Network for Studies on Pensions, Aging and Retirement

Netspar THESES

*Tom Kamps*

# Predicting Value and Size Returns

## Can Short-Term Value and Size Premiums Be Predicted?

MSc Thesis 2014-058

# PREDICTING VALUE AND SIZE RETURNS

CAN SHORT-TERM VALUE AND SIZE PREMIUMS BE PREDICTED?

T.C. KAMPS
ANR: 132182

JULY 31ST, 2014

MASTER THESIS
**PUBLIC VERSION**

SCHOOL OF ECONOMICS AND MANAGEMENT
DEPARTMENT OF FINANCE

SUPERVISOR TILBURG UNIVERSITY:
PROF. DR. J.J.A.G. DRIESSEN

SUPERVISOR KEMPEN&CO:
DR. R.M. SALOMONS

## MANAGEMENT SUMMARY


The economic and societal goal of asset management is to allocate assets most efficiently, which implies: providing capital to entities that are able to use it as a resource to create highest economic value added, while managing the volatility of these returns. A widely used way to do this is via factor investing: allocation based upon other measures than price (or market capitalization); often fundamentals. Factor investing provides an opportunity to reap more return, without increasing the volatility of a portfolio.

In order to capture value and size returns efficiently, timing is important and to that extent, predictors of factor returns are valuable. This thesis describes possible predictors of these returns; provides their economic rationales; tests whether they do predict factor returns; and concludes with recommendations to asset managers.

This research project can be seen as a scientific empirical test of indicators that Kempen Capital Management (KCM) already uses to invest in the value and size factor. Furthermore, it is an attempt to increase the predictability of the value and size returns using macro-economic, market, and behavioral factors KCM did not use yet. Figure 1 shows the framework of this thesis with the paths used to translate observations in the domains of macro economy, the financial market, and investor behavior to predicting the returns of SMB and HML portfolios.

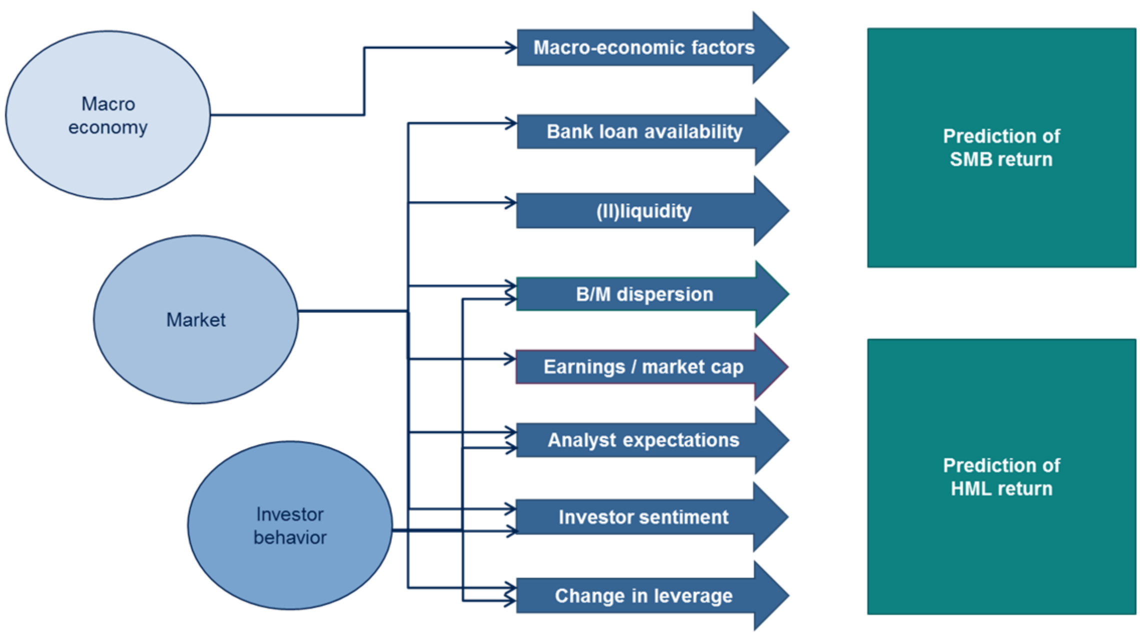



**Figure 1: Framework of thesis; prediction of SMB and HML returns based upon macro, market and behavioral measures.**

Using a wide range of datasets to either obtain predicting variables as well as performing a cross-validation with multiple markets, the eight factors shown in Figure 1 are tested. This is done using linear regressions. Various robustness checks are performed to validate results.

The regression tests and validation show (i) dispersion, (ii) absolute investor sentiment, and (iii) analyst surprise to be significant predictors for HML return. SMB return is predicted by (i) dispersion, (ii) bank loan availability, and (iii) liquidity of small stocks. However, as European and U.S. data do not always show consistent results, analyst surprise and bank loan availability are not used in prediction. Liquidity of small stocks only shows strong results for portfolios that have strong weights in small caps. Absolute investor sentiment was not an ex-ante hypothesis; using this as predictor would bring a selection bias. The linear regression model based upon these estimates is for respectively HML and SMB 1-quarter future return:

$$HMLreturn_{t+1} = 0.06 * BMDispHML - 0.05 + \varepsilon$$
$$SMBreturn_{t+1} = 0.16 * BMDispSMB - 0.02 + \varepsilon$$

These models show $R^2$ values of respectively 6% and 25%, meaning that in the dataset used, that percentage of variance in the portfolio return is being predicted by these models.

Using this model on the historical data gives excess returns of respectively 1.3% and 0.7% per year over respectively regular long HML and long SMB returns. Since these results are validated with the same data, the external validity of the model presented is limited; Figure 2 below shows an example of the difference in cumulative return.

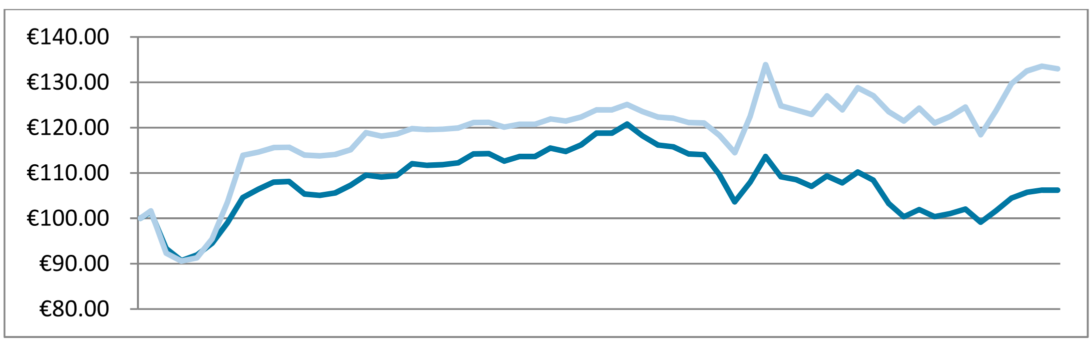



**Figure 2: Indication of difference in portfolio return over a 15 year time-span, with the upper line using dispersion as predictor and the lower line being a regular buy and hold portfolio.**

The following remarks are the key messages of this thesis to asset managers:

- Book-to-market dispersion shows predictive power for HML and SMB returns;
- Liquidity predicts returns of very small firms;
- The recovery phase in the economic cycle shows higher returns for SMB;
- A prediction horizon of two years seems optimal, but depends on the strategy used;
- Value returns seem to be driven by investor behavior;
- HML and SMB portfolios are volatile, also when predictors are used.

Although the methodology has been constructed to provide most robust and valid results, and validation has been done in many ways, it does not provide guarantees about future returns or volatility.

# PREFACE

This report is the result of the research to complete my master Finance at Tilburg University, a master thesis about a discovery of predictors of returns of specific portfolios, the so called value and size portfolios. This research project was a great opportunity to get to know the asset management industry, a sector that works on highest level of a fundamental economic problem: allocation.

As a student with a background in industrial engineering, I have not often looked beyond individual companies or supply chains. The internship at Kempen Capital Management not only offered a view on investing, but even more on macro-economics and the implications on strategic investment decisions. Starting with very limited knowledge on this topic, this internship was a great opportunity to broaden my knowledge in this field.

Although I did not need the KCM environment on the analyses I performed, the insights on the actual investment process provided me the tools to translate the theoretical findings to practical and implementable recommendations. This gave me the possibility to write something relevant for both academics and practitioners.

I am very thankful to the allocation team, Roelof, Ivo, Florian, Luuk and Ruth, for introducing me to the broad field of knowledge they cover, for involving me in the challenges they are facing and for feedback and support they provided me with. Next to the asset allocation team, I want to thank the other employees of KCM who helped me, including but not limited to the trading desk (Philip in particular) for helping on getting good estimates of transaction costs, and the members of value and size project team, as their reports provided me a good starting point.

I also need to thank the Statalist and several other online available resources to support Stata users, which was of much value in the merging of datasets and the data analysis. Although I still have the opinion that Stata uses a rather inconsistent language; provides unclear error messages; and is an instable software program, it is one of the few software packages that provides the functionality to perform the required data analysis.

Last but not least, I want to thank the core fixed income desk. Although they had not much to do with the content of my project, I made grateful use of the empty desk at their team. I would like to thank them for their kind hospitality, the insights in the fixed income investing and pleasant days I worked there.

Tom Kamps

(In the train headed to) Amsterdam, July 2014

*The author declares that the text and work presented in this thesis is original and that no sources other than those mentioned in the text and its references have been used in creating this thesis.*

## TABLE OF CONTENTS

# 1 INTRODUCTION

The economic and societal goal of asset management is to allocate assets most efficiently, which implies: providing capital to entities that are able to use it as a resource to create highest economic value added, while managing the volatility of these returns. The information required to manage assets efficiently is the expected asset return and risk, yet it is hard, if possible, to predict these accurately.

In 1964, the capital asset pricing model (CAPM) was introduced; a theoretical framework that predicts asset returns based on their correlation with market risk (Sharpe, 1964). The most important concept behind this model is that investors are able to diversify their assets to minimize idiosyncratic risk (firm-specific risk) and only be exposed to market risk (which influences all firms in the market), therefore they should only be compensated for the exposure to market risk. Although this model is often used both in theory and practice, Fama and French showed that it does not explain stock returns well (1992). They demonstrated though that value and size (Carhart, 1997) have a (as well statistical as economical) significant explanatory power in the cross-sectional prediction of stock returns. This means that small firms (the 50% smallest market capitalization) yield higher returns than one would expect based on their correlation with the market risk; the same holds for stocks that have a low price relative to their fundamentals (the lowest 30% P/B ratios perform better than the highest 30%).

Although these studies provide evidence that small stocks and value stocks on average pay off in a long period, they do not always in shorter periods. The outperformance of value by growth stocks in the 90's is only one of the many examples that proves that value does not pay off in any interval. Finding indicators of short-term returns on small stocks and value stocks is beneficial for investors. It is, however, even more important to find out whether this investment strategy is sustainable, as it seems to be an anomaly. Little research has been done in investigating the drivers of small and value stock excess returns on a shorter term. In this thesis project, I am going to find indicators to predict these excess returns, focusing on developed economies and the European economy in particular. Current literature has a strong U.S. focus and studies in particular the long term returns; a European study on short term returns has not been performed earlier.

The next chapter will elaborate on the research that has been done on factor investing, either in long-term as well as short return predictions and more theoretical foundations underlying these returns. Chapter 3 states my research questions and hypotheses and in chapter 4 the methodology of testing them. Results will be presented in chapter 5 and chapter 6 will conclude with a reflection on my research as well as recommendations for future research and for asset managers to predict the value and size returns.

## 2 THEORETICAL FRAMEWORK AND LITERATURE REVIEW

In the past 40 years, many scholars have investigated the factor phenomenon. This chapter explains factor investing and gives an overview of the existing factors. Furthermore, it provides argumentation on the factors and countries that are selected in this study and concludes with a literature review on the factors selected: value and size.

### 2.1 FACTOR INVESTING

A wide-known theoretical model that should explain returns of assets is the Capital Asset Pricing Model (CAPM) developed by Sharpe (1964). This model explains returns of assets based upon their correlation with market risk, with as rationale that an investor cannot diversify away market risk. When for example the returns of an asset are highly cyclical, investors should be compensated more for this asset than for an asset that does not correlate with global economy, even if the volatility of the latter is higher. The equation below states the CAPM-formula:

$$E(r_A) = r_f + \beta_A(E(r_m) - r_f)$$

*with $r_f$ representing the risk-free return, $r_m$ the return of the market, $r_A$ the return of asset A, and β the covariance of the return of A with the market over the variance of the market return:* $\beta_A = \frac{Cov(r_i, r_p)}{Var(r_m)}$.

Although the CAPM is widely used in both theory and practice, it appeared a bad predictor of stock returns. Fama and French demonstrated that higher value (book value over market value) and smaller size yielded higher returns than CAPM estimates (1992). They not only demonstrated the explanatory power of value and size, but also rejected the CAPM, in the sense that beta did not have any economically significant explanatory power in their regressions.

This rejection of the CAPM model provided an opportunity for other variables to explain asset returns, so called ‘factors’ or ‘smart beta’ (an umbrella term for investing using other than market capitalization weights) (Financial Times Lexicon, 2014). Next to value and size, momentum was introduced (Carhart, 1997), which predicts high returns based on recent high returns. Later also quality (based on fundamental analysis) and low-vol (low volatility) became popular.

## 2.2 RATIONALES OF FACTOR INVESTING

As several researches demonstrated over the past decades, factor investing can create excess returns. Although factor investing is widely accepted, the theoretical foundation and empirical results differ for each factor. In this paragraph I will elaborate about the findings on the five factors: value, size, momentum, low-vol, and quality. Figure 3 shows the introduction of these factors in scientific literature and on the market as ETFs.

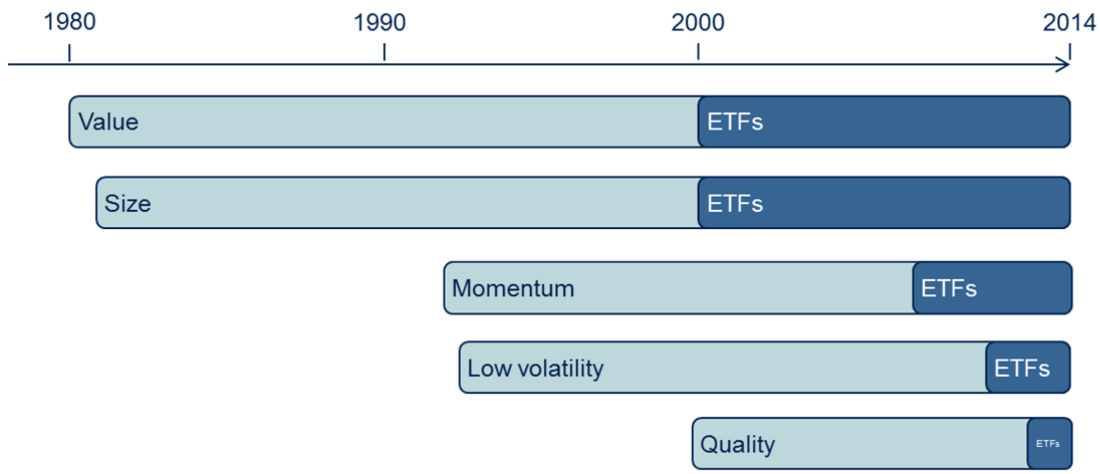



**Figure 3: Scientific introduction of factor and start of ETF trading upon factor, (Fama & French, 1992), (Jegadeesh & Titman, 1993), (Black, et al., 1972), (Dechow, et al., 2001). ETF introduction estimated based upon ETF database data (ETF Database, 2014).**

One of the best known explanations of the value effect is given by Lakonishok, Shleifer, and Vishny, who test whether the value returns (a high minus low portfolio; HML) come from a fundamentally riskier strategy or from the suboptimal behavior of the typical investor (Lakonishok, et al., 1994). They find that the extrapolation of historical returns drives the HML returns; investors tend to overpay for firms that have recently shown high growth and overdiscount weak performance. Barberis and Huang provide an alternative model, also using a behavioral explanation for the value return, but via loss aversion (2001). They find support for loss aversion and narrow framing on individual stock returns as basis for the value premium.

The size effect has been discovered in early 80's by Banz (1981); finding a positive return between small stocks versus large stocks. The effect is generally explained by two arguments. The first is that smaller stocks are less liquid and that investors should be compensated for this illiquidity; an empirical demonstration by Amihud (2002) shows only that it gives a partial explanation. Next to the illiquidity, also higher idiosyncratic risk is given as argument (since as well investors in small firms as the firms themselves might not be able to diversify away risk). Finally, take-out premiums mainly exist at small firms; investors in large firms often do not benefit directly from takeovers.

The momentum factor was introduced by Jegadeesh and Titman (1993), who studied the profitability of a strategy that sold poorly performing stocks and bought well performing stocks, and by Carhart (1997), who investigated the performance of mutual funds. Jegadeesh and Titman find that a portfolio of winners minus losers performs well in the first 12 months, but after this period the effect dissipates. They attribute this anomaly to underreaction by investors. Carhart shows the same effect for mutual funds (instead of individual stocks). The momentum factor is, as well as the value factor, potentially explained by expectations about irrational behavior of investors (Jegadeesh & Titman, 1993).

The low-volatility investment strategy was first documented by Black, Jensen and Scholes, and there are at this moment four rationales that could explain this factor (Black, et al., 1972), (Haugen & Heins, 1975). These are: leverage aversion; investor behavior; analysts' optimism; and the delegated agency model. The leverage aversion means that many investors are over-allocating to high beta stocks, because they are leverage constrained. Investors might be risk-seeking in high volatility stocks (a gambling preference), implying a premium for risky stocks. Analysts are more optimistic (high growth forecasts) about high-volatility stocks, which pushes up prices (Hsu & Li, 2013). The delegated agency model states that portfolio managers tend to under-allocate to low-volatility stocks, attributable to a constraint on leverage (Baker, et al., 2011).

Finally, investing in a quality factor is based upon a fundamental analysis; firms that have a better market position or stronger accounting figures can yield abnormal returns (Dechow, et al., 2001), (Abarbanell & Bushee, 1998). The construction of a quality portfolio can be done in various ways, depending on the quality measures chosen. Abarbanell and Bushee use accounting measures like capital expenditures and gross margins. Piotroski also performs a similar study, trying to enhance the returns of a value portfolio using accounting measures (2000).

## 2.3 ARGUMENTATION ON FACTORS SELECTED

This study focusses on value and size returns, not on the factors momentum, low-volatility, and quality. Main arguments for this decision are the longer history of the value and size factors and the extent to which a strategy on yielding factor returns can be implemented on a strategic level.

Investing in value has a strong foundation; is studied for several decades and is market-tested for several years. Subsequently it can be expected that the anomaly will persist although its discovery is made public.

The same argumentation to choose for value also holds for the size factor. The effect is known for a long period and there exists a theoretical foundation to explain the size effect. This long history increases the likelihood that there exists a persisting size effect in the market. Next to the high likelihood of the existence of the effect, investigating the size effect has implementation benefits to KCM, since it has a good track record in small cap investing. This enhances the degree in which an implementation of a small cap investment strategy can yield the size premiums (Kempen & Co, 2014).

Momentum has already been implemented by KCM (Kempen & Co, 2014), but since transaction costs are rather high, a strategic asset allocation fully focused on the momentum factor might be expensive. Momentum uses technical analysis to determine stock returns for an interval of three to twelve months. When a large amount of capital is allocated via a momentum strategy, transaction costs would be high, in particular when compared with other factor strategies, which have longer holding periods.

Low-volatility is winning popularity recently, but is theoretically less founded than value and size. Next to this, many of the rationales of low-volatility also explain the value premium (the behavioral argument as well as the analyst' optimism argument). This claim is supported by empirical evidence; Li, Sullivan and Garcia-Feijóo show that the yields of low-vol portfolios are lower than what is widely believed and attribute part of its return to a high value part (Li, et al., 2014). The recent gain in attention on low-vol poses the question whether this anomaly will persist; the first low volatility ETF was released in 2011 and low volatility firms are recently in high demand (Fabian, 2013). The value, size, and quality factor have been investigated since 1980's; low-vol, on the contrary, has only got attention recently, it is therefore not tested in the market; the size and value premiums are. Finally, low-vol implicitly assumes that volatility among stocks remains constant, which can be doubted as Figure 4 illustrates: only two third of the low-vol stocks will stay low-vol stocks and even fewer low-vol stocks were a low-vol stock in the past.

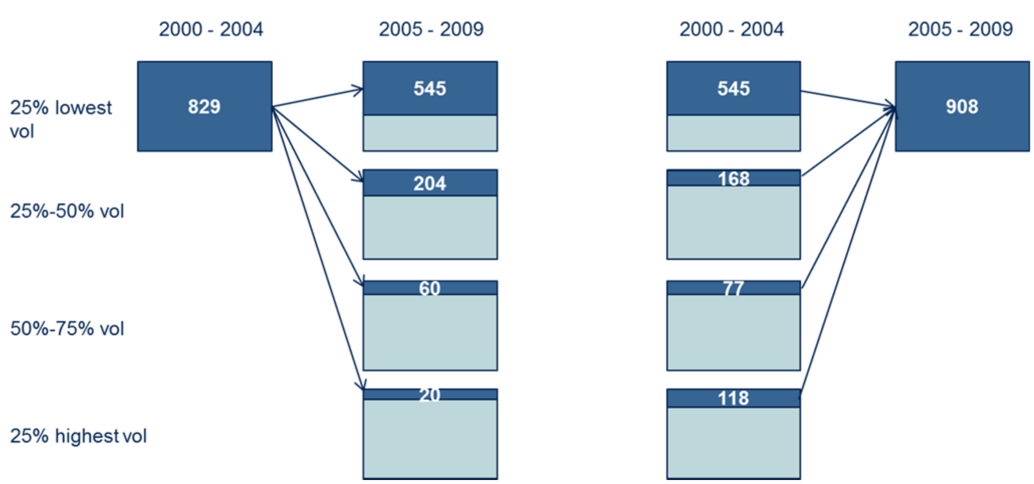



**Figure 4: Quartiles of volatility in U.S. stocks over 2000-2009 and transitions. Data source: CRSP.**

Investing in the quality factor can be done in many different ways. In most occasions, it is based upon accounting figures, and has in that sense some intersection with the value factor (which compares prices to a fundamental). Subsequently both portfolios are fundamental weighted. This intersection, combined with the longer market testing of the value portfolio, makes the prediction of the value factor more relevant to investigate than the quality factor.

Table 1 below summarizes the findings mentioned in this paragraph.

**Table 1: Summary of rationales and arguments on factor selection**

| | Rationales of excess return | Market tested | Investable on strategic level |
|---|---|---|---|
| Value | Extrapolation of historical results, loss aversion by investors | ++ | Good |
| Size | Illiquidity of small stocks, idiosyncratic risk by diversification restrictions of small-cap investors and firms, less informed investors | ++ | Good |
| Momentum | Underreaction and overreaction of investors | + | Fair |
| Low-vol | Leverage aversion, gambling preference of investors, analysts' optimism, delegated agents | +/- | Good |
| Quality | Better market position, stronger financial position of firm | - | Good |

Concluding, the factors value and size are selected because they are investable on a strategic level; have a strong theoretical foundation; and are tested in the market.

## 2.4 ARGUMENTATION ON REGIONS SELECTED

Most cross-sectional regressions on stock returns are performed on U.S. data. This will probably have two main reasons; the first is that many researchers on this topic are based in the U.S. and U.S. data has good availability compared with European market data and even more compared with EM and Japan. In order to both extend the literature and provide recommendations that can be implanted in the investable universe of Kempen & Co, I focus mainly on the European market. In order to perform a robust validation, I use other markets to either validate models constructed on European data (or to construct models that are validated with European data). Because of data availability, this other market will be mainly U.S., but also Asia where possible. The latter might be preferred since there is less correlation between the Asian (emerging market) stocks and the developed market stocks, but has a drawback in a potential lack of data.

## 2.5 LITERATURE ON VALUE RETURNS

As definition of value, I choose to use an HML portfolio. High book value or high earnings versus market value are called "value stocks" in the remainder of this thesis, but "relatively cheap stocks" would have probably been a better name. The returns of this HML portfolio, consisting of U.S. stocks, are provided in Figure 5 below (method of construction is provided in chapter 4).

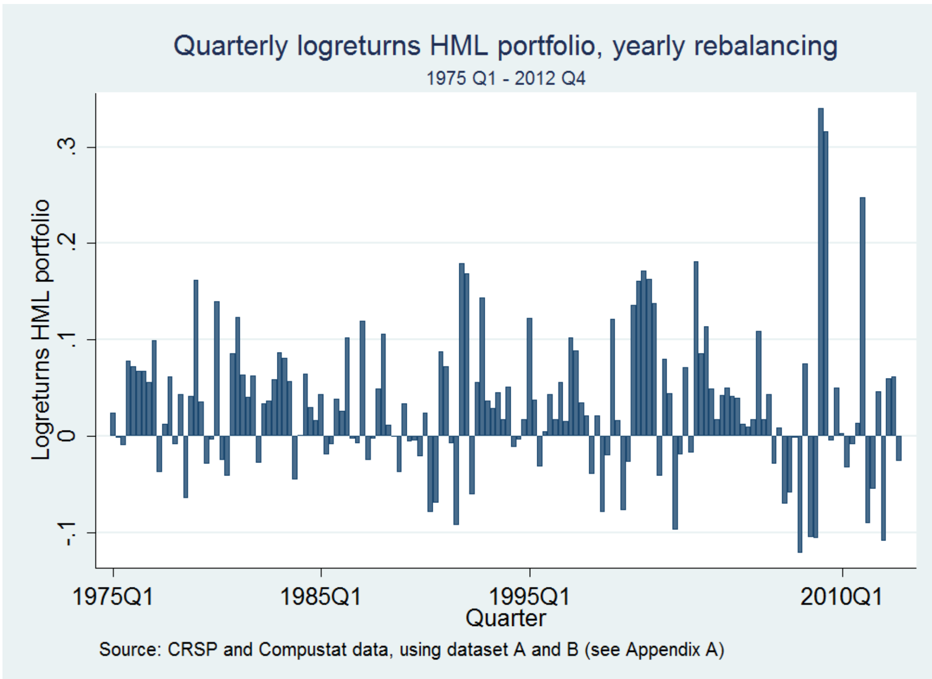



**Figure 5: Value outperformance is relatively stable, with volatility around bubbles.**

Value premiums on long periods are investigated intensively; Basu (1983) shows that NYSE stocks with low P/E ratios (high E/P), thus value stocks, earned higher risk adjusted returns than stocks with high P/E ratios. He used a multivariate statistical model with portfolios of stocks grouped by P/E ratio, corrected for beta. After Basu, several researchers demonstrated the value effect in different markets and periods.

Fama and French confirmed the value effect in U.S. stocks and later in international stocks (1992), (1998), (2012) using a linear regression with value, beta, and size. Using this methodology they could empirically reject the CAPM model. They however study, as well, a long period. Asness et al. find value returns in different markets (U.S., UK, Europe, and Japan) and asset classes (stocks, currencies, bonds, and commodities) (2013). Finally, Malkiel and Jun (2009) demonstrate the value effect in Chinese stocks and use a nonparametric method of portfolio construction. Existence of the value effect is thus demonstrated for many different markets and time periods, reducing the likelihood of statistical fluke.

The explanations of the value effect can be structured by a reward to underlying risk factors (Fama & French, 1992) and behavioral phenomena: extrapolation of historical results; loss aversion (Lakonishok, et al., 1994); and narrow framing (Barberis & Huang, 2001).

Lakonishok, Shleifer, and Vishny test whether the HML returns come from a fundamentally riskier strategy or from the suboptimal behavior of the typical investor (Lakonishok, et al., 1994). They find that the extrapolation of historical returns drives the HML returns; investors tend to overpay for firms that have recently shown high growth and overdiscount weak performance.

Barberis and Huang (2001) give loss aversion and narrow framing as explanation of the cross-sectional return of value stocks. Loss aversion was introduced by Kahneman and Tversky (1979) who argue that people “get utility from gains and losses in wealth, rather than from absolute levels”. Narrow framing means that investors take gains and losses of individual stocks into account when making decisions. They find that the cross section of individual stocks contains a substantial value premium.

Finally, Robert Novy-Marx (2013) finds that gross profits-to-assets can be used to improve value returns; this ratio has about the same predictive power as book to market. The rationale he provides for the effect on this ratio is that it allocates funds to companies that use assets most efficiently.

## 2.6 LITERATURE ON SIZE RETURNS

In 1981, the first notion of the size returns was made in scientific literature by Rolf Banz, who studied stocks on the NYSE using a linear regression model, and found that very small stocks outperform average and large stocks. The difference in returns between average stocks and large stocks is not significant, neither is the effect present in any interval in time; it does however hold for long periods (Banz, 1981). Figure 6 illustrates that the SMB effect is not existent in any time interval by showing the portfolio returns of an SMB portfolio (construction method is provided in chapter 4).

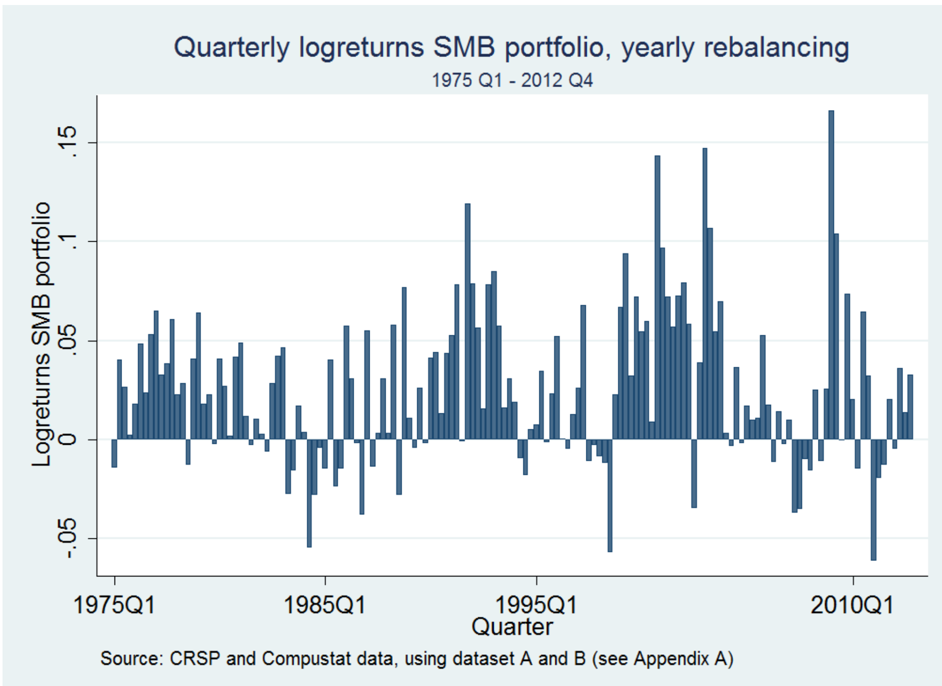



**Figure 6: Size outperformance is cyclical, with more variability in recent years.**

In the decades following the publication of Banz's work, scholars both test his results for other markets and try to find rationales to explain the effect he found. A few years later, Keim and Stambaugh extend Banz's work by looking at a broader universe of stocks and introducing predictors to explain these results (1986). Fama and French test the effect on both U.S. and European market in several papers using cross-sectional regressions of stock returns (1992), (1998), (2004), (2012). They find that the effect holds for both markets, but a large part of the size premium comes from microcaps that are often non-investable.

Van Dijk (2011) provides an overview of empirical studies on the size premium on different markets and periods. These studies show the existence of the size effect in many markets (developed and emerging) and many time periods.

The explanations of the size effect can be structured in four categories: a premium for illiquidity risk (Amihud, 2002) a premium for idiosyncratic risk (Fu, 2009), (Hou & Moskowitz, 2005) and investor behavior (Merton, 1987).

Amihud (2002) finds that illiquidity affects positively ex ante stock returns. He uses the daily ratio of absolute stock return to dollar volume as measure of illiquidity. Since illiquidity affects in particular small stocks, it could be an explanation for the size premium. Goyenko, Holden and Trzcinka (2009) have tested various liquidity measures and find strong measures for liquidity.

Fu (2009) uses an exponential GARCH model to estimate expected idiosyncratic volatilities and expected returns. Given that small cap firms have less diversified owners and that those firms are often concentrated in both a geographical as sector orientation, this could (partially) explain the size premium. In addition, Hou and Moskowitz (2005) show that market frictions partially explain the size effect via idiosyncratic risk, which only is priced for among most delayed firms. They use CRSP data from ‘63 to ’01.

Since small firms are more dependent of external financing than larger firms, they might experience more problems in getting the financing for investments. When bank loan spreads are high, fewer new projects can be started, possibly resulting in lower growth in future earnings. Credit spread would then predict high future SMB returns. Hwang et al. find evidence for this explanation using credit spread as a proxy for the non-diversifiable part of stock risk (2010). In particular the difference in the increase in default premium is a good predictor for SMB returns (the difference between returns on AAA and BBB rated bonds); SMB yields high returns when there is a decrease in default premiums.

Merton explains the size premium by investors that lack information of a part of the firms in the investable universe (1987). Less well-known stocks then could provide higher returns, simply because many (smaller) investors unintentionally exclude them from their portfolios.

Critics on the existence of the size effect attribute the effect found to either statistical fluke (Black, 1993), (MacKinlay, 1995), (Lo & MacKinlay, 1990), (Shumway & Warther, 2002) or the value effect explaining (part of the) size premium (Basu, 1983).

Since researchers explain the size premium ex post, variation in datasets and periods give the possibility of data snooping or a hindsight bias (Black, 1993), (MacKinlay, 1995), (Lo & MacKinlay, 1990). Another selection bias is created when one does not incorporate delisting of stocks. Shumway and Warther (2002) tested the size effect on NASDAQ stocks including the strong negative returns of delisted companies and found that the size effect to be insignificant.

Basu looks mainly to E/P effects as predictor of stock returns and finds that high E/P outperforms low E/P. The small cap excess returns disappear when he controls for the E/P ratio, suggesting that part of the size premium is actually explained by a value premium (Basu, 1983). This evidence is expanded by Berk (1997).

The January effect, stating that most of the SMB returns are obtained in that month, which still is present in SMB returns (Haugen & Jorion, 1996) is left out of the scope of this study, since it does not influence decision making about an SMB portfolio in the scope of a year, neither does it explain the size effect. Furthermore, there are several explanations about SMB returns that remain the same over smaller time intervals, like the competitive environment of the firms (Hou & Robinson, 2006). These factors are therefore excluded from this study, since they will not contribute to a prediction of the size effect

In recent years, Zakamulin (2013) investigated the forecasting of the short term size premium using macro-economic variables, like inflation and T-bill rate. He shows that an active managed portfolio can generate a sustainable, positive alpha that is statistically and economically significant. This result is in line with earlier work of Van Dijk (2011), arguing that the size effect still exists in stock markets and that investors can capture.

Only little has been written on the predictability of short term premiums on value and size. Malkiel and Jun (2009) investigate the value effect for Chinese stocks and find that value portfolio’s outperform growth portfolio’s by comparing portfolio’s constructed using a rank weighting method. The behavior shown in the Figure 6 above seems to indicate a cyclical behavior in small stock excess returns, as is concluded by Malkiel and Jun (2009).

## 3 RESEARCH QUESTIONS AND HYPOTHESES

The goal of this study is to provide an answer to the question:

**Can macro-economic or market factors predict short term future returns of a small minus big (SMB) and value minus growth (HML) portfolio?**

Since this is a broad question, I will formulate several subquestions. These subquestions will be splitted into two parts: size and value. There are correlations between small market capitalization and value (low P/B value) (Fama & French, 1992). This can be explained by the market capitalization being an important part in differentiating stocks between growth and value as well as between small and big. Although there is a (positive) correlation between size and growth, these effects are studied separately. By knowing the drivers of both effects, it is possible to find return indicators. Studying both effects simultaneously would make this discovery harder.

The short term is, in this thesis, defined as a period between 3 and 60 months. For periods shorter than three months, data availability is limited (since I am using fundamentals in portfolio construction). Furthermore, predictability in a period shorter than a quarter is expected to be low and it is unattractive to perform major changes on the strategic allocation on a monthly basis. Five years is chosen as an upper limit; predictors over a longer time period are irrelevant when looking for a tool to actively manage strategic allocation.

The first paragraph of this chapter will discuss value, the second size. Per question will be stated what is tested (hypotheses), why it is tested (rationale), and how it is tested (data sources en short description of method). Because methodology is sometimes overlapping for different research questions, methodology is extensively described in chapter 4. All hypotheses are tested using Newey-West adjusted errors, this choice is explained in paragraph 4.4.

## 3.1 PREDICTING HML RETURNS

Since HML returns are assumed to be driven by investor behavior, my research questions are focused on capturing indicators of that behavior or market data that can influence that behavior. High total market capitalization; high multiple dispersion; high investor sentiment; exceeding analyst expectations; low analyst consensus and a high change in leverage can all indicate a strong optimism of investors and subsequently a high payoff of HML.

### 3.1.1 HIGH MARKET CAPITALIZATION TO FUNDAMENTALS

**Does high P/E of market indicate a short term HML return?**

A high P/E implicates a high growth estimate of the market relative to the discount rate. The actual question is to determine whether the P/E is on a high level. This does not seem to be easy; mainly because earnings can change a lot and especially in recession years these can be very low, leading to high P/E, but not to higher prices; 2009 demonstrated extremely high P/E's, but not extremely high prices. A high P/E can indicate irrational high expectations, which could result in negative returns of growth companies and higher returns of value companies, and thus a positive HML return. The inverse, E/P, would then be negatively related to HML returns.

*Hypothesis 1: The ratio of earnings to prices is negatively related to future HML return.*

Data sources that will be used for market capitalization are CRSP (U.S.) and FactSet (EU); fundamentals are obtained using Compustat (U.S.) and FactSet (EU). As independent variable earnings/price is used, since it reduces the effect of negative and low earnings. Furthermore a risk-free rate is used to include future expectations of an appropriate discount rate.

The detailed methodology is provided in chapter 4; results are given in paragraph 5.1.1.

### 3.1.2 MULTIPLE DISPERSION

**Does P/B dispersion predict high HML returns?**

This question is in line with those of Kalesnik and Beck (2014), Malkiel and Jun (2009) and Arnott (2009). Kalesnik and Beck (2014) use P/B dispersion to predict HML return, stating that dispersion in P/B should indicate future HML return when assumed that the value factor is driven by irrational investor behavior; the potential increase in value is simply larger when the dispersion increases. Assumed that the difference in P/B values tend to get to an equilibrium; a high dispersion among P/B values would indicate a positive HML return.

Malkiel and Jun (2009) and Arnott (2009) used P/E instead of P/B. I will use book value to market value because book value of equity is more stable positive; negative earnings suggest a negative price when a P/E is used. Book value therefore seems a stronger anchor to evaluate whether a price is high or low. Furthermore, earnings can be manipulated more easily by, for example, changing the depreciation method.

P/B dispersion is, given data availability, a practical measure; free cash flow over total enterprise value is a better predictor for future performance (Gray & Vogel, 2012), but this measure has a worse data availability. Therefore I will initially use P/B as measure; this can later be enhanced using free cash flow and total enterprise value data.

*Hypothesis 2: P/B dispersion is positively related to future HML return.*

Data sources that will be used for market capitalization are CRSP (U.S.) and FactSet (EU), book value of equity is obtained using Compustat (U.S.) and FactSet (EU). Portfolios are constructed in line with Fama and French (1992); the methodology of Malkiel and Jun (2009) is hard to implement (since it assigns large weights to very small companies). As independent variable the dispersion between B/M of companies is used (the inverse of price to book).

The detailed methodology is provided in chapter 4; results are given in paragraph 5.1.2.

### 3.1.3 SENTIMENT INDICATOR

**Does investor sentiment predict high HML returns?**

Under the assumption that investor behavior is a main driver of HML return and given that investor sentiment can indicate cross-sectional stock return (Baker & Wurgler, 2003), this measure could have predictive value. If HML returns can be explained by irrational investor behavior, then HML would pay off during periods in which investor behave more irrational. In those periods they would be more loss-averse and are more inclined to extrapolate historical returns, leading to over- and undervalued assets which will give positive HML return when mispriced stocks are corrected. As investor sentiment might be a proxy to measure the irrational behavior of investors, it could predict positive future HML return.

*Hypothesis 3: Investor sentiment is positively related to future HML return.*

Data sources that will be used for an investor sentiment index is the investor sentiment data of Wurgler (2011), using the methodology of Baker and Wurgler (2003). Since the required data for a European replication is unavailable and the time-window of this project does not accommodate a replication for the European market, the series of Wurgler will be used for both tests. As independent variable the investor sentiment index is used.

The detailed methodology is provided in chapter 4; results are given in paragraph 5.1.3.

### 3.1.4 PERFORMANCE RELATIVE TO ANALYST EXPECTATIONS

**Does exceeding the analyst expectations provide a long term positive HML return?**

Exceeding expectations of the market would logically immediately lead to an increase in prices. These companies then could become high P/E; reflecting extrapolation of high earnings in the past by investors. When the investors' expectations finally are not met, a decline in growth stocks will occur and HML will yield a positive return. Bartov, Givoly and Hayn study the effect of exceeding analyst expectations to individual stock returns and find a positive quarterly return (Bartov, et al., 2002).

*Hypothesis 4: The exceeding of analyst expectations is positively related to future HML return.*

The data source that will be used for analyst expectations is I/B/E/S; which contains analyst surprises of both U.S. and international stocks. As independent variable, the difference between the actual and expected value is used, aggregated to portfolio level.

The detailed methodology is provided in chapter 4; results are given in paragraph 5.1.4.

### 3.1.5 ANALYST CONSENSUS

**Does high analyst consensus indicate low future HML returns?**

High P/E stocks are priced high because the market beliefs their earnings will increase in the coming years. The long term HML returns provide evidence that investors suffer overconfidence. However, the short term performance of HML fluctuates; indicating that either investors are not always overconfident or the awareness that they are overconfident has a (non-constant) lag. In line with Diether, Malloy, and Scherbina (2002), it is postulated that analyst dispersion in earnings forecast could work as a proxy for the dispersion of beliefs among investors. Dispersion in beliefs would then indicate overconfidence by (part of the) investors. However, Diether et al. find a counterintuitive result; in their cross-sectional regressions high analyst dispersion indicates low future HML returns (instead of high) (Diether, et al., 2002).

*Hypothesis 5: Analyst dispersion is positively related to future HML return.*

The data source that will be used for analyst consensus is I/B/E/S; which contains analyst recommendations of both U.S. and international stocks. As independent variable analyst consensus is used based upon buy, hold, and sell recommendations.

The detailed methodology is provided in chapter 4; results are given in paragraph 5.1.5.

### 3.1.6 CHANGE IN LEVERAGE

**Does a change in the leverage of firms lead to higher HML returns?**

Leverage itself is one of the Fama & French factors, and is found to not predict stock returns (Fama & French, 1992). Piotroski (2000) finds the same result, studying on the value of accounting fundamentals to predict stock returns. However, the change in leverage influences the relative size (and volatility) of the residual equity claim; it leads to more risk, higher required returns and thus higher expectations of future returns. These expectations of higher returns could lead to an overvaluation of stocks; which could result in positive future HML returns. Likewise should the difference in change in leverage between the both parts of the portfolio indicate positive HML return.

*Hypothesis 6: The change in leverage of (growth) companies is positively related to future HML return.*

The data sources that will be used for change in leverage are Compustat (U.S.) and FactSet (EU). Both databases contain fundamental information of firms. As independent variable the change in leverage is used (computed per company, aggregated to portfolio level).

The detailed methodology is provided in chapter 4; results are given in paragraph 5.1.6.

## 3.2 PREDICTING SMB RETURNS

An often heard explanation of the size premium is the compensations for additional risk; small firms seem to be more exposed to the economic cycle and are less liquid than large firms. Furthermore, several researchers suggested that investors in small firms are rewarded for idiosyncratic risk, since the investors are often less diversified. I will investigate four predictors for the SMB portfolio return: macro-economic factors, dispersion in multiples, bank loan availability, and illiquidity.

### 3.2.1 MACRO-ECONOMIC FACTORS

**Do European macro-economic factors indicate SMB return?**

The SMB returns show a correlation with economy; in particular in recovery periods, small caps outperform large caps and underperform large caps in recessions, as several periods in Figure 7 (CRSP data, construction described in Appendix C.4) illustrate. As small companies are often less diversified, they could depend more on the economic cycle and will therefore show a positive relation with expectations on the development of the economy. Then, a leading indicator of the economy, such as the U.S. CFNAI (Chicago Fed National Activity Index) would give a positive relation with future SMB return. Trading upon this strategy requires, next to a strong relation between macro-economic factors and SMB returns, a good indicator of European growth. Estimation of the phase within the economic cycle could be based upon a European equivalent of the CFNAI; but such an index does not (yet) exist. One could use the same construction procedure of the CFNAI based upon Eurostat or European country-based statistical services.

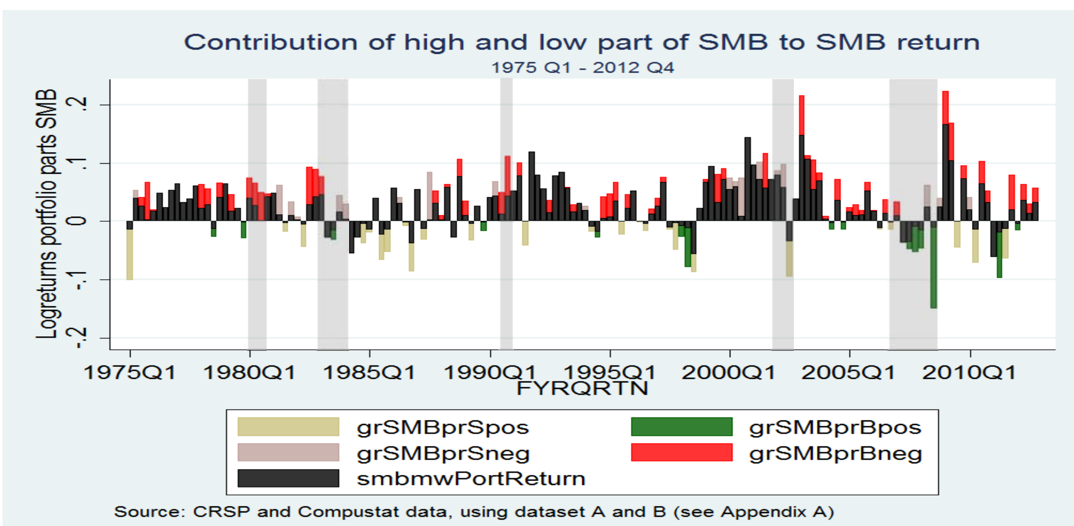



Figure 7: Contribution of 'small' and 'minus big' part to SMB portfolio, showing that in recession periods (gray shaded), SMB return mainly comes from outperformance of small caps to big caps (when both returns are negative). Black bars indicate the SMB return, the dark (red and green) bars indicate the contribution of the short (large cap part), light bars indicate the contribution of the long (small cap part). For example, in the first quarter, returns are negative, this is due to an outperformance of large to small and both were positive. The second quarter shows an outperformance of small to large when both are negative.

*Hypothesis 7: Macro-economic factors are positively related to future SMB return.*

Data sources that will be used for economic indicator are the Chicago Federal Reserve Bank (U.S.) and Eurostat (EU). As independent variable the CFNAI will be used for the U.S. market and a likewise constructed measure for the European market.

The detailed methodology is provided in chapter 4; results are given in paragraph 5.2.1.

### 3.2.2 DISPERSION IN MULTIPLES

**Does P/B dispersion between small and large stocks predict SMB returns?**

As the dispersion in multiples could predict HML portfolio future returns, it can also be used in the SMB portfolio, since small caps are often value stocks. Berk (1997) showed that the SMB return disappears when other measures than market capitalization are used. When prices of small firms and large firms differ much, it might be probable that they converge when the value effect is an important driver of this return (as explained in 3.1.2). This reasoning is in line with Kalesnik and Beck (2014) and they demonstrated an effect in the SMB return based upon price to book dispersion.

*Hypothesis 8: P/B dispersion between small and big firms is positively related to future SMB return.*

Data sources that will be used for market capitalization are CRSP (U.S.) and FactSet (EU), book value of equity is obtained using Compustat (U.S.) and FactSet (EU). As independent variable the dispersion between B/M of companies is used (the inverse of price to book).

The detailed methodology is provided in chapter 4; results are given in paragraph 5.2.2.

### 3.2.3 AVAILABILITY OF BANK LOANS

**Does the availability of bank loans indicate future SMB return?**

Since small firms are more dependent of external financing than larger firms, they might experience more problems in getting the financing for investments; when bank loan spreads are high, fewer new projects can be started, possibly resulting in lower growth in future earnings. Low bank loan spread would then predict high future SMB returns. As bank loan spread data is not widely available; the magnitude of bank loans is used (the sum of the corporate loans at the balance sheet of all U.S. banks). This is a more direct measure of bank loan availability.

*Hypothesis 9: The availability of bank loans is positively related to future SMB return.*

Data sources that will be used for the availability are the FDIC and ECB lending survey. As independent variable the volume (dollar or euro) of outstanding corporate loans is used.

The detailed methodology is provided in chapter 4; results are given in paragraph 5.2.3.

### 3.2.4 ILLIQUIDITY OF SMALL STOCKS

**Does an illiquidity index indicate future SMB return?**

Investors in small stocks should be compensated for illiquidity; thus a liquidity measure would have a correlation with the excess return of small stocks over large stocks when the liquidity of the first is lower (Goyenko, et al., 2009). Due to data availability, I use measures that use daily data; intraday liquidity data is not widely available. The Holden measure, a measure using serial correlation and price clustering, seems to be an accurate proxy of the spread. (Goyenko, et al., 2009). However, the Holden measure is relatively hard to construct and does not measure the price impact, which is relevant for trading a stock, therefore the Amihud measure is used (Amihud, 2002).

*Hypothesis 10: There exists an economical and statistical significant positive relationship between the Amihud illiquidity measure and future SMB return.*

Data sources that will be used to compute the Amihud measure are CRSP (U.S.) and FactSet (EU), portfolios are constructed using CRSP|Compustat (U.S.) and FactSet (EU) data. The Amihud measure is computed following the method described in his article (Amihud, 2002, p. 34), but with aggregation per quarter instead of per year (formula given in Table 3 on page 32).

The detailed methodology is provided in chapter 4; results are given in paragraph 5.2.4.

## 3.3 INTEGRAL PREDICTION OF HML & SMB

In order to implement the different predictors found in this study in an investment strategy, information about the relations between the predictors is valuable. Therefore the correlations between the factors found will be investigated using a correlation table and multiple linear regression.

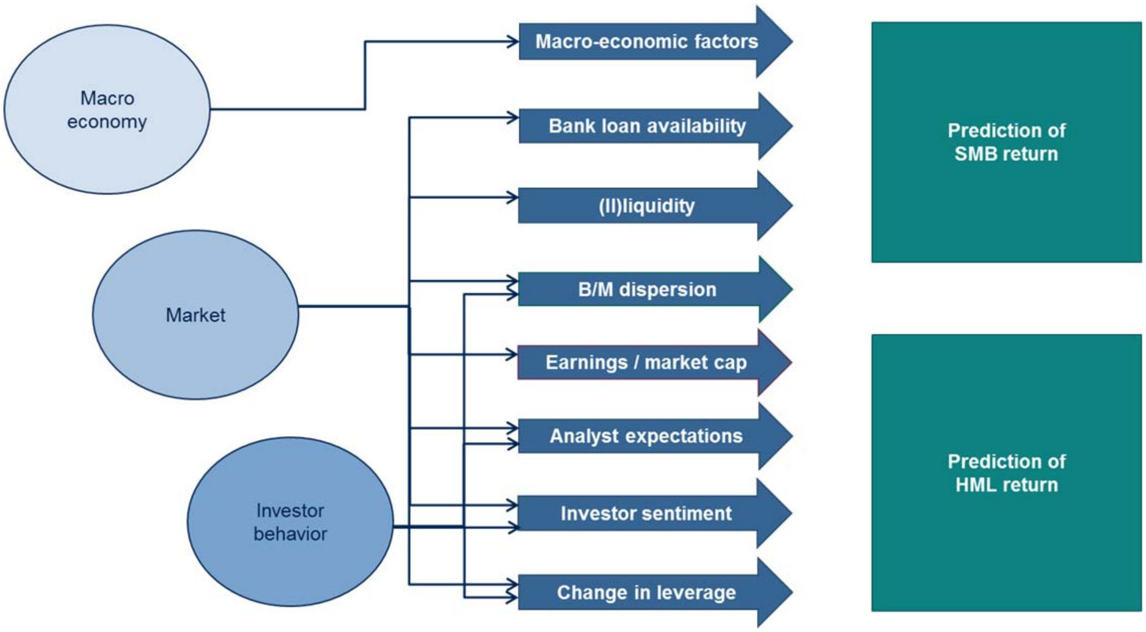



**Figure 8: Framework of thesis; prediction of SMB and HML returns based upon macro, market and behavioral measures.**

### 3.3.1 INVESTIGATE CORRELATIONS BETWEEN FACTORS FOUND

In order to find a model that predicts HML and SMB returns based on multiple predictors, a multiple regression model with Newey-West corrected error terms is build using predictors that are significant in the univariate regressions. First, correlations between the predictors are calculated and longer time period predictions are given, in order to select a set of variables that provides a statistical significant multivariate model. Second, several multiple regression models are tested to find the strongest predictive model. Results are provided in paragraph 5.3.

### 3.3.2 TRANSLATION TO PORTFOLIO CONTEXT

A good prediction of SMB and HML returns might be theoretically possible, but trading upon those predictors and yielding high portfolio returns is a next challenge; how can be traded upon these results? In order to investigate this, portfolios within the investable universe of KCM are compared with the portfolios tested. Using backtesting on this universe, the expected return, volatility, and Sharpe ratio of a timed HML and SMB portfolio are calculated on difference prediction horizons.

# 4 METHODOLOGY

The methodology to investigate the research questions, outlined in the previous chapter, is being described in detail in this chapter. The first paragraph describes the datasets used, why these are used, and how they are merged. The second paragraph elaborates about extreme and missing values in the data and the portfolio construction procedures. Third, the procedures of the construction of independent variables are described and the last paragraph of this chapter describes the statistical procedures to test the hypotheses.

## 4.1 DATA PREPROCESSING

To test the different hypotheses, various databases have been used to obtain the required data. The dependent variables (HML and SMB portfolio return) are obtained using merged CRSP and Compustat data (dataset B), combined with dataset A to deal with delistings. The fundamental data of t-2 is used, as the information gets to the market later. Daily CRSP data (dataset H) is used to calculate illiquidity; I/B/E/S data is used for analyst consensus and surprise (dataset I). Finally, datasets D, E, F, and G contain per-quarter data about respectively the risk-free rate; investor sentiment (Wurgler, 2011) and the CFNAI (Federal Reserve Bank of Chicago, 2014). Appendix A provides more information about the datasets used. Data between different CRSP databases is merged using company names as firm identifier (COMPNAME), I/B/E/S data is merged using the ticker symbol and the one-dimensional datasets D, E, and G only required the time variables. Figure 9 below illustrates this merging procedure.

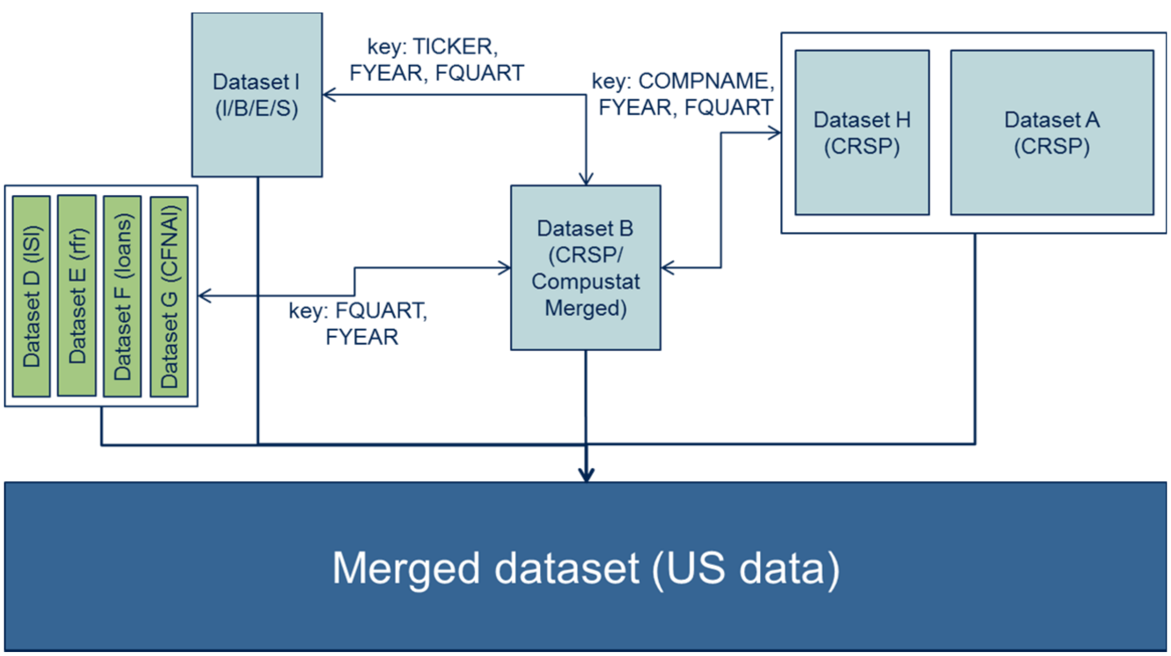



Figure 9: Dataset merging procedures U.S. data

The European data is merged in a likewise manner. There existed some issues on this merging process, as FactSet and I/B/E/S use different company names and tickers. The U.S. investor sentiment index is also used on the European dataset, this has two reasons. The first is that the index is constructed using various variables on which EU data is not good available and gathering this information does not fit within the time window of this thesis project. Second, as many investors are diversified among developed markets, the investor sentiment of the U.S. market could be a good proxy for the European market. The liquidity of the market is constructed by using daily FactSet data of prices and trading volumes.

The construction of the European Activity Index is done in the same way as with the CFNAI (the first principle component of a number of series of macro-economic indicators), but the signs of the weights of the indicators are reversed. In contrary to the CFNAI, I used many indicators with negative relation to economy (like inventory levels and unemployment). This procedure led to an index that correlated relatively well with the GDP growth of EU countries (0.8 on a 1-quarter lag), and is leading, as is illustrated in Figure 10.

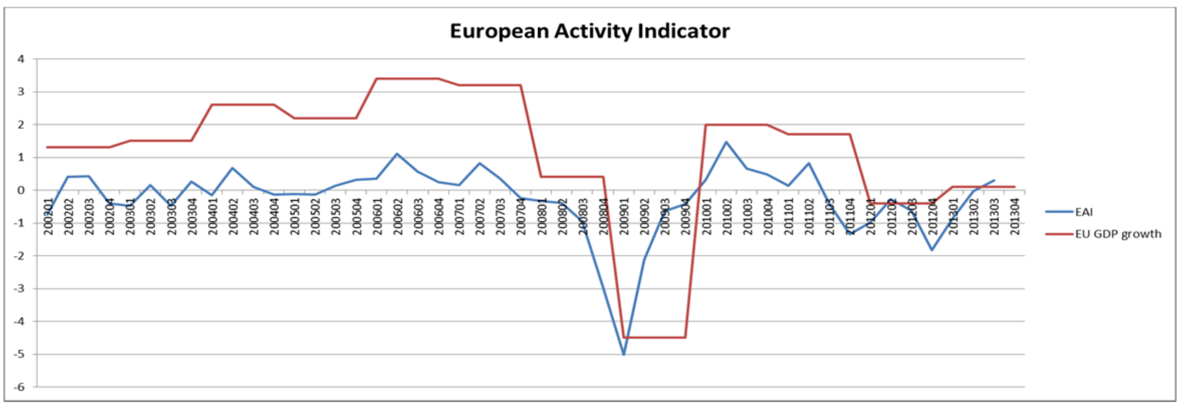



**Figure 10: European activity indicator and GDP growth 2002-2013**

The source for U.S. bank loan data was FDIC data on assets of banks, for Europe, ECB data is used (bank survey data). Ideally, one would use borrowing data of corporations instead of lending data of banks, but without access to these databases (Markit, LoanConnector), the other side of the transaction is used.

The merging procedure for European data is illustrated in Figure 11 on the next page; some datasets are used for both U.S. and EU data when they provide information for both U.S. and European stocks. Different are the following datasets: C (price and fundamental stock information from FactSet) to calculate portfolio returns; dataset L (prices of stocks and trade volumes) to calculate illiquidity; dataset N (that links dataset C to I using Bloomberg stock information). Finally, datasets J, K, and M contain per-quarter information about respectively the risk free rate, European Activity Index and European bank loan data. Appendix A provides more information about these datasets.

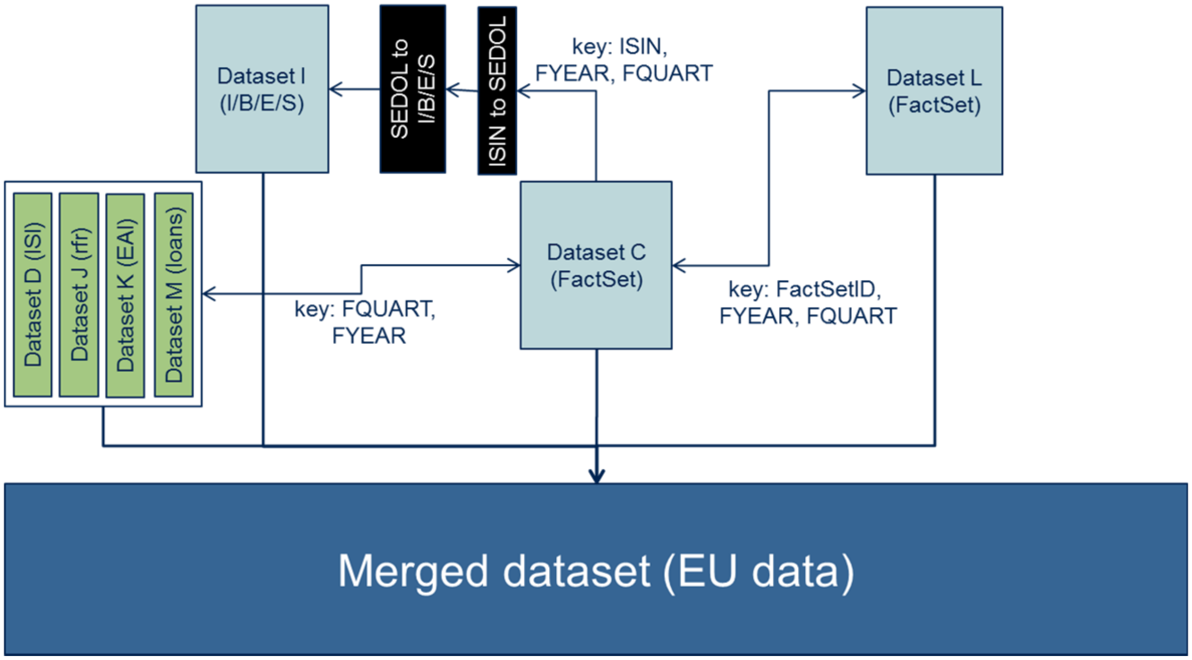



**Figure 11: Dataset merging procedures EU data**

## 4.2 DATA PREPARATION

Within the scope of data preparation, it is verified whether the combined raw data (the result of the previous step) is reliable and complete. This is done by extreme values and missing value analysis. On all used variables, extreme values analysis is performed. Furthermore, missing value analysis has been performed on the dataset that had large amounts of missing data (dataset B). Furthermore, the step of data preparation entails the creation of portfolios, the dependent variable in the later analyses. Although the datasets contain values that deviate much from the mean, it does not seem to be errors in the data. For example, Fannie Mae and General Motors show extremely low values of total common equity, but these are real observations. Because the extreme values are part of the investable universe, they are included in further analysis. Errors in CRSP and Compustat were investigated by Robert Bennin in 1980, concluding that both CRSP and Compustat have few errors since 1975 (Bennin, 1980).

Since the dataset is too large to perform a full missing value analysis, a random sample is drawn of 5% of the cases. This results in a subset of around 50,000 cases. The Little's MCAR $X^2$ test is significant, suggesting that the missing values are not missing at random, but have a systematic component. Although this is in line with reason, it might reveal a possible selection bias. This missing data (over 200,000 observations) cannot be gathered within the time scope of this research project. Since the removal of observations with missing values is based on an ex ante test to exclude a security from a portfolio, removing them will not give major data selection issues (in the contrary to ex post missing removal). Though this method is often applied in earlier studies, for example by Fama and French (1992), the risk of a selection bias is present. Table 25 and Table 26 in Appendix B provide the results of the missing value analysis performed with SPSS. The SPSS syntax for this procedure is provided in Appendix C under C.1. On the European dataset, no missing value analysis is performed as missing values were negligible.

Since only a part of the companies is in the database for the entire sample period, not dealing with this delisting issue would give a survivorship and delisting bias. In line with Li, Sullivan, and Garcia-Feijóo (2014), I use a delisting return of -55% for NASDAQ listed companies and -30% for NYSE and Amex stocks when delisting is performance-related. Using the delisting information in the CRSP database, the dataset obtained from Compustat and CRSP is merged with this delisting data using company names, this is illustrated in Figure 9. The CRSP dataset with delisting and exchange information is described as dataset 'A' in Appendix A.

HML and SMB portfolios can be constructed in several ways. It is most common to use a double sort, in line with Fama and French (1995), but in the number of fractions and the weighting within the portfolios, methods differ. I will use, in line with Fama and French, three fractions in value and two in size and a market-weighted portfolio (in order to compensate for the effect of micro-caps within the sample). I will validate my results using respectively six and ten fractions in value and size all using market-weighted portfolios; furthermore I verify my results using portfolios with equal weights in line with the methodology of Fama and French in their earlier work (Fama & French, 1992). Finally I use book-to-market as measure for value and market capitalization as measure for size. The latter is an often used variable; on value, P/E is also a common metric. However, in case of negative earnings, P/E cannot be interpreted well. Since book value is often positive, I prefer to use this above earnings.

Portfolios returns are calculated on a quarterly basis as many independent variables are based upon quarterly available data (fundamentals, bank loan data). When, for example, monthly data is used for predicting HML returns based upon a price-to-book dispersion (3.1.2), the prediction would use to many degrees of freedom since the underlying book value of equity does not change in the model (due to a lack of data), while the market capitalization does. The use of monthly data might therefore overestimate the reliability of the predictor.

## 4.3 CREATION OF VARIABLES

The several constructs mentioned in chapter 3, which could predict HML and SMB portfolio returns, have to be translated to measurable variables. Table 3 provides an overview of the independent variables used to answer my research questions. The argumentation to choose for these measures is explained below.

Table 3, on the following pages, shows for each independent variable the name; the method of computation; the method of aggregation to portfolio level; the mean; and the standard deviation. In addition to the mean and standard deviation, histograms of all independent variables are provided in Appendix D. In some instances, the method of aggregation is not provided, for example at "EP level of market", in these instances the computed variable is already a quarterly series. Leverage difference is, for example, calculated per company and therefore has to be aggregated to portfolio level to predict HML or SMB return. Since the effects of illiquidity and leverage difference are larger when market-capitalization is larger, market weights have been used. Since the relation of the impact of analyst surprise and consensus towards market capitalization is less obvious, the arithmetic mean is used here. In the formulas in Table 3, symbols are used that are explained in Table 2 below.

**Table 2: Explanations of symbols used**

| Symbol | Explanation |
|---|---|
| $\mathbf{x_{c,t}}$ | Measure "x" of company c at time t |
| $\mathbf{x_t}$ | Measure "x" at time t |
| $\mathbf{n}$ | Number of time periods (quarters) in dataset |
| $\mathbf{\lvert C_t \rvert}$ | Number of companies in dataset at time t |
| $\mathbf{\lvert C_{t,high} \rvert}$ | Number of companies in dataset at time t in high portfolio port, likewise for low, small, and big |
| $\mathbf{m(x)}$ | Median of measure x |
| $\boldsymbol{\sigma}\mathbf{(x)}$ | Standard deviation of x |
| $\mathbf{dqrt_t}$ | Number of days in quarter t |
| $\boldsymbol{\Delta_{t-1,t}}\mathbf{(x)}$ | Value of x at t minus value at t-1 (increment in x between t-1 and t) |
| $\mathbf{\lvert x \rvert}$ | Absolute value of x (not in case of C, see above) |

Table 3a: Independent variables used

| Name | Method of computation | Method of aggregation | Mean[1] (U.S. data) | SD[1] (U.S. data) |
|---|---|---|---|---|
| **1. EP level of market** | $cEPlevelMarket_t = \frac{\sum_{c=1}^{\|C_t\|} earnings_{c,t}}{\sum_{c=1}^{\|C_t\|} cMCAP_{c,t}}$ | | 0.003 | 0.067 |
| **2. Perpetual earnings value** | $cPerpEarnVal_t = \frac{cEPlevelMarket_t}{(RFR_t + 0.04)}$ | | 0.095 | 0.195 |
| **3. Dispersion between high and low** | $cBMdispHML_t = \frac{\sum_{c}^{\|C_{t,high}\|} cBM_{c,t}}{\|C_{t,high}\|} - \frac{\sum_{c}^{\|C_{t,low}\|} cBM_{c,t}}{\|C_{t,low}\|}$ | | 1.205 | 0.289 |
| **4. Dispersion around median (broad)** | $cBMdispBroad_t = \frac{\sum_{c=1}^{\|C_t\|} 1_{m(cBM_t)*0.6<cBM_{c,t}<1.4*m(cBM_t)}}{\|C_t\|}$ | | 0.425 | 0.057 |
| **5. Dispersion around median (narrow)** | $cBMdispNarrow_t = \frac{\sum_{c=1}^{\|C_t\|} 1_{m(cBM_t)*0.8<cBM_{c,t}<1.2*m(cBM_t)}}{\|C_t\|}$ | | 0.220 | 0.036 |
| **6. Standard deviation of B/M** | $sdBM_t = \sigma_t(BM_{c,t})$ | | 0.003 | 0.067 |

To use a high P/E as a predictor for HML return, total earnings are taken over the total market capitalization; E/P is used to deal with possible negative earnings, which would otherwise affect the measure worse. Assuming an equity premium of 4%, the perpetual value of the earnings of a company would be the earnings divided by the risk-free rate plus 0.04.

Dispersion is measured in four ways: the difference between the two parts of HML (and SMB) is a measure that Kalesnik and Beck use (2014). The next two are based upon the measure of Malkiel and Jun (2009) (dispersion around the median). The formula in Table 3a shows an indicator function (which is one when a company is within the bandwidth around the median and zero otherwise).The fourth measure, standard deviation, seems statistically the most common measure of dispersion.

[1] Both mean and standard deviation are calculated based upon the aggregate value.

Table 3b: Independent variables used

| Name | Method of computation | Method of aggregation | Mean (U.S. data) | SD (U.S. data) |
|---|---|---|---|---|
| **7. Investor sentiment index (ISI)** | Obtained from Jeffrey Wurgler (2011), original series used. | | 0.056 | 0.905 |
| **8. Absolute ISI** | $absISI_t = \|ISindex_t\|$ | | 0.632 | 0.648 |
| **9-11. Analyst surprise (high, low, and total)** | $cSurpComp_{c,t} = \frac{(actual_{c,t} - surprise_{c,t}) * CSHOQ_{c,t}}{cMCAP_{c,t}}$ | Arithmetic mean per portfolio part and total | -0.023<br>-0.001<br>-0.007 | 0.025<br>0.008<br>0.008 |
| **12-14. Analyst consensus (high, low, and total)** | $cConsComp_{c,t} = \left(\left(\frac{buypct_{c,t}}{100}\right)^2 + \left(\frac{selpct_{c,t}}{100}\right)^2 + \left(\frac{holdpct_{c,t}}{100}\right)^2 - 0.33\right) * \frac{numrec - 1}{numrec}$ | Arithmetic mean per portfolio part and total | 0.339<br>0.398<br>0.374 | 0.036<br>0.073<br>0.053 |
| **15-17. Change in leverage (high, low, and difference)** | $cmwLeverageDiff_{c,t} = \Delta_{t-1,t}\left(\frac{LLTQ_{c,t} + LTCQ_{c,t}}{ATQ_{c,t}}\right)$ | Market-weighted average per portfolio part and the difference between high and low | 0.0029<br>-0.0002<br>0.0031 | 0.0061<br>0.0048<br>0.0066 |

The investor sentiment index is used unaltered. The absolute investor sentiment index is chosen since HML seems high when investor sentiment takes extreme values (which can indicate irrational investing behavior and thus high HML return, irrespective whether it is positive or negative sentiment).

For analyst surprise, the difference between expected and real earnings per share is multiplied with the number of shares and divided by the market capitalization to make it comparable between companies. Consensus is computed using a sum of squares method and corrected for the number of recommendations (in case of one recommendation, one cannot really speak of consensus, it rather is a lack of data). While other aggregation methods are mainly market cap driven, for the analyst variables an arithmetic mean is taken. Since analyst coverage is biased towards large caps, using market weights over these variables would give large caps even more weight.

Change in leverage is calculated as the change over one quarter of total debt over total assets, which is a straightforward measure of leverage.

Table 3c: Independent variables used

| Name | Method of computation | Method of aggregation | Mean (U.S. data) | SD (U.S. data) |
|---|---|---|---|---|
| **18. CFNAI (EAI)** | Obtained from Chicago Federal Reserve Bank, original series used. | | -0.130 | 0.897 |
| **19. PI** | | | -0.014 | 0.305 |
| **20. EUH** | | | -0.007 | 0.294 |
| **21. CH** | | | -0.021 | 0.161 |
| **22. SOI** | | | -0.030 | 0.158 |
| **23. Dispersion between small and big** | $cBMdispSMB_t = \frac{\sum_{c=1}^{\lvert C_{t,small}\rvert} cBM_{c,t}}{\lvert C_{t,small}\rvert} - \frac{\sum_{c=1}^{\lvert C_{t,big}\rvert} cBM_{c,t}}{\lvert C_{t,big}\rvert}$ | | 0.303 | 0.124 |
| **24. Normalized bank loans** | $bankLoanNorm_t = \frac{bankLoanLNdiff_t - \mu(bankLoanLNdiff)}{\sigma(bankLoanLNdiff)}$ | | 0.000 | 1.000 |
| **25-27. Amihud illiquidity measure (small, big, and difference)** | $cAmiComp_{c,t} = \frac{\sum_{day=1}^{dqrt_t} \frac{\lvert retx_{c,t}\rvert}{vol_{c,t} * prc_{c,t}}}{dqrt_t}$ | Market-weighted average per portfolio part and the difference between small and big | $1.1*10^{-6}$<br>$8.9*10^{-9}$<br>$1.1*10^{-6}$ | $9.3*10^{-7}$<br>$7.2*10^{-9}$<br>$9.2*10^{-7}$ |

The economic series were used in its original, unrevised, form as published by the Federal Reserve Bank of Chicago (2014).

The dispersion between small and big is created in the same manner as dispersion between high and low, using the method of Kalesnik and Beck (2014).

The bank loans are the normalized differences on the bank loan data, in order to remove long-term trends like economic growth.

Finally, for illiquidity, the Amihud measure (2002) has been used, calculating the daily price change over the dollar trade volume aggregated to quarters. Since it is an easy and rather good measure (Goyenko, et al., 2009), Amihud is chosen.

## 4.4 LINEAR REGRESSION TESTS

Since both portfolio returns and predictors are serially correlated (Figure 12 and Figure 13 show serial correlation of the HML and SMB portfolio returns), the cross-sectional regressions need to be robust against this effect. This is done by using Newey-West adjusted errors. The number of lags used is 20, corresponding to five years. Figure 12 and Figure 13 show that the serial correlation is mainly present in the past few quarters, so five years seems to be a safe choice. The regression assumptions are tested using normal probability plots, which is in line with Miller (1997) who prefers this method above Kolmogorov-Smirnov tests, as it gives more insight and could better deal with fat tails.

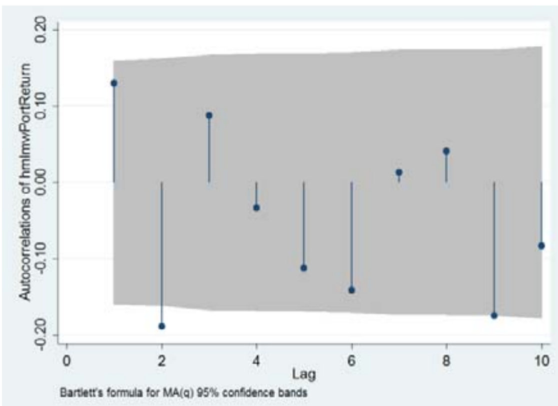



Figure 12: Autocorrelation HML portfolio return

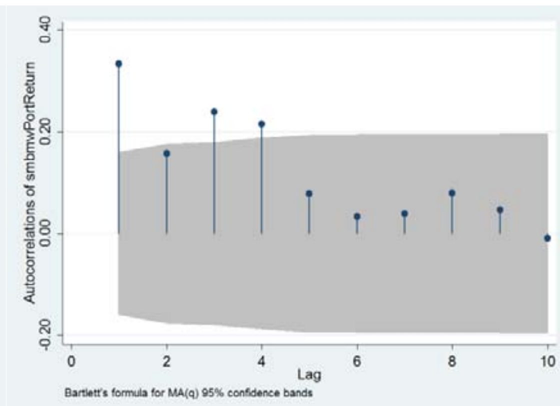



Figure 13: Autocorrelation SMB portfolio return

Next to the regressions, autocorrelation tests are performed to find mean-reverting behavior. These tests are performed using a portmanteau white noise test and an augmented Dickey-Fuller root test.

Integral prediction is performed by estimating correlations between predictors and by multiple regression models (with Newey-West adjusted errors). The results of these tests are provided in chapter 5.

# 5 EMPIRICAL RESULTS

This chapter provides the results for the hypotheses tested, starting with the test of HML returns (paragraph 5.1), followed by SMB (5.2). It concludes with integral prediction of HML and SMB returns (5.3), robustness tests (5.4), and a translation to asset allocation (5.5). In the first paragraphs only market-weighted HML and SMB portfolios are reported as they are better investable in a passive strategy. Later, the results are checked for other portfolio definitions (5.4.1).

## 5.1 EMPIRICAL RESULTS ON HML TESTS

Table 4 below shows the linear regression results on the hypotheses 1 to 6. The following subparagraphs elaborate about these results.

**Table 4: Univariate regression results HML tests (with Newey-West adjusted error terms). The numbers indicate the independent variables used, as mentioned in Table 3 on page 32. Significant regressions are made bold; they are respectively (i) B/M dispersion between high and low; (ii) absolute investor sentiment; and (iii) analyst surprise on 'high' portfolio part. The asterisks indicate significance levels, respectively 90%, 95% and 99% for 1, 2 and 3 asterisks.**

| | | Market cap to fundamentals | | Multiple dispersion | | | | Investor sentiment | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| US data | Coeff. (b) | -1.3068 | -0.4952 | **0.0644** | -0.0749 | -0.1373 | **0.0905** | 0.0029 | **0.0185** |
| | t-value | -1.18 | -2.11** | **2.40**** | -0.82 | -0.95 | **1.87*** | 0.45 | **2.20**** |
| | $R^2$ | 2.5% | 16.8% | **6.2%** | 0.3% | 0.4% | **2.6%** | 0.1% | **2.8%** |
| | b*s | | | **1.86%** | | | **1.21%** | | **1.20%** |
| EU data | Coeff. (b) | -0.0321 | 0.0018 | **0.0387** | -0.2001 | -0.4023 | **0.0532** | 0.0076 | **0.0234** |
| | t-value | -0.15 | 0.11 | **2.53**** | -1.01 | -1.04 | **2.36**** | 0.58 | **1.81*** |
| | $R^2$ | 0.0% | 0.0% | **2.6%** | 1.3% | 2.2% | **1.6%** | 0.6% | **3.7%** |
| | b*s | | | **1.01%** | | | **0.79%** | | **1.25%** |

| | | Analyst surprise | | | Analyst consensus | | | Change in leverage | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| US data | Coeff. (b) | **-0.6364** | 0.1396 | -1.7649 | -0.0255 | 0.0791 | 0.0613 | 0.4624 | -0.8125 | 0.8108 |
| | t-value | **-1.80*** | 0.11 | -1.72* | -0.09 | 0.84 | 0.42 | 0.29 | -0.49 | 0.77 |
| | $R^2$ | **3.2%** | 0.0% | 2.6% | 0.0% | 0.4% | 0.1% | 0.1% | 0.3% | 0.5% |
| | b*s | **1.59%** | | 1.42% | | | | | | |
| EU data | Coeff. (b) | **-0.2682** | 0.2939 | -0.6560 | -0.1608 | -0.1677 | -0.3139 | -1.4892 | -3.0361 | 1.5911 |
| | t-value | **-0.89** | 1.3 | -1.06 | -0.58 | -0.62 | -1.21 | -0.64 | -2.40** | 3.93*** |
| | $R^2$ | **3.6%** | 1.4% | 4.2% | 0.5% | 0.6% | 1.9% | 1.5% | 11.9% | 4.7% |
| | b*s | **1.11%** | | 1.19% | | | | | | |

### 5.1.1 HIGH MARKET CAPITALIZATION RELATIVE TO FUNDAMENTALS

As described in subparagraph 3.1.1, a high market capitalization to fundamentals could indicate a high future HML return.

*$H_0$ 1: The ratio of earnings to prices is not negatively related to future HML return.*

*$H_1$ 1: The ratio of earnings to prices is negatively related to future HML return.*

As Table 4 shows, the ratio of earnings to market capitalization (1) does not predict HML returns for both samples. The perpetual earnings value, as described in chapter 4, does only show statistical significance for the U.S. dataset, but the European results are not in line at all. Subsequently, the null hypotheses cannot be rejected.

### 5.1.2 MULTIPLE DISPERSION TO PREDICT HML RETURN

As described in subparagraph 3.1.2, a price-to-book multiple could predict HML return.

*$H_0$ 2: P/B dispersion is not positively related to future HML return.*

*$H_1$ 2: P/B dispersion is positively related to future HML return.*

Table 4 shows the results of a linear regression, using the difference between the book-to-price difference of the value and growth portfolio as predictor (3). For both the U.S. and European data sets, it shows significant results at 95% confidence level. The variance that is explained by book-to-market dispersion differs for both datasets. It explains more variance in the U.S. sample (illustrated by Figure 14, Figure 15 and Figure 16 on the following pages), but the economic significance is high in both cases. The null hypothesis can therefore be rejected.

The standard deviation in dispersion (6) also has a high predictive value, but due to a high correlation (0.89) with the dispersion between high and low, combined with the higher statistical and economic significance of the latter, it is not valuable to use it in predictive models when dispersion between high and low can be used.

The other two predictors, based upon dispersion around the median, seem to perform worse as HML predictors; however, the sign of the relation is in line with the economic rationale. Fewer observations around the median indicate higher returns.

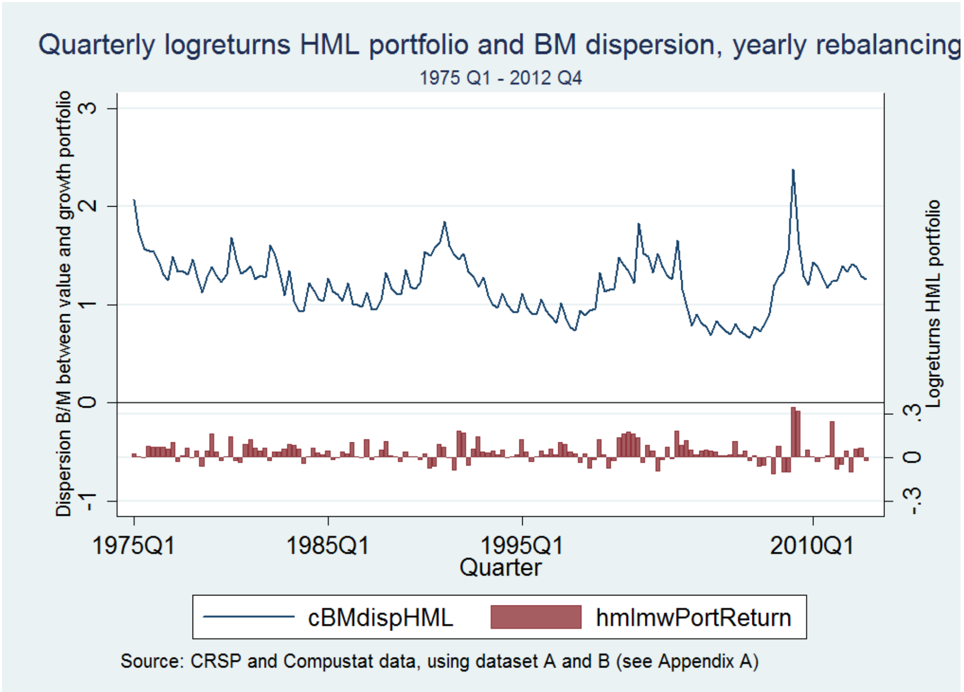



**Figure 14: Quarterly log returns HML portfolio and B/M dispersion between value and growth portfolio**

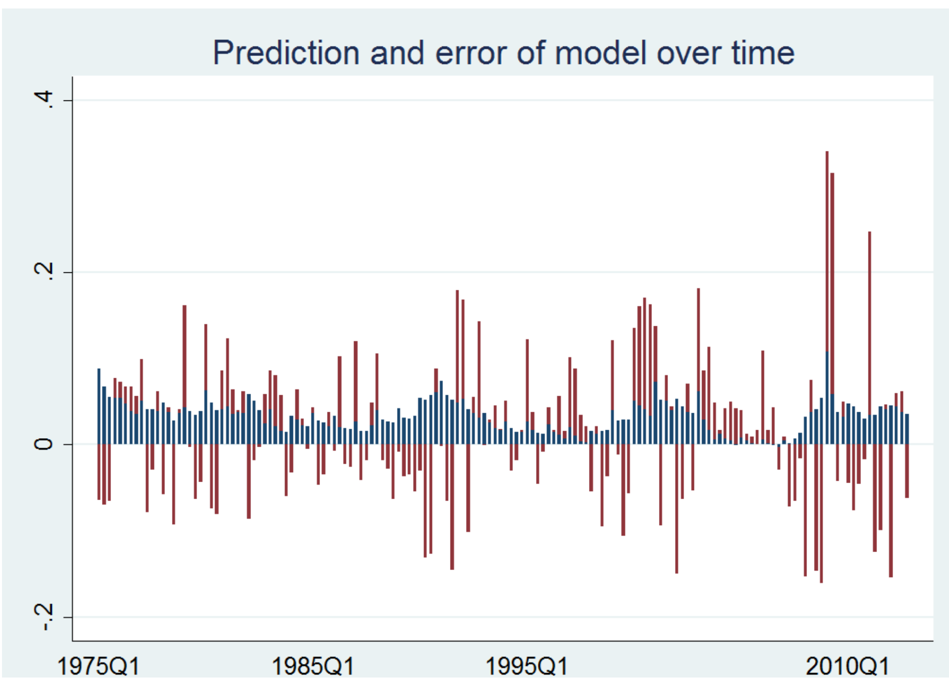



**Figure 15: Prediction and error of model with B/M dispersion as predictor of HML return on U.S. data**

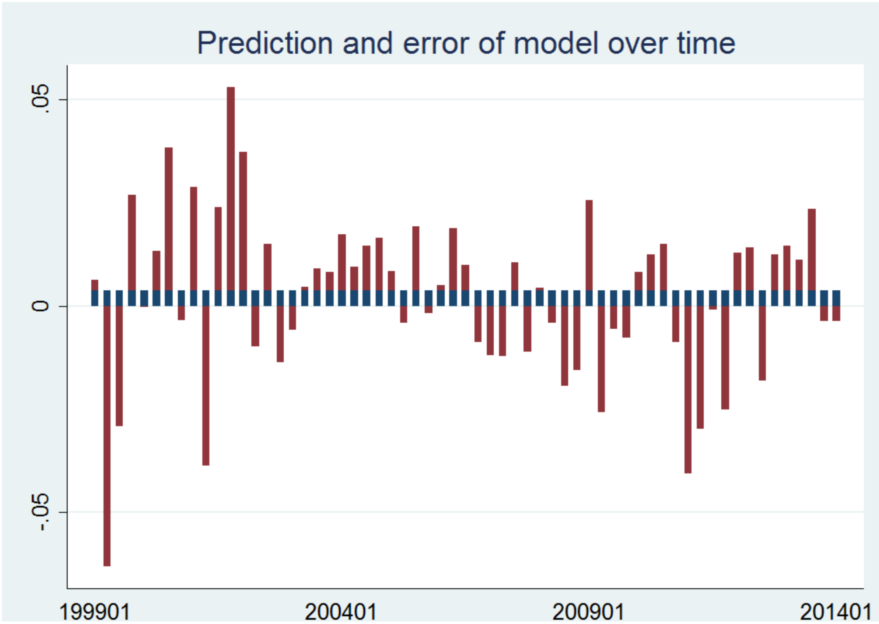



**Figure 16: Prediction and error of model with B/M dispersion as predictor of HML return on EU data**

#### 5.1.3 INVESTOR SENTIMENT TO PREDICT HML RETURN

As described in subparagraph 3.1.3, investor sentiment could indicate a high future HML return.

*$H_0$ 3: The investor sentiment index is not positively related to future HML return.*

*$H_1$ 3: The investor sentiment index is positively related to future HML return.*

As Table 4 shows, investor sentiment (7), the original series, does not show predictive power for HML return. However, when the investor sentiment index is compared to the HML return, one can observe that in extreme high sentiment, going short low pays off as well as going long high in extreme low sentiment (shown in HML returns are positive (as shown in Figure 17 and Figure 18). Therefore, I tested whether absolute investor sentiment would predict HML return, and it does with high statistical and economic significance. Of course, there is a selection bias present on the U.S. data. The European dataset is slightly significant at 90% level. As the index is based upon U.S. sentiment indicators, a European sentiment indicator might reveal better predictive power.

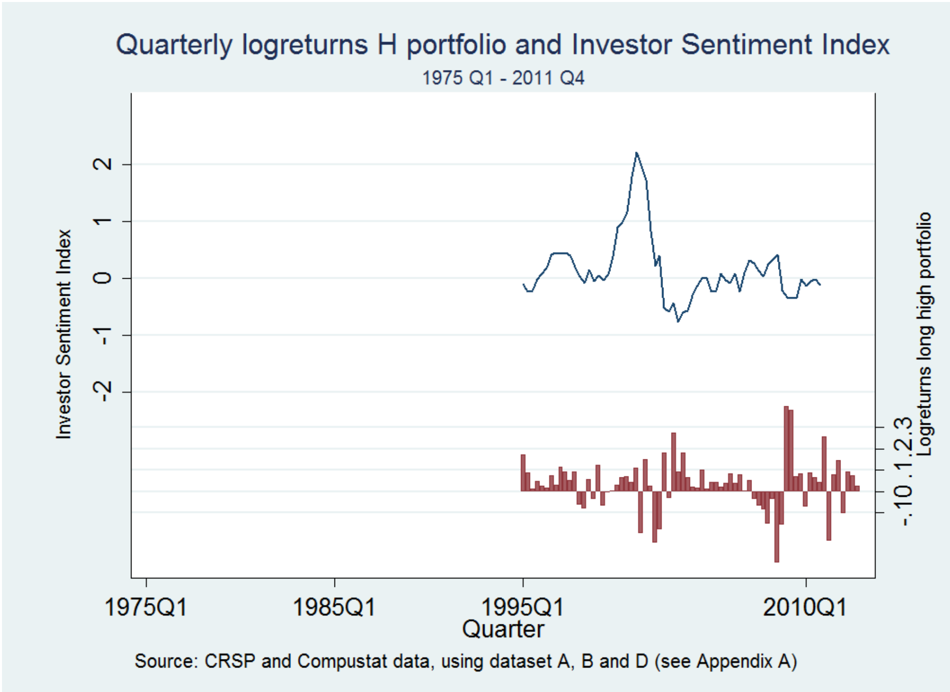



**Figure 17: Quarterly log returns long high portfolio and investor sentiment index**

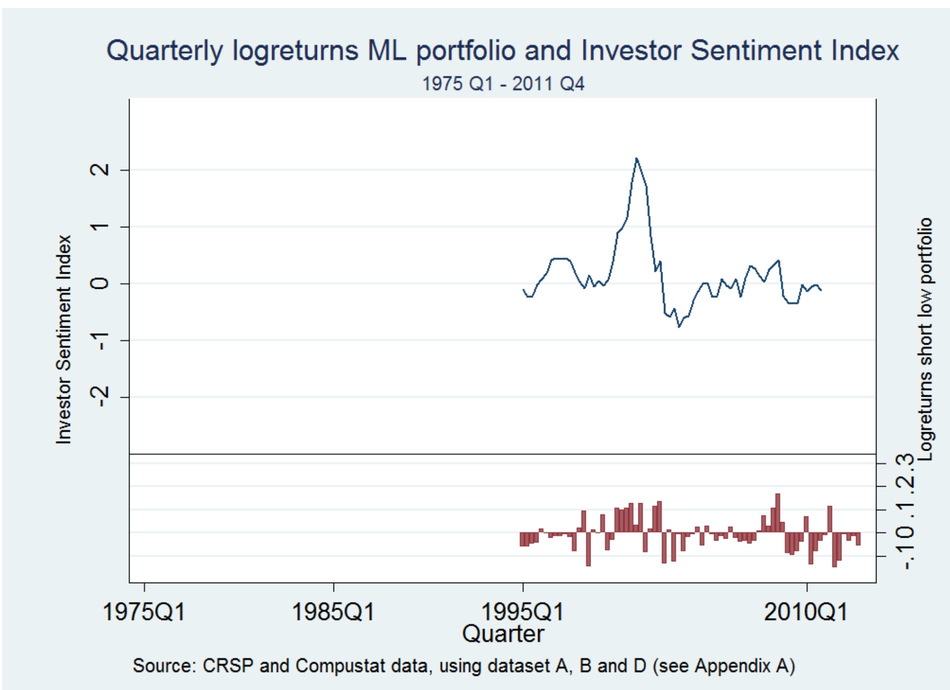



**Figure 18: Quarterly log returns short low portfolio and investor sentiment index**

Furthermore, Table 5 and Table 6 on the following page show that the absolute ISI has (low) predictable power on both the high and low part of the HML portfolio; high sentiment predicts both returns of the high book-to-price companies as well as negative returns to the low portfolio (the table shows the result of the short position on 'low'). This supports the idea that extreme sentiment predicts a moment that valuations of both high and low converge. The hypothesis that the unadjusted index predicts HML returns is rejected though.

**Table 5: Regression result on absolute investor sentiment to predict 1-quarter future return of the HML high part**

| Prediction of 'high' part of HML market-weighted return using absolute investor sentiment index, U.S. data | | | | |
|---|---|---|---|---|
| | Coefficient | t-value | $R^2$ | b*s |
| abs ISI | 0.00856 | 1.58 | 0.3% | 0.1% |

**Table 6: Regression result on absolute investor sentiment to predict 1-quarter future return of the HML low part (result of the short position of 'low')**

| Prediction of 'minus low' part of HML market-weighted return using absolute investor sentiment index, U.S. data | | | | |
|---|---|---|---|---|
| | Coefficient | t-value | $R^2$ | b*s |
| abs ISI | 0.0100 | 1.25 | 0.1% | 0.1% |

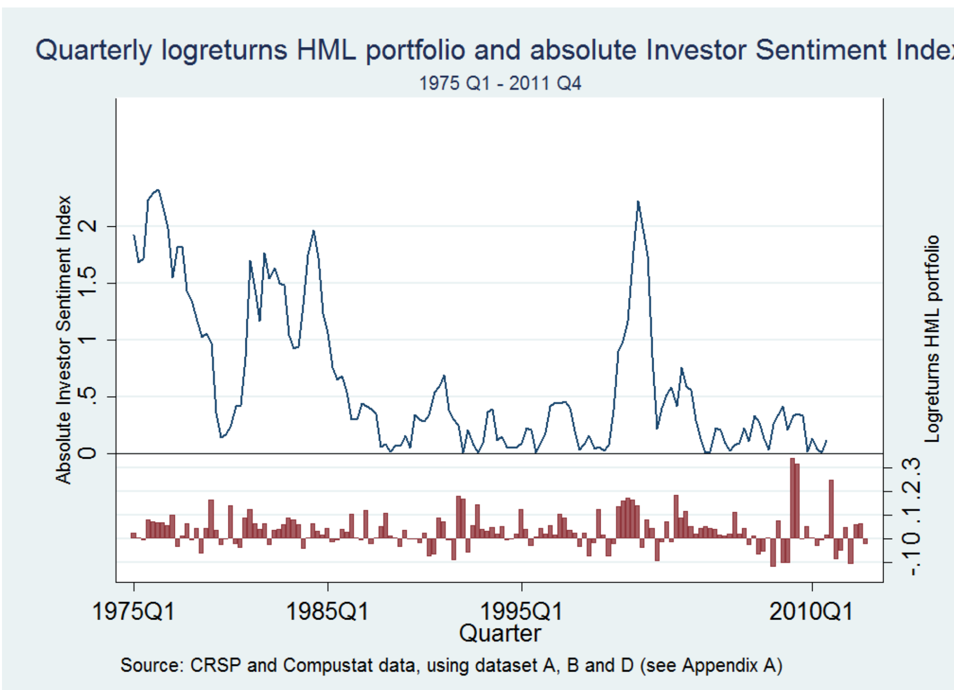



**Figure 19: Quarterly log returns HML portfolio and absolute investor sentiment index**

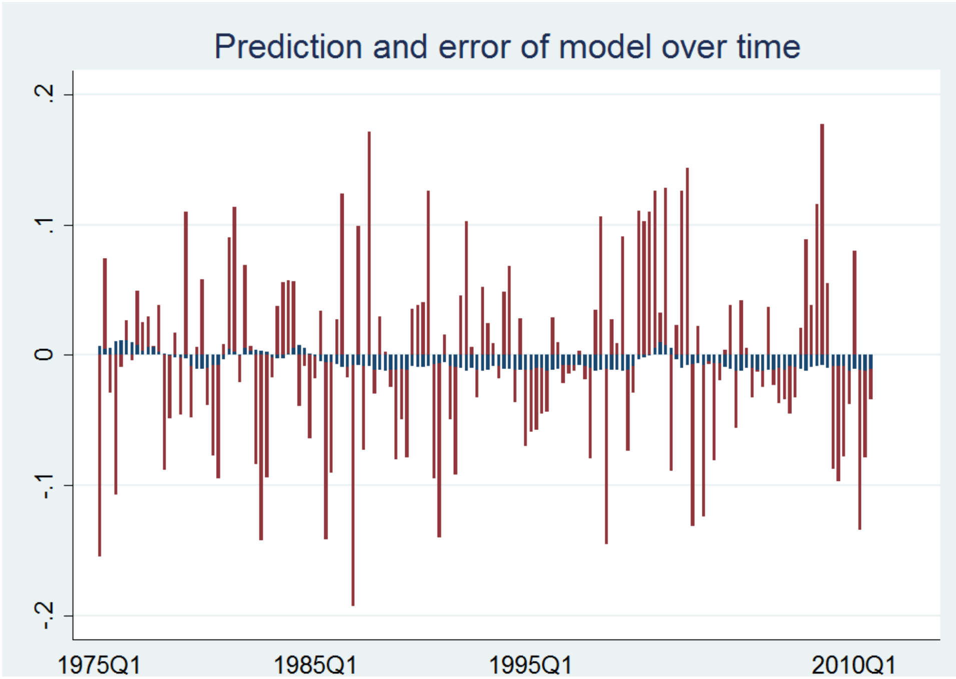



**Figure 20: Prediction and error of model with absolute investor sentiment as predictor of HML return on U.S. data**

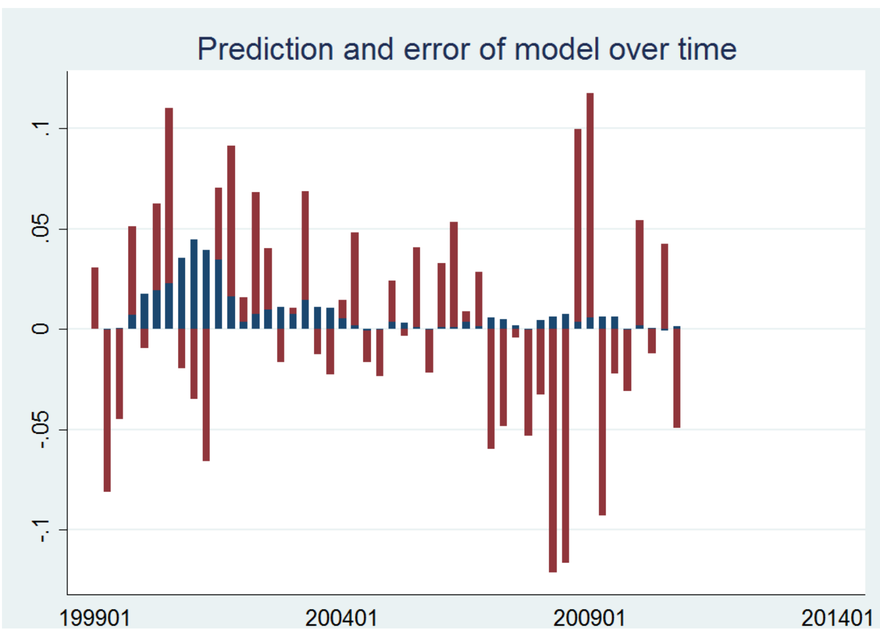



**Figure 21: Prediction and error of model with absolute investor sentiment as predictor of HML return on EU data**

#### 5.1.4 PERFORMANCE RELATIVE TO ANALYST EXPECTATIONS TO PREDICT HML RETURN

As described in subparagraph 3.1.4, the performance relative to analyst expectations could indicate a high future HML return.

*$H_0$ 4: The exceeding of analyst expectations is not positively related to future HML return.*

*$H_1$ 4: The exceeding of analyst expectations is positively related to future HML return.*

Analyst surprise seems a predictor for U.S. HML return, both via the 'high' part of the portfolio (9) as on total market (11). As the original hypothesis is stated for the extrapolation on growth stocks, a negative sign for this effect on value stocks would be in line with the rationale. As Table 4 shows, the European dataset does not show statistical significant results. This could be due to lower analyst coverage of the securities within the FactSet dataset. Furthermore, the European dataset is smaller in two dimensions; it contains about a sixth of the number of companies of the U.S. dataset and has a timespan that is smaller than half that of the U.S. sample; this would contribute to a lower statistical significance. There is not enough evidence to reject the null hypothesis, but the findings are support for the alternative hypothesis.

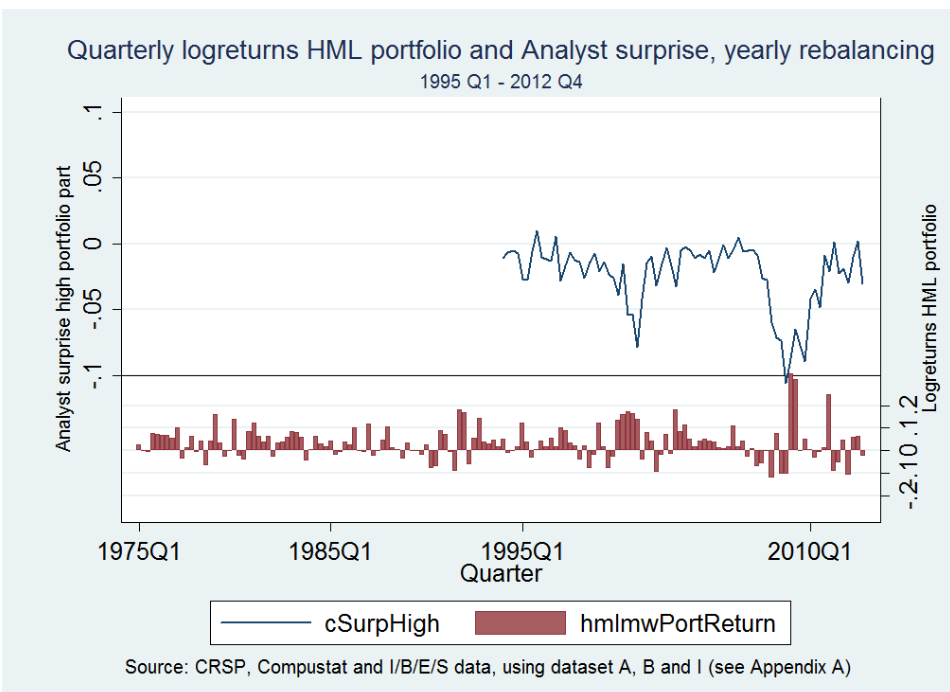



**Figure 22: Quarterly log returns HML portfolio and analyst surprise on the high portfolio part**

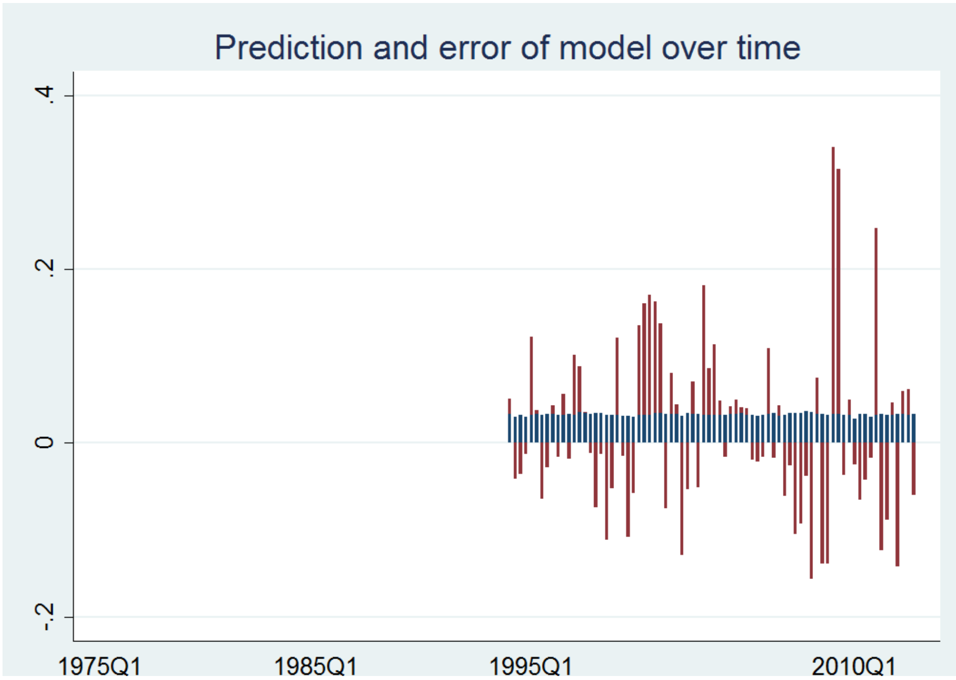



**Figure 23: Prediction and error of model with analyst surprise as predictor of HML return on U.S. data**

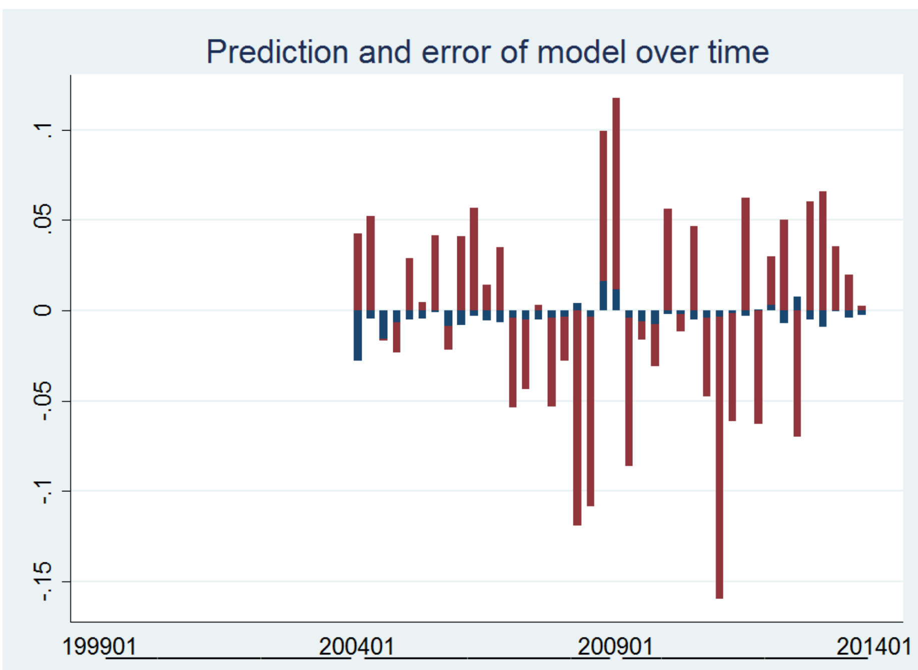



**Figure 24: Prediction and error of model with analyst surprise as predictor of HML return on EU data**

#### 5.1.5 ANALYST CONSENSUS TOT PREDICT HML RETURN

As described in subparagraph 3.1.5, analyst consensus could indicate a high future HML return.

*$H_0$ 5: Analyst consensus is not negatively related to future HML return.*

*$H_1$ 5: Analyst consensus is negatively related to future HML return.*

Table 4 shows at the regressions 12-14 (analyst consensus on respectively high, low, and total) no strong results, furthermore, coefficients at the European dataset show different signs of the coefficients, which implies that it is hard to attribute predictive power of analyst consensus that could work in both markets. Subsequently, the null hypothesis cannot be rejected.

#### 5.1.6 CHANGE IN LEVERAGE TO PREDICT HML RETURN

As described in subparagraph 3.1.6, analyst consensus could indicate a high future HML return.

*$H_0$ 6: The change in leverage of companies is not positively related to future HML return.*

*$H_1$ 6: The change in leverage of companies is positively related to future HML return.*

As Table 4 shows at regressions 15-17 (high, low, and the difference between the two portfolio parts), change in leverage gives mixed results. For the U.S. dataset, it has no predictive power. The European data shows a significant result, but the sign is not in line with expectations. These results lead to no rejection of the null hypothesis.

## 5.2 EMPIRICAL RESULTS ON SMB TESTS

Table 7 below shows the linear regression results on the hypotheses 7 to 10. The following subparagraphs elaborate about these results.

**Table 7: Univariate regression results SMB tests (with Newey-West adjusted error terms). The numbers indicate the independent variables used, as mentioned in Table 3 on page 32. Significant regressions are made bold; they are respectively B/M dispersion between small and big; bank loan availability; and Amihud illiquidity based on both the small portfolio part and the difference between high and low. The asterisks indicate significance levels, respectively 90%, 95% and 99% for 1, 2 and 3 asterisks.**

| | | Macro-economic variables | | | | |
|---|---|---|---|---|---|---|
| | | 18 | 19 | 20 | 21 | 22 |
| US data | Coeff. (b) | -0.0068 | -0.0135 | -0.0180 | -0.0467 | -0.0244 |
| | t-value | -1.33 | -1.02 | -1.06 | -1.54 | -0.78 |
| | $R^2$ | 2.5% | 1.1% | 1.8% | 3.7% | 1.0% |
| | b*s | 0.6% | | | | |
| EU data | Coeff. (b) | -0.0027 | | | | |
| | t-value | -1.00 | | | | |
| | $R^2$ | 0.8% | | | | |
| | b*s | 0.3% | | | | |

| | | Multiple dispersion | Bank loan availability | Illiquidity of market | | |
|---|---|---|---|---|---|---|
| | | 23 | 24 | 25 | 26 | 27 |
| US data | Coeff. (b) | **0.1568** | **-0.0185** | **4194** | -273881 | **4238** |
| | t-value | **5.38*** ** | **-8.05*** ** | **1.62** | -0.36 | **1.64** |
| | $R^2$ | **24.7%** | **18.7%** | **1.0%** | 0.3% | **1.0%** |
| | b*s | **1.9%** | **1.9%** | **0.4%** | | **0.4%** |
| EU data | Coeff. (b) | **0.1000** | **0.0052** | **96265** | 66487 | **33471** |
| | t-value | **3.76*** ** | **1.78*** | **5.52*** ** | 0.66 | **4.49*** ** |
| | $R^2$ | **8.9%** | **5.1%** | **15.2%** | 2.0% | **7.8%** |
| | b*s | **1.0%** | **0.5%** | **1.2%** | | |

### 5.2.1 MACRO-ECONOMIC FACTORS TO PREDICT SMB RETURN

As described in subparagraph 3.2.1, macro-economic factors indicators could predict a high future SMB return.

*$H_0$ 7: Macro-economic indicators are not positively related to future SMB return.*

*$H_1$ 7: Macro-economic indicators are positively related to future SMB return.*

Macro-economic variables have a low economic significance, next to a low statistical significance, as is presented in regressions 19-22 in Table 7. The t-value of 1.33 could indicate a lack of data (as it is a company-independent series, there are relatively few data points), as it is within 20% confidence level. However, economical significance remains low, giving this variable low value in predicting SMB return. Only regression 21 gives significant result at 10% confidence for the U.S. data, but this result is not convincing. Therefore the null hypothesis cannot be rejected.

When not the economic index itself is used, but the four phases of the economic cycle instead (expansion, slow-down, recession and recovery), an SMB portfolio performs better in periods of recovery. This effect, using a two-sample t-test, is found significant for the U.S. market at 95% confidence, but not at the European dataset. This could be attributable to a lack of data: in about 15 years there could be too few economic cycles. The difference between the other three periods is not statistically significant.

### 5.2.2 MULTIPLE DISPERSION TO PREDICT SMB RETURN

As described in subparagraph 3.2.2, dispersion in multiples could indicate a high future SMB return.

*$H_0$ 8: P/B dispersion is not positively related to future SMB return.*

*$H_1$ 8: P/B dispersion is positively related to future SMB return.*

As regression 23 in Table 7 shows, both for the European and U.S. dataset, the dispersion between the high and low parts of the portfolio seems a good predictor for SMB portfolio return, both statistically as economically. As the composition of the SMB portfolio has many companies in common with the HML portfolio, the result found could be driven by the HML dispersion. This does, however, not explain the weaker significance for the HML sample. One of the possible explanations is that rebalancing on SMB uses newer data (only market capitalization) than HML (book value of several months ago); it therefore is stronger related to the cheapest stocks. Given significance in both statistical as economic sense, the null hypothesis is rejected.

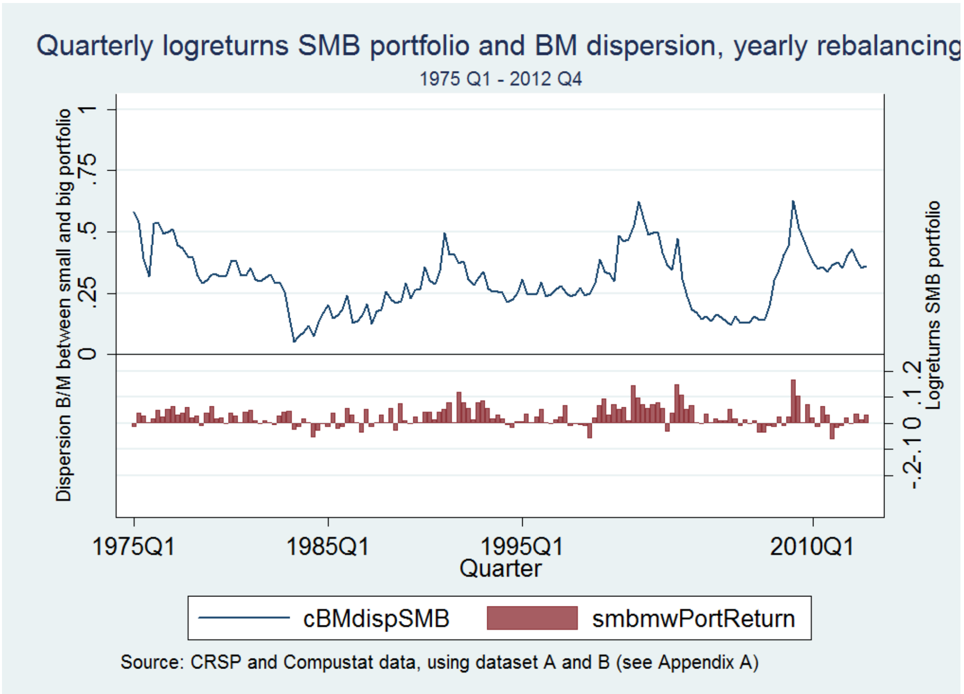



**Figure 25: Quarterly log returns SMB portfolio and B/M dispersion between small and big portfolio**

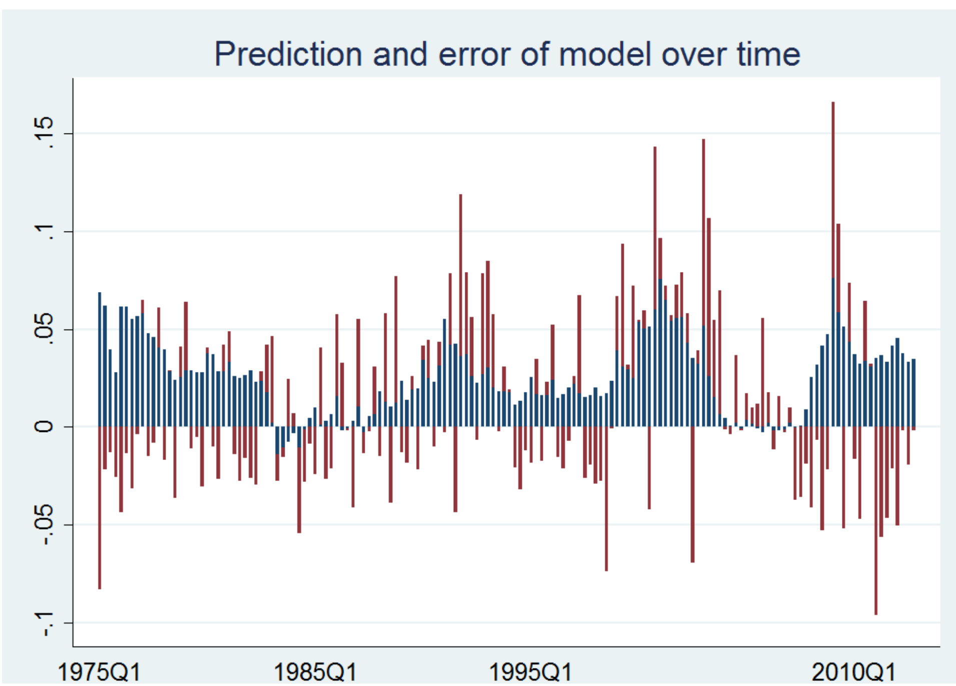



**Figure 26: Prediction and error of model with B/M dispersion as predictor of SMB return on U.S. data**

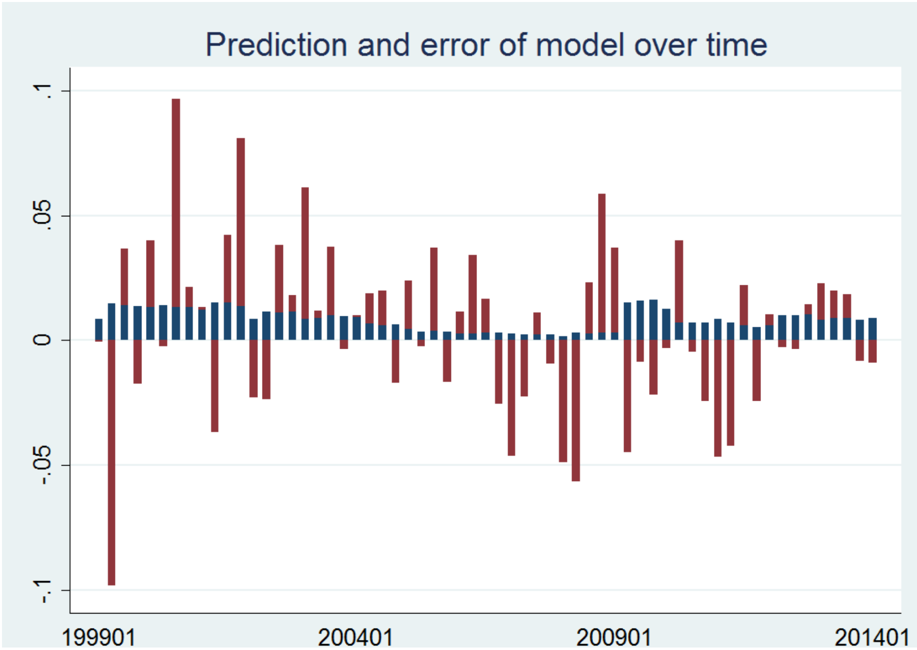



**Figure 27: Prediction and error of model with B/M dispersion as predictor of SMB return on EU data**

### 5.2.3 AVAILABILITY OF BANK LOANS

As described in subparagraph 3.2.3, the availability of bank loans could indicate a high future SMB return.

*$H_0$ 9: The availability of bank loans is not positively related to future SMB return.*

*$H_1$ 9: The availability of bank loans is positively related to future SMB return.*

As Table 7 shows that the availability of bank loans is a significant predictor for the U.S. dataset (regression 24); the European dataset shows less significance. The negative sign could be explained by a lagged effect (for example, the three-year lagged effect of bank loans is also statistical significant) (Figure 28). The effect of the loans in stock returns might be lagged; for example when projects are financed using debt, it may take several years before they generate returns.

It is unlikely that the availability of bank loans does relate to SMB returns, but the magnitude, direction and the lag of the effect are not yet clear. The null hypothesis can therefore not yet be rejected.

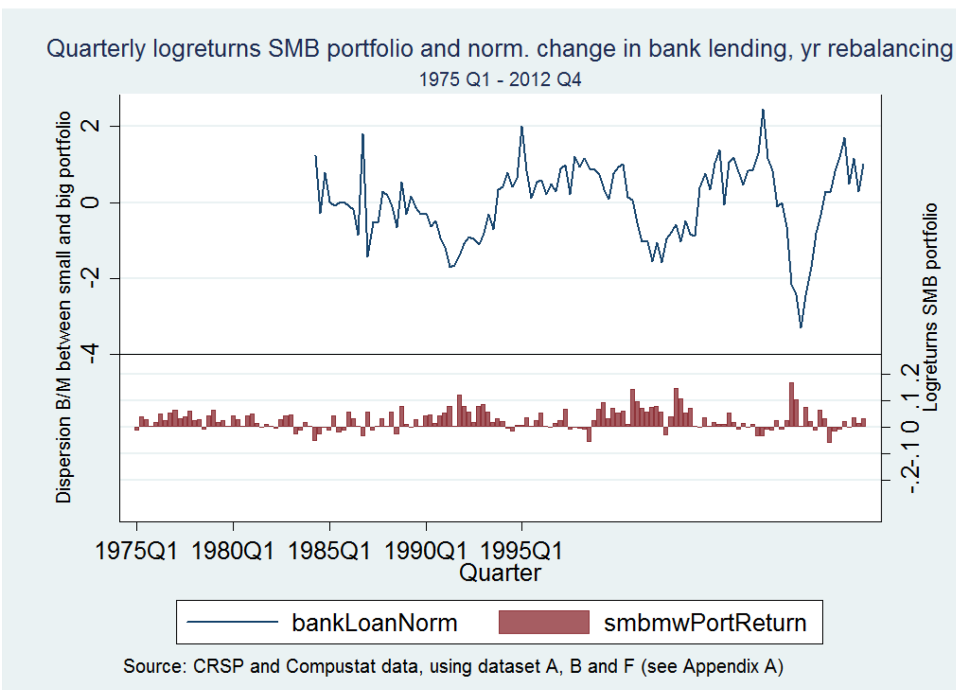



**Figure 28: Quarterly log returns SMB portfolio and bank loan availability (normalized)**

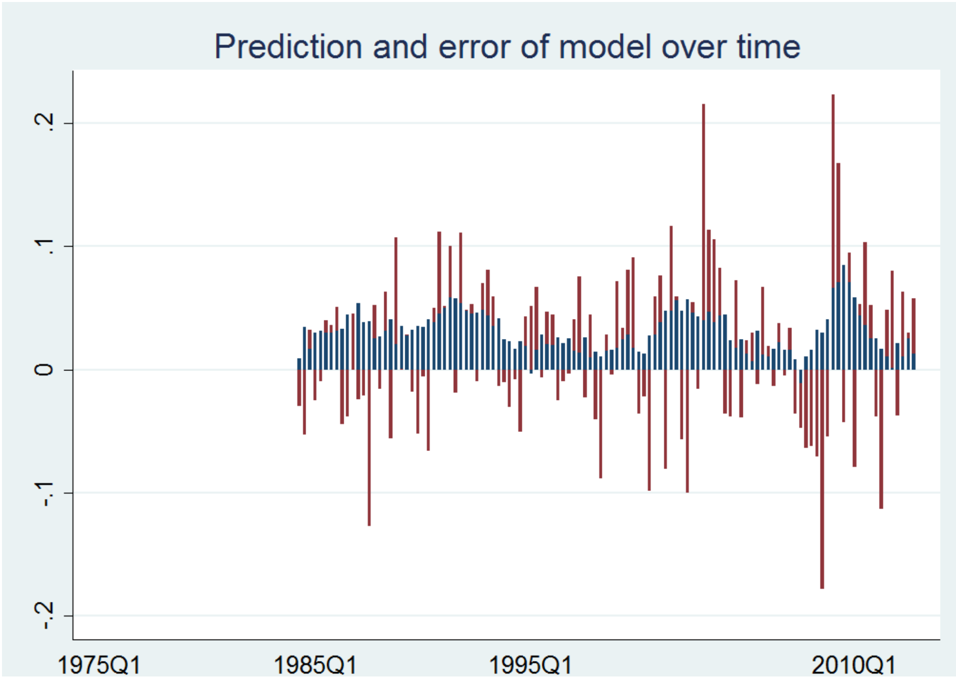



**Figure 29: Prediction and error of model with bank loan availability as predictor of SMB return on U.S. data**

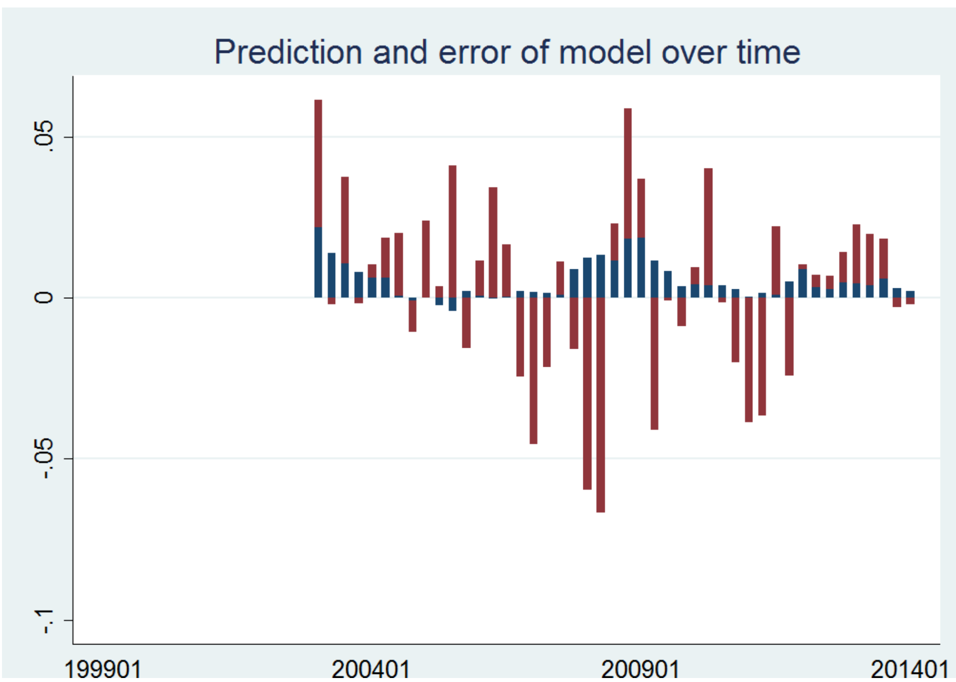



**Figure 30: Prediction and error of model with bank loan availability as predictor of SMB return on EU data**

### 5.2.4 ILLIQUIDITY TO PREDICT SMB RETURN

As described in subparagraph 3.2.4, illiquidity could indicate a high future SMB return.

*$H_0$ 10: Illiquidity (measured by the Amihud measure) is not positively related to future SMB return.*

*$H_1$ 10: Illiquidity (measured by the Amihud measure) is positively related to future SMB return.*

As Table 4 shows, illiquidity is a statistical weak predictor on the U.S. market, but strong on the EU (and it has a good economic significance), both using an illiquidity measure based upon the small part of the portfolio (25) as the illiquidity of small minus big (27). Big companies are often very liquid; therefore the total illiquidity is driven by the illiquidity of the small part. These results support the idea that SMB returns are (among others) driven by an illiquidity premium. In Table 8, the prediction result on the small part of the SMB portfolio is given based upon the Amihud measure of the portfolio part; this shows a strong (and significant) positive relationship, in line with expectations. The impact of illiquidity on the U.S. SMB portfolio as a whole is rather low, but on the small part side, prediction could be used.

As the European sample show high economical and statistical significant results and the U.S. approaches this significance, the null hypothesis can be rejected. It should however be stated that illiquidity only might explain returns of a particular (small) segment of companies, and not always for an entire SMB composition.

**Table 8: Regression result on illiquidity index (Amihud) of small caps to predict 1-quarter future small-cap part of SMB using illiquidity of small-cap part of portfolio**

| Prediction of 'small' part of SMB market-weighted return using Amihud illiquidity index of 'small' part of portfolio, U.S. data | | | | |
|---|---|---|---|---|
| | Coefficient | t-value | $R^2$ | b*s |
| cAmihudSmall | 6995 | 2.70*** | 1.6% | 0.6% |

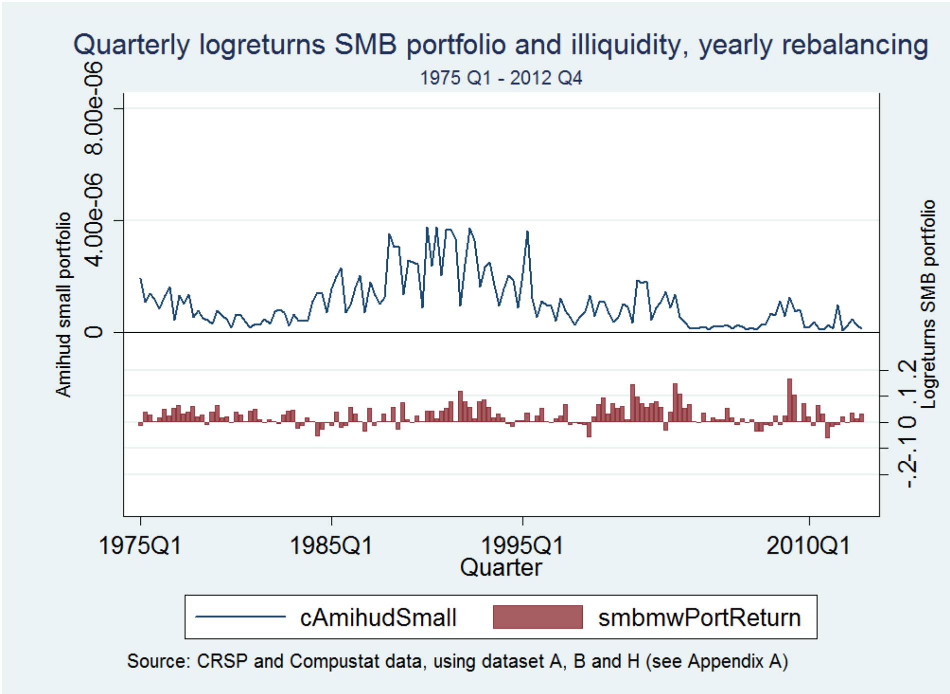



**Figure 31: Quarterly log returns SMB portfolio and Amihud illiquidity of the small portfolio part**

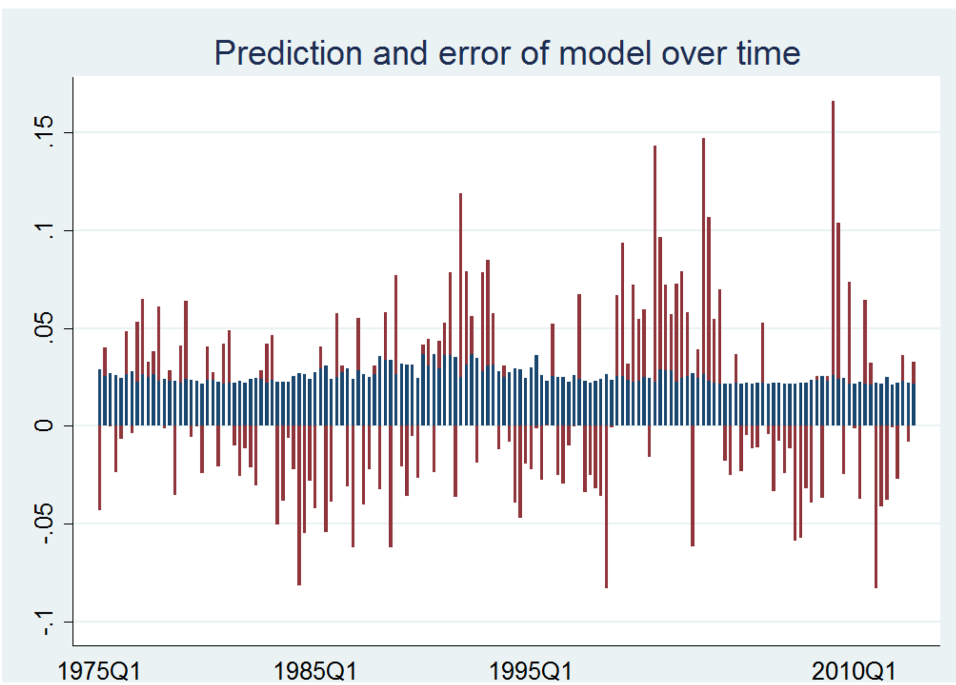



**Figure 32: Prediction and error of model with Amihud illiquidity as predictor of SMB return on U.S. data**

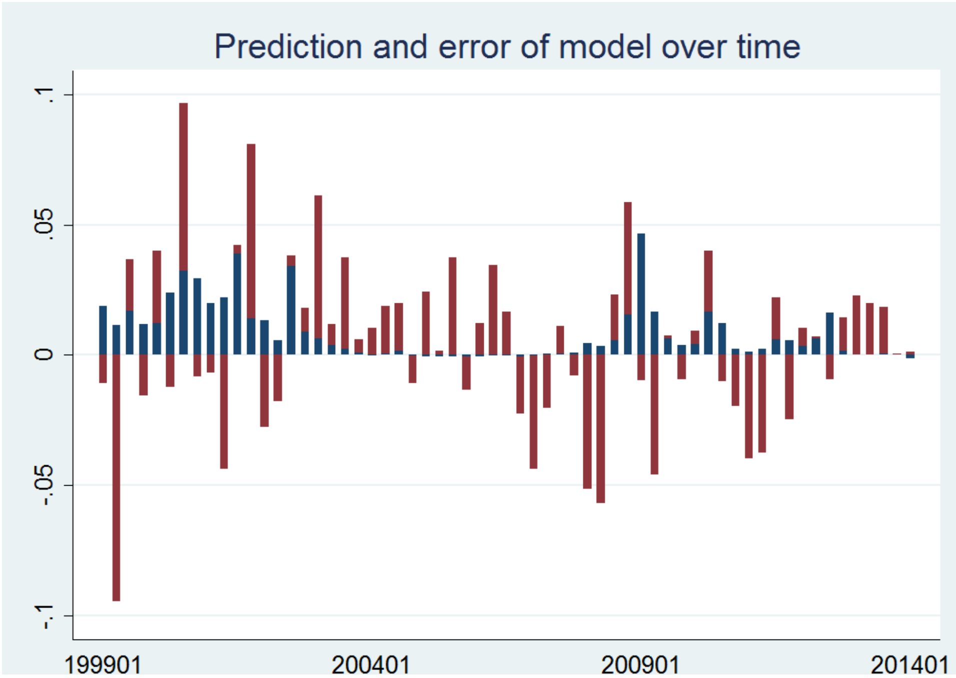



**Figure 33: Prediction and error of model with Amihud illiquidity as predictor of SMB return on EU data**

## 5.3 INTEGRAL PREDICTION OF HML AND SMB RETURNS

This paragraph provides prediction horizons for both HML and SMB portfolios, showed in Figure 34 and Figure 35 below. Confidence intervals for the $R^2$ are calculated using the two-sided 95% student-t statistics and the variance of $R^2$ as found by Wishart (Wishart, 1931), (Olkin & Finn, 1995). Next, correlations between the factors are provided as well as integral models for HML and SMB prediction at 1-quarter future horizon.

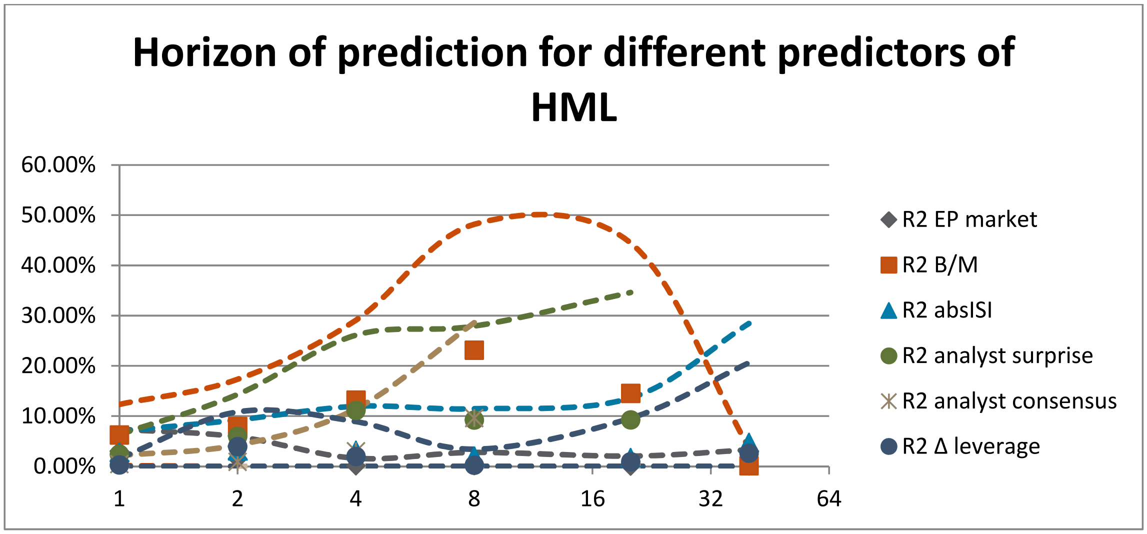



**Figure 34: Percentage of variance of HML return explained by different predictors over varying horizons, dashed lines indicate (upper) 90% confidence levels bounds of $R^2$.**

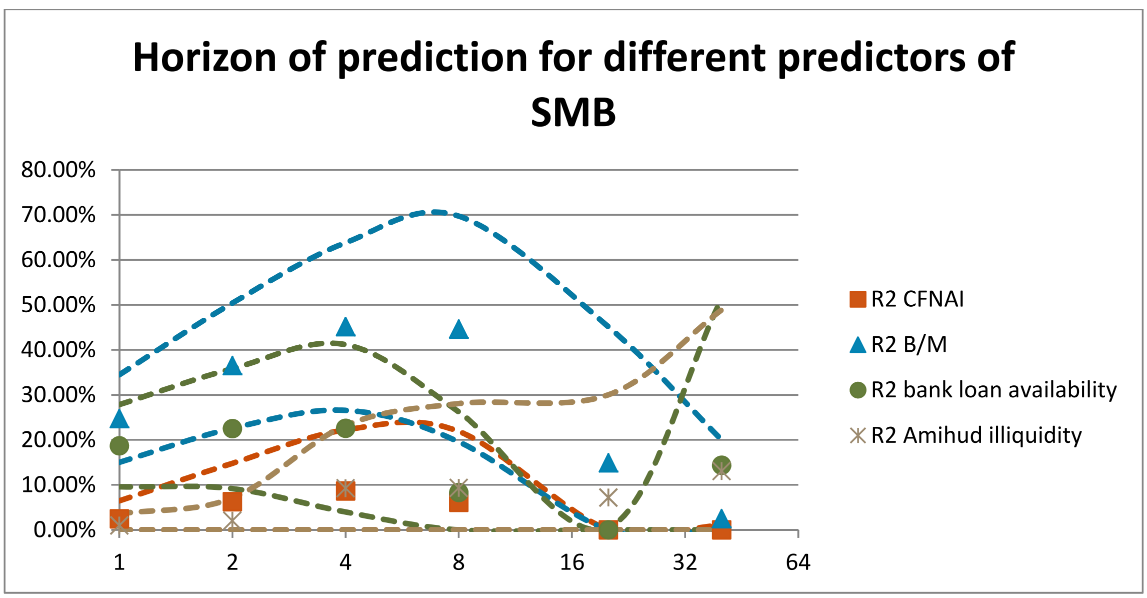



**Figure 35: Percentage of variance of SMB return explained by different predictors over varying horizons, dashed lines indicate (upper) 90% confidence levels bounds of $R^2$.**

For both HML and SMB predictors, the optimal horizon seems two years.

Table 9 and Table 10 below show the correlations between the different predictors for respectively the HML and SMB portfolio. The dispersion in book-to-market to absolute investor sentiment and analyst surprise seems relatively high, which for sentiment can be explained (extreme sentiment drives dispersion), the relation with surprise is less clear. SMB predictor correlations are weaker.

**Table 9: Correlations between predictors HML portfolio**

| | cBMdispHML | absISI | cSurpHigh |
|---|---|---|---|
| cBMdispHML | 1.0000 | | |
| absISI | 0.4466 | 1.0000 | |
| cSurpHigh | -0.6873 | -0.2431 | 1.0000 |

**Table 10: Correlations between predictors SMB portfolio**

| | cBMdispSMB | bankLoanNorm | cAmihudSmall |
|---|---|---|---|
| cBMdispSMB | 1.0000 | | |
| bankLoanNorm | -0.4573 | 1.0000 | |
| cAmihudSmall | 0.1064 | -0.3118 | 1.0000 |

Table 11 below provides the regression results for different multivariate regressions for the prediction of future HML returns.

**Table 11: Multiple regressions predicting 1- quarter future HML return**

| | HML prediction | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| BMdispHML | 0.0619 | 0.1050 | | **0.1128** |
| | 1.90* | 1.83* | | **1.26** |
| \|IS index\| | 0.0070 | | 0.0518 | **0.0286** |
| | 0.76 | | 3.93** | **1.59** |
| SurpHigh | | 0.2151 | -0.4193 | **0.4505** |
| | | 0.32 | -1.65 | **0.53** |
| constant | -0.0451 | -0.0795 | 0.0058 | **-0.0883** |
| | -1.23 | -1.46 | 0.79 | **-1.17** |
| $R^2$ | 8.1% | 11.6% | 11.5% | **19.7%** |

The best 2-variables predicting models use analyst surprise combined with either dispersion or sentiment. Regression 4 seems the best using dispersion, sentiment, and analyst surprise to predict future HML return with an $R^2$ of 19.7%.

Table 12 below provides the regression results for different multivariate regressions for the prediction of future SMB returns.

**Table 12: Multiple regressions predicting 1-quarter future HML return**

| | SMB prediction | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| **BMdispSMB** | **0.1586** | 0.1552 | | 0.1582 |
| | **4.68*** ** | 5.26*** | | 4.75*** |
| **Bank loans normalized** | **-0.0098** | | -0.0191 | -0.0102 |
| | **-3.74*** ** | | -6.95*** | -3.67*** |
| **AmihudSmall** | | 1875.59 | -1754.02 | -1003.41 |
| | | 0.64 | -0.54 | -0.24 |
| **constant** | **-0.0199** | -0.0236 | 0.0284 | -0.0186 |
| | **-2.22*** ** | -2.71*** | 3.60*** | -1.78* |
| **$R^2$** | **34.3%** | 24.9% | 16.4% | 34.3% |

Most variables are highly significant in the four regression above; bank loans seems the most valuable predictor, and also dispersion is also significant in the multivariate model, regression 1 is regarded as the best prediction model. Subsequently, the best prediction models, based upon the regressions above, are described by the following equations:

$$HMLreturn_{t+1} = 0.11 * BMDispHML + 0.03 * absISI + 0.45 * SurpHigh - 0.09 + \varepsilon$$

$$SMBreturn_{t+1} = 0.16 * BMDispSMB - 0.01 * BankLoans - 0.02 + \varepsilon$$

However, only the dispersion variables and illiquidity are variables which is attributed predictive power. In the combined model, the sign of illiquidity is on the wrong side. Therefore only book-to-market would be recommended to use in predicting HML and SMB returns, unless SMB contains much small stocks (and thus uses and alternative definition), then illiquidity is useful as well. This would lead to the following univariate models:

$$HMLreturn_{t+1} = 0.06 * BMDispHML - 0.05 + \varepsilon$$

$$SMBreturn_{t+1} = 0.16 * BMDispSMB - 0.02 + \varepsilon$$

## 5.4 ROBUSTNESS OF RESULTS

In the methodology used to test the hypotheses, specific choices have been made regarding the portfolio construction, the stock universe and the time periods selected. This paragraph shows the regression results if HML and SMB portfolios are constructed using other definitions and describes the implications of these results. First, both market- and equalweighted portfolios are tested with the Fama-French definitions; two doublesorts with respectively six and ten quantiles and the original definition with another rebalancing moment. Furthermore, I test on random subsamples of the dataset. Finally, I perform the tests on a part of the time period, to investigate whether the effects are consistent over time. Next to the tests on different definitions and composition of the portfolios, also sensitivity tests on changes on independent variables are performed using 10% and 20% winsorizing; this reduces extreme values, which could have a high impact on the results.

### 5.4.1 DIFFERENT PORTFOLIO DEFINTIONS

Table 13 and Table 14 show the robustness of the regression results when other portfolio definitions are used. The book-to-market dispersion seems sensitive for the weights within the portfolio, this could suggest that the returns very 'high' companies cannot be predicted well by the BM dispersion or that the returns are driven by very 'low' companies within the portfolio.

**Table 13: Regression results for different portfolio definitions of HML, U.S. data, using Newey-West adjusted errors**

| | HML BM dispersion | | | HML absolute ISI | |
|---|---|---|---|---|---|
| | Market-weighted | Equal-weighted | | Market-weighted | Equal-weighted |
| FF | **2.40**** | 1.14 | FF | **2.20**** | 1.36 |
| | **6.2%** | 3.4% | | **2.8%** | 1.5% |
| 6x6 | 2.22** | 2.04** | 6x6 | 1.90* | 1.24 |
| | 6.6% | 7.4% | | 2.6% | 1.2% |
| 10x10 | 2.56** | 3.09*** | 10x10 | 1.24 | 1.25 |
| | 7.2% | 10.4% | | 1.7% | 1.4% |

| | HML analyst surprise | |
|---|---|---|
| | Market-weighted | Equal-weighted |
| FF | **-1.80*** | -4.33*** |
| | **3.2%** | 8.0% |
| 6x6 | -2.79*** | -5.20*** |
| | 4.6% | 10.5% |
| 10x10 | -3.16*** | -4.86*** |
| | 7.6% | 9.3% |

Analyst' surprise regarding 'high' companies seems to be strong for all portfolios. Investor sentiment does not seem to be a robust predictor for different portfolios, espacially when weigths on the extreme buckets increase.

The SMB regression results (Table 14) show robustness for the predictor bank loan availability, but illiquidity and book-to-market dispersion seem to be not robust for other portfolio definitions. One of the explanations for this sensitivity could be a value effect within the SMB return; the market-weighted Fama-French definition could correlate strong with the market-weighted HML portfolio definitions.

**Table 14: Regression results for different portfolio definitions of SMB, U.S. data, using Newey-West adjusted errors**

| | SMB BM dispersion | | | SMB bank loan availability | |
|---|---|---|---|---|---|
| | Market-weighted | Equal-weighted | | Market-weighted | Equal-weighted |
| FF | **5.38*** ** | 1.92* | FF | **-8.05*** ** | -3.05*** |
| | **24.7%** | 4.1% | | **18.7%** | 9.2% |
| 6x6 | 1.78* | 0.24 | 6x6 | -4.72*** | -3.74*** |
| | 5.3% | 0.1% | | 10.7% | 8.3% |
| 10x10 | 1.37 | -0.64 | 10x10 | -3.07*** | -3.70*** |
| | 2.1% | 0.4% | | 6.3% | 5.1% |

| | SMB illiquidity | |
|---|---|---|
| | Market-weighted | Equal-weighted |
| FF | **1.62** | 2.20** |
| | **1.0%** | 5.5% |
| 6x6 | 0.82 | -0.82 |
| | 0.3% | 0.7% |
| 10x10 | -0.37 | -2.47** |
| | 0.0% | 4.0% |

In this thesis, I rebalance at the start of January, but rebalancing in other periods should not give other results. Therefore I test with rebalancing at the start of July. The results are shown in Table 15 below. This shows that the rebalancing moment has no major impact on the results, surprise shows though insignificance and illiquidity is significant.

**Table 15: Regression results for different portfolio rebalancing moment, using Newey-West adjusted errors.**

| | HML BM dispersion | HML absolute ISI | HML analyst surprise | SMB BM dispersion | SMB bank loan availability | SMB illiquidity |
|---|---|---|---|---|---|---|
| t-value (January) | 2.40** | 2.20** | -1.80* | 5.38*** | -8.05*** | 1.62 |
| t-value (July) | 2.22** | 2.96*** | -1.08 | 3.64*** | -5.74*** | 2.16** |

### 5.4.2 TESTS ON SUBSAMPLES

The testing on subsamples is performed on 30 samples of 1500 randomly chosen companies. This is done in two steps: first a list of unique companies in the dataset (around 21,500) is sorted randomly and then the first 1500 observations are matched with the original datafile (in order to have a file with the entire period for each company). All companies in the subsamples are unique, i.e. no standard bootstrapping procedure is used. This is done, because the interpretation of the predictor variable would make more sense (if one company is selected multiple times in one dataset, one increase the weight within the portfolio with the same multiple, but the weighting of i.e. illiquidity would not be straightforward). The number of 1500 is chosen to be able to create portfolios at each quarter (on average, there are 400 companies per quarter in the subsample, but this varies much over time). Results are provided in Table 16 below.

**Table 16: Statistical significance of regression tests on 30 subsets of U.S. data**

| | HML BM dispersion | HML absolute ISI | HML analyst surprise | SMB BM dispersion | SMB bank loan availability | SMB illiquidity |
|---|---|---|---|---|---|---|
| t-value (original) | 2.40** | 2.20** | -1.80* | 5.38*** | -8.05*** | 1.62 |
| % sign at >99% | 27% | 20% | 17% | 53% | 83% | 3% |
| % sign at >95% | 50% | 40% | 43% | 80% | 90% | 10% |
| % sign at >90% | 70% | 50% | 57% | 93% | 97% | 13% |
| % sign at <90% (insignificant) | 30% | 50% | 43% | 7% | 3% | 87% |

These results show that the investor sentiment index as well as the dispersions and bank loans are strong predictors; for investor sentiment and bank loans, this could be explained by the predictor that does not depend on the company sample. The four others are calculated based upon the set of companies, and most of these show lower significance figures, although they still are quite strong. Analyst surprise is 90%-significant for half of the samples, which still is a relatively good basis to hold on to a rejection of the hypothesis.

Illiquidity shows results that are doubtfull, since only 10% of the subsamples show significant results at 95%-level. One of the explanations at illiquidity is that illiquidity of small caps varies strongly; a subset that contains relatively few smallcaps will have a predictor that is relatively unbalanced. As a manual winsorizing has taken place on these tests, the results are insignificant here. In relation to the results of Table 14, where the Amihud measure has a higher significance for smaller size buckets, the problem of too few small caps in the subsample might be present.

### 5.4.3 DIFFERENT TIME PERIODS

The time period is split at the first quarter of 1995, resulting in the periods 1975-1994 and 1995-2012. For both periods the tests are done, results are provided in Table 17. This shows strong results for the SMB predictors dispersion and bank loan availability, but the HML predictors seem not time-window independent; only absolute investor sentiment is highly significant in both time-windows. Analyst surprise could not be tested as the I/B/E/S series do not cover the first period. HML dispersion shows lower significicance in the first period; this could be attributed to a low performance of HML over this period and little variation in dispersion, but Figure 14 does not support this.

**Table 17: Regression results for different time periods using U.S. data**

| | | HML BM dispersion | HML absolute ISI | HML analyst surprise | SMB BM dispersion | SMB bank loan availability | SMB illiquidity |
|---|---|---|---|---|---|---|---|
| Original | t-value | 2.40** | 2.20** | -1.80* | 5.38*** | -8.05*** | 1.62 |
| | $R^2$ | 6.2% | 2.8% | 3.2% | 24.7% | 18.7% | 1.0% |
| 1975-1994 | t-value | 1.99** | 1.85* | No data | 3.20*** | -5.43*** | 1.47 |
| | $R^2$ | 1.7% | 1.4% | No data | 19.7% | 29.6% | 1.5% |
| 1995-2012 | t-value | 3.00*** | 4.47*** | -1.78* | 6.11*** | -6.20*** | 1.50 |
| | $R^2$ | 11.1% | 9.6% | 3.0% | 28.8% | 18.6% | 4.1% |

### 5.4.4 WINSORIZED INDEPENDENT VARIABLES

Table 18 shows the results with winsorized independent variables; if extreme values have a major contribution to the results, they should dissapear when winsorized. Winsorizing at investor sentiment and bank loan availability is done on time-series, other variables are winsorized at the root (company level). The results at 10% winsorizing show that the predictors are still significant (except for illiquidity). Only when winsorized at 20%, signicance is reduced, for the dispersion variables this can be explained easily; when dispersion is reduced, the predictive value is restricted in one dimension, which should result in less explanetory power. The increase in power at 10% winsorizing (for SMB) could be attributed to removed extreme values that have a relatively high impact on the predictor. Illiquidity is driven by a relatively small group of fewer liquid small caps, as shown by Table 8; removing many low liquidity values removes the driver of this predictor (as is already done in the original regression), but the results found here point towards another direction; having too many extreme values that distort a possible predictor. Analyst surprise predicting is probably driven by the relatively big surprises, which are removed when winsorized.

**Table 18: Regression results with winsorized independent variables using U.S. data**

| | | HML BM dispersion | HML absolute ISI | HML analyst surprise | SMB BM dispersion | SMB bank loan availability | SMB illiquidity |
|---|---|---|---|---|---|---|---|
| Original | t-value | 2.40** | 2.20** | -1.80* | 5.38*** | -8.05*** | 1.62 |
| | $R^2$ | 6.2% | 2.8% | 3.2% | 24.7% | 18.7% | 1.0% |
| 10% Winsorized | t-value | 2.95*** | 2.00** | -2.19** | 6.54*** | -6.37*** | 1.25 |
| | $R^2$ | 5.9% | 2.8% | 4.3% | 22.7% | 17.6% | 1.1% |
| 20% Winsorized | t-value | 2.39** | 2.08** | -1.46 | 4.16*** | -5.46*** | 2.39** |
| | $R^2$ | 2.5% | 3.9% | 1.2% | 15.9% | 15.7% | 3.0% |

## 5.5 TRANSLATION TO ASSET ALLOCATION

*This paragraph is omitted in the public version of this thesis.*

# 6 DISCUSSION AND RECOMMENDATIONS

In this chapter I will reflect upon the methods used to test the hypothesis; provide recommendations for future research on this topic; and conclude with recommendations to Kempen Capital Management.

## 6.1 STATISTICAL ISSUES IN FINANCIAL MARKET CONTEXT

The sample size, used in testing the hypothesis, might be too small. As returns are studied on a quarterly basis, the number of quarters within selected the time period is not that large (around 150 quarters for U.S. data). When mean-reverting behavior over several years is expected, the time period should be much larger; this data is not (yet) available though. Active feedback on predictors (or premiums) found may also influence future returns, which worsens the sample size issue.

Next to the sample size issue, the cross-validation with European market might not be independent, as globalization led to correlated stock markets. This increases the exposure to a selection bias. A comparison to an emerging market might be better (as correlation should be less), but data availability is an issue there.

## 6.2 VALUE AND SIZE RETURNS

Several of the hypotheses have been rejected and this subsequently gives support for the underlying ideas. The strong results found for dispersion as predictor of both HML and SMB returns supports irrational investor behavior as explanation of the value effect, in line with Lakonishok et al. (1994). Furthermore, the weak support from the investor sentiment index and analyst surprise also points into this direction.

Some empirical results did not provide enough evidence to reject hypotheses. As growth companies perform better when interest rates are low; a link between rates and HML would then be logical. Furthermore, a 10-year T-bill rate might be a better benchmark than a 5y-5y swap as also short term rates influence the discount of future cash flows of those companies (but not the perpetual value of cash streams, on which the long term rate would have a stronger influence). In the early 80’s, both interest rates and value returns are high and in the past few years, interest rates and value returns are low; it does however not seem to be among the most important drivers of HML return. Next to the lack of supporting data; as rates are not well predictable, this could influence the predictive power of rates, possibly explaining why rates of t-1 do not explain return of HML at t; perhaps only, not ex-ante known, rates of period t do.

## 6.3 SUSTAINABILITY OF PREDICTORS FOUND

One of the most important questions, after the discovery of these predictors, is whether they will persist over time. This is only testable ex post, but a robustness test over multiple periods might give some support to statements about the sustainability of the predictors.

Dispersion and analyst surprise are based upon investor behavior; both loss aversion and extrapolation of historical results have influenced investor behavior in the past. Unless investors are going to invest fully rational, these effects are likely to persist. Dispersion could reveal differences in estimates or assumptions of market participants; as CAPM assumes equal assumptions of all participants; book-to-market dispersion could reveal a degree of behavioral influence in a price, creating positive alpha returns.

Bank loan availability and illiquidity are market factors. As illiquidity imposes a risk (liquidity risk), it makes sense that investors are rewarded for the risk, as the illiquidity (of small caps) is correlated between firms. As capital markets probably will develop even further, access to credit might increase, giving companies also access to credit in periods of economic downturn.

## 6.4 INDEPENDENT VARIABLES ON EUROPEAN DATA

The European independent variables are less widely available and some that are used in this thesis are constructed in less comprehensive ways than their U.S. counterparts; like the European Activity Index and bank loan data. Furthermore, some datasets have little data about European stocks compared with the U.S. ones (like I/B/E/S) and the Investor Sentiment Index has been used for both U.S. and European data, it is although constructed on U.S. data. The construction of independent variables that are valid representing the European market would give better tools for testing the hypotheses.

Finally, one could argue whether the European market can be compared well with its U.S. counterpart. Not only market access to firms is higher in the U.S., also the common monetary policy of Europe, and the different country policies can distort results. Furthermore, the lending system in the U.S. is market driven, where Europe has a bank driven lending system. These differences can have a big impact of the validity of identical predictors for both markets; which could result in insignificant validation results, where a U.S. effect may exist.

## 6.5 RECOMMENDATIONS TO KEMPEN CAPITAL MANAGEMENT

*This paragraph is omitted in the public version of this thesis.*

## REFERENCES


Abarbanell, J. S. & Bushee, B. J., 1998. Abnormal Returns to a Fundamental Analysis Strategy. *The Accounting Review,* 73(1), pp. 19-45.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets,* pp. 31-56.

Arnott, R. D., Hsu, J. & Moore, P., 2005. Fundamental Indexation. *Financial Analysts Journal,* pp. 83-99.

Arnott, R. D., Li, F. & Sherrerd, K. F., 2009. Clairvoyant Value and the Value Effect. *The Journal of Portfolio Management,* pp. 12-26.

Asness, C. S., Moskowitz, T. J. & Pedersen, L. H., 2013. Value and Momentum Everywhere. *The Journal of Finance,* pp. 929-985.

Baker, M., Bradley, B. & Wurgler, J., 2011. Benchmarks as Limits to Arbitrage: Understanding the Low Volatility Anomaly. *Financial Analysts Journal,* 67(1), pp. 40-54.

Baker, M. & Wurgler, J., 2003. Investor Sentiment and the cross-section of stock returns. *working paper NYU Stern Department of Finance,* pp. 1-34.

Banz, R. W., 1981. The relationship between returns and market value of common stocks. *Journal of Financial Economics,* pp. 3-18.

Barberis, N. & Huang, M., 2001. Mental Accounting, Loss Aversion, and Individual Stock Returns. *The Journal of Finance,* LVI(4), pp. 1247-1292.

Barron, O. E., Stanford, M. & Yu, Y., 2005. *Further Evidence on the Relation between Analysts' Forecast Dispersion and Stock Returns,* University Park: Smeal College of Business Administration.

Bartov, E., Givoly, D. & Hayn, C., 2002. The rewards to meeting or beating earnings expectations. *Journal of Accounting and Economics,* pp. 173-204.

Basu, S., 1983. The relationship between earnings' yield, market value and return for nyse common stocks: further evidence. *Journal of Financial Economics,* pp. 129-156.

Bennin, R., 1980. Error Rates in CRSP and COMPUSTAT: A Second Look. *The Journal of Finance,* XXXV(5), pp. 1267-1271.

Berk, J. B., 1997. Does Size Really Matter?. *Financial Analysts Jouranl,* 53(5), pp. 12-18.

Black, F., 1993. Beta and Return. *The Journal of Portfolio Management,* 20(1), pp. 8-18.

Black, F., Jensen, M. & Scholes, M., 1972. The Capital Asset Pricing Model: Some Empirical Tests. *Studies in the Theory of Capital Markets.*

Carhart, M. M., 1997. On Persistence in Mutual Fund Performance. *Journal of Finance,* pp. 57-82.

Damodaran, A., 2008. *Equity Risk Premiums (ERP): Determinants, Estimation and Implications,* New York: Stern School of Business.

Daniel, K. & Titman, S., 1997. Evidence on the Characteristics of Cross Sectional Variation in Stock Returns. *The Journal of Finance,* pp. 1-33.

De Jong, F. & Driessen, J., 2013. *The Norwegian Government Pension Fund's Potential for Capturing Illiquidity Premiums,* Tilburg: Netspar.

Dechow, P. M., Hutton, A. P., Meulbroek, L. & Sloan, R. G., 2001. Short-sellers, fundamental analysis and stock returns. *Journal of Financial Economics,* 61(1), pp. 77-106.

Diether, K. B., Malloy, C. J. & Scherbina, A., 2002. Differences of Opinion and the Cross Section of Stock Returns. *The Journal of Finance,* pp. 2113-2141.

Dimson, E., Marsh, P. & Staunton, M., 2008. The Worldwide Equity Premium: A Smaller Puzzle. In: *Handbook of the equity risk premium.* Amsterdam: Elsevier, pp. 467-514.

ETF Database, 2014. *ETF Database.* [Online]
Available at: http://etfdb.com/
[Accessed 17 April 2014].

Fabian, D., 2013. *Low Volatility Funds in High Demand.* [Online]
Available at: http://www.nasdaq.com/article/low-volatility-funds-in-high-demand-cm249032
[Accessed 15 April 2014].

Fama, E. F. & French, K. R., 1992. The cross-section of expected stock returns. *The Journal of Finance,* pp. 427-465.

Fama, E. F. & French, K. R., 1995. Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance,* L(1), pp. 131-155.

Fama, E. F. & French, K. R., 1998. Value versus Growth: The International Evidence. *The Journal of Finance,* pp. 1975-1999.

Fama, E. F. & French, K. R., 2004. The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives,* pp. 25-46.

Fama, E. F. & French, K. R., 2012. Size, value, and momentum in international stock returns. *Journal of Financial Economics,* pp. 457-472.

Federal Reserve Bank of Chicago, 2014. *CFNAI data series 1967-2013,* Chicago: Federal Reserve Bank of Chicago.

Ferson, W. E. & Harvey, C. R., 1999. Conditional Variables and the Cross Section of Stock Returns. *The Journal of Finance,* pp. 1325-1360.

Financial Times Lexicon, 2014. *Definition of smart beta.* [Online]
Available at: http://lexicon.ft.com/Term?term=smart-beta
[Accessed 11 June 2014].

Fu, F., 2009. Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics,* pp. 24-37.

Goyal, A., 2012. Empirical cross-sectional asset pricing: a survey. *Financial Markets and Portfolio Management,* pp. 3-38.

Goyenko, R. Y., Holden, C. W. & Trzcinka, C. A., 2009. Do Liquidity Measures Measure Liquidity?. *Journal of Financial Economics,* pp. 153-181.

Gray, W. R. & Vogel, J., 2012. Analyzing Valuation Measures: A Performance Horse Race over the Past 40 Years. *Journal of Portfolio Management,* pp. 112-121.

Haugen, R. A. & Heins, J., 1975. Risk and Rate of Return on Financial Assets. *Journal of Financial and Quantitative Analysis,* 10(5), pp. 775-784.

Haugen, R. A. & Jorion, P., 1996. The January Effect: Still There after All These Years. *Financial Analysts Journal,* pp. 27-31.

Hou, K., Andrew, K. G. & Kho, B. C., 2011. What Factors Drive Global Stock Returns?. *Review of Financial Studies,* pp. 2527-2574.

Hou, K. & Moskowitz, T. J., 2005. Market Frictions, Price Delay, and the Cross-Section of Expected Returns. *The Review of Financial Studies,* pp. 981-1020.

Hou, K. & Robinson, D. T., 2006. Industry Concentration and Average Stock Returns. *The Journal of Finance,* pp. 1927-1956.

Hsu, J. & Li, F., 2013. Low-Volatility Investing. *Index Investing,* 4(2), pp. 66-72.

Hwang, Y.-S., Min, H.-G., Kim, H. & Kim, B.-H., 2010. Using the credit spread as an option-risk factor: Size and value effects in CAPM. *Journal of Banking & Finance,* 34(12), pp. 2995-3009.

Jegadeesh, N. & Titman, S., 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance,* XLVIII(1), pp. 65-91.

Jiang, H. & Sun, Z., 2011. *Dispersion in Beliefs among Active Mutual Funds and the Cross-Section of Stock Returns,* Rotterdam: Erasmus University.

Kahneman, D. & Tversky, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica,* 47(2), pp. 263-291.

Kalesnik, V. & Beck, N., 2014. *Factor Timing,* Newport Beach: Research Affiliates.

Keim, D. B. & Stambaugh, R. F., 1986. Predicting returns in the stock and bond market. *Journal of Financial Economics,* pp. 357-390.

Kempen & Co, 2014. *DIG follow up: quantitative data & implementation,* Amsterdam: Kempen & Co.

Kempen & Co, 2014. *Value/size DIG maart,* Amsterdam: Kempen & Co.

Lakonishok, J., Shleifer, A. & Vishny, R. W., 1994. Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance,* pp. 1541-1578.

Liew, J. & Vassalou, M., 2000. Can book-to-market, size and momentum be risk factors that predict economic growth?. *Journal of Financial Economics,* pp. 221-245.

Li, X., Sullivan, R. N. & Garcia-Feijoo, L., 2014. The Limits to Arbitrage and the Low-Volatility Anomoly. *Financial Analysts Journal,* 70(1), pp. 52-63.

Lo, A. W. & MacKinlay, A. C., 1990. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies,* 3(3), pp. 431-467.

MacKinlay, A. C., 1995. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics,* 38(1), pp. 3-28.

Malkiel, B. G., 2003. The Efficient Market Hypothesis and Its Critics. *Journal of Econonomic Perspectives,* pp. 59-82.

Malkiel, B. & Jun, D., 2009. The “value” effect and the market for Chinese stocks. *Emerging Markets Review.*

Merton, R. C., 1987. A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance,* XLII(3), pp. 483-510.

Miller, R. G., 1997. *Beyond ANOVA: Basics of Applied Statistics.* 1st ed. London: Chapman & Hall.

Novy-Marx, R., 2013. The Other Side of Value: The Gross Profitability Premium. *Journal of Financial Economics,* pp. 1-28.

Olkin, I. & Finn, J. D., 1995. Correlations Redux. *Quantitative Methods in Psychology,* 118(1), pp. 155-164.

Pastor, L. & Stambaugh, R. F., 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy,* pp. 642-685.

Piotroski, J. D., 2000. Value Investing: The Use of Historical Financial Statement Information to Seperate Winners from Losers. *Journal of Accounting Research,* pp. 1-41.

Russell Investments, 2014. *Russell 2000(r) Index: 30 years of Small Cap,* Washington: Russell Investments.

Sharpe, W. F., 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance,* pp. 425-442.

Shumway, T. & Warther, V. A., 2002. The Delisting Bias in CRSP's Nasdaq Data and Its Implications for the Size Effect. *The Journal of Finance,* LIV(6), pp. 2361-2379.

Van Dijk, M. A., 2011. Is size dead? A review of the size effect in equity returns. *Journal of Banking & Finance,* pp. 3263-3274.

Weisstein, E. W., 2014. *Reversion to the Mean.* [Online]
Available at: http://mathworld.wolfram.com/ReversiontotheMean.html
[Accessed 11 July 2014].

Wishart, J., 1931. The mean and second moment coefficient of the multiple correlation coefficient in samples from a normal population. *Biometrika,* Volume 22, pp. 353-361.

Wurgler, J., 2011. *Investor Sentiment Data,* New York: NYU Stern.

Zakamulin, V., 2013. Forecasting the size premium over different time horizons. *Journal of Banking & Finance,* pp. 1061-1072.

## APPENDIX A: DESCRIPTION OF DATASETS USED

Table 19 shortly summarizes the datasets used, which are explained in detail on the following pages.

**Table 19: Summary of datasets**

| Dataset code | Content |
|---|---|
| **A** | CRSP data on U.S. stock delistings and exchanges |
| **B** | CRSP data on U.S. stock prices and fundamentals |
| **C** | FactSet data on EU stock prices and fundamentals |
| **D** | Investor sentiment index series of Jeffrey Wurgler (2011) |
| **E** | Bloomberg data on risk-free rate US |
| **F** | FDIC balance sheet data on U.S. bank loans to corporations |
| **G** | CFNAI data series |
| **H** | CRSP data on daily stock returns and trade volumes |
| **I** | I/B/E/S data on analyst consensus and surprise |
| **J** | Bloomberg data on risk-free rate EU |
| **K** | Eurostat data on leading economic indicators of EU |
| **L** | FactSet data on daily stock returns and trade volumes |
| **M** | ECB Lending Survey Statistics data on bank loans EU |
| **N** | Bloomberg data to link ISIN to U.S. tickers (to link ‘C’ to ‘I’) |

### DATASET A: CRSP DATA ON U.S. STOCK DELISTINGS AND EXCHANGES

Dataset A contains delisting and exchange data of U.S. firms listed on AMEX, NYSE or NASDAQ. The date range used is based upon the data range that is effectively available on dataset B (which is 1970 as starting year).

**Table 20: WRDS data request on dataset A**

| | |
|---|---|
| Libraries/Data Sets | crspa/msf / |
| Frequency/Date Range | mon / Jan 1969 - Dec 2013 |
| Search Variable | CUSIP |
| Input Codes<br>all item(s) | `-all-` |
| Conditional Statements | n/a |
| Output format/Compression | dta / |
| Variables Selected | CUSIP NCUSIP COMNAM TICKER PERMCO SHRCD SHRCLS ISSUNO EXCHCD HEXCD SHROUT SHRENDDT SHRFLG DLSTCD NWPERM NEXTDT DLAMT DLPRC DLPDT DLRET DLRETX |
| Extra Variables and Parameters Selected | |

## DATASET B: CRSP DATA ON U.S. STOCK PRICES AND FUNDAMENTALS

Dataset B contains stock price data and company fundamentals of U.S. firms listed on AMEX, NYSE or NASDAQ. The entire date range available for this database is used (1961-2013).

**Table 21: WRDS data request on dataset B**

| | |
|---|---|
| Libraries/Data Sets | crspa/ccmfundq / |
| Frequency/Date Range | qtr / Jan 1961 - Dec 2013 |
| Search Variable | CUSIP |
| Input Codes<br>all item(s) | -all- |
| Conditional Statements | n/a |
| Output format/Compression | dta / |
| Variables Selected | CONSOL INDFMT DATAFMT POPSRC DATAFQTR DATACQTR CURCDQ COSTAT CONM TIC CUSIP CIK EXCHG FYR FIC DATAFQTR ATQ CEQQ CSH12Q CSHFDQ CSHOQ CSHPRQ LCTQ LLTQ TEQQ CIY EPSFIY EPSFXY EPSPIY EPSPXY NIY OEPSXY OIADPY OIBDPY REVTY SALEY DVPSXQ MKVALTQ PRCCQ |
| Extra Variables and Parameters Selected | C INDL STD |

Sources: CRSP, Center for Research in Security Prices. Graduate School of Business, The University of Chicago 1960-2013. Used with permission. All rights reserved. www.crsp.uchicago.edu. Compustat database. Standard & Poor's (2014). Retrieved 1st of May, 2014.

## DATASET C: FACTSET DATA ON EU STOCK PRICES AND FUNDAMENTALS

Dataset C contains stock price data and company fundamentals of European firms within the MSCI universe. The entire date range available for this database is used (1998-2013).

| | |
|---|---|
| Variables used: | Ticker<br>Company name<br>Quarterly stock return<br>Market Capitalization<br>Price-to-Book ratio |

Source: FactSet. Copyright 2014 FactSet Research Systems Inc. All rights reserved. Retrieved 3rd of June, 2014.

## DATASET D: INVESTOR SENTIMENT INDEX SERIES

Dataset D contains the Investor Sentiment data for the U.S. as defined by Baker and Wurgler (2003), obtained from NYU Stern (Wurgler, 2011). The measure is based upon the principal first component of the number and first-day returns on IPO's; closed-end fund discount; equity share in new issues; dividend premium and NYSE share turnover.

Source: Investor Sentiment Data of Jeffrey Wurgler, 2011., New York: NYU Stern. Retrieved 9th of May, 2014.

## DATASET E: BLOOMBERG DATA ON RISK-FREE RATE US

Dataset E contains the 5-year forward libor interest rate swap, starting five year from now (a so-called 5yr/5yr USD interest swap). Date range from 1995 until 2013 is used (the entire period that was available in Bloomberg).

Source: Bloomberg LP, 1995-2013. Retrieved 19th of May, 2014.

## DATASET F: FDIC BALANCE SHEET DATA ON BANK LOANS TO CORPORATIONS US

Dataset F contains bank loan data of U.S. banks, obtained from the FDIC. The amount of total bank loans to corporations is used as proxy of availability of bank loans.

Source: Federal Deposit Insurance Corporation, 1984-2012. Retrieved 17th June, 2014.

## DATASET G: CFNAI DATA SERIES

Dataset G contains the Chicago Fed National Activity Index, a leading indicator for the U.S. economy. Date range from 1975 until 2013 is used (the entire period of dataset B).

Source: Federal Reserve Bank of Chicago . Retrieved 20th May, 2014.
http://chicagofed.org/digital_assets/others/research/data/cfnai/cfnai_data_series.xlsx

## DATASET H: CRSP DATA ON DAILY STOCK RETURNS AND TRADE VOLUMES

Dataset H contains stock return and trade volume data price data of U.S. firms listed on AMEX, NYSE or NASDAQ. The date range selected is in line with the effective date range of dataset C (1975-2013).

**Table 22: WRDS data request on dataset H**

| | |
|---|---|
| Data Request ID | 33d3931c72e06f51 |
| Libraries/Data Sets | crspa/dsf / |
| Frequency/Date Range | day / 01Jan1970 - 31Dec2013 |
| Search Variable | CUSIP |
| Input Codes<br>all item(s) | -all- |
| Conditional Statements | n/a |
| Output format/Compression | dta / |
| Variables Selected | CUSIP NCUSIP COMNAM TICKER PERMCO PRC VOL RETX |
| Extra Variables and Parameters Selected | |

Source: CRSP, Center for Research in Security Prices. Graduate School of Business, The University of Chicago 1960-2013. Used with permission. All rights reserved. www.crsp.uchicago.edu. Retrieved 4th June, 2014.

## DATASET I: I/B/E/S DATA ON ANALYST CONSENSUS AND SURPRISE

Dataset I contain analyst estimate data obtained from I/B/E/S (International Brokers' Estimates System). The full date range available is selected for dataset I (1992-2014).

**Table 23: WRDS data requests on dataset I, the first containing consensus, the second analyst surprise**

| | |
|---|---|
| Data Request ID | 85411d6e6aa9d0d4 |
| Libraries/Data Sets | ibes/recdsum / |
| Frequency/Date Range | / 19Nov1993 - 21Feb2014 |
| Search Variable | OFTIC |
| Input Codes<br>all item(s) | -all- |
| Conditional Statements | n/a |
| Output format/Compression | dta / zip |
| Variables Selected | OFTIC TICKER CUSIP CNAME BUYPCT HOLDPCT SELLPCT MEANREC MEDREC STDEV NUMUP NUMDOWN NUMREC USFIRM |
| Extra Variables and Parameters Selected | |
| | |
| Data Request ID | 0237497cdadb8b02 |
| Libraries/Data Sets | ibes/surpsum / |
| Frequency/Date Range | / 22Apr1992 - 21Feb2014 |
| Search Variable | OFTIC |
| Input Codes<br>all item(s) | -all- |
| Conditional Statements | n/a |
| Output format/Compression | dta / zip |
| Variables Selected | OFTIC TICKER PYEAR PMON ACTUAL ANNDATS SUESCORE SURPMEAN SURPSTDEV USFIRM |
| Extra Variables and Parameters Selected | |

Source: International Brokers' Estimates System via Thomson Reuters, 2014. Retrieved 6th June, 2014.

## DATASET J: BLOOMBERG DATA ON RISK-FREE RATE EU

Dataset E contains the 5-year forward interest rate swap, starting five year from now (a so-called 5yr/5yr EUR interest swap). Date range from 1999 until 2014 is used (the entire period that was available in Bloomberg).

Source: Bloomberg LP, 1995-2013. Retrieved 19th May, 2014.

## DATASET K: EUROSTAT DATA ON LEADING ECONOMIC INDICATORS OF EU

Dataset K contains an index that is constructed using macro-economic indicators based on Eurostat data.

| | |
|---|---|
| Date: | 28-05-2014 |
| Series: | Unemployment rate<br>Total employment (resident population concept - LFS)<br>Employment (15 to 64 years)<br>Employment rates<br>Active population<br>Activity rates<br>Youth unemployment ratio 15-24<br>Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply<br>Manufacturing series<br>Electricity, gas, steam and air conditioning supply<br>Mining of metal ores<br>Net primary income transfers with the rest of the world<br>Net national income at market prices<br>Intention to buy a car within the next 12 months<br>Purchase or build a home within the next 12 months<br>Home improvements over the next 12 months<br>Assessment of current production capacity<br>Duration of production assured by current order-books<br>New orders in recent months<br>Export expectations for the months ahead<br>Current level of capacity utilization (%) |
| Country: | 15 EU countries |

Source: Eurostat database, 2014. Last update on 28-05-2014. Retrieved 12th June, 2014.

The construction of the European Activity Index is provided in chapter 4.

### DATASET L: FACTSET DATA ON DAILY STOCK RETURNS AND TRADE VOLUMES

Dataset L contains stock return and trade volume data price data of European firms within the MSCI universe. The date range selected is in line with the effective date range of dataset C (1998-2013).

Source: FactSet. Copyright 2014 FactSet Research Systems Inc. All rights reserved. Retrieved 28th June, 2014.

### DATASET M: ECB LENDING SURVEY STATISTICS DATA ON BANK LOANS EU

Dataset M contains data about the money supply of European Banks, inquired via the ECB lending survey. The full available data range has been used (2003-2014).

Source: Eurostat, European Lending Survey, 2003-2014. Retrieved 17th June, 2014.

### DATASET N: BLOOMBERG DATA TO LINK ISIN TO SEDOL

In a first instance I tried to connect tickers of Bloomberg to I/B/E/S official tickers, but as Bloomberg uses local tickers and I/B/E/S mainly U.S. tickers, this did not work out. I found out that linking using SEDOL worked better, and Bloomberg has been used to link ISINs to SEDOL codes. Finally this procedure led to a dataset that linked the core identifiers of I/B/E/S to the one of FactSet (the I/B/E/S-ticker to the FactSet code).

Source: Bloomberg LP, 1995-2013. Retrieved 26th June, 2014.

## APPENDIX B: RESULTS MISSING VALUE ANALYSIS

**Table 24: Univariate statistics all scale variables, entire dataset**

**Univariate Statistics**

| | N | Mean | Std. Deviation | Missing | | No. of Extremes[a] | |
|---|---|---|---|---|---|---|---|
| | | | | Count | Percent | Low | High |
| Date | 997117 | 19935214,01 | 117211,289 | 0 | ,0 | 39030 | 0 |
| Assets | 832104 | 4728,4810 | 52179,22400 | 165013 | 16,5 | 0 | 5199 |
| CEQQ | 855059 | 802,5441 | 4872,38663 | 142058 | 14,2 | 32 | 11419 |
| #shares (to calc 12m moving EPS) | 922437 | 61,4978 | 308,49401 | 74680 | 7,5 | 0 | 13005 |
| #shares (to calc dil. EPS) | 391167 | 115,2596 | 468,93941 | 605950 | 60,8 | 0 | 7325 |
| CSHOQ | 916173 | 57,3590 | 299,09450 | 80944 | 8,1 | 0 | 11539 |
| #shares (to calc b. EPS) | 928290 | 62,0142 | 313,92319 | 68827 | 6,9 | 0 | 12987 |
| Current liabilities | 691035 | 445,0941 | 2879,27139 | 306082 | 30,7 | 0 | 9139 |
| LT liabilities | 631083 | 620,3954 | 3510,45267 | 366034 | 36,7 | 0 | 11069 |
| BV equity | 91073 | 2853,9381 | 12327,96686 | 906044 | 90,9 | 2 | 1963 |
| Compr. Inc. | 79208 | 191,1723 | 1252,94333 | 917909 | 92,1 | 121 | 1314 |
| EPSPIY | 834309 | ,8050 | 41,07769 | 162808 | 16,3 | 27 | 211 |
| EPS dil. ex | 833357 | ,7859 | 33,63458 | 163760 | 16,4 | 34 | 198 |
| EPS b. inc. | 854314 | ,7816 | 33,29912 | 142803 | 14,3 | 40 | 243 |
| EPS b. inc. | 854630 | ,7759 | 33,21903 | 142487 | 14,3 | 34 | 206 |
| Net Income | 588530 | 73,9013 | 757,94414 | 408587 | 41,0 | 874 | 6720 |
| EPS dil. from operations | 365567 | 1,0705 | 53,00549 | 631550 | 63,3 | 14 | 116 |
| OI after dep. | 777780 | 136,5196 | 1069,46461 | 219337 | 22,0 | 172 | 8809 |
| OI before dep. | 426709 | 264,1732 | 1695,01036 | 570408 | 57,2 | 48 | 5951 |
| Revenue | 535228 | 1253,8258 | 7677,33004 | 461889 | 46,3 | 1 | 7560 |
| Sales/Turnover | 587553 | 1207,5001 | 7472,56665 | 409564 | 41,1 | 1 | 8304 |
| Div per share | 984748 | ,1065 | ,53802 | 12369 | 1,2 | 0 | 3904 |
| MCAP | 141923 | 3613,4003 | 15709,87001 | 855194 | 85,8 | 0 | 2692 |
| PRCCQ | 980399 | 27,6311 | 878,59997 | 16718 | 1,7 | 0 | 168 |
| cMCAP | 900932 | 1710,6748 | 10248,83391 | 96185 | 9,6 | 0 | 13002 |
| cPB | 835240 | 38,9801 | 31478,91248 | 161877 | 16,2 | 0 | 4 |

a. Number of cases outside the range (Mean - 2*SD, Mean + 2*SD).

**Table 25: EM Correlations between all scale variables in dataset downloaded in 5% random sample**

EM Correlations[a,b]

| | var2 | var15 | CEQQ | var17 | var18 | CSHOQ | var20 | var21 | var22 | var23 | var24 | EPSPIY | var26 | var27 | var28 | var29 | var30 | var31 | var32 | var33 | var34 | var38 | var39 | PRCCQ | cMCAP | cPB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| var2 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | |
| var15 | ,071 | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| CEQQ | ,126 | ,555 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| var17 | ,162 | ,435 | ,708 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| var18 | ,162 | ,438 | ,693 | ,991 | 1 | | | | | | | | | | | | | | | | | | | | | |
| CSHOQ | ,163 | ,445 | ,703 | ,995 | ,993 | 1 | | | | | | | | | | | | | | | | | | | | |
| var20 | ,162 | ,442 | ,701 | ,996 | ,994 | ,998 | 1 | | | | | | | | | | | | | | | | | | | |
| var21 | ,067 | ,996 | ,522 | ,419 | ,421 | ,427 | ,425 | 1 | | | | | | | | | | | | | | | | | | |
| var22 | ,067 | ,997 | ,499 | ,402 | ,407 | ,413 | ,411 | ,993 | 1 | | | | | | | | | | | | | | | | | |
| var23 | ,038 | ,537 | ,852 | ,704 | ,689 | ,700 | ,698 | ,524 | ,498 | 1 | | | | | | | | | | | | | | | | |
| var24 | ,070 | ,176 | ,484 | ,195 | ,173 | ,167 | ,171 | ,150 | ,137 | ,221 | 1 | | | | | | | | | | | | | | | |
| EPSPIY | ,004 | ,047 | ,229 | ,000 | -,001 | -,001 | -,001 | ,007 | ,016 | -,144 | ,533 | 1 | | | | | | | | | | | | | | |
| var26 | ,005 | ,047 | ,228 | ,000 | -,001 | -,001 | -,001 | ,007 | ,016 | -,144 | ,532 | 1,000 | 1 | | | | | | | | | | | | | |
| var27 | ,004 | ,047 | ,229 | ,000 | -,001 | -,001 | -,001 | ,007 | ,016 | -,144 | ,533 | 1,000 | 1,000 | 1 | | | | | | | | | | | | |
| var28 | ,005 | ,047 | ,229 | ,000 | -,001 | -,001 | -,001 | ,007 | ,016 | -,144 | ,532 | 1,000 | 1,000 | 1,000 | 1 | | | | | | | | | | | |
| var29 | ,049 | ,115 | ,461 | ,240 | ,209 | ,206 | ,208 | ,102 | ,083 | ,335 | ,785 | ,075 | ,074 | ,075 | ,074 | 1 | | | | | | | | | | |
| var30 | ,004 | ,047 | ,229 | ,000 | ,000 | ,000 | ,000 | ,007 | ,016 | -,143 | ,526 | ,999 | ,998 | ,998 | ,998 | ,065 | 1 | | | | | | | | | |
| var31 | ,090 | ,498 | ,669 | ,497 | ,492 | ,495 | ,496 | ,476 | ,472 | ,592 | ,520 | ,088 | ,088 | ,089 | ,089 | ,599 | ,089 | 1 | | | | | | | | |
| var32 | ,099 | ,454 | ,717 | ,554 | ,545 | ,550 | ,551 | ,432 | ,425 | ,670 | ,490 | ,060 | ,060 | ,060 | ,060 | ,561 | ,060 | ,951 | 1 | | | | | | | |
| var33 | ,111 | ,346 | ,642 | ,493 | ,482 | ,488 | ,489 | ,347 | ,309 | ,601 | ,380 | ,097 | ,097 | ,097 | ,097 | ,402 | ,097 | ,663 | ,751 | 1 | | | | | | |
| var34 | ,111 | ,346 | ,642 | ,493 | ,482 | ,488 | ,489 | ,347 | ,309 | ,601 | ,380 | ,097 | ,097 | ,097 | ,097 | ,402 | ,097 | ,663 | ,751 | 1,000 | 1 | | | | | |
| var38 | -,041 | ,017 | ,042 | ,020 | ,019 | ,020 | ,020 | ,015 | ,017 | ,046 | ,004 | ,005 | ,005 | ,005 | ,005 | ,008 | ,003 | ,057 | ,053 | ,042 | ,042 | 1 | | | | |
| var39 | ,077 | ,335 | ,775 | ,757 | ,743 | ,749 | ,749 | ,318 | ,287 | ,699 | ,455 | ,170 | ,170 | ,170 | ,170 | ,435 | ,170 | ,638 | ,674 | ,606 | ,606 | ,044 | 1 | | | |
| PRCCQ | ,012 | ,049 | ,239 | ,000 | -,001 | -,001 | -,001 | ,007 | ,016 | -,176 | ,528 | ,949 | ,949 | ,948 | ,949 | ,058 | ,949 | ,081 | ,047 | ,094 | ,094 | ,001 | ,177 | 1 | | |
| cMCAP | ,135 | ,337 | ,778 | ,762 | ,748 | ,754 | ,754 | ,320 | ,289 | ,697 | ,456 | ,169 | ,169 | ,169 | ,169 | ,435 | ,169 | ,639 | ,675 | ,608 | ,608 | ,040 | ,998 | ,176 | 1 | |
| cPB | ,004 | -,001 | -,001 | -,001 | -,001 | -,001 | -,001 | -,001 | -,001 | -,011 | -,002 | ,000 | ,000 | ,000 | ,000 | ,000 | ,000 | -,001 | -,001 | -,001 | -,001 | ,005 | -,002 | ,000 | -,001 | 1 |

a. Little's MCAR test: Chi-Square = 42487,589, DF = 8013, Sig. = ,000

b. The EM algorithm failed to converge in 500 iterations.

**Table 26: EM Correlations between used variables in model in 5% random sample**

**EM Correlations[a]**

| | cMCAP | cPB | PRCCQ | EPSPIY | CEQQ | CSHOQ |
|---|---|---|---|---|---|---|
| cMCAP | 1 | | | | | |
| cPB | -,001 | 1 | | | | |
| PRCCQ | ,176 | ,000 | 1 | | | |
| EPSPIY | ,169 | ,000 | ,948 | 1 | | |
| CEQQ | ,783 | -,001 | ,242 | ,232 | 1 | |
| CSHOQ | ,754 | -,001 | -,001 | ,000 | ,700 | 1 |

a. Little's MCAR test: Chi-Square = 350,320, DF = 37, **Sig. = ,000**

# APPENDIX C: SUMMARY OF CODE

## C.1 SPSS SYNTAX TO PERFORM MISSING VALUE ANALYSIS

```
DATASET ACTIVATE  DataSet3.
DATASET COPY  DataSubSet.
DATASET ACTIVATE  DataSubSet.
FILTER OFF.
USE ALL.
SAMPLE  .05.
EXECUTE.
DATASET ACTIVATE  DataSet3.

SET WORKSPACE = 506349.

DATASET ACTIVATE DataSubSet.
MVA VARIABLES=var2 var15 CEQQ var17 var18 CSHOQ var20 var21 var22 var23 var24 EPSPIY
var26 var27
    var28 var29 var30 var31 var32 var33 var34 var38 var39 PRCCQ cMCAP cPB
  /EM(TOLERANCE=0.001 CONVERGENCE=0.0001 ITERATIONS=500).

DATASET ACTIVATE DataSubSet.
MVA VARIABLES=cMCAP cPB PRCCQ EPSPIY CEQQ CSHOQ
  /MAXCAT=25
  /EM(TOLERANCE=0.001 CONVERGENCE=0.0001 ITERATIONS=500).
```

## C.2 STATA SYNTAX TO PREPROCES DATA

Stata file “Data preprocessing” contains code for the merging of datasets for both the U.S. and European data. File merge procedures are described in Figure 9 and Figure 11 in paragraph 4.1. Basically, this code file uses 82.3 and 17.5 million observations to create datasets of 737 and 73 thousand observations. One of the major issues in the preprocessing was sorting, matching and memory handling (Stata simply cannot handle 80 million observations with 1.5 GB of memory). It uses a support file (Support.do) for a part of the data compression required on illiquidity data.

The output consist of two preprocessed data files; quarterly observations per company, sometimes multiple observations per company per quarter when multiple analyst recommendations were given in that quarter.

## C.3 STATA SYNTAX TO PREPARE DATA

Stata file “Data preparation” contains code for the description of input variables, testing for extreme values and the generation of portfolios. Portfolio generation is a rather time-intensive task, due to the large number of observations; the “xtile” command takes some time. Therefore this code is split from the variable generation part (which uses this code, but is changed more during the project than this file).

In the output, the quarterly portfolio returns and portfolio composition are added.

## C.4 STATA SYNTAX TO GENERATE VARIABLES

Stata file “Variable generation” contains code for the generation of independent variables that are used in the regressions. Variable construction methods are provided in paragraph 4.3 in Table 3.

The output consists of two data files with all 27 independent variable series and the portfolio returns per quarter, individual company records still are present in the dataset.

## C.5 STATA SYNTAX TO GENERATE RESULTS

Stata file “Generating results” contains code for the regressions and manages output (saves tables and figures in the appropriate files).

The input datasets are filtered from individual company records, leaving one observation per quarter and all regression tests, tables and figures are produced (and logged) in this code file. It heavily relies on a support file (Support.do), in which standardized Newey-West regression procedures are programmed (together with relevant output statistics, figures and tables).

## C.6 OTHER STATA SYNTAX

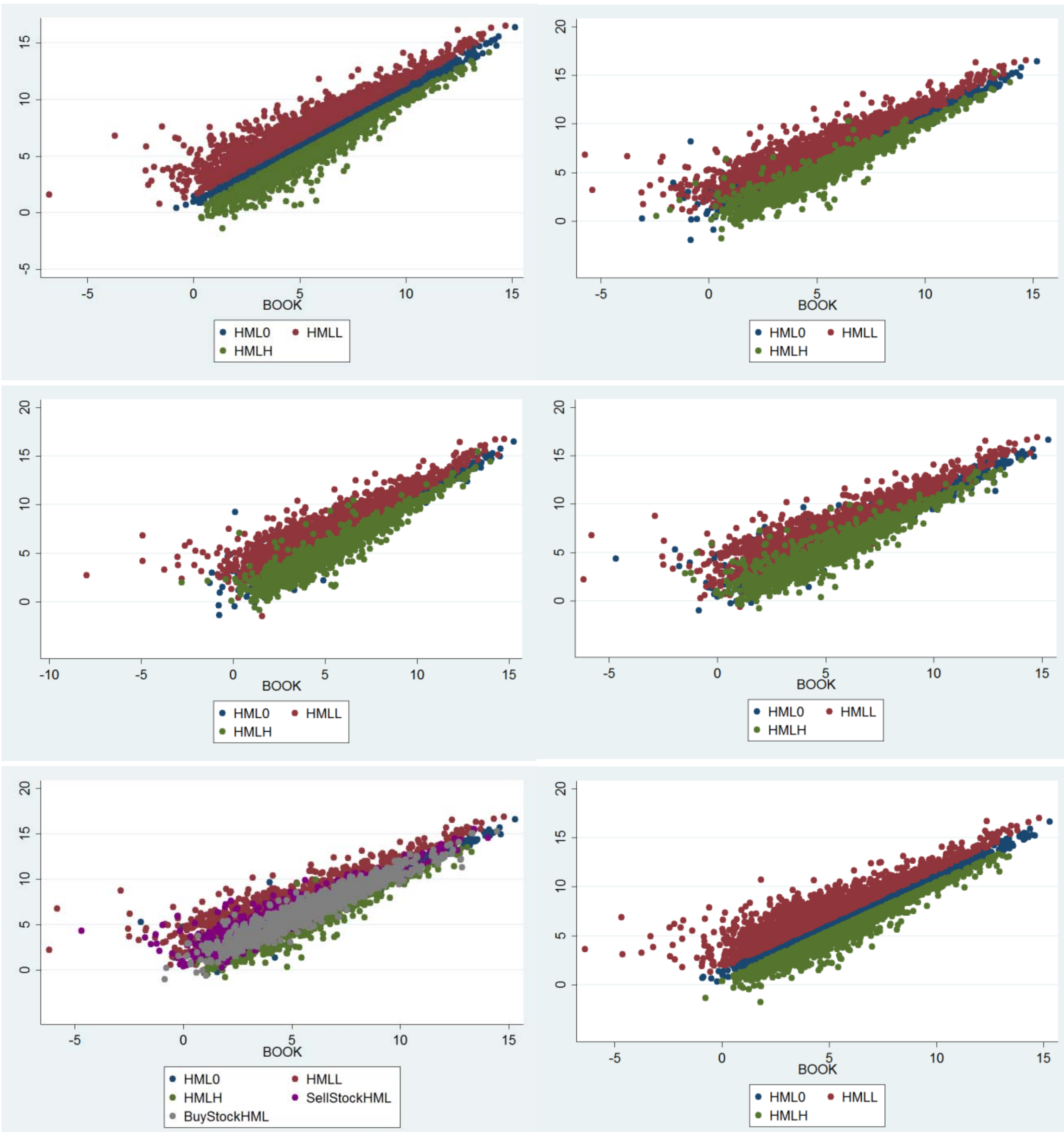
Next to the four most important code files described above, some adapted versions are used for the creation of different portfolios and winsorized variables. “Full execution.do” executes the different code files in the right order.

## APPENDIX D: DATA DESCRIPTION

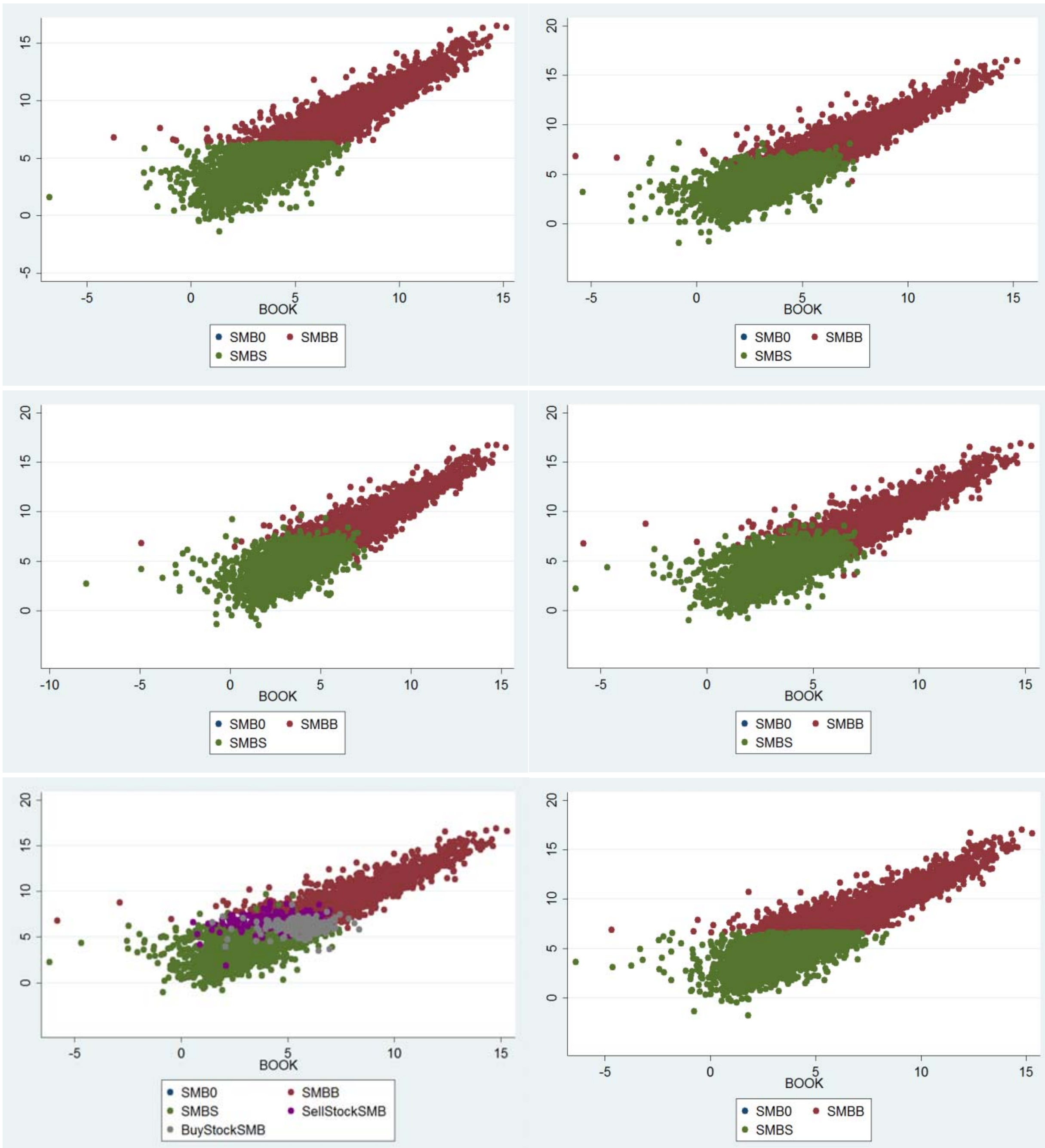
This appendix shows various plots to extend the data description in chapter 4.

### D.1 DESCRIPTION OF HML AND SMB IN STOCK UNIVERSE

The following six scatterplots indicate the composition of an HML portfolio. On the y-axis the market capitalization is given on a 2-log scale in millions; on the x-axis the book value is given (also 2-log scale in millions). Red indicates short positions, green long positions, blue are neutral positions (companies not in the portfolio). Represented are respectively the four quarters of 1995, the rebalancing at the end of the 1995 and the first quarter of 1996.

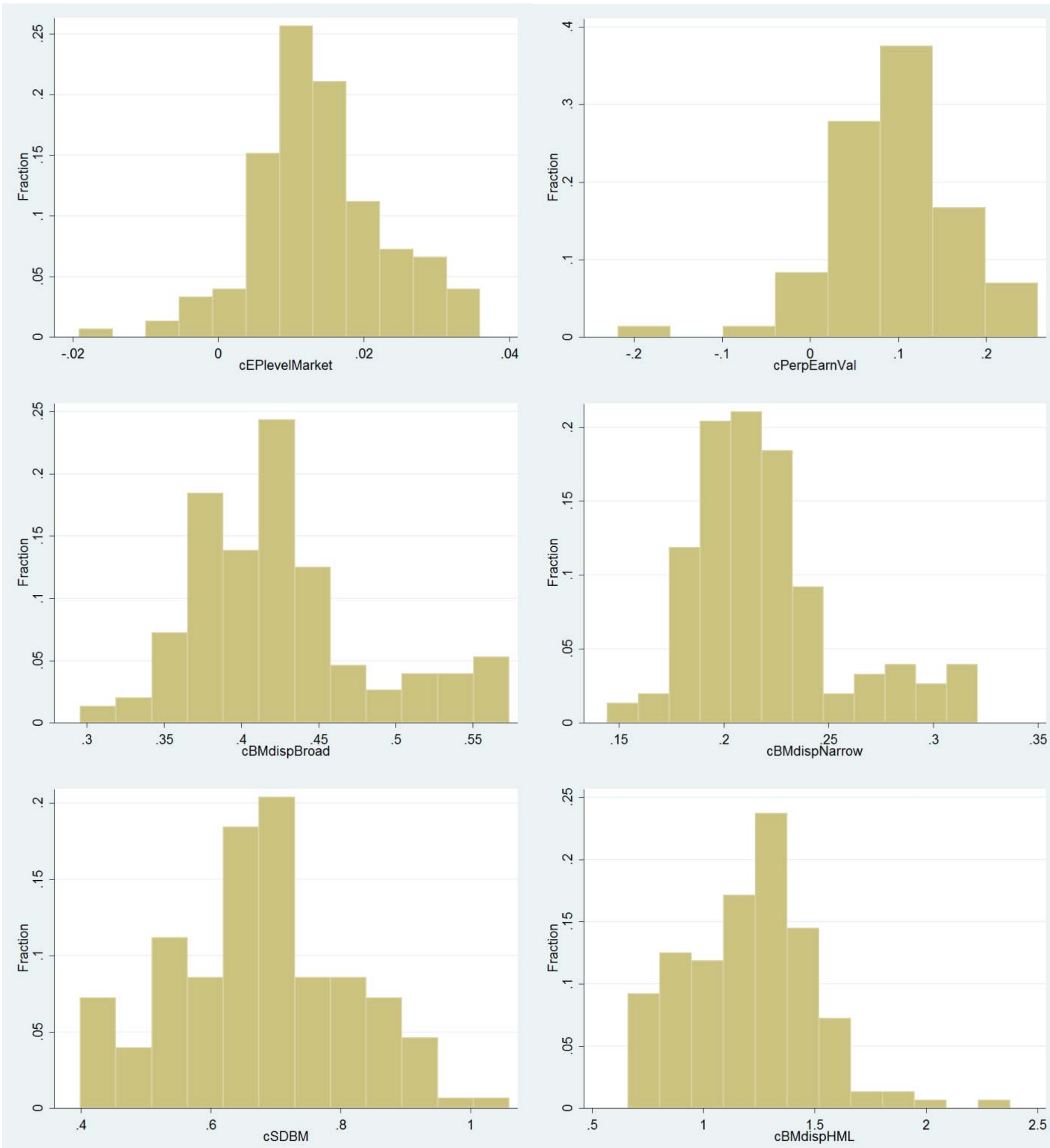


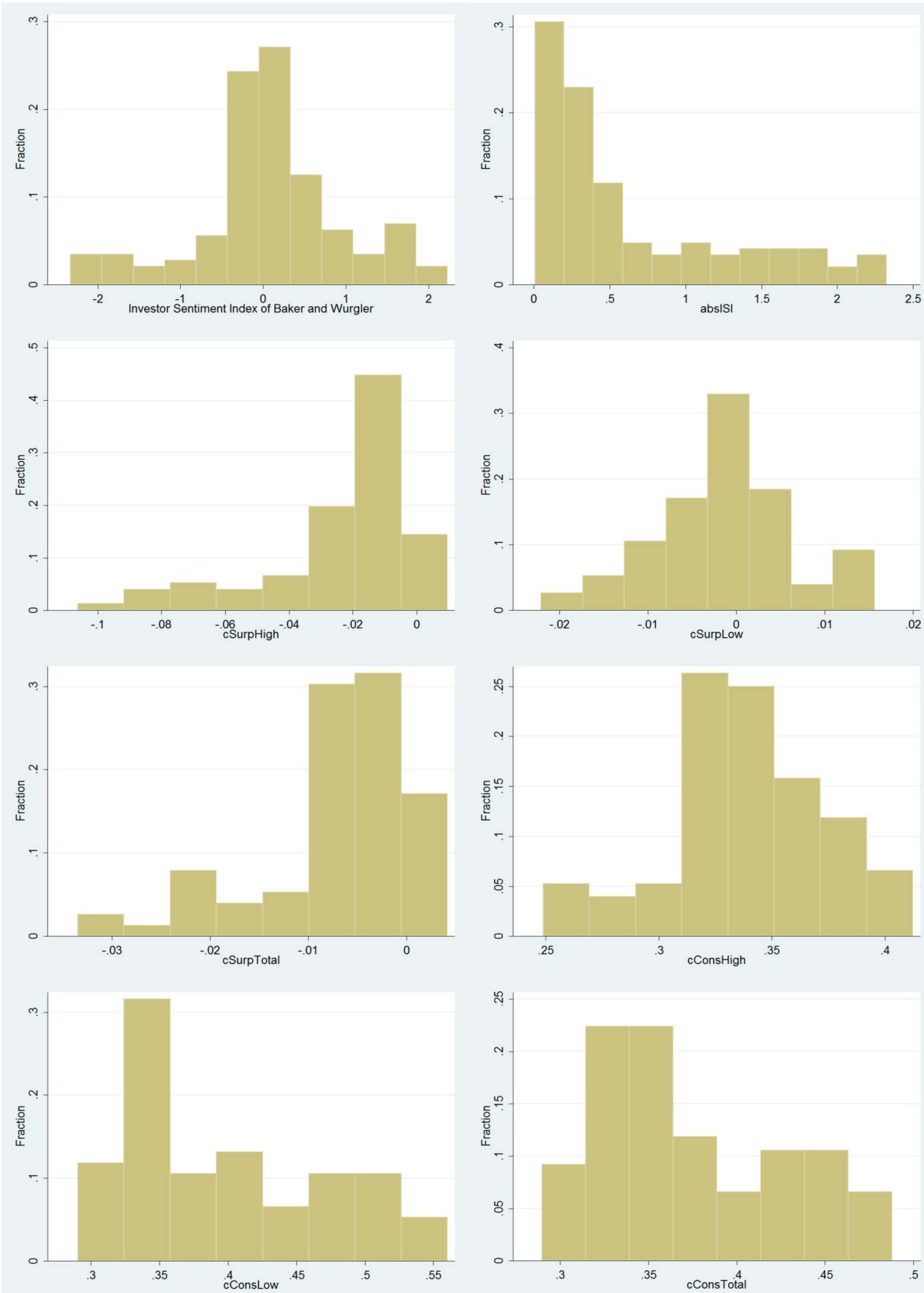
The following six scatterplots indicate the same for the SMB portfolio compositions using the same format as described above (1995 Q1-Q4, rebalancing after Q4 and 1996 Q1).

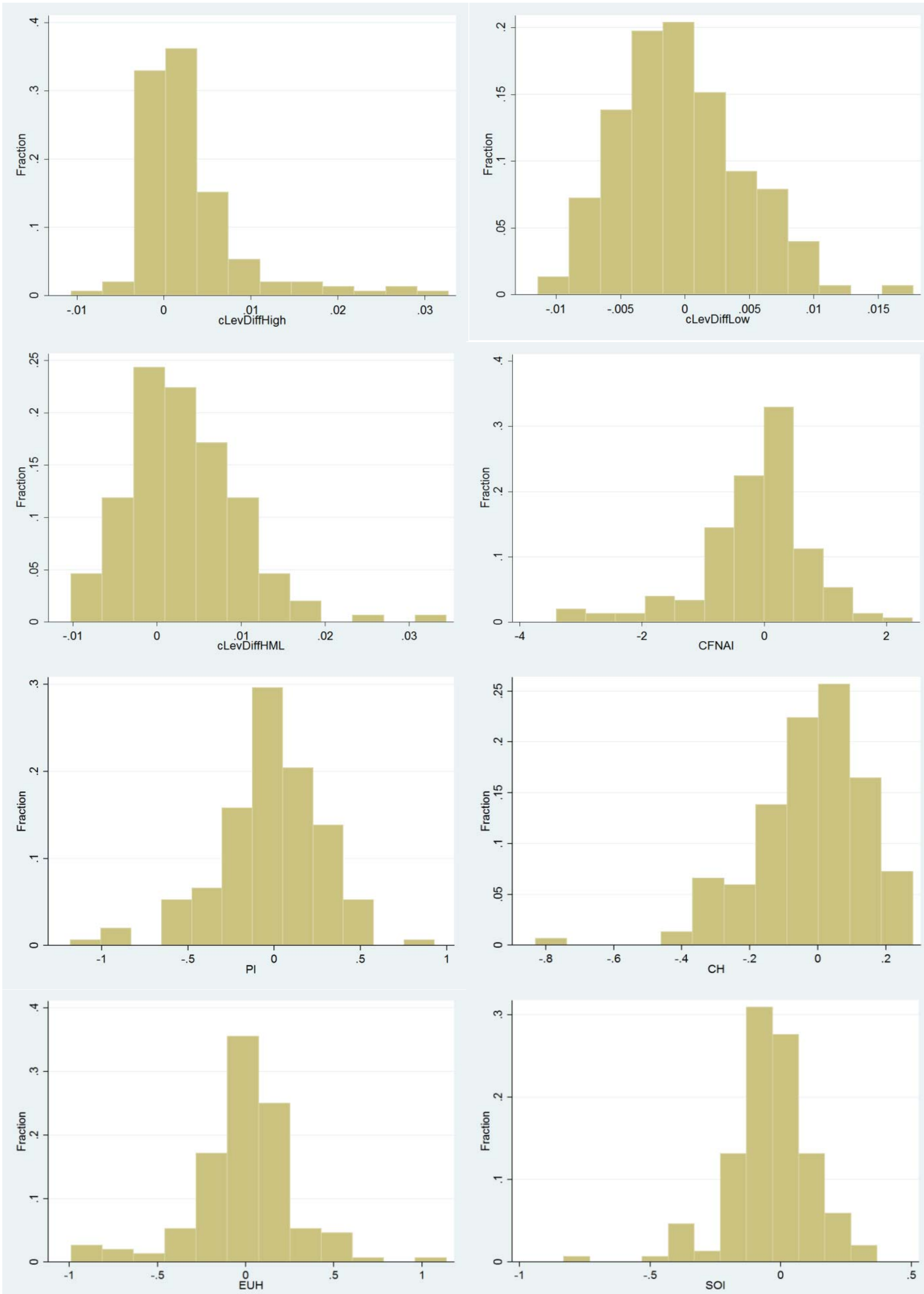


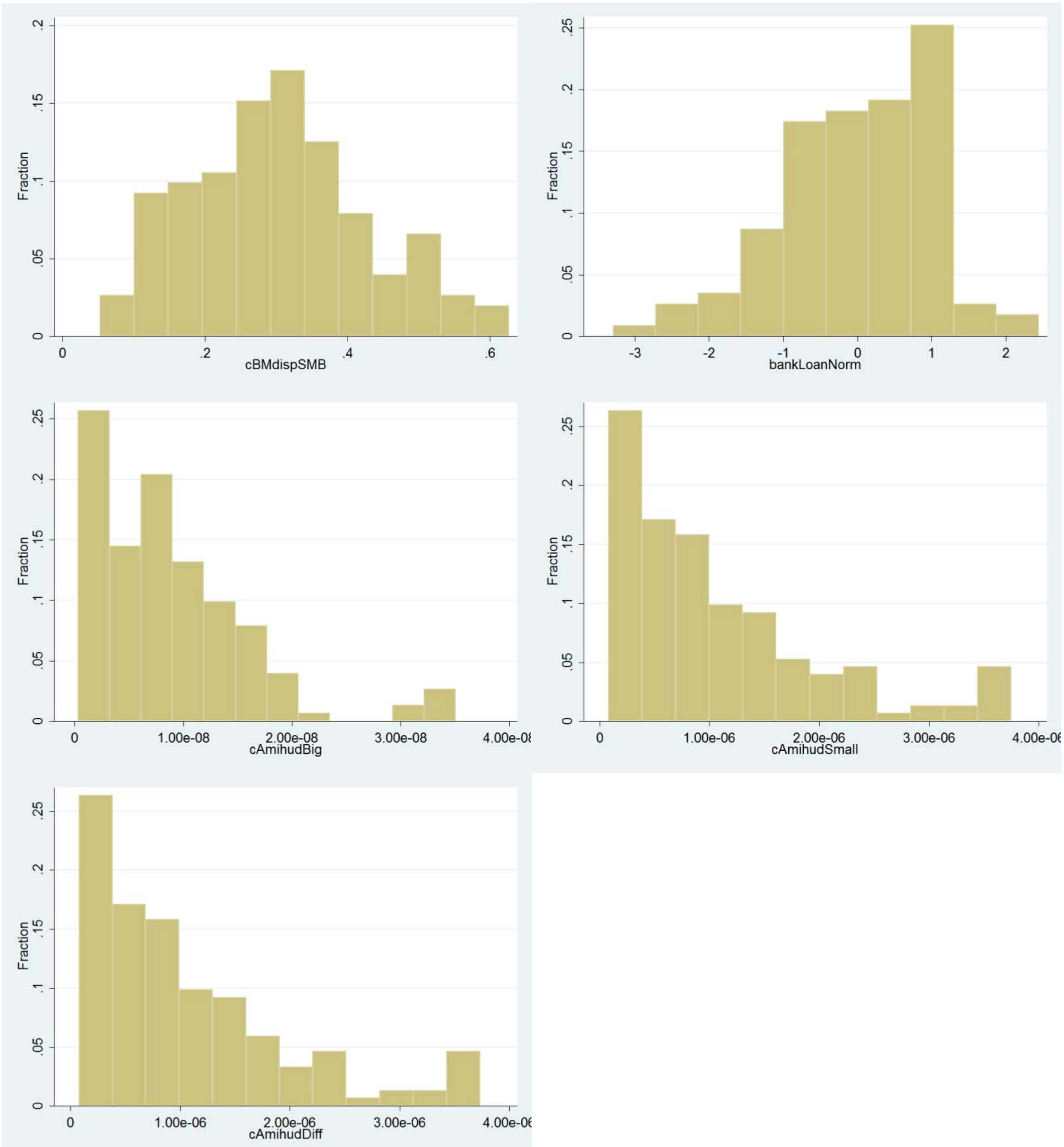
## D.2 HISTOGRAMS OF INDEPENDENT VARIABLES

The following 27 figures indicate the distribution of the values of the variables used in the regressions and described in Table 3a: Independent variables used.




Fraction
Investor Sentiment Index of Baker and Wurgler
absISI
cSurpHigh
cSurpLow
cSurpTotal
cConsHigh
cConsLow
cConsTotal

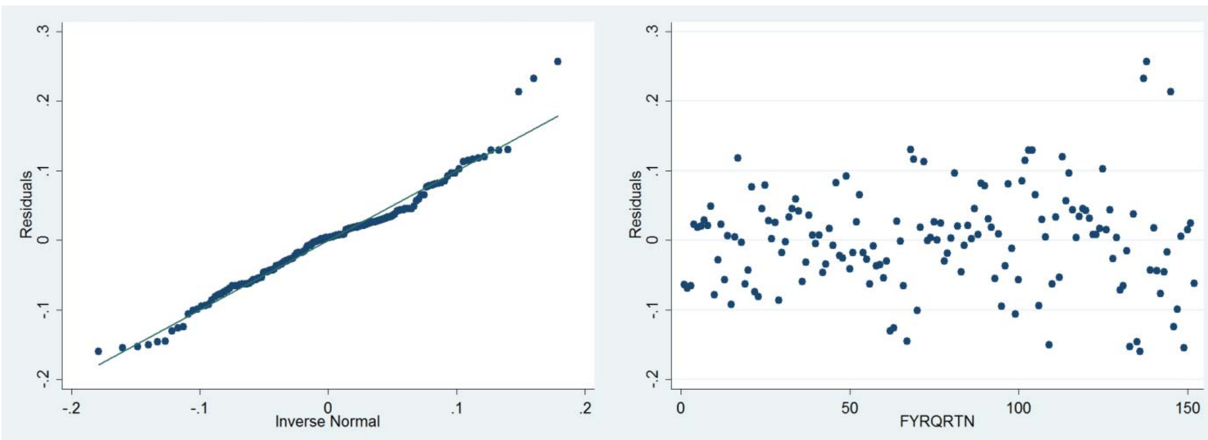

Fraction
cLevDiffHigh
cLevDiffLow
cLevDiffHML
CFNAI
PI
CH
EUH
SOI

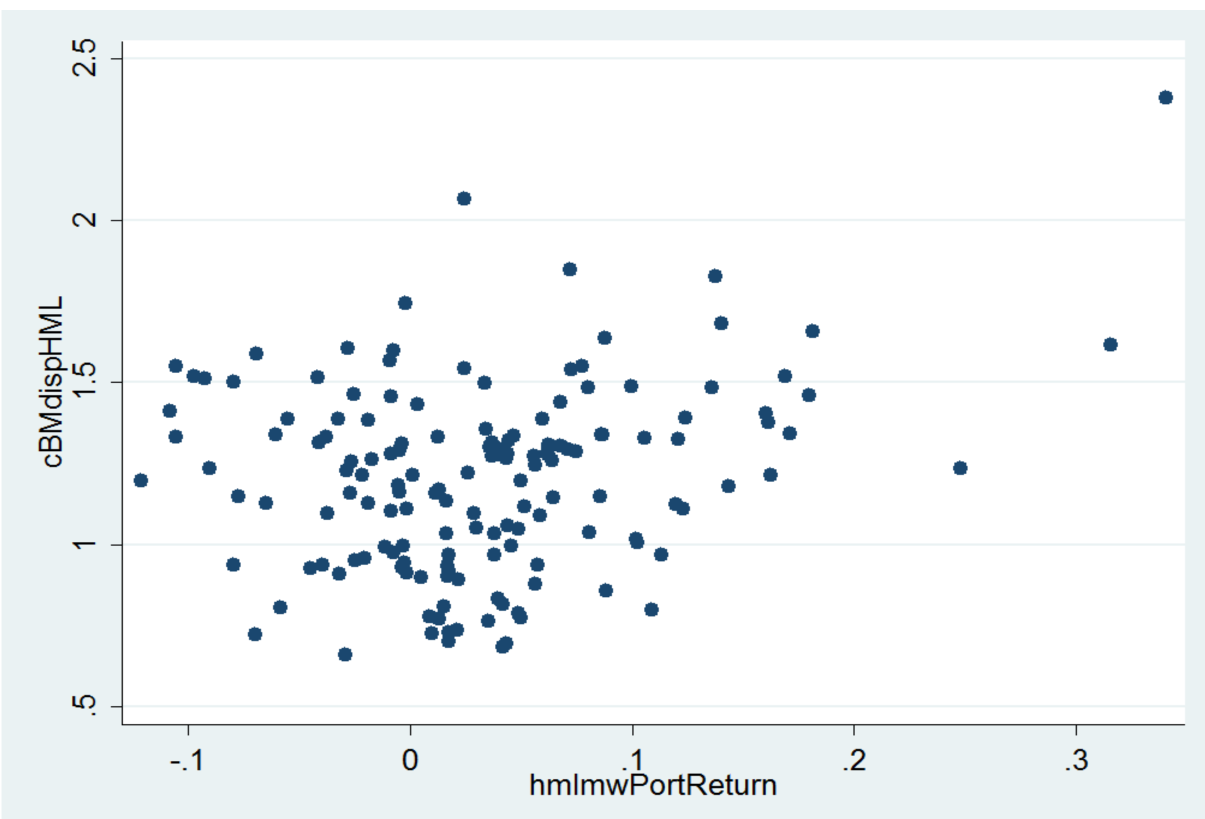

Fraction
cBMdispSMB
0
.05
.1
.15
.2
.4
.6
Fraction
bankLoanNorm
.25
-3
-2
-1
1
2
Fraction
cAmihudBig
1.00e-08
2.00e-08
3.00e-08
Fraction
cAmihudSmall
1.00e-06
2.00e-06
3.00e-06
Fraction
cAmihudDiff

## APPENDIX E: SCATTERPLOTS UNIVARIATE REGRESSIONS

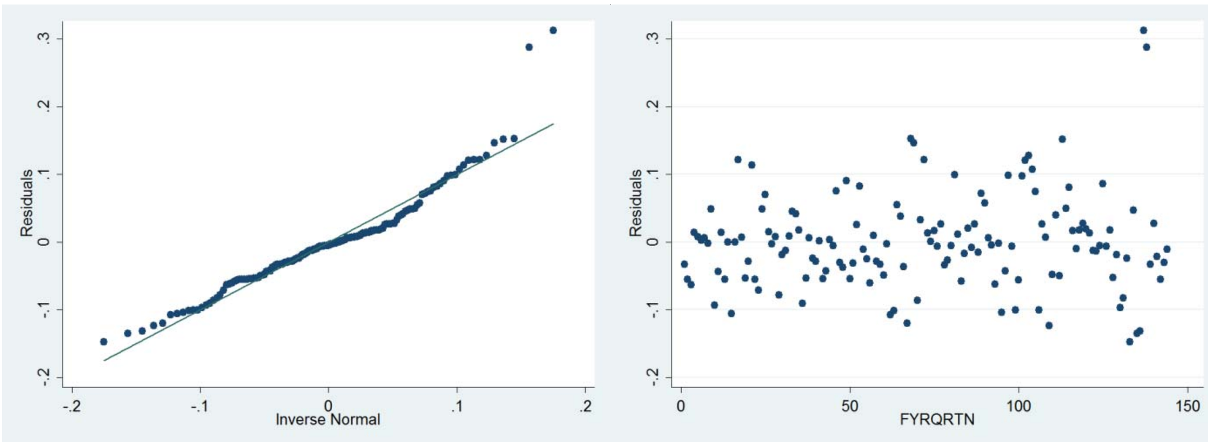
This appendix shows various scatterplots to illustrate the empirical results described in chapter 5. For each variable there are provided three plots, left above shows the residuals compared to a normal distribution (to test normality); right above shows the residuals over time (to test homogeneity of error terms); the plot below shows the portfolio return vs independent variable (to illustrate the correlation).

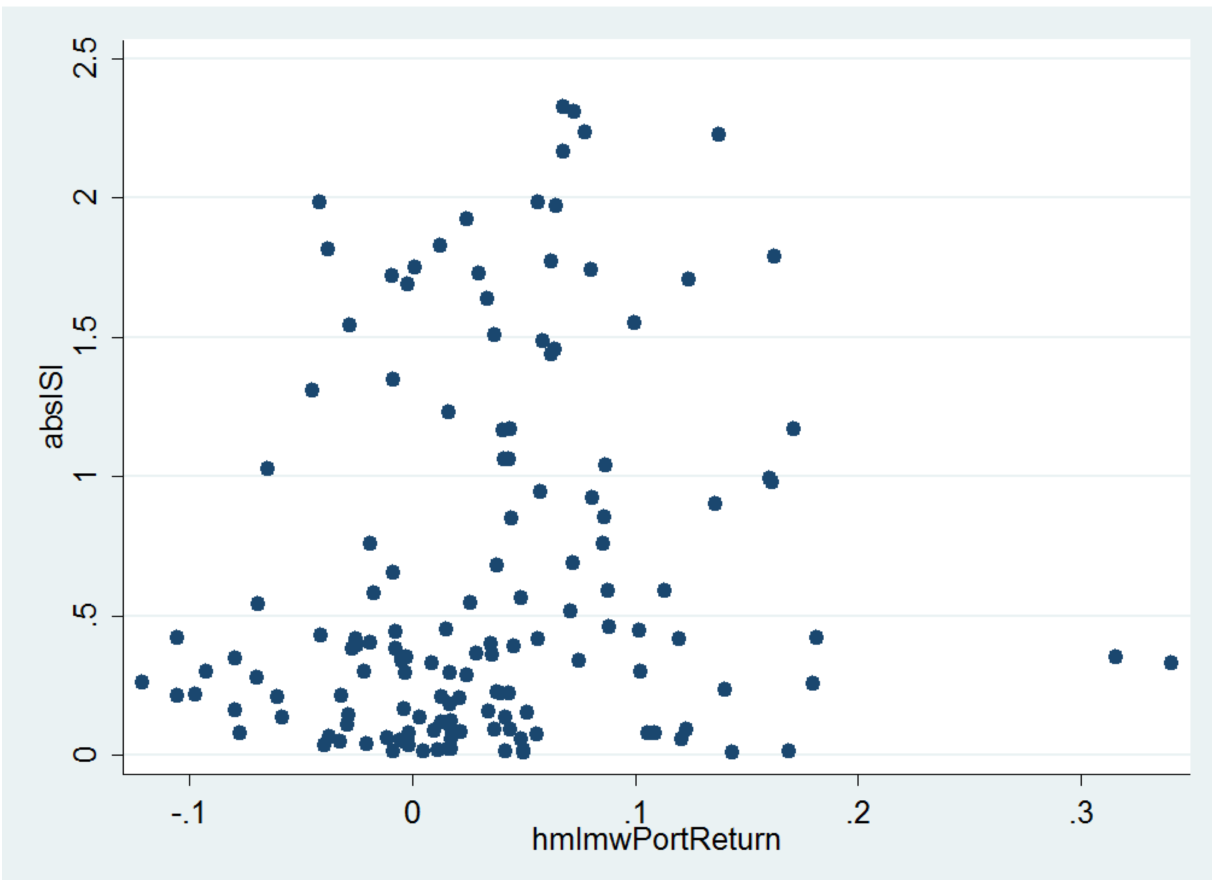
### E.1 MULTIPLE DISPERSION TO PREDICT HML RETURN



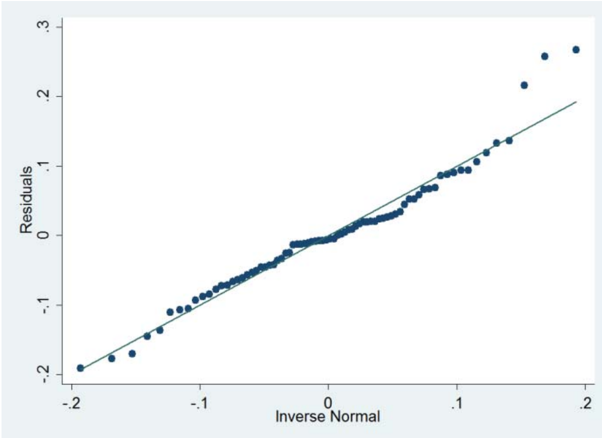



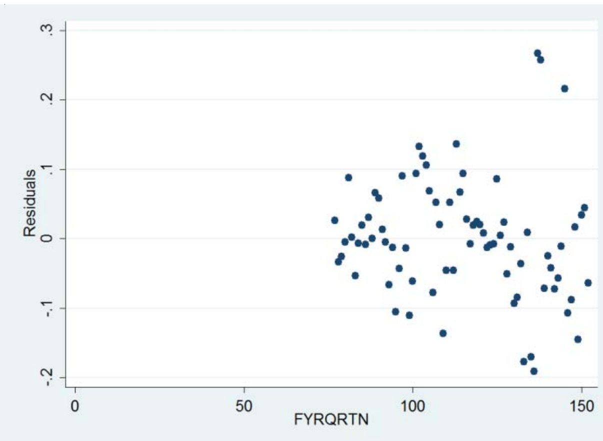
## E.2 ABSOLUTE INVESTOR SENTIMENT TO PREDICT HML RETURN

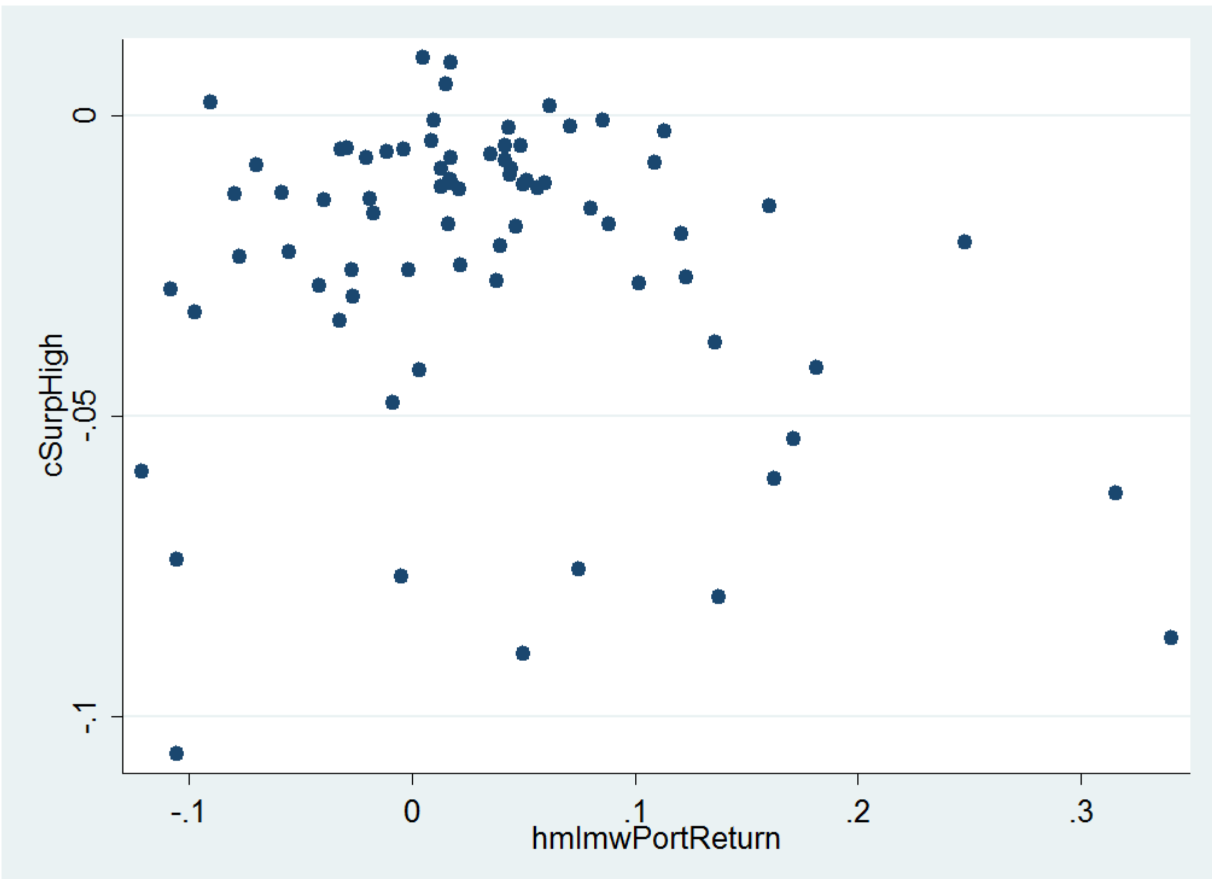


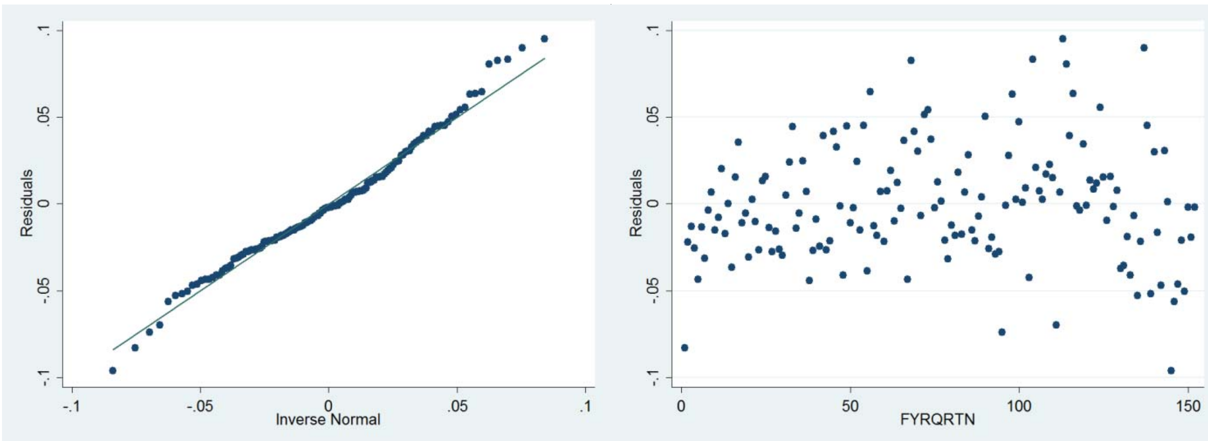
## E.3 ANALYST SURPRISE TO PREDICT HML RETURN

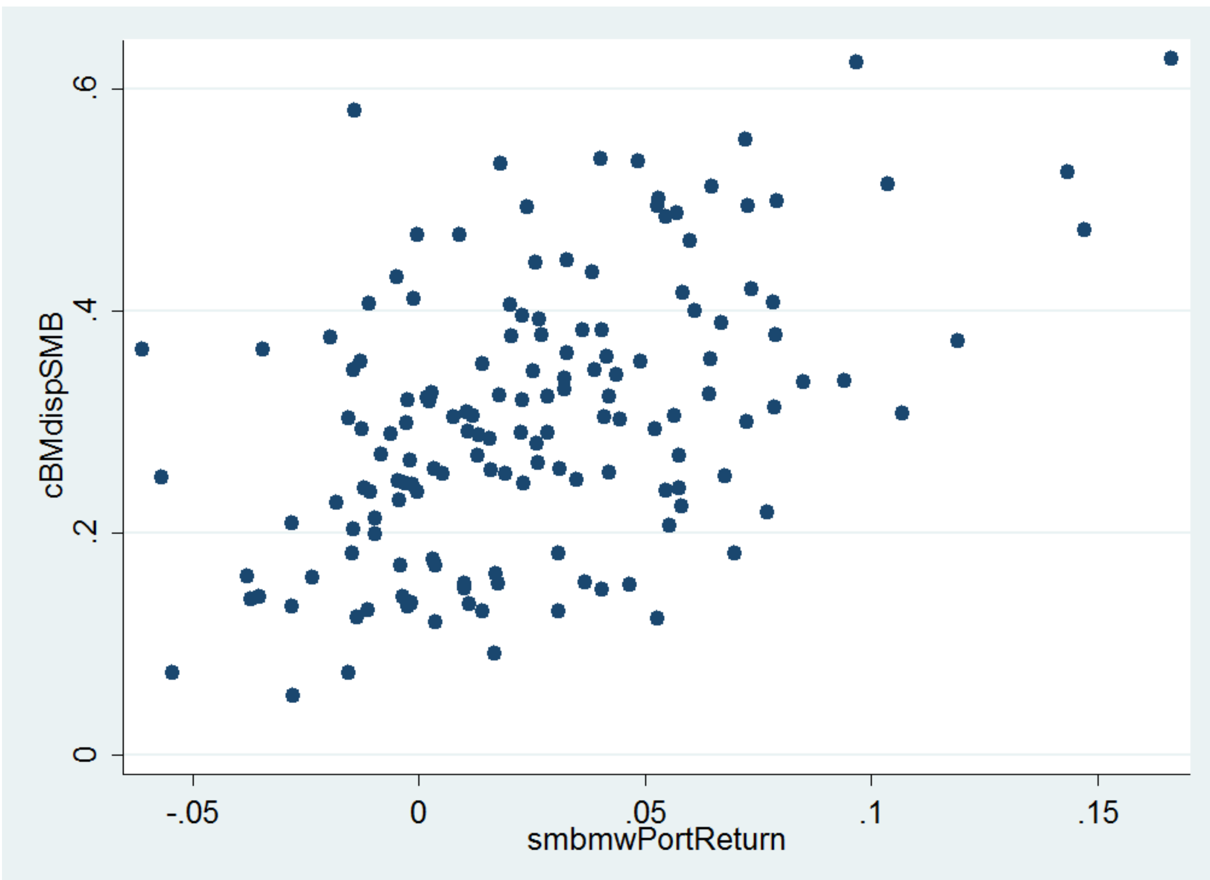


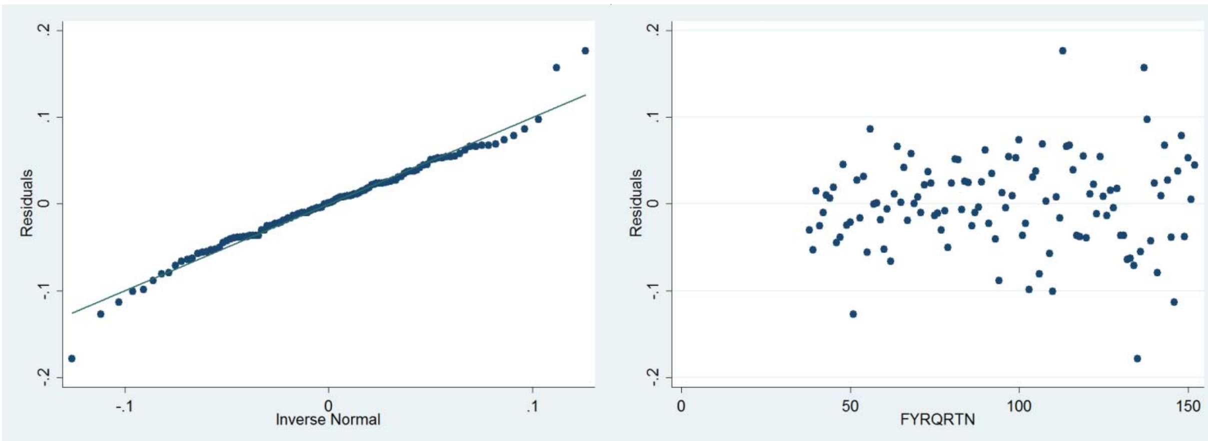



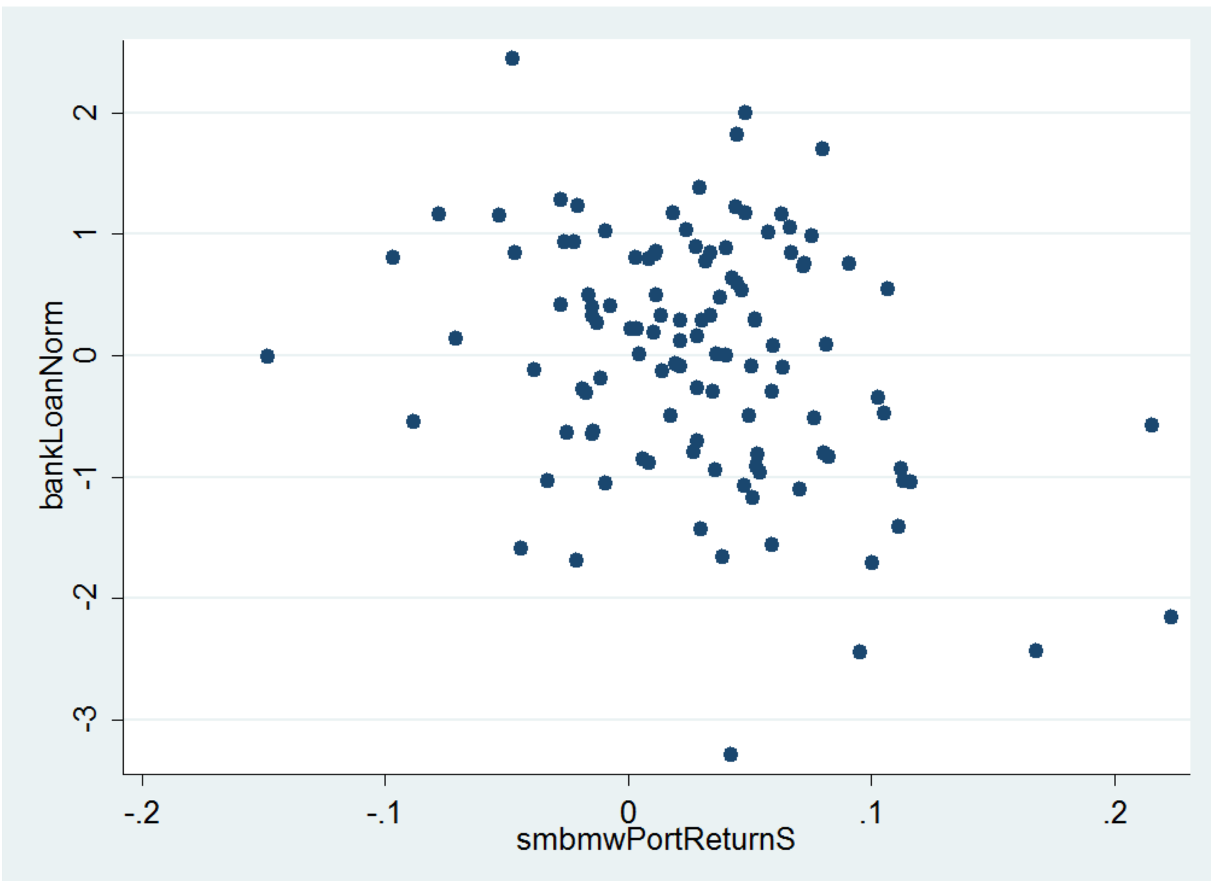
## E.4 MULTIPLE DISPERSION TO PREDICT SMB RETURN



## E.5 BANK LOAN AVAILABILITY TO PREDICT SMB RETURN



## E.6 AMIHUD ILLIQUIDITY TO PREDICT SMB RETURN