.49 (5)	0.42 (9)	0.53 (3)	0.51 (4)	0.44 (8)	0.45 (7)	0.53 (2)	0.58 (1)
5. Maximum Drawdown										
EV/Sales	-61.4%	-47.3%	-52.6%	-60.0%	-61.5%	-60.8%	-60.8%	-58.5%	-65.8%	-66.7%
EV/EBITDA	-61.4%	-61.9%	-60.1%	-67.5%	-61.9%	-55.6%	-62.3%	-49.9%	-62.2%	-61.3%
P/CF	-61.4%	-45.2%	-49.2%	-43.3%	-58.0%	-51.4%	-56.9%	-50.9%	-59.4%	-52.0%
P/B	-61.4%	-54.5%	-57.5%	-59.2%	-56.2%	-50.7%	-59.2%	-55.6%	-45.9%	-53.8%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

This sample presents average portfolio annualized returns of 16.2%, which is still significantly higher than the benchmark's but is 60 basis points lower than for the previous sample. EV/Sales still produces the greatest return compared to the other multiples as it reaches 17.9%. Portfolio 7 for EV/Sales, EV/EBITDA and P/CF display the best return for the respective multiples and, compared to the peer selection accuracy test, strategy 7 is in the top 33%. However, not all multifactor portfolios are the best performing in terms of return. For example, G in P/CF performs well in terms of return, but its mean absolute percentage errors were the highest for different strategies.

Alpha has decreased slightly overall for this sample, as the average is now 0.51% compared to 0.55% for the previous sample. The statistics exhibit a lack of significance for ROE, G and ROIC in EV/EBTIDA that were previously able to be explained by the model. However, EV/Sales and P/CF still display significant abnormal returns across all strategies, though the degree of significance has declined for some of the portfolios. This indicates that using more peers in emerging markets does not offer a degree of stabilization, but rather adds some outliers to the data. The result is less accuracy in the absolute percentage errors, particularly for P/CF, and overall lower alpha.

The addition of more peers has not had any effect on the average beta as it remained unchanged from the previous sample. However, the volatility of the betas illustrates that there is an underlying difference. For example, for EV/Sales in this sample the st. dev. is 1.97% of the betas across the 9 strategies compared to the previous sample, EV/EBITDA is 2.43% vs 2.6%, and P/CF is 1.50% vs 2.36%. Only P/B has slightly higher variation in the beta. This points to the idea that using more peers will spread risk out when determining undervalued investments. Furthermore, the maximum drawdown also reflects this observation because the mean drawdown is -56.5%, 0.2% lower relative to the previous sample. While the differences are still relatively small, there are some benefits to using 12 peers.

Lastly, the Sharpe ratio presents a small decrease in risk adjusted returns as the mean for this sample is 0.53 compared to 0.56 calculated previously. As risk is slightly lower for these portfolios, the return also is proportionally lower. Relative to the benchmark the Sharpe ratio is still considerably higher, and may be a result of small unprofitable firms negatively influencing the benchmark, whereas these same firms would be excluded from the model due to their characteristics.

8.3.3. Average & 6 Peer Sample

This backtest outlines the results of the portfolios from the MSCI Emerging Markets strategy using average selection fundamentals and 6 peers.

Table 8.3.3 – Returns and Performance

(MXEF, average, 6 peers)

(Strategy)	**MXEF**	**ROE** (1)	**TA** (2)	**G** (3)	**ROIC** (4)	**WACC** (5)	**ROE TA** (6)	**ROE TA G** (7)	**ROIC TA G WACC** (8)	**ROE TA G WACC ROIC** (9)
1. Annualized Return										
EV/Sales	10.1%	18.2% (8)	18.3% (7)	20% (4)	20.6% (2)	20.7% (1)	18.4% (6)	17.9% (9)	20.1% (3)	19.9% (5)
EV/EBITDA	10.1%	13.5% (9)	14.5% (8)	17.4% (2)	14.9% (7)	18.5% (1)	15.4% (5)	15.1% (6)	15.8% (4)	16.8% (3)
P/CF	10.1%	15% (8)	16.6% (6)	17.3% (3)	14.5% (9)	15.1% (7)	16.6% (5)	18.2% (2)	17.3% (4)	18.7% (1)
P/B	10.1%	16.4% (3)	13.9% (8)	15.2% (5)	16.3% (4)	13.9% (9)	17.2% (2)	13.9% (7)	14.7% (6)	17.8% (1)

2. Alpha										
EV/Sales		0.64% ***	0.65% ***	0.78% ***	0.82% ***	0.83% ***	0.66% ***	0.61% ***	0.77% ***	0.76% ***
EV/EBITDA		0.3%	0.38% *	0.58% ***	0.42% **	0.67% ***	0.45% **	0.42% **	0.48% **	0.54% ***
P/CF		0.42% **	0.56% ***	0.6% ***	0.39% **	0.45% **	0.56% ***	0.67% ***	0.61% ***	0.69% ***
P/B		0.51% **	0.36% *	0.43% **	0.5% **	0.34%	0.56% ***	0.33%	0.39% *	0.6% ***
3. Beta										
EV/Sales	1.00	0.94	0.98	0.95	0.95	0.95	0.98	1.01	0.98	1.01
EV/EBITDA	1.00	0.92	0.94	0.93	0.87	0.92	0.93	0.94	0.92	0.93
P/CF	1.00	0.86	0.84	0.85	0.86	0.83	0.84	0.84	0.84	0.88
P/B	1.00	0.97	0.86	0.93	0.94	0.92	0.99	0.95	0.95	0.99
4. Sharpe Ratio										
EV/Sales	0.26	0.61 (6)	0.58 (8)	0.67 (3)	0.70 (1)	0.70 (2)	0.59 (7)	0.55 (9)	0.66 (4)	0.63 (5)
EV/EBITDA	0.26	0.41 (9)	0.45 (8)	0.58 (2)	0.49 (5)	0.64 (1)	0.49 (6)	0.48 (7)	0.51 (4)	0.55 (3)
P/CF	0.26	0.51 (8)	0.58 (6)	0.63 (3)	0.48 (9)	0.52 (7)	0.59 (5)	0.68 (1)	0.62 (4)	0.66 (2)
P/B	0.26	0.51 (4)	0.44 (6)	0.48 (5)	0.53 (3)	0.42 (8)	0.53 (2)	0.41 (9)	0.44 (7)	0.56 (1)
5. Maximum Drawdown										
EV/Sales	-61.4%	-51.4%	-63.3%	-59.7%	-60.4%	-57.8%	-59.2%	-64.8%	-58.0%	-64.3%
EV/EBITDA	-61.4%	-62.1%	-63.2%	-57.4%	-63.2%	-57.6%	-62.3%	-62.6%	-56.6%	-56.5%
P/CF	-61.4%	-50.9%	-60.2%	-53.3%	-63.4%	-63.6%	-57.6%	-51.5%	-53.1%	-56.2%
P/B	-61.4%	-46.0%	-60.6%	-57.8%	-50.8%	-67.8%	-51.1%	-66.7%	-63.8%	-54.3%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The first sample using average fundamentals results in similar annualized returns as the non-average sample with 6 peers because the return is 16.8% on average. EV/Sales clearly outperforms the other multiples in terms of annualized returns because the average over the 9 portfolios is 19.4%, while P/CF has the 2nd highest average of 16.6%. As previously mentioned, peer selection accuracy is highest for the multifactor strategies which is more prevalent for P/CF and P/B returns, as the EV multiples only reflect better returns for multifactor portfolios in a couple instances.

The average alpha for this sample is 0.55%, the same as for the non-average portfolios. EV/Sales shows a high level of significance as each portfolio is significant at the 1% level. The degree of positive alpha for the various portfolios shows that the model is able to actively generate return in excess of the market, which is not explained by normal benchmark movements.

EV/EBITDA and P/CF were the only strategies to show significant improvement as given by the Wilcoxon signed rank test, so it is interesting that EV/Sales has the highest returns because the high return would mean that the relative valuation of the base companies is also converging greatly. Therefore, it would be expected that the Wilcoxon test showed significant results for this multiple, but theoretically the relative value could be increasing greatly and surpassing the average peer group, thus diverging from the average. This does not appear to be the case though after inspecting the change in relative valuation for the base companies and their respective peer groups. Therefore, it is fair to assume that the reason for not displaying significant improvement is most likely due to peers not actually being closely comparable.

The average beta of the portfolios is 0.92, illustrating that the portfolios are on average less exposed to market fluctuations. P/CF is the least exposed to the benchmark as it has a beta of 0.85, whereas EV/Sales has it at 0.97. Since beta is lower than the market's, then the volatility and risk of the portfolios are also predicted to be lower than the market's. The maximum drawdown reaffirms this theory as the maximum loss of -58.6% on average is below that of the market (-61.4%).

Lastly, the Sharpe ratio of 0.56 on average is 115% greater than the benchmark ratio. Portfolio 9 for the different multiples has the highest Sharpe ratio on average (0.61), which ties in with the peer selection analysis as strategy 9 was consistently the most accurate.

8.3.4. Average & 12 Peer Sample

This backtest outlines the results of the portfolios from the MSCI Emerging Markets strategy using average selection fundamentals and 12 peers.

Table 8.3.4 – Returns and Performance

(MXEF, average, 12 peers)

										ROE
									ROIC	**TA**
								ROE	**TA**	**G**
							ROE	**TA**	**G**	**WACC**
	MXEF	**ROE**	**TA**	**G**	**ROIC**	**WACC**	**TA**	**G**	**WACC**	**ROIC**
(Strategy)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Annualized Return										
EV/Sales	10.1%	19.5% (2)	19.4% (3)	18.1% (7)	18.1% (9)	19.7% (1)	18.8% (5)	18.1% (8)	19.1% (4)	18.3% (6)
EV/EBITDA	10.1%	12.2% (9)	13% (8)	15.3% (5)	15.7% (3)	16.4% (1)	14.4% (6)	15.6% (4)	16.2% (2)	14.1% (7)
P/CF	10.1%	15.9% (6)	18.2% (1)	15.4% (9)	16.3% (4)	15.6% (8)	16.1% (5)	16.5% (3)	16.7% (2)	15.7% (7)
P/B	10.1%	15.1% (5)	16.2% (3)	14.6% (6)	18.4% (1)	15.5% (4)	13.5% (9)	16.2% (2)	14.5% (7)	14.3% (8)
2. Alpha										
EV/Sales		0.75% ***	0.74% ***	0.64% ***	0.64% ***	0.77% ***	0.69% ***	0.64% ***	0.71% ***	0.65% ***
EV/EBITDA		0.2%	0.26%	0.43% **	0.47% **	0.51% ***	0.37% *	0.44% **	0.5% ***	0.34% *
P/CF		0.5% ***	0.67% ***	0.46% **	0.52% ***	0.48% **	0.52% **	0.54% ***	0.55% ***	0.48% **
P/B		0.41% *	0.52% **	0.39% *	0.66% ***	0.45% **	0.3%	0.49% **	0.37% *	0.36% *
3. Beta										

EV/Sales	1.00	0.92	0.96	0.94	0.97	0.92	0.96	0.99	0.98	1.00
EV/EBITDA	1.00	0.95	0.94	0.94	0.91	0.92	0.93	0.96	0.94	0.94
P/CF	1.00	0.82	0.83	0.86	0.86	0.84	0.85	0.86	0.87	0.87
P/B	1.00	0.98	0.87	0.92	0.95	0.96	0.97	0.97	0.98	0.96
4. Sharpe Ratio										
EV/Sales	0.26	0.66 (2)	0.64 (3)	0.59 (6)	0.57 (7)	0.67 (1)	0.62 (4)	0.56 (9)	0.61 (5)	0.57 (8)
EV/EBITDA	0.26	0.34 (9)	0.38 (8)	0.49 (4)	0.50 (3)	0.54 (1)	0.45 (6)	0.49 (5)	0.53 (2)	0.43 (7)
P/CF	0.26	0.57 (3)	0.67 (1)	0.53 (9)	0.57 (5)	0.54 (7)	0.55 (6)	0.57 (4)	0.58 (2)	0.53 (8)
P/B	0.26	0.45 (6)	0.54 (2)	0.45 (5)	0.61 (1)	0.48 (4)	0.38 (9)	0.51 (3)	0.42 (8)	0.42 (7)
5. Maximum Drawdown										
EV/Sales	-61.4%	-44.5%	-59.7%	-62.0%	-63.2%	-56.0%	-61.7%	-66.0%	-62.1%	-68.0%
EV/EBITDA	-61.4%	-70.4%	-65.4%	-59.8%	-59.0%	-56.3%	-62.8%	-61.9%	-57.5%	-62.2%
P/CF	-61.4%	-47.9%	-57.3%	-53.7%	-50.8%	-57.9%	-53.7%	-55.7%	-53.7%	-56.1%
P/B	-61.4%	-57.6%	-54.2%	-60.9%	-48.9%	-56.1%	-66.2%	-54.7%	-62.7%	-62.6%

In section 1 & 4, ranks of Annualized Returns and Sharpe Ratios across the 9 strategies are given in brackets

Significance levels are denoted for Alphas: * is the 10% level, ** is the 5% level, *** is the 1% level

Source: Authors' findings

The final sample of the emerging markets index generates an average annualized return of 16.3%, which is lower compared to the two 6 peer samples but 10 basis points higher than the non-average test with 12 peers. EV/Sales continues to display the highest average return, this time at 18.8% and EV/EBITDA the lowest at 14.8%. The returns of portfolio 5 in the two enterprise value multiples beat the other strategies but in the peer selection WACC was the 2nd least accurate.

Average alpha for this sample is about 0.52% which is a bit lower than in the 6-peer sample. Overall, peer selection is slightly more accurate for the 12-peer sample as absolute percentage errors are lower, but this appears to have a negative impact on absolute returns.

The average beta is the same as for the previous 3 samples in emerging markets. Compared to the 6-peer test, standard deviation of EV/EBITDA is slightly lower, 1.14% versus 1.90%, in addition to P/B which is now 3.20% compared to 3.90% from earlier. Risk based on market exposure does not seem to be lower relative to the samples with fewer peers, and beta is very similar as well. Average maximum drawdown is also identical at -58.6% which further supports the observation of no clear improvement to portfolio riskiness. The single factor portfolios tend to have a lower maximum expected loss, 57.1%, relative to the multifactor strategies, -60.5%. Considering the strategies that use multiple fundamentals are supposed to be more accurate in selecting peers, a disadvantage of this may be that the undervalued firms are typically within the same sectors what hurts a portfolio diversifications.

Unsurprisingly, risk adjusted return is slightly lower in this sample, 0.53 on average. Returns and alphas are slightly lower as well, while the level of risk is still the same as for previous tests. This signals a lower return per unit of risk taken. While the Sharpe ratio is lower, the difference is still quite small but it shows that for emerging markets, the use of averages and more peers does not necessarily correspond to better investment performance.

8.3.5. MXEF Returns and Performance Conclusion

The results in performance section for the portfolios based on the MSCI Emerging Markets index highlighted the returns had an investor used the model historically. The backtest illustrates the performance of the different strategies in relation to the benchmark, where statistics present measures of risk and return.

Annualized return of the different portfolios indicates a clear trend of excess returns relative to the benchmark. The returns of the four different samples range between 16.3% and 16.8% on average compared to the market of 10.1%. While the total return of MSCI Emerging Markets index is quite high, precisely, 285.5%, the return of 16.3% over 14 years for our portfolios gives 728.2%. An outperformance of 442.6% is exceptionally high and from investors' perspective, such a return would likely be a valuable asset in their overall portfolios, especially since this return has outperformed developed markets.

Alpha corresponds to the returns because the absolute returns displayed are over 0.50% on average between the various portfolios. This level of abnormal return illustrates that the model is creating significant value and generating return that cannot be explained by returns in the benchmark. Alpha and return exhibit clear conclusion that EV/Sales and P/CF perform best in terms of generating portfolio return. Relating this conclusion to the peer selection accuracy results, the absolute percentage errors illustrate that accuracy is lowest for those two multiples. While we would expect that more accurate peer selection would lead to improved portfolio performance, greater differences in relative valuation between the base company and peers also mean that the base company could be proportionally more undervalued and could therefore have large market returns.

The beta results show that the model creates portfolios that are typically less exposed to the market because the average beta is about 0.92. Since returns are quite high compared to the index, a beta under 1 is surprising, but asserts how great the portfolio performance is. Outperformance and less market exposure illustrates that a portfolio captures excessive upside return, while simultaneously minimizing the downside. The average beta for P/CF is about 0.85 and is the lowest of the different multiples. P/CF shows significant relative valuation mean reversion based on the Wilcoxon test which may explain why the strategy performs better than EV/Sales which is not significant.

Another measure of risk, maximum drawdown, presents the maximum amount an investor would have lost based on historical return profile of the portfolio. The non-average versus average samples show slightly different results, where the two non-average samples show a loss of 56.6% on average compared to 58.6% for the two average samples. This could be a result of some of the average fundamentals being outdated, since emerging markets are characterized by rapid growth. Applying 12 peers to the model also does not yield improvement to riskiness overall. However, when looking at ROE the maximum drawdown is lower for EV/EBITDA and P/B for the 6-peer sample. Relative to peer selection, 6 peer sample is also slightly less accurate which provides more potential upside to the undervalued companies.

Lastly, the Sharpe ratio shows the risk adjusted return which is highly correlated with the performance measurements previously discussed. For the non-average tests, the portfolios under P/CF perform the best at around 0.61, whereas EV/Sales is about 0.58. On the other hand, for the average tests EV/Sales has the highest Sharpe ratio at about 0.62, while it is 0.58 for P/CF. Since the non-average tests have slightly lower absolute percentage errors and lower risk based on maximum drawdown, it is fair to assume the non-average tests are better performing. Furthermore, EV/Sales is considered the least accurate multiple of the four chosen because it is the least detailed.

In conclusion, the strategies tested in this market perform very well in terms of return and risk relative to the benchmark. Since emerging markets are likely less efficient, the model is able to generate higher alpha.

8.4. Portfolio Performance – Conclusion

This previous section outlined the risk and return of portfolios across the three different markets; S&P 500, STOXX Europe 600, and MSCI Emerging Markets. Within each market we examined the return and risk profile versus the benchmark and identified key differences between the different markets.

The portfolios based in the U.S. are characterized by a total return of 445% compared to the benchmark of 222%. This outperformance is higher than the outperformance for the European index, where the average portfolio yielded 358% in total return relative to the benchmark of 161%. The clear leader in return is in the emerging markets, where the portfolios gave 751% compared to the benchmark of 286%. Initially, this illustrates that the model generates higher return in emerging markets compared to developed markets.

The trend continues for alpha, where emerging markets capture more absolute return than the developed strategies. This is likely due to market efficiency, because emerging companies tend to be less covered by analysts, so there is more opportunity to seize pricing and valuation inefficiencies. Alpha for emerging markets is generally unchanged between the different samples, and based on the multiples like EV/Sales and P/CF it is consistently the best performing. These two multiples have the greatest variation based on the absolute percentage errors which shows that the base companies are less similar in terms of relative valuation. Portfolios in the U.S. are characterized by the lowest alpha of the three markets, and the analysis clearly shows that the samples using average fundamentals yielded a higher absolute return. The same is also true for the 12 peer samples relative to using 6 peers. For Europe, P/B exhibited the most significant portfolios and especially for strategies including ROE, ROIC and growth. With respect to absolute percentage errors, P/B presented the most accurate peer selection for most strategies, so this multiple shows performance which corresponds peer selection.

The beta presents different characteristics of volatility and risk compared to the respective benchmark for the various markets. Specifically, the two developed market portfolios were historically about 20% more exposed to market movements as the beta is around 1.20. On the other hand, the MSCI EM strategies described an average portfolio beta of 0.92 which is about 8% less exposed to the market swings. However, just because beta in EM is lower than the benchmark does not mean the portfolios are less risky than those of developed market counterparts because the benchmarks are also inherently different. For example, when looking at the maximum drawdowns for the various benchmarks which are a result of the financial crisis, it can be seen that U.S. has the lowest expected loss of 50.9%, with Europe being 3% higher and EM the entire 10.5% higher. This illustrates that the model selects investments that minimize risks within EM, but does the opposite for developed markets. Despite developed market portfolios being compared to a benchmark with lower risk, the average expected loss for EM is actually lowest at 57.6% relative to 58.6% for the U.S. and 61.8% for Europe. Another interesting statistic for drawdown shows that the non-average 12-peer sample tends to yield the lowest drawdown of 57.9% on average across all markets, whereas the 6-peer sample gives the highest percentage, 61.1%. The two average samples have an average drawdown of 59.0% and 59.3% for 6 and 12 peers, respectively. This illustrates a more robust risk profile as there is less variation.

Finally, the Sharpe ratio describes the return relative to the risk taken. Initial observation offers the conclusion that there is not a clear result of better risk-adjusted returns for multi-factor portfolios. The overall Sharpe ratio for U.S. is 0.54, 0.49 for Europe, and 0.55 for EM on average between the different portfolios. Though returns are higher in EM than in S&P 500 with a similar level of risk in terms of drawdown, the Sharpe ratios are quite close because the excess return in EM is affected by a higher risk-free rate. Lastly, relating the risk adjusted return to peer selection, EV/Sales exhibits the lowest Sharpe ratio overall and is consistently the multiple with the highest absolute percentage errors and variation of errors.

9. Discussion

Model

To begin with, we will discuss the underlying assumptions of our model as those were the main drivers behind our peer selection and investment decision making. Peer selection using the SARD approach is guided by the idea that the five fundamentals (ROE, ROIC, WACC, TA, G) actually represent characteristics that are applicable to selecting peers. Had we selected different fundamentals for the model, the outcome could have led to another conclusion. Moreover, we have not directly considered factors like tax rates or accounting policies that are specific to individual countries, sectors or firms. For instance, our profitability measures are all after tax which could lead to a biased assessment of the value of a firm's core operations. Furthermore, differences in accounting policies can skew the results of otherwise identical firms, thus undermining the model's ability to select peers. While companies in the U.S. adhere to GAAP and European firms follow IFRS, the members of emerging market index are subject to varied standards and potentially less stringent market ethics. Furthermore, the quality of regulation or presence of corruption can have a significant impact on how and what financials are reported.

There are several potential pitfalls of using a ranking system to select peers. For example, if many of the fundamentals are quite close to one another on paper, they will still be assigned an individual rank that can influence the sum of absolute ranked differences between firms. This means a firm may be similar to another company percentage wise, but because their ranks are very different, the model would not identify them as close peers. The opposite is also true, as two firms could have very different financial measures but based on the dataset there could be a gap resulting in them being assigned similar ranks. Another consideration of the dataset is the exclusion of a company in a given year if one of the fundamentals was unavailable. This meant that the same dataset was used for each strategy and allowed for a fair comparison. However, we could have included more firms for some strategies if all required data those specific strategies was available, thus expanding the amount of observations.

The multiples applied to select undervalued firms were either strictly positive or likely to be positive for a given firm. As previously explained, the income statement based multiples use numbers that account for limited amount of or no expenses, in order to have more observations. To an extent, this is a drawback because investors are often interested in the bottom line, which represents earnings available to shareholders. On the other hand, more positive multiples meant fewer observations had to removed and resulted in a larger dataset, which is important for statistical analysis.

The underlying assumption of investing in undervalued companies is subject to a major risk. We believe that the peer groups represent the relative value of the base companies, which is why we have an expectation of mean reversion. The risk of this philosophy is that the base company might be undervalued for a specific reason, which is not accounted for in the peer selection. As a result, the relative value would not converge, and no excess return would be earned. However, we might be in trouble even if the model picks truly undervalued stocks based on our setup. In case of a market downturn, the value of both base company and its peers can decrease. The selected investments would then yield negative returns. However, we would have still beaten the benchmark if mean reversion had actually happened. Another point to mention is the method of rebalancing, as it implicitly results in buying the losers and selling the winners over the past month. The popular quantitative strategy of momentum investing does the opposite of this. It tries to capitalize on investing in upward trending stocks while shorting downward trending prices. However, this is fundamentally a different strategy and we still believe the model selects undervalued firms which should appreciate in price.

Implementation costs were largely ignored because trading costs have significantly decreased over time. While this is true for developed markets where many market makers are active, smaller exchanges in emerging markets can have higher transactions costs and larger spreads. These are difficult to estimate and can vary drastically between investors but are still important to keep in mind.

Peer selection

Analyzing and comparing absolute percentage errors for the different samples in SPX, SXXP and MXEF show very slight changes in accuracy between samples of varying peer group sizes and type of fundamentals. Some tests illustrated that the samples designed to be robust had lower APE's by a couple percentage points, whereas other tests resulted in the opposite. This suggests that peer selection is quite stable in the U.S., European and emerging markets, and that larger peer groups plus average fundamentals just reaffirm the "less" robust tests. Therefore, it is difficult to claim that robustness is created through the methods used to test this idea. The result also shows that adding 6 more peers, which are fundamentally farther away from the base company, does not impact the absolute percentage error. This means the firms that are closest in terms of SARD are not necessarily the closest in terms of relative valuation. An observation that clearly reoccurs throughout the different samples is the improvement of the absolute percentage errors for strategies that use multiple factors. As we intended, the increase in the amount of selection criteria covering growth, profitability and risk led to more similar relative valuations, i.e. increased peer selection accuracy.

The Wilcoxon test shows positive convergence for all strategies, which supports the underlying investment theory. It is important to note that the test presented focuses on the 20 most undervalued firms, which likely explains why the convergence test is positive. If we had run the test for the whole sample, then we would probably have observed variation in the trend signs. P/CF and EV/EBITDA stand out in the test as they are consistently significant around the 5% level. In terms of relative valuation, EV/Sales results in the largest valuation differences between the base company and its peers. Although P/CF is also quite different for Europe and emerging markets, it shows significant convergence suggesting that the base companies were actually relatively undervalued. A potential explanation for the significance of P/CF is related to the cash flow statement, which is typically considered to be least subject to manipulation of the three accounting statements. Generally, it illustrates a clearer picture of how a company funds their operations and where cash movements are coming from. Particularly in emerging markets where regulatory agencies may not have as many resources to combat fraud and corruption, cash flows can provide a more representative picture of financial health. P/B had consistently the lowest absolute percentage errors, which might explain the lack of significant mean reversion. Based on the absolute percentage errors for P/B, the results support the model by exhibiting that the peer groups have similar relative valuations to the base companies. This argument primarily stems from the comparison with other multiples tested.

We have constrained our tests to two peer group sizes, and both yielded nearly identical results. This means the optimal peer group size is unknown. A better sample size may be different from the ones we have used in the model, and may also vary between markets. The choice of 20 most undervalued firms can also be reconsidered. This decision makes sense from a diversification standpoint, however varying the sample size could have provided different results for the Wilcoxon tests.

Another key consideration of the peer selection model and ultimately the performance is the industry that the selected stocks are in. The model is built on the idea that firms with similar growth, profitability and risk characteristics should be valued similarly regardless of sector. However, industries tend to perform better or worse in different stages of the business cycle because sectors are impacted differently by economic contractions and expansions. This means that our model risks being overweight select industries that are systematically undervalued because of economic factors that make these sectors appear cheap.

Performance

The performance of the various portfolios differed between markets, as presented in the analysis. Developed markets experienced market betas that were on average higher than the benchmarks. The returns were higher as well, likely a result of beta driving the return as suggested by CAPM. The added risk exhibited by more market exposure is also present in the maximum drawdown because portfolios in the U.S. and Europe have had larger losses historically compared to their benchmarks. Based on the efficient market hypothesis, more excess return is a result of a greater risk such that higher risk-adjusted return is not possible. For the S&P 500 portfolios, the SR is above the market suggesting the market is not completely efficient. On the other hand, alpha is rarely significant above 5% which indicates that abnormal return greater than 0 is not a result of the strategy. This, in its turn, implies that markets are quite efficient in the U.S. when benchmarked against our investment model. Strategies within the STOXX Europe index yield slightly higher risk adjusted returns relative to the benchmark and also exhibit more significant alpha. However, it is still difficult to clearly state that markets are not efficient in Europe since much of the excess return can be attributed to a higher risk.

The emerging market strategies exhibit very different characteristics compared to their developed counterparts. Alpha is significant in most cases, risk-adjusted return is high, and beta is lower than the benchmark. This clearly conveys the notion that markets are not efficient in emerging markets, because we were able to systematically generate excess return with a lower risk profile. Firms in emerging markets are fundamentally related to their respective economies and the future country development. Many of the firms in the index are smaller compared to the developed markets, which often means less available information and coverage of these firms. We believe this is a central factor in driving returns from the model.

Another important consideration when discussing the returns of firms in emerging markets, is the currency exposure. All the returns are given in USD, so if the dollar appreciates from the time of investment to selling the position, then the position would increase correspondingly. This means that the returns could also have been driven by currency fluctuations. The Bloomberg Dollar Spot Index, which measures the value of the dollar against both developed and emerging market currencies, increased by 14.67% over our investment period. An annualized return of 1.03% has definitely had a positive impact on the returns, however, the benchmark return was subject to the same currency volatility.

Peer selection analysis illustrated that accuracy was relatively unchanged for the different samples. Contrary to this conclusion, we see slightly improved risk-adjusted return and alpha for the 3-year average and 12-peer samples in developed markets. Firms in these indices are already established and have a solid track record, so using the robustness measures we reduce the impact of outliers in the data, whether that may be financial data or peer outliers. The results from the MSCI EM index present the opposite conclusion, as non-average and 6-peer samples perform best. The underlying assumption for developed markets is that firms are established, but this does not necessarily hold in emerging markets as their economies are characterized by high growth and more drastic developments. Therefore, it does not make sense to compare firms using average numbers since firms can be fundamentally different from where they were several years ago. We expected more peers to provide a degree of stability, but the features are diverse across geographies and economies so adding peers resulted in selecting firms that are essentially different.

Lastly, the multi-factor strategies do not decisively correspond to better performance. While some strategies offer the conclusion that performance is improved, we cannot confirm that returns are driven by better peer selection from multi-factor strategies. Though peer selection accuracy improves, the firms that are most undervalued may be more likely to operate within similar industries. If a whole sector is undervalued, then the model could be unconsciously overweight firms from the cheaper valued industry that is underperforming.

10. Conclusion

In this paper we cover a number of topics, from backtesting to peer selection and performance measurement across developed and emerging markets. The main intention of this research was to create a proprietary model for portfolio selection which would outperform the benchmark. In order to answer this fundamental question, we had to investigate how accurate our model was at selecting peers because it was the foundation for buying undervalued stocks.

Firstly, we analyzed the accuracy of peer selection through the absolute percentage errors of the base company relative to the peer group. We find that strategies using multiple factors to identify peer groups tend to have lower valuation errors than the single factor tests. However, there is no clear tendency of multi-factor portfolios outperforming the single factor strategies.

Another aspect of the testing was robustness, where we investigated how 3-year average valuation fundamentals and larger peer groups impacted peer selection and performance. The results of these samples for peer selection are not compelling, because the accuracy is essentially unchanged for all the markets. On the other hand, the S&P 500 and STOXX Europe 600 portfolios exhibit improved performance in the samples that were designed to be robust. In contrast, portfolio performance for MSCI Emerging Markets is superior for samples that use smaller peer groups and non-average financial measures.

Thirdly, the results differed between the three markets tested and in particular developed markets presented contrasting performance compared to emerging markets. MSCI EM clearly illustrated the highest and most significant alpha, as a result of greater annualized returns and lower risk relative to the benchmark. Risk-adjusted returns for the S&P 500 and MSCI EM are quite similar despite EM having higher total returns, as their volatility and risk-free rates are also larger. The U.S. based portfolios have the least significant alpha's on average. This result is related to the efficient market hypothesis because active investing in an efficient market should not generate any abnormal return due to informational efficiency. However, the results of EM portfolios suggest that the market does not reflect all available information as our investment strategy is able to consistently outperform the index.

In conclusion, our investment model manages to consistently construct portfolios that outperform the market, though with varying degrees of confidence. This proposition allows us to confirm the hypothesis, thus recognizing the statement that the investment strategies beat the benchmark in terms of risk and return. As an investor or investment manager, it is important to consider that the conclusions reached were based on historic performance, which does not guarantee future results.

Appendices

Appendix 1

```
capture log close
log using FFlog, replace


*** Preparing the data
*** REMEMBER TO REMOVE DUPLICATE COMPANIES (MULTIPLE SHARE CLASSES) AND DATA THAT IS
#N/A - DO IT IN EXCEL
use "SPX_2003.dta", clear // this command loads the data set and clears out any data which might
have been present
//browse

drop if evsales<0 | evebitda<0 | pcf<0 | pb<0 | evsales_nxt<0 | evebitda_nxt<0 | pcf_nxt<0 |
pb_nxt<0

forvalues m=1/1 {

gen company=_n

drop roe3yr roic3yr wacc3yr ta3yr g3yr



******************************************************************************
*** STEP 1 - FINDING PEER COMPANIES

rename roe var1
rename roic var4
rename wacc var5
rename ta var2
rename g var3



//rank the companies
egen r_var1=rank(-var1)
egen r_var2=rank(-var2)
egen r_var3=rank(-var3)
egen r_var4=rank(-var4)
egen r_var5=rank(-var5)



forvalues j=1/6 {    //CHANGE TO 12 LATER

gen peer`j'=.
```

```
gen SARD`j'=.

}


forvalues i=1/1000 {  //CHANGE NUMBER BASED ON AMOUNT OF COMPANIES IN THE INDEX


//Loop for 1st factor (roe)
gen var1`i'=r_var1 if company==`i'
egen rank_var1`i'=sum(var1`i')
drop var1`i'


//Loop for 2nd factor (ta)
gen var2`i'=r_var2 if company==`i'
egen rank_var2`i'=sum(var2`i')
drop var2`i'


//Loop for 3rd factor (g)
gen var3`i'=r_var3 if company==`i'
egen rank_var3`i'=sum(var3`i')
drop var3`i'


//Loop for 4th factor (roic)
gen var4`i'=r_var4 if company==`i'
egen rank_var4`i'=sum(var4`i')
drop var4`i'


//Loop for 5th factor (wacc)
gen var5`i'=r_var5 if company==`i'
egen rank_var5`i'=sum(var5`i')
drop var5`i'


//taking difference of ranks, then summing in new variable
gen var1_diff=abs(r_var1-rank_var1`i')
gen var2_diff=abs(r_var2-rank_var2`i')
gen var3_diff=abs(r_var3-rank_var3`i')
gen var4_diff=abs(r_var4-rank_var4`i')
gen var5_diff=abs(r_var5-rank_var5`i')


gen rnkd_diff_`i' = var2_diff   //CHANGE var1 TO var2 OR OTHER
//gen rnkd_diff_`i' = var1_diff+var2_diff
//gen rnkd_diff_`i' = var1_diff+var2_diff+var3_diff
//gen rnkd_diff_`i' = var1_diff+var2_diff+var3_diff+var4_diff
```

```
//gen rnkd_diff_`i' = var1_diff+var2_diff+var3_diff+var4_diff+var5_diff

drop var1_diff var2_diff var3_diff var4_diff var5_diff

drop rank_var1`i' rank_var2`i' rank_var3`i' rank_var4`i' rank_var5`i'



egen rnkd_peer`i'=rank(rnkd_diff_`i')

replace rnkd_peer`i'=. if rnkd_peer`i'>7 //REMEMBER TO CHANGE TO 13 (I.E. ADD 1 TO THE NUMBER OF PEERS)

gen peercomps`i'= company if rnkd_peer`i' !=. & rnkd_peer`i' !=1
gen peerscore`i'=rnkd_diff_`i' if rnkd_peer`i' !=. & rnkd_peer`i' !=1
egen newvar`i'=rank(peercomps`i') if peerscore`i' !=.



forvalues j=1/6 {   //CHANGE TO 12 LATER

gen company`j'=.
gen score`j'=.

replace company`j'=peercomps`i' if newvar`i' == `j'
egen company_`j'=sum(company`j')
replace score`j'=peerscore`i' if newvar`i' == `j'
egen score_`j'=sum(score`j')

replace peer`j'=company_`j' if company==`i'
replace SARD`j'=score_`j' if company==`i'

drop company`j' score`j' company_`j' score_`j'
}

//Dropping excess variables
drop rnkd_diff_`i' rnkd_peer`i' peercomps`i' peerscore`i' newvar`i'

}

drop name var1 var2 var3 var4 var5 r_var1 r_var2 r_var3 r_var4 r_var5



*******************************************************************************
*** STEP 2 - FINDING WHICH COMPANIES TO BUY
```

```
rename evsales mult1
rename evebitda mult2
rename pcf mult3
rename pb mult4

rename evsales_nxt nxtmult1
rename evebitda_nxt nxtmult2
rename pcf_nxt nxtmult3
rename pb_nxt nxtmult4

gen avg_mult_1 =. // EV/SALES
gen avg_mult_2 =. // EV/EBITDA
gen avg_mult_3 =. // P/CF
gen avg_mult_4 =. // P/B

gen nxtavg_mult_1 =. // EV/SALES
gen nxtavg_mult_2 =. // EV/EBITDA
gen nxtavg_mult_3 =. // P/CF
gen nxtavg_mult_4 =. // P/B

forvalues i=1/1000 {   //CHANGE NUMBER BASED ON AMOUNT OF COMPANIES IN THE INDEX

forvalues j=1/6 {   //CHANGE TO 12 LATER

*** MULTIPLE 1
gen peers`j'=peer`j' if company == `i'
egen peers_`j'=sum(peers`j')
drop peers`j'

gen mult_1_`j' = mult1 if company == peers_`j'
egen mult1_`j' = sum(mult_1_`j')

gen nxtmult_1_`j' = nxtmult1 if company == peers_`j'
egen nxtmult1_`j' = sum(nxtmult_1_`j')

drop peers_`j'
drop mult_1_`j'
drop nxtmult_1_`j'


*** MULTIPLE 2
gen peers`j'=peer`j' if company == `i'
egen peers_`j'=sum(peers`j')
drop peers`j'

gen mult_2_`j' = mult2 if company == peers_`j'
egen mult2_`j' = sum(mult_2_`j')

gen nxtmult_2_`j' = nxtmult2 if company == peers_`j'
egen nxtmult2_`j' = sum(nxtmult_2_`j')
```

```
drop peers_`j'
drop mult_2_`j'
drop nxtmult_2_`j'

*** MULTIPLE 3
gen peers`j'=peer`j' if company == `i'
egen peers_`j'=sum(peers`j')
drop peers`j'

gen mult_3_`j' = mult3 if company == peers_`j'
egen mult3_`j' = sum(mult_3_`j')

gen nxtmult_3_`j' = nxtmult3 if company == peers_`j'
egen nxtmult3_`j' = sum(nxtmult_3_`j')

drop peers_`j'
drop mult_3_`j'
drop nxtmult_3_`j'

*** MULTIPLE 4
gen peers`j'=peer`j' if company == `i'
egen peers_`j'=sum(peers`j')
drop peers`j'

gen mult_4_`j' = mult4 if company == peers_`j'
egen mult4_`j' = sum(mult_4_`j')

gen nxtmult_4_`j' = nxtmult4 if company == peers_`j'
egen nxtmult4_`j' = sum(nxtmult_4_`j')

drop peers_`j'
drop mult_4_`j'
drop nxtmult_4_`j'

}



//****CHANGE NUMBER OF MULTIPLES TO NUMBER OF PEERS

gen avg1 = (mult1_1 + mult1_2 + mult1_3 + mult1_4 + mult1_5 + mult1_6)/6
gen avg2 = (mult2_1 + mult2_2 + mult2_3 + mult2_4 + mult2_5 + mult2_6)/6
gen avg3 = (mult3_1 + mult3_2 + mult3_3 + mult3_4 + mult3_5 + mult3_6)/6
gen avg4 = (mult4_1 + mult4_2 + mult4_3 + mult4_4 + mult4_5 + mult4_6)/6

/*
gen avg1 = (mult1_1 + mult1_2 + mult1_3 + mult1_4 + mult1_5 + mult1_6 + mult1_7 + mult1_8 +
mult1_9 + mult1_10 + mult1_11 + mult1_12)/12
```

```
gen avg2 = (mult2_1 + mult2_2 + mult2_3 + mult2_4 + mult2_5 + mult2_6 + mult2_7 + mult2_8 +
mult2_9 + mult2_10 + mult2_11 + mult2_12)/12
gen avg3 = (mult3_1 + mult3_2 + mult3_3 + mult3_4 + mult3_5 + mult3_6 + mult3_7 + mult3_8 +
mult3_9 + mult3_10 + mult3_11 + mult3_12)/12
gen avg4 = (mult4_1 + mult4_2 + mult4_3 + mult4_4 + mult4_5 + mult4_6 + mult4_7 + mult4_8 +
mult4_9 + mult4_10 + mult4_11 + mult4_12)/12
*/

replace avg_mult_1 = avg1 if company==`i'
replace avg_mult_2 = avg2 if company==`i'
replace avg_mult_3 = avg3 if company==`i'
replace avg_mult_4 = avg4 if company==`i'


drop mult1_1  mult1_2  mult1_3  mult1_4  mult1_5  mult1_6
drop mult2_1  mult2_2  mult2_3  mult2_4  mult2_5  mult2_6
drop mult3_1  mult3_2  mult3_3  mult3_4  mult3_5  mult3_6
drop mult4_1  mult4_2  mult4_3  mult4_4  mult4_5  mult4_6

/*
drop mult1_1  mult1_2  mult1_3  mult1_4  mult1_5  mult1_6  mult1_7  mult1_8  mult1_9  mult1_10
mult1_11  mult1_12
drop mult2_1  mult2_2  mult2_3  mult2_4  mult2_5  mult2_6  mult2_7  mult2_8  mult2_9  mult2_10
mult2_11  mult2_12
drop mult3_1  mult3_2  mult3_3  mult3_4  mult3_5  mult3_6  mult3_7  mult3_8  mult3_9  mult3_10
mult3_11  mult3_12
drop mult4_1  mult4_2  mult4_3  mult4_4  mult4_5  mult4_6  mult4_7  mult4_8  mult4_9  mult4_10
mult4_11  mult4_12
*/

drop avg1 avg2 avg3 avg4


//FOR NXT MULT
//****CHANGE NUMBER OF MULTIPLES TO NUMBER OF PEERS

gen nxtavg1 = (nxtmult1_1 + nxtmult1_2 + nxtmult1_3 + nxtmult1_4 + nxtmult1_5 + nxtmult1_6)/6
gen nxtavg2 = (nxtmult2_1 + nxtmult2_2 + nxtmult2_3 + nxtmult2_4 + nxtmult2_5 + nxtmult2_6)/6
gen nxtavg3 = (nxtmult3_1 + nxtmult3_2 + nxtmult3_3 + nxtmult3_4 + nxtmult3_5 + nxtmult3_6)/6
gen nxtavg4 = (nxtmult4_1 + nxtmult4_2 + nxtmult4_3 + nxtmult4_4 + nxtmult4_5 + nxtmult4_6)/6

/*
gen avg1 = (mult1_1 + mult1_2 + mult1_3 + mult1_4 + mult1_5 + mult1_6 + mult1_7 + mult1_8 +
mult1_9 + mult1_10 + mult1_11 + mult1_12)/12
gen avg2 = (mult2_1 + mult2_2 + mult2_3 + mult2_4 + mult2_5 + mult2_6 + mult2_7 + mult2_8 +
mult2_9 + mult2_10 + mult2_11 + mult2_12)/12
gen avg3 = (mult3_1 + mult3_2 + mult3_3 + mult3_4 + mult3_5 + mult3_6 + mult3_7 + mult3_8 +
mult3_9 + mult3_10 + mult3_11 + mult3_12)/12
gen avg4 = (mult4_1 + mult4_2 + mult4_3 + mult4_4 + mult4_5 + mult4_6 + mult4_7 + mult4_8 +
mult4_9 + mult4_10 + mult4_11 + mult4_12)/12
```

```
*/

replace nxtavg_mult_1 = nxtavg1 if company==`i'
replace nxtavg_mult_2 = nxtavg2 if company==`i'
replace nxtavg_mult_3 = nxtavg3 if company==`i'
replace nxtavg_mult_4 = nxtavg4 if company==`i'


drop nxtmult1_1  nxtmult1_2  nxtmult1_3  nxtmult1_4  nxtmult1_5  nxtmult1_6
drop nxtmult2_1  nxtmult2_2  nxtmult2_3  nxtmult2_4  nxtmult2_5  nxtmult2_6
drop nxtmult3_1  nxtmult3_2  nxtmult3_3  nxtmult3_4  nxtmult3_5  nxtmult3_6
drop nxtmult4_1  nxtmult4_2  nxtmult4_3  nxtmult4_4  nxtmult4_5  nxtmult4_6

/*
drop mult1_1  mult1_2  mult1_3  mult1_4  mult1_5  mult1_6  mult1_7  mult1_8  mult1_9  mult1_10
mult1_11  mult1_12
drop mult2_1  mult2_2  mult2_3  mult2_4  mult2_5  mult2_6  mult2_7  mult2_8  mult2_9  mult2_10
mult2_11  mult2_12
drop mult3_1  mult3_2  mult3_3  mult3_4  mult3_5  mult3_6  mult3_7  mult3_8  mult3_9  mult3_10
mult3_11  mult3_12
drop mult4_1  mult4_2  mult4_3  mult4_4  mult4_5  mult4_6  mult4_7  mult4_8  mult4_9  mult4_10
mult4_11  mult4_12
*/

drop nxtavg1 nxtavg2 nxtavg3 nxtavg4

}

//Finds % of under or overvalued CURRENT YEAR
gen diff_mult1 = (mult1/avg_mult_1)-1
gen diff_mult2 = (mult2/avg_mult_2)-1
gen diff_mult3 = (mult3/avg_mult_3)-1
gen diff_mult4 = (mult4/avg_mult_4)-1

//Ranking the % of under or overvalued
egen r_diff_mult1=rank(diff_mult1)
egen r_diff_mult2=rank(diff_mult2)
egen r_diff_mult3=rank(diff_mult3)
egen r_diff_mult4=rank(diff_mult4)



//Finds % of under or overvalued CURRENT YEAR for full sample
gen sample_mult1 = abs((avg_mult_1/mult1)-1)
gen sample_mult2 = abs((avg_mult_2/mult2)-1)
gen sample_mult3 = abs((avg_mult_3/mult3)-1)
gen sample_mult4 = abs((avg_mult_4/mult4)-1)

egen sErr_avg_mult1 = mean(sample_mult1)
egen sErr_avg_mult2 = mean(sample_mult2)
```

```
egen sErr_avg_mult3 = mean(sample_mult3)
egen sErr_avg_mult4 = mean(sample_mult4)

egen sErr_med_mult1 = median(sample_mult1)
egen sErr_med_mult2 = median(sample_mult2)
egen sErr_med_mult3 = median(sample_mult3)
egen sErr_med_mult4 = median(sample_mult4)

//gen total_rank = r_diff_mult1 + r_diff_mult2 + r_diff_mult3 + r_diff_mult4   //CHANGE: INCLUDE THIS WHEN USING SEVERAL MULTIPLES AT THE SAME TIME

//Finds the companies to buy
//egen rank_of_tot=rank(total_rank)   //CHANGE: INCLUDE THIS WHEN USING SEVERAL MULTIPLES AT THE SAME TIME
//gen buy_these=1 if rank_of_tot <= 20   //CHANGE: INCLUDE THIS WHEN USING SEVERAL MULTIPLES AT THE SAME TIME

gen buy_these=1 if r_diff_mult`m' <= 20   //CHANGE mult1 TO mult2 OR OTHER WHEN DEALING WITH ONE MULTIPLE ONLY

//Finds the companies to short
//egen inv_rank_of_tot=rank(-rank_of_tot)   //CHANGE: INCLUDE THIS WHEN USING SEVERAL MULTIPLES AT THE SAME TIME

//egen inv_rank_of_tot=rank(-r_diff_mult1)   //CHANGE mult1 TO mult2 OR OTHER WHEN DEALING WITH ONE MULTIPLE ONLY
//gen short_these=1 if inv_rank_of_tot <= 20

//Keeps the companies that we will buy or short
drop if buy_these ==. //& short_these ==.

***for Wilcoxon test
//Absolute error current year
gen err_mult1 = abs((mult1/avg_mult_1)-1)
gen err_mult2 = abs((mult2/avg_mult_2)-1)
gen err_mult3 = abs((mult3/avg_mult_3)-1)
gen err_mult4 = abs((mult4/avg_mult_4)-1)

//Absolute error NEXT YEAR
gen nxterr_mult1 = abs((nxtmult1/nxtavg_mult_1)-1)
gen nxterr_mult2 = abs((nxtmult2/nxtavg_mult_2)-1)
gen nxterr_mult3 = abs((nxtmult3/nxtavg_mult_3)-1)
gen nxterr_mult4 = abs((nxtmult4/nxtavg_mult_4)-1)

//Find mean and median of our 20 sample of absolute % errors
egen s20erravg_1 = mean(err_mult1)
egen s20erravg_2 = mean(err_mult2)
egen s20erravg_3 = mean(err_mult3)
egen s20erravg_4 = mean(err_mult4)
```

```
egen s20errmed_1 = median(err_mult1)
egen s20errmed_2 = median(err_mult2)
egen s20errmed_3 = median(err_mult3)
egen s20errmed_4 = median(err_mult4)

egen nxt20erravg_1 = mean(nxterr_mult1)
egen nxt20erravg_2 = mean(nxterr_mult2)
egen nxt20erravg_3 = mean(nxterr_mult3)
egen nxt20erravg_4 = mean(nxterr_mult4)

egen nxt20errmed_1 = median(nxterr_mult1)
egen nxt20errmed_2 = median(nxterr_mult2)
egen nxt20errmed_3 = median(nxterr_mult3)
egen nxt20errmed_4 = median(nxterr_mult4)

//Wilcoxon Test
gen bef_aft1 = err_mult1-nxterr_mult1
gen bef_aft2 = err_mult2-nxterr_mult2
gen bef_aft3 = err_mult3-nxterr_mult3
gen bef_aft4 = err_mult4-nxterr_mult4

gen absbef_aft1 = abs(bef_aft1)
gen absbef_aft2 = abs(bef_aft2)
gen absbef_aft3 = abs(bef_aft3)
gen absbef_aft4 = abs(bef_aft4)

egen rWil_1 = rank(absbef_aft1)
egen rWil_2 = rank(absbef_aft2)
egen rWil_3 = rank(absbef_aft3)
egen rWil_4 = rank(absbef_aft4)



keep ticker mult`m' nxtmult`m' SARD1 SARD2 SARD3 SARD4 SARD5 SARD6 avg_mult_`m'
nxtavg_mult_`m' err_mult`m' nxterr_mult`m' bef_aft`m' rWil_`m' sErr_avg_mult`m'
sErr_med_mult`m' s20erravg_`m' s20errmed_`m' nxt20erravg_`m' nxt20errmed_`m'

}
```

Appendix 2

	Weight	Default Spread
China	34.1%	0.79%
South Korea	15.4%	0.56%
Taiwan	12.1%	0.68%
India	8.8%	2.15%
Brazil	8.8%	3.39%
South Africa	6.6%	2.48%
Russia	4.4%	2.82%
Mexico	3.3%	1.35%
Thailand	2.2%	1.80%
Malaysia	2.2%	1.35%
Indonesia	2.2%	2.15%
Total	**100.0%**	**1.37%**

References

Ahmed, P., & Nanda, S. (2005). Performance of Enhanced Index and Quantitative Equity Funds. *The Financial Review,40*(4), 459-479. doi:10.1111/j.1540-6288.2005.00119.x

Becker, Y. L., & Reinganum, M. R. (2018). *The Current State of Quantitative Equity Investing. CFA Institute Research Foundation*.

Bryman, A. (2012). *Social research methods (4th ed.).* Oxford University Press.

Chan, E. P. (2013). *Algorithmic trading: Winning strategies and their rationale*. Hoboken, NJ: John Wiley et Sons.

Damodaran, A. (2001). *The Dark Side of Valuation: Valuing Old Tech, New Tech and New Economy Companies* (1st ed.). Upper Saddle River, NJ, USA: Financial Times Prentice Hall.

Damodaran, A. (2006). *Damodaran on Valuation: Security Analysis for Investment and Corporate Finance* (2nd ed.). New Jersey: John Wiley & Sons.

Damodaran, A. (2011). *Little book of valuation: How to value a company, pick a stock, and profit*. Hoboken, NJ: Wiley.

Damodaran, A. (2019, January). *Country Default Spreads and Risk Premiums.* Retrieved from http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/ctryprem.html

Dittmann, I., & C. Weiner. (2005). *Selecting Comparables for the Valuation of European Firms.* SFB 649 Discussion Paper 2005-002

Fama, E. F., & French, K. R. (1993). *Common risk factors in the returns on stocks and bonds.* Journal of Financial Economics,33(1), 3-56. doi:10.1016/0304-405x(93)90023-5

Frazzini, A., & Pedersen, L. H. (2014). *Betting against beta.* Journal of Financial Economics,111(1), 1-25. doi:10.1016/j.jfineco.2013.10.005

Hooke, J. C. (2010). *Security analysis and business valuation on Wall Street: A comprehensive guide to today's valuation methods* (2nd ed.). Hoboken, NJ: Wiley.

Israel, R., Jiang, S., & Ross, A. (2017). *Craftsmanship Alpha: An Application to Style Investing.* The Journal of Portfolio Management. doi:10.3905/jpm.2017.2017.1.075

Knudsen, J. O., Kold, S., & Plenborg, T. (2017). *Stick to the Fundamentals and Discover Your Peers.* Financial Analysts Journal,73(3), 85-105. doi:10.2469/faj.v73.n3.5

LaMorte, W. (2017, May 4). *Nonparametric Tests: Wilcoxon Signed Rank Test.* Retrieved from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Nonparametric/BS704_Nonparametric6.html

Lee, C.M.C., Ma, P. and Wang, C.C.Y. (2014). *Search-Based Peer Firms: Aggregating Investor Perceptions through Internet Co-Searches.* Journal of Financial Economics.

Lie, E., & Lie, H. J. (2002). *Multiples Used to Estimate Corporate Value.* Financial Analysts Journal,58(2), 44-54. doi:10.2469/faj.v58.n2.2522

Massa, A. (2017, March 9). *Payment for Order Flow*. Retrieved from https://www.bloomberg.com/quicktake/payment-for-order-flow

MSCI Emerging Markets Index. (n.d.). Retrieved from https://www.msci.com/emerging-markets

Munk, C. (2017). *Financial Markets and Investments* (Copenhagen Business School). Retrieved from https://sites.google.com/view/clausmunk/home

Nel, S., Bruwer, W., & Roux, N. L. (2014). *An emerging market perspective on peer group selection based on valuation fundamentals.* Applied Financial Economics,24(9), 621-637. doi:10.1080/09603107.2014.894629

Oyedele, A. (2016). *Here's the strategy that has 'worked really well' for Goldman's quant investing group* | Markets Insider. Retrieved from https://markets.businessinsider.com/news/stocks/heres-the-strategy-that-has-worked-really-well-for-goldmans-quant-investing-group-1001494609

Pandey, A. (2012). *Cash Flow is King in Relative Valuation: A Myth or Reality.* Asia-Pacific Journal of Management Research and Innovation,8(4), 441-449. doi:10.1177/2319510x13481903

Pedersen, L. H. (2015). *Efficiently inefficient: How smart money invests and market prices are determined*. Princeton: Princeton University Press.

Petersen, C. V., Plenborg, T., & Kinserdal, F. (2017). *Financial statement analysis: Valuation - Credit analysis - Performance evaluation*. Bergen: Fagbokforlaget.

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2016). *Research methods for business students* (7th ed.). New York: Pearson.

Sraders, A. (2019). *How Does Robinhood Make Money?* Retrieved from https://www.thestreet.com/investing/how-does-robinhood-make-money-14856528

Statman, M. (2004). *The Diversification Puzzle.* Financial Analysts Journal,60(4), 44-53. doi:10.2469/faj.v60.n4.2636

STOXX® Europe 600. (n.d.). Retrieved February 24, 2019, from https://www.stoxx.com/index-details?symbol=SXXP

S&P Dow Jones Indices, S&P 500. (n.d.). Retrieved from https://us.spindices.com/indices/equity/sp-500

Veal, A. J. (2011). *Research methods for leisure and tourism* (4th ed.). Harlow: Pearson Education Limited.

Viebig, J., Poddig, T., & Varmaz, A. (2008). *Equity valuation: Models from leading investment banks*. Chichester: John Wiley.

Wigglesworth, R. (2018). *Quant hedge funds set to surpass $1tn management mark.* Retrieved from https://www.ft.com/content/ff7528bc-ec16-11e7-8713-513b1d7ca85a

Wall, E. (2018). *The difference between an Emerging and Frontier Market.* Retrieved from https://www.morningstar.in/posts/47141/difference-emerging-frontier-market.aspx

Zuckerman, G., & Hope, B. (2017). *The Quants Run Wall Street Now.* Retrieved from https://www.wsj.com/articles/the-quants-run-wall-street-now-1495389108