UNIVERSITY OF LJUBLJANA
FACULTY OF ECONOMICS

MASTER'S THESIS

A STOCK PICKING ALGORITHM BASED ON A FINANCIAL ANALYSIS USING THE PRINCIPAL COMPONENT METHOD AND KALMAN FILTERING

Ljubljana, July 2017 LUKA VIDOVIČ

AUTHORSHIP STATEMENT

The undersigned Luka Vidovič, a student at the University of Ljubljana, Faculty of Economics, (hereafter: FELU), author of this written final work of studies with the title A Stock Picking Algorithm Based on a Financial Analysis Using the Principal Component Method and Kalman Filtering, prepared under supervision of prof. dr. Igor Lončarski

DECLARE

1. this written final work of studies to be based on the results of my own research;

2. the printed form of this written final work of studies to be identical to its electronic form;

3. the text of this written final work of studies to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the FELU's Technical Guidelines for Written Works;

4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;

5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the FELU in accordance with the relevant FELU Rules;

6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;

7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;

8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the FELU Study Information System;

9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;

10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, 13. 7. 2017 Author's signature: ___________________

TABLE OF CONTENTS

LIST OF TABLES

LISTE OF FIGURES

INTRODUCTION

Investment decisions in financial markets are inherently linked to an estimation of asset's value. Although asset's fundamental data and market data are, in principle, universally available, their interpretation is a subjective process. On average, however, markets correctly interpret all available information to obtain asset's intrinsic value, i.e. the current market price. This notion is the central tenet of the efficient market hypothesis (hereinafter: EMH). To that end, the valuation process serves as a medium to capture and interpret the most current information, detect over or undervalued assets, and correspondingly adjust the market price.

This thesis focuses on the **subject** of equity valuation of publicly traded companies. Practitioners developed and advocate many approaches in equity valuation that generally fall under either relative valuation approach or present value models. The former is based on peer analysis of market multiples, while the latter focuses on a detailed projection of company's future cash flows. The empirical studies show that active investors have a capacity to successfully apply such techniques to pick outperforming stocks and can earn excess returns (Wermers, 2000; Cohen, Polk, & Silli, 2010).

The **purpose** of the thesis is to present an alternative equity valuation approach, which combines the comparative market view embedded in the relative valuation (hereinafter: RV) and bottom-up fundamentals approach of the discounted cash flow (hereinafter: DCF) valuation. Both RV and DCF valuation are based on financial statement analysis. As such, the company's intrinsic value should be reflected in the financial statements. Moreover, previous studies demonstrated that financial statement data do provide valuable signals for a financial analyst and can be successfully exploited in stock-picking investment strategies (Ou & Penman, 1989; Piotroski, 2000; Hirshleifer, Hou, & Teoh, 2009; Lallemand & Strauss, 2016). E.g., Fama and French (2015) also incorporated operating profit in their five-factor asset pricing model to explain a greater cross-section of stock returns. A systematic analysis of financial statements thus has a potential to provide an advantageous basis for equity valuation. However, selection and aggregation of financial ratios can be challenging, and previously authors focused on designing a summary metric to aggregate financial statement data.

The **motivation** of the thesis is to systematically and efficiently use a large amount of financial statement data for the purpose of stock valuation. The **main objective** is to design a coherent approach to utilize financial statement data for stock valuation and propose an active investment strategy. The central **hypothesis** is thus the following: financial statements reflect company's intrinsic value and can thus be used to determine mispriced stock and applied in investment strategy to yield superior return. To that end, we form the following **research questions**. Firstly, how can financial statement data be aggregated in a systemic and economically meaningful manner? Secondly, how to use financial statement data in stock valuation? And lastly, how does such active investment strategy based on financial analysis perform compared to a simple index portfolio?

To address the presented issues, we introduce principal component analysis (hereinafter: PCA) to analyze financial statement data at a company level. In this manner, we transform financial statement data into new variables with a distinct economic notion, i.e. factors, which can be considered as value drivers reflecting profitability, growth prospects, and risk.

We then apply regression analysis where the derived factors are used as the explanatory variables and selected multiple (e.g., P/E) as the dependent variable. Based on the obtained regression equation we are able to calculate implied price multiple. The latter is presumed to reflect the company's intrinsic value based on company's fundamental data reflected in the financial statements. In that manner, we are able to determine over and undervalued companies. The deviations from observed market values are, according to EMH, believed to revert to their intrinsic value in the near future. We utilize such reasoning and propose an investment strategy, which exploits such mispriced stocks.

The investment strategy is then further refined by actively incorporating all available historical data using the Kalman filtering technique. The Kalman filter enables dynamic and optimal adjustment of model parameters to present data while simultaneously taking into account past performance of the model.

We apply the proposed investment strategy in the analysis of the S&P 100 Index, which represents an applicable market universe and compare it to a simple buy-and-hold index portfolio using a number of risk-adjusted indicators. In the analysis, we consider the period from March 2001 to March 2017.

The thesis is structured as follows. Chapter 1 gives a brief summary of the efficient market hypothesis (EMH). In Chapter 2 we summarize different approaches to valuation and present a literature review of financial statement indicators for stock selection. Chapter 3 gives a more detailed description of the methods used in the empirical study (regression analysis, principal component analysis, and Kalman filter algorithm). Chapter 4 describes the framework of the empirical study. First, the selected investment universe and the choice of relevant financial variables are presented. The following methodology sections outline the steps in the empirical study. Chapter 5 presents the determination of financial statement factors, accompanied by a comprehensive analysis of factor structure and stability. In Chapter 6, we apply the derived factors to calculate implied price multiples and determine undervalued stocks. This analysis is then utilized in selected investment strategies. We compare the applied investment strategies using a number of risk-adjusted measures. The last chapter summarizes the main findings and concludes the thesis.

1 EFFICIENT MARKET HYPOTHESIS

1.1 Definition and implications

The efficient market hypothesis (EMH) refers to the market's capacity to process information swiftly and rationally, resulting in market prices which are unbiased estimates of value (Fama, 1976). The extent of market efficiency is highly relevant for investment managers and analysts who seek to exploit market inefficiencies to capture profitable trading opportunities.

In a highly effective market, a passive investment strategy will always dominate an active investment strategy because of lower costs (such as information seeking costs and transaction costs). However, if some form of inefficiency does exist in the market, active investment can achieve a superior return (i.e. higher than expected return given asset's risk) and outperform passive strategy on a risk-adjusted basis. Nevertheless, any pricing discrepancies (i.e. inefficiencies) are eventually arbitraged away through trading. As such, EMH can be seen as a self-correcting mechanism. In efficient markets, such self-correcting processes can happen rather quickly. For example, on New York Stock Exchange (hereinafter: NYSE) the adjustment time to new information is between 5 and 60 minutes (Chordia, Roll, & Subrahmanyam, 2005).

In the context of EMH, it is important to distinguish between the market value and the intrinsic value (or true value). The market value is simply the price at which trades are executed. The intrinsic value, however, is the price an asset would have if investors had complete information regarding asset's characteristics. In the example of a stock market, the implications of the EMH are the following (Damodaran, 2002):

- Stock prices can deviate from intrinsic value for an extended period. However, the EMH only requires that deviations are random;
- Since market prices are random, they are also uncorrelated with any observable variable;
- Randomness of market prices implies no active strategy should be able to consistently outperform passive investment strategy over the extended period;
- In an efficient market, the expected return is consistent with asset's riskiness over the long term. However, a short-term return may deviate from this expected return.

1.2 Forms of market efficiency

The concept of EMH is intimately related to information available to investors, which are subsequently reflected in the market price. In that respect, Fama (1970) defined three forms of market efficiency: weak, semi-strong, and strong. The three forms reflect the type of information incorporated in the market price.

In a weak-form efficient market, market prices reflect all past market data. This suggests active investment based on technical analysis of historical trading patterns cannot achieve a superior return. A semi-strong-form efficient market is weak form efficient and additionally assumes market prices reflect all publicly available information. This includes past market prices, financial statements, news reports, and other market data. In a semi-strong efficient market, prices adjust quickly to new information and investors should not be able to profit from analyzing publicly available information. A strong-form efficient market reflects all available information, public as well as private. Thus no investor, including insiders, would be able to consistently earn a superior return.

The observed type of market efficiency varies through time, by the type of the market, and across geographical regions. A number of factors promote and impede a higher degree of market efficiency, such as:

- Market participants: A larger number of market participants contributes to more efficient processing of new information;
- Information availability: Fair, orderly, and efficient dissemination of information (news, financial disclosures) contributes to the integrity of the market and promotes market efficiency;
- Limits to trading: Any trading restrictions mute the process of price discovery, maintain arbitrage opportunities, and thus impede market efficiency;
- Transaction costs: Transaction costs are incurred in the process of price discovery and market efficiency is limited by such costs;
- Information-acquisition costs: Incorporation of new information in market prices involves gathering and analyzing information, and thus incurs a cost. Prices must thus offer a return to information acquisition (Grossman & Stiglitz, 1980). In an efficient market, active investing cannot earn superior return after deducting such costs.

1.3 Studies of market efficiency

The studies of market efficiency test if a specific investment strategy has a potential to yield superior return. Consequently, a test of market efficiency is simultaneously also a test of expected return model. Potential superior return can thus be attributed to market inefficiency or can merely suggest that expected returns model is incomplete and does not account for all relevant risk factors (Damodaran, 2002).

Tests of weak-form market efficiency examine serial correlation in prices, which would imply a predictable pattern. Overall, the evidence shows developed markets are weak-form efficient (Bessembinder & Chan, 1998; Fifield, Power, & Sinclair, 2005). However, some studies indicate developing markets are not fully weak-form efficient (Chen & Li, 2006; Fifield et al., 2005; Mobarek, Mollah, & Bhuyan, 2008). To test semi-strong market efficiency studies apply event-based studies and portfolio-based studies. Most research support the semi-strong efficiency hypothesis for developed markets, however, markets in developing countries may not be fully semi-strong efficient (Gan, Lee, Hwa, & Zhang, 2005). The tests of strong-form market efficiency show that superior return can be achieved by trading based on material nonpublic information and thus confirm markets are not strong-form efficient (Jaffe, 1974; Rozeff & Zaman, 1988).

Despite many demonstrations of market efficiency, researchers identified apparent exceptions to the efficient market hypothesis. Those market anomalies are examples of changes in market prices that cannot be explained by any newly obtained information in the market. Numerous examples of inconstancies are reported in the literature, which mainly refers to time series anomalies or anomalies based on company characteristics. Examples of time series anomalies include, e.g., January effect (Gultekin & Gultekin, 1983; Haugen & Lakonishok, 1988; Roll, 2010), Weekend effect (Gibbons & Hess, 1981), and other days-of-the-week anomalies (Jacobs & Levy, 1987). Anomalies based on company characteristics include Size effect (Banz, 1981; Keim, 1983) and Value effect (Basu, 1977; Capaul, Rowley, & Sharpe, 1993; Chan, Hamao, & Lakonishok, 1991; Rosenberg, Reid, & Lanstein, 1985).

More recent studies suggest many anomalies can be sufficiently explained using updated methodology. E.g., the January effect is not persistent and does not produce abnormal returns after appropriate risk adjustments (Kim, 2006). Similarly, the size effect and value effect disappear when additional risk factors are included in the analysis (Fama & French, 1995).

Nevertheless, empirical studies show that stock picking can earn an excess return. Wermers (2000) analyzed the performance of mutual funds and concluded that stocks held in their portfolios outperform the market by 1.3 percent per year. However, when adjusting funds' performance for expenses and transaction costs, the excess return disappears, in line with the EMH. Cohen et. al. (2010) also reports that active portfolio managers have the ability to pick outperforming stocks and find that ex-ante best "idea stocks" earn a 6% excess return. Active portfolio managers thus continue to seek novel opportunities to exploit market inefficiencies, while accounting for appropriate risk factors in the process.

2 APPROACHES TO VALUATION

Investment decisions in financial markets are inherently linked to an estimation of asset's value. Prudent investment analysis takes into account all publicly available information, such as market data, company data, industry and economic data, as well as the political and regulatory environment. As such, the valuation process serves as a medium to capture and interpret the most current information, detect over or undervalued assets, and correspondingly adjust the market price. It is a necessary process in well-functioning markets and facilitates semi-strong market efficiency. Although such analysis is costly, it has a potential to create comparative advantages and in turn generate superior returns (Brealey, 1986).

Although asset's fundamental data and market data are, in principle, universally available, their interpretation is a subjective process. This is reflected in the choice of valuation approach, the inclusion of data, and assumptions underpinning future cash flows. Therefore, a successful investment decision process extends beyond the point value estimate, includes the understanding of the sources of asset's value and their manifestation through the valuation model into the derived asset price.

Historically, the two major categories of equity valuation models are present value models (or discounted cash flow models; DCF models) and relative valuation (RV). The former is based on a detailed projection of company's future cash flows and an estimation of appropriate discount factors, while the latter focuses on a peer analysis of market price multiples. New approaches continue to explore novel methods to process information more efficiently, gain comparative advantages to discover pricing discrepancies, and earn a superior return in the process. (Damodaran, 2012; Rosenbaum & Pearl, 2013; Pinto, Henry, Robinson, Stowe, & Wilcox, 2015)

2.1 Present value models

Present value models are based on the concept that financial assets are acquired with the purpose of capturing future financial benefits. Thus, the intrinsic value of the company is equal to the present value of expected generated cash flows and can be expressed as

$$V_0 = \sum_{t=1}^{\infty} \frac{CF_t}{(1+r)^t} \tag{1}$$

where V_0 is the value of the share today, at $t = 0$. CF_t is the expected cash flow at time t and r is discount rate reflecting a required rate of return or, equivalently, the riskiness of the corresponding cash flow. Hence the model is also referred to as the discounted cash flow (DCF) model.

The DCF models are applied in two ways. The first approach is to value only the equity stake by estimating the cash available to be distributed to shareholders. The example of which is the Dividend discount model (hereinafter: DDM), which specifies cash flows as dividends only and is in its simplest form represented by the Gordon growth model (Gordon, 1962).

The second approach focuses on the entire free cash flow available to all stakeholders, i.e. beside shareholders also to bondholders and preferred stockholders. The two approaches differ in applied cash flows and discount rates but, if applied consistently, yield the same estimate of equity value.

To perform the DCF valuation analyst needs to estimate all the cash flows during the company's lifetime, their timing, and appropriate discount rates. These estimates are based on company's fundamental data and are combined with analyst's views and assumptions. While DCF valuation can be time-consuming due to a large scope of information to be processed and evaluated, it is also highly flexible and can incorporate analyst's insights. Since DCF valuation is based on company's fundamentals, the valuation is more rigorous and less influenced by the market sentiment and perception. A process of DCF valuation also helps the investor understand the company's business and drivers of value and associated risks. However, the amount of data required makes DCF valuation very sensitive to assumptions and hence susceptible to error. Namely, the input data is inherently noisy, estimates are subjective and can also be manipulated. (Damodaran, 2012; Rosenbaum & Pearl, 2013; Pinto, Henry, Robinson, Stowe, & Wilcox, 2015)

Since DCF models estimate the intrinsic value and market prices can independently deviate from estimated intrinsic value, no company may necessary be determined as under or over valued. The DCF valuation is thus better suited for long-term investors who believe markets will eventually correct these mistakes and prices will revert to its true, intrinsic value.

2.2 Relative valuation

In relative valuation (RV), the value of an asset is derived by comparing market prices of similar (or comparable) assets. The proponents of relative valuation argue that the intrinsic value is near impossible to estimate and hence market prices reflect the true value. The relative value of equity is thus determined by comparing the market prices of a selected peer group, standardized using a common variable. Additionally, other variables may be used to control for differences within the peer group.

In practice, investors use a number of different multiples and compare them both cross-sectionally and versus the time series. To that end, the stock prices can be standardized using a number of accounting metrics, which can be categorized based on a common variable used as:

- earnings multiples: Price-to-Earnings (P/E), P/E-to-growth (PEG), Value-to-Cash flow, Value-to-EBITDA, etc.;
- book value multiples: Price-to-Book (P/B), Value-to-Book, Value-to-Replacement cost, etc.;
- revenue multiples: Price-to-Sales (P/S), Value-to-Sales, etc.;
- sector-specific multiples: Value per Customer, Value per unit of product, etc.

When relative valuation is used, a chosen price multiple is compared to similar companies in the market. The underlying rationale for this method is the law of one price, i.e. identical assets should have the same price. However, selecting truly comparable companies is a challenging and a highly subjective process. It is thus important to understand how multiples are related to the fundamentals such as, e.g., risk, growth rates, payout ratio, profitability, and control for these differences in the analysis. Analysis using fundamentals expresses, multiples in terms of company's fundamentals used in DCF modeling. The alternative approach is a multiple regression analysis where relevant company specifics are used as the explanatory variables and selected price multiple as the dependent variable.

Relative valuation implicitly assumes markets are overall correct but can make mistakes for individual securities. Relative valuation requires much less information at the company level than DCF and relies more heavily on the markets. Consequently, RV more closely reflects market sentiment and perception. This can be advantageous for relative-value investment strategies, as some percentage of companies will always be marked as undervalued. However, such companies may still be overvalued on an intrinsic value basis, in turn implying an overall overvalued market. The RV is better suited for investors with shorter horizon who manage portfolio relative to the benchmark and operate in a market with a large number of priced securities. (Damodaran, 2012; Rosenbaum & Pearl, 2013; Pinto, Henry, Robinson, Stowe, & Wilcox, 2015)

2.3 Financial statement indicators

In addition to a traditional RV and DCF valuation, a number of variations were proposed as alternatives to identify deviations of observed market prices from intrinsic values and to design a corresponding investment strategy to exploits such investment opportunities.

Both RV and DCF valuation are, in essence, based on financial statement analysis. In that manner, alternative usages of financial statements were proposed to identify investment opportunities. The central tenet here is that company's intrinsic value is reflected in the financial statements and can thus be extracted using some selected technique to identify potential mismatch with the market prices.

E.g., Ou and Penman (1989) formulated a summary value from a set of financial ratios as an indicator of future earnings. They demonstrate that financial statements contain additional information not reflected in the market prices and can be successfully exploited in stock selection. Similarly, Lev and Thiagarajan (1993), and Abarbanell and Bushee (1997) show financial statement data can provide valuable signals for financial analyst and provides an advantageous basis for an investment strategy. Piotroski (2000) and Mohanram (2005) also designed a summary metric to aggregate financial statement data, which then serves as a valuable guidance in an investment strategy to earn an excess return. Consistent with these findings, Hirshleifer et. al. (2009) find that company-level accruals and cash flows provide a strong stock return prediction signals. E.g., aggregate accruals are a strong positive time series predictor, while cash flows is a negative predictor of total stock return (Hirshleifer et al., 2009). Fama and French (2015) also included operating profit as an important explanatory variable in their five-factor asset pricing model. Lallemand and Strauss (2016) successfully applied the analysis of industry-level data and accounting ratios (e.g., gross profits, earnings, investments, aggregate accruals) in portfolio allocation. More evolved mathematical approaches, such as e.g. neural networks, were also applied to manipulate financial statements data and provide the basis for relative stock valuation (Emir, Dinçer, & Timor, 2012; Olson & Mossman, 2003; Zahedi & Rounaghi, 2015).

Such financial statement valuation approach has potential to incorporate a large amount of information across many companies systematically. However, selection and aggregation of financial ratios can be challenging. The relative importance of specific ratio may differ across industries (e.g. manufacturing, internet services, or banking sector). Additionally, differences in accounting standards may adversely affect empirical results. E.g., Biscarri and Espinosa (2008) demonstrated that Fama – French three-factor model is accounting-specific and works best if the data are homogeneous in terms of accounting standards.

3 METHODS

3.1 Regression analysis

The regression analysis is a statistical process to estimate the relationship between the dependent variable and independent variables. Arguably, the linear regression model is most used form of regression analysis and is stated in its generic form as

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \cdots + x_{Ki}\beta_K + \varepsilon_i \quad (2)$$

where y is the dependent or explained variable, x is independent or explanatory variables, and β are unknown model parameters. The error term ε represents the unobserved randomness in the model, which can be attributed to measurement errors, omitted variables, or idiosyncratic effects.

The linear regression model consists of a set of underlying assumptions (Greene, 2012):

- Linearity: A linear relationship between y and x_i;
- Full rank: The independent variables are linearly independent;
- Exogeneity of the independent variables: The error terms are random and are not a function of independent variables. The expected value of error term is thus zero;
- Homoscedasticity and no autocorrelation: The error terms are not correlated, exhibit constant variance, and covariance matrix of the errors is diagonal;
- Exogenously generated data: The independent variables are non-stochastic;
- Normal distribution of errors: The error terms are distributed with zero mean and constant variance.

The unknown model parameters β are usually estimated using ordinary least square (hereinafter: OLS) technique. In short, the OLS method minimizes the sum of squared residuals, i.e. the unexplained variance of y. The explained variance of y is attributed to the variation of independent variables. The share of explained variance in total variance of y is referred to as the coefficient of determination, R^2, and serves as a measure of goodness of fit. A more extensive overview of regression models can be found elsewhere (Greene, 2012).

Regression analysis is useful in relative valuation. Namely, relative valuation is conceptually applied in the context of comparable companies, which are usually interpreted as companies in the same line of business. However, this is not a fixed constraint, and relative valuation can be expanded to a wider sector or the market using a multiple regression. This increases and, in turn, also diversifies the number of companies in the analysis. In a multiple regression, the differences can be accounted for by selecting appropriate proxies for risk, growth, and profitability as independent variables and a selected price multiple as the dependent variable. In this manner, the concept of comparable companies can be expanded and thus relates to the comparability in terms of relevant value drivers and not in terms of a business segment. A larger number of companies in comparison can lead to a more precise analysis and also enables detection of over or undervalued sectors compared to the entire market.

E.g., Cragg and Malkiel (1968) regressed the P/E ratio against a wider stock market for the period from 1961 to 1965 using a sample of 185 companies. As dependent variables, they used the growth rate in earnings, the earnings payout ratio, and the beta. The study was later reproduced by Damodaran (2012) for the period from 1987 to 1991 and again from 2000 to 2011, and using a much larger sample of 1600 companies. The results indicate that regression coefficients and R^2 vary considerably year to year. E.g., the reported R^2 values range between 0.32 and 0.94. This range was attributed to earnings volatility, which is then translated into the volatility of the P/E ratio. The volatility of the regression coefficients offers an interesting insight into how the market is pricing, e.g., risk and growth during different periods. The regressions using P/B ratio and P/S ratio yield similar results, however, the volatility of R^2 is somewhat lower (Damodaran, 2012).

However, the regression methodology does have some drawbacks. The independent variables are often correlated, and the presence of multicollinearity in the model results in higher standard errors, unstable and unreliable coefficients, and unintuitive signs of coefficients. The regression implies linear dependency of chosen multiple and the fundaments, which is limiting and may not be appropriate. Additionally, the regression may not be reliable for extended periods since the fundamental drivers of price may change (Damodaran, 2012).

3.2 Principal component analysis (PCA)

The principal component analysis (PCA) is a multivariate statistical technique of dimension reduction. The PCA is applied to a larger dataset consisting of several inter-correlated variables and allows a reduction of the number of variables with a very limited loss of information. In this context, the information refers to the variance of the data. Additionally, PCA allows an analysis of structural correlations between variables and is thus a useful tool in exploratory data analysis.

The property of the PCA to decompose correlated data into distinct uncorrelated sources makes PCA a useful tool in financial markets. The technique was applied e.g. to find common factors in bond returns (Driessen, Nijman, & Melenberg, 2000; Perignon, Smith, & Villa, 2007), derive measures of systemic risk (Billio, Getmansky, Lo, & Pelizzon, 2012; Kritzman, Li, Page, & Rigobon, 2010), and to study cross-market correlations (Billio et al., 2012; Fenn et al., 2011; Zheng, Podobnik, Feng, & Li, 2012). The PCA is also used to identify major risk components in a stock market (Kim & Jeong, 2005), determine the return-generating factors and exploit their diversification properties (Fung & Hsieh, 1997, 2002; Rudin & Morgan, 2006), and in portfolio management to construct optimal, principal portfolios (Meucci, 2010; Partovi & Caputo, 2004).

A comprehensive overview of the PCA and its derivation can be found elsewhere (Abdi & Williams, 2010; Jolliffe, 2002). In short, the PCA is mathematically defined as an orthogonal linear transformation of a coordinate system and can be performed using the singular value decomposition (hereinafter: SVD) (Jolliffe, 2002).

Let $\mathbf{X}$ be the data matrix of $n \times m$, where n is the number of observations and m is the number of variables. We assume, the columns of $\mathbf{X}$ are centered and have a mean of zero.

Alternatively, the data can easily be transformed accordingly. In addition, it is customary to normalize each variable to unit norm. This is particularly the case when variables vary in scale and are expressed in different units. The corresponding correlation matrix **C** can be decomposed using the SVD as

$$C = VLV^T \tag{3}$$

where **V** is $m \times m$ matrix contains eigenvectors, and **L** is $m \times m$ diagonal matrix of respective eigenvalues λ_i, sorted in decreasing order on the diagonal.

The eigenvectors represent the orthogonal set of principal axes or principal directions. The corresponding principal components are obtained as a projection of the original data **X** onto the principal axes

$$F = XV \tag{4}$$

Matrix **F** can thus be interpreted as a transformation of variables into the space given by the principal axes, and each principal component is expressed as a linear combination of the original variables. The eigenvalues represent the variance of the corresponding principal component, and the square root of eigenvalue $\sqrt{\lambda_i}$ represent the volatility of the principal component *i*.

To reduce the dimensionality of the data from *m* to $k<m$, only *k* largest singular values, and corresponding principal axes are considered. Thus the principal axes matrix **V** is truncated at *k*, resulting in a truncated $n \times k$ principal components matrix $\mathbf{F}_k$ given by

$$F_k = XV_k \tag{5}$$

However, the number of an optimal number of principal components *k* is not known a priori and has to be selected only after the SVD using empirical rules (Jolliffe, 2002). The Cattell's Scree graph (Cattell, 1966) plots singular values in their consecutive order. In theory, the corner or "elbow" point of the Scree graph where the slope changes from a steeper to a flatter region determines the optimal number of principal components. The optimal number of principal components *k* is equal to the rank of the singular value before the elbow point. The second approach is so-called Kaiser's rule (Kaiser, 1961) which states that only principal components with singular values greater than one should be retained (for standardized data). In this manner, only principal components explaining a larger share of variance than the variance of the original variables are taken into account. Both Kaiser's rule and Scree graph criteria weight between ease of use by limiting the number of principal components and tendency to retain as much information (i.e. total variance) as possible.

After the number of principal components is selected, the components are often rotated in order to facilitate the interpretation. The rotation is performed in the subspace of retained principal axes and thus does not influence the variance explained by the corresponding principal components. The most widely used is the orthogonal varimax rotation (Kaiser, 1958). The rotation preserves orthogonally of the principal axes and returns a linear combination of the original variables such that the variance of the squared loadings is maximized. This rotation results in a solution where each principal component is strongly associated with only a limited number of original variables, thereby enabling easier interpretation of each principal component.

3.3 Kalman Filter (KF)

The Kalman filter (hereinafter: KF) (Kalman, 1960) is a recursive algorithm that applies Bayes' theorem to estimate the future state of a system and estimate unobservable, time-varying parameters of the model. The algorithm uses past model estimates and current observations to produce optimal updated parameters, while also taking into account imprecisions and random noise in the observations. As such, it is a self-adapting approach efficiently utilizes all available data and enables a dynamic adaptation of model parameters to the changes in the environment.

The Kalman filter is widely applied in engineering and increasingly also in many aspects of finance to obtain dynamic estimates of model parameters (Harvey, 1991; Wells, 1995). The technique was used, e.g., in factor models to study futures prices in energy and agricultural markets (Manoliu & Tompaidis, 2002; Sørensen, 2002; Cortazar & Naranjo, 2006), modeling expected returns (Conrad & Kaul, 1988; Pastor & Stambaugh, 2009; Rytchkov, 2012), and estimating stochastic volatility models (Pennacchi, 1991; Barndorff-Nielsen & Shephard, 2002; Racicot & Théoret, 2010). The Kalman filter was also used extensively in modeling time-varying betas in factor models such as Capital asset pricing models (hereinafter: CAPM) and Arbitrage pricing theory (hereinafter: APT) (Wells, 1995). These studies indicate that models with betas estimated using Kalman filter approach outperform constant or rolling window regression betas estimated using OLS (Lie, Brooks, & Faff, 2000; Robert W. Faff, Hillier, & Hillier, 2000; Faff & Brooks, 1998; Ebner & Neumann, 2005; Choudhry & Wu, 2008; Zhang & Choudhry, 2016).

Conceptually, the Kalman filter consists of two steps. In the prediction step, the current state of the system, i.e. parameter values, is estimated based on prior values. In the update step, this estimate is improved using current observation. A comprehensive derivation and application of Kalman filter can be found elsewhere (Hamilton, 1994).

We apply a linear Kalman filter in an example of a factor model, where regression parameters β and α are unknown parameters, relating two observable variables as (Chan, 2013):

$$y_t = \alpha_t + \beta_t x_t + \varepsilon_t \tag{6}$$

where ε is the Gaussian process noise.

The relation can be easily transformed in matrix notation

$$y_t = \begin{bmatrix} 1 & x_t \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \varepsilon_t \tag{7}$$

and thus generalized in a matrix form as

$$y_t = x_t B_t + R_t \tag{8}$$

This is denoted the measurement equation of Kalman filter, where matrix B represents the state of the system and consists of unobservable, hidden parameters. In term of Kalman Filter, matrix **x** represent the observation model, relating the state of the system and observable variables. The matrix **R** represents the observation (or measurement) Gaussian white noise.

The Kalman filter assumes that the state of the system at time *t*, i.e. hidden parameters, is a linear function of the state of the system at time *t*-1, and can be thus expressed as

$$B_t = F_t B_{t-1} + Q_t \tag{9}$$

This is denoted as the state transition equation of Kalman filter. The matrix **F** is the state transition model, and matrix **Q** represents the process noise with zero mean multivariate normal distribution.

Hereafter, we assume the current state of the system is best approximated by the prior state of the system or, more specifically, regression parameters remain unchanged. The identity matrix **F** = **I** thus gives the state transition model.

The corresponding state prediction, or a priori state estimate, at time *t* given knowledge at time *t*-1 (denoted as $t|t-1$) can thus expressed as

$$B_{t|t-1} = I_t B_{t-1|t-1} \tag{10}$$

and the corresponding measurement prediction $\hat{y}_t$ equals

$$\hat{y}_t = x_t B_{t-1|t-1} \tag{11}$$

Let matrix **P** be an error covariance matrix of state estimates, representing the estimated accuracy of unobservable parameters. Then the a priori state error covariance prediction at time *t* given knowledge at time *t*-1 equals

$$P_{t|t-1} = I_t P_{t-1|t-1} I_t^T + Q_t \tag{12}$$

and the measurement prediction error can be expressed as

$$S_t = x_t P_{t|t-1} x_t^T + R_t \tag{13}$$

By defining matrix **K** as a Kalman filter gain at time t equal to

$$K_t = P_{t|t-1} x_t^T S_t^{-1} = \frac{P_{t|t-1} x_t^T}{x_t P_{t|t-1} x_t^T + R_t} \tag{14}$$

an updated, a posteriori, state estimate at time *t* given all available knowledge at time *t* is expressed as

$$B_{t|t} = B_{t|t-1} + K_t \left(y_t - x_t B_{t|t-1} \right) \tag{15}$$

where y_t is an actual measurement at time *t*. The updated error covariance matrix of state estimates equals

$$P_{t|t} = P_{t|t-1} - K_t x_t P_{t|t-1} \tag{16}$$

The updated state estimate $B_{t|t}$ is constructed as a linear combination of a priori state estimate and weighted measurement innovation, defined as the difference between the actual measurement y_t and the measurement prediction $\hat{y}_t$. The Kalman gain **K** serves as a corresponding weight.

If measurement error **R** approaches zero, corresponding to highly accurate measurement, the gain **K** increases, and Kalman filter weights the innovation term more heavily $\left(\lim_{R_t \to 0} K_t = x_t^T \right)$. If the measurement is highly unreliable, the gain **K** approaches zero and

Kalman filter effectively disregards the new measurement $\left(\lim_{R_t \to \infty} K_t = 0\right)$. Similarly, if a priori state error covariance approaches zero, corresponding to a highly trusted estimate of unknown parameters, the Kalman filter is reduced to zero $\left(\lim_{P_{t|t-1} \to 0} K_t = 0\right)$. Likewise, the Kalman gain **K** also minimizes a posterior, updated error covariance $P_{t|t}$, thus improving the estimate of unknown parameters.

In a practical implementation of Kalman filter, the measurement error **R** and the process noise **Q** have to be provided. They can be selected a priori, measured using some other technique, or determined from data using more advanced statically approaches (Rajamani & Rawlings, 2009).

4 EMPIRICAL FRAMEWORK

The following sections describe the framework of the empirical study. The first part presents the selected investment universe and the choice of relevant financial statement indicators. The following methodology sections outline the steps in the empirical study. First, the original data are transformed and aggregated using principal component analysis. Next, the derived principal components are applied in valuation and investment process, using the regression analysis coupled with Kalman filtering. The complete data analysis and empirical study are implemented in Matlab (Mathworks Inc., Massachusetts, USA), while some statistical tests were performed in SPSS (IBM Corp., New York, USA).

4.1 Data

4.1.1 Investment universe

The empirical data universe consists of the S&P 100 Index (Bloomberg ticker: OEX Index) as a proxy for tradable securities, and a generic 1-year U.S. Treasury notes as a proxy for a short-term risk-free investment. In the analysis, we consider the period from March 2001 to March 2017. We note that the selected period spans a full business cycle, including period of strong growth 2002-2007, financial and subprime mortgage crisis 2007-2009, global recovery and expansion from 2009 onward.

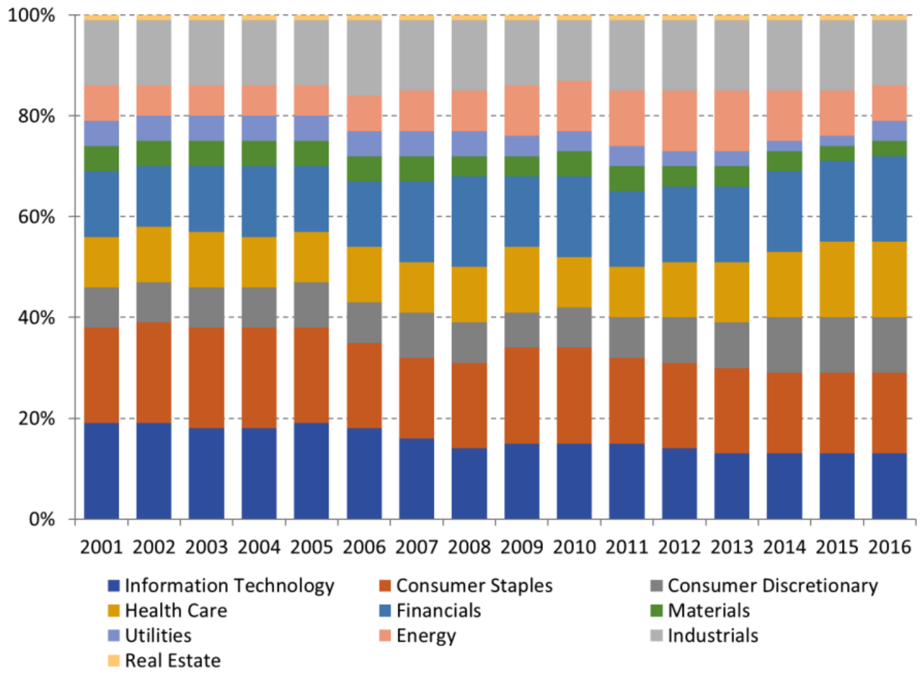
The generic 1-year U.S. Treasury Index (Bloomberg ticker: USGG12M Index) is comprised of generic United States on-the-run government notes and is updated in line with the underlying benchmark securities (Bloomberg, 2017). The U.S. Treasury discontinued auction of the 52-week notes as a regular series on February 27th, 2001 (Bloomberg, 2017). The U.S. Treasury resumed auction on June 3rd, 2008. To fill the gap in the series, we extrapolated the yield-to-maturity for 1-year index using generic 6-month

U.S. Treasury Index (Bloomberg ticker: USGG6M Index) and a generic 2-year U.S. Treasury Index (Bloomberg ticker: USGG2YR Index).

The S&P 100 Index is float-adjusted market cap weighted index and represents a subset of 100 companies selected from the S&P 500 Index. The companies included in the Index should be among the larger and more stable companies (i.e. blue chip companies) and allocated across multiple industry groups (classified according to the Global Industry Classification Standard, GICS). The S&P 100 Index is designed to reflect the U.S. equity markets and thus also the U.S. economy. The Index is rebalanced quarterly, after the close on the third Friday of the quarter-ending month. (S&P Dow Jones Indices, 2017a, 2017b)

Our sample consists of annual data collected each fourth Friday in March. This cut-off date approximately corresponds to the period when fourth quarter results are being announced, and updated accounting ratios become available on Bloomberg. The cut-off date also just exceeds the S&P 100 rebalancing date, thus reflecting the most current U.S. blue chip companies.

Figure 1. S&P 100 Index - Sector breakdown



Note. Breakdown is based on GICS sectors (as determined on April 4th, 2017). The annual cut-off date is fourth Friday in March.

Source: Bloomberg, *Bloomberg Terminal*, 2017.

Each year at the cut-off date, current constitutes of the S&P 100 Index are recorded. Over the period March 2001 to March 2017 a total of 168 companies constituted the Index on at least one of the cut-off dates. The full list of companies is provided in the Appendix. All companies compile financial statements according to the Generally accepted accounting principles (hereinafter: US GAAP) (Financial accounting standards board, 2017). This ensures a level of consistency in a company's financial statements and facilitates the comparison of financial information across different companies.

The sector structure of S&P 100 Index is depicted in Figure 1. The structure of the index by sector is relatively stable during the analyzed period 2001-2016, although a slight shift from sectors Materials, Consumer Staples, and Information Technology towards sectors Consumer Discretionary, Health Care, and Financials is present.

4.1.2 Financial statement data

We selected 40 accounting ratios and collected them for each company constituting the S&P 100 Index (A detailed list of companies is provided in the Appendix). The ratios were selected based on multiple criteria. First, the variable should be comparable among companies and should be either size independent or reflect appropriate adjustments for the company size. Next, the variable should be available for the majority of companies (at least 90%). The variables were selected to reflect different aspects of business, e.g., capital structure, cash flow dynamics, income generation, expansions and investment activities, value volatility, etc. Additionally, previous studies provided valuable guidance on suitable financial ratios to be included in the analysis (Abarbanell & Bushee, 1997; Hirshleifer et al., 2009; Lallemand & Strauss, 2016; Lev & Thiagarajan, 1993; Mohanram, 2005; Ou & Penman, 1989; Piotroski, 2000). At the same time, we opted to limit the size of the dataset in order to keep the analysis operationally and analytically manageable. The list of selected accounting ratios is given in Table 1. A more detailed description for each accounting ratio is provided in the Appendix.

Additionally, we collected five price multiples to be later used in the valuation part of the study (i.e., P/E, P/B, P/FCF, P/S, P/EBITDA). Equity prices for each company, the value of S&P 100 Index, and yield-to-maturity for 1-year U.S. Treasury notes were collected weekly from March 30th, 2001 to April 4th, 2017. The data was collected using appropriate amendment as reported by Bloomberg (2017):

- normal cash adjustments (regular cash, interim, 1st interim, 2nd interim, 3rd interim, 4th interim, 5th interim, income, estimated, partnership distribution, final, interest on capital, distribution, prorated),
- abnormal cash adjustments (special cash, liquidation, capital gains, long-term capital gains, short-term capital gains, memorial, return of capital, rights redemption, miscellaneous, return premium, preferred rights redemption, proceeds/rights, proceeds/shares, proceeds/warrants),

- and capital changes (spin-offs, stock splits/consolidations, stock dividend/bonus, rights offerings/entitlement).

Table 1. List of accounting ratios

	Name	Short description
1	ASSET_GROWTH	Assets - 1 Year Growth
2	BETA_RAW_OVERRIDABLE	Overridable Raw Beta
3	BVPS_GROWTH	BVS - 1 Year Growth
4	CAP_EXPEND_RATIO	Capital Expenditure Ratio
5	CASH_FLOW_GROWTH	Cash Flow - 1 Yr Growth
6	COM_EQY_TO_TOT_ASSET	Common Equity to Tot Assets
7	CONT_INC_GROWTH	Continuing Income - 1 Yr Growth
8	EBIT_MARGIN	Trailing 12M EBIT Margin
9	EBIT_YR_GROWTH	EBIT - 1 Yr Growth
10	EBITDA_GROWTH	EBITDA Growth Year over Year
11	EBITDA_MARGIN	Trailing 12M EBITDA Margin
12	EMPL_GROWTH	Employees - 1 Year Growth
13	EPS_GROWTH	EPS - 1 Yr Growth
14	FREE_CASH_FLOW_MARGIN	Free Cash Flow Margin
15	GROSS_MARGIN	Gross Margin
16	GROWTH_IN_CAP	Capital - 1 Year Growth
17	INC_TAX_EXP_YR_GROWTH	Income Tax Expenses - 1 Yr Growth
18	NET_DEBT_TO_CASHFLOW	Net Debt to Cashflow
19	NET_INC_GROWTH	Net Income - 1 Yr Growth
20	NET_WORTH_GROWTH	Net Worth - 1 Year Growth
21	NORMALIZED_PROFIT_MARGIN	Normalized Profit Margin
22	NORMALIZED_ROE	Normalized ROE
23	OPER_INC_GROWTH	EBIT Growth Year over Year
24	OPER_MARGIN	Operating Margin
25	PROF_MARGIN	Profit Margin
26	REINVEST_EARN_YR_GROWTH	Reinvested Earnings - 1 Yr Growth
27	RETURN_COM_EQY	Return on Common Equity
28	RETURN_ON_ASSET	Return on Assets
29	RETURN_ON_CAP	Return on Capital
30	RETURN_ON_INV_CAPITAL	Return on Invested Capital
31	SALES_GROWTH	Revenue Growth Year over Year
32	TOT_DEBT_TO_EBITDA	Total Debt to EBITDA
33	TOT_DEBT_TO_TOT_ASSET	Total Debt to Total Assets
34	TOT_DEBT_TO_TOT_CAP	Total Debts to Total Capital
35	TOT_DEBT_TO_TOT_EQY	Total Debt to Total Equity
36	VOLATILITY_180D	Volatility 180 Day
37	VOLATILITY_260D	Volatility 260 Day
38	VOLATILITY_90D	Volatility 90 Day
39	WACC	Weighted Average Cost of Cap
40	WORK_CAP_GROWTH	Working Capital - 1 Yr Growth

Source: Bloomberg, *Bloomberg Terminal*, 2017.

The data was later checked for outliers and erroneous data. The corresponding adjustments made are described in the methodology section.

All the data were collected from Bloomberg Terminal (Bloomberg, 2017) using the Bloomberg Excel Add-in on April 4th, 2017 and are stated in USD. The complete dataset for 168 companies covering the period from March 2001 to March 2017 constitutes of approximately 130,000 historical accounting ratios and approximately 150,000 equity and index prices.

4.2 Methodology

We split the collected dataset into two periods. We consider the first three years of our dataset (March 2001 to March 2003) as an initial in-sample testing period to formulate our methodology. The remaining sample period (March 2004 to March 2017) is used as an out-of-sample period to test the proposed investment strategy.

4.2.1 PCA of financial statement data

The obtained historical accounting ratios represent our initial dataset. The ratios are drawn from interconnected financial statements and are thus mutually related while also expressing similar concepts from different perspectives. E.g., the profitability of the company can be expressed using return on equity, return on assets, return on capital, etc. Similarly, growth can be seen to manifest through sales growth, assets growth, or net income growth. By using multiple accounting ratios, we avoid selection bias and concerns of relying too heavily on a single financial indicator. Also, since accounting ratios are inherently interdependent, by using multiple ratios, we can lower the susceptibility of the analysis to unconventional accounting adjustments either due to too conservative or too aggressive accounting practices.

To obtain a more manageable number of variables we first apply a cross-sectional principal component analysis (PCA) in combination with rotation. This allows us to represent the dataset in a more parsimonious form using a limited number of significant and independent variables (i.e. principal component). The PCA also allows us to demonstrate structural correlations within the initial dataset. Due to the large size of the original dataset, we presume a partial loss of information in the process is reasonable in light of potential benefits and will not impede further analysis.

We apply the PCA using the historical data from our in-sample testing period (2001 to 2003). In year t, the corresponding data matrix $\mathbf{X}$ consists of 40 variables (i.e. the number of accounting ratios) and 100 observations (i.e. number of companies). Since variables vary in scale and are expressed in different units, we first transform the individual variables into normalized values to calculate the correlation matrix.

Here we note that the PCA implicitly assumes a multivariate normal distribution, i.e. mean and variance are sufficient statistics to describe the data. Namely, the PCA becomes progressively less efficient when data exhibits more significant nonlinearities. Some ratios (e.g., TOT_DEBT_TO_TOT_EQY, TOT_DEBT_TO_EBITDA, RETURN_ON_CAP) in our dataset do exhibit non-normal distributions, which are generally somewhat similar to the lognormal distribution. To that end, we apply a preliminary transformation of the data to obtain a more symmetric distribution prior to calculating the correlation matrix. In this manner, mean and variance suffice to describe the distribution of the data. Moreover, since data can also exhibit negative values (e.g., RETURN_ON_CAP), a common logarithmic transformation is not feasible. Thus, we apply a cube root transformation function which effectively penalizes large values, like logarithmic function, and is applicable for non-positive values, unlike a logarithmic function, while also maintains the argument sign (Cox, 2011; Miles, Stokes, Vieli, & Cox, 2013).

In order to obtain a robust PCA output, it is important to resolve issues regarding outliers and missing data, which can adversely affect the PCA. Outliers are atypical data points that have an unusually large influence on the estimated model parameters, such as mean and variance. In practice, however, it is challenging to distinguish between a proper data point as a result of pure probability and erroneous data. Moreover, no consensus definition of an outlier exists (Barnett & Lewis, 1994). The bottom line is that an outlier is not consistent with the rest of the dataset and should be omitted to avoid large interference with estimates of model parameters.

One of the more common approaches is dropping the first and last few percentiles of the data. However, the choice of percentile is highly arbitrary, and optimal percentile depends on a specific data sample. Here we use a more structured approach to flag and omit possible outliers. We apply Shapiro-Wilk parametric hypothesis test of normality for each variable (Shapiro & Wilk, 1965). The Shapiro-Wilk null hypothesis states that normality is a reasonable assumption regarding the population distribution for a given random sample data and chosen significance level α. If the p-value is less than α, the null hypothesis is rejected. In other words, this implies the random sample data are statistically significantly different from the normal distribution at chosen significance level. It was previously shown that Shapiro-Wilk test performs better than comparable Kolmogorov–Smirnov test or Lilliefors test (Razali & Wah, 2011).

For each variable, we apply Shapiro-Wilk test at significance level α=0.001% using previously already standardized values. We use rather small α, since we do not require a strict normality for our data and only use the test to detect outliers. If the null hypothesis is rejected, we proceed by omitting the data point with highest absolute value and declaring it an outlier. The remaining data is then again standardized, and the normality test is repeated. The procedure is then iterated until hypothesis cannot be rejected. In this manner, we are able to successfully detect outliers and on average exclude 3% of the dataset and a maximum of 11% of data per variable.

We then perform the PCA, where previously determined outliers are replaced with mean values, i.e. zero. The missing values are treated similarly and are also replaced with the sample mean. The validity of the PCA is confirmed by the Kaiser-Meyer-Olkin (hereinafter: KMO) test for sampling adequacy (Cerny & Kaiser, 1977). Next, we determine a suitable number of principal components k using both Scree graph (Cattell, 1966) and Kaiser's rule (Kaiser, 1961). The k principal components are then rotated using the varimax rotation (Kaiser, 1958). By examining weights of original variables for each principal component, we can assign an economic rationale for the obtained principal components. In this manner, we transform financial statement data into new variables with a distinct economic notion, which are used as a factor in the subsequent valuation step.

4.2.2 Company valuation

The derived principal components represent company's fundamental data as reflected in the financial statements. We relate the principal components to company's value using the regression analysis, where selected price multiple is the dependent variable.

Prior to the regression analysis, the price multiples are transformed using the logarithmic transformation. We also address the issue of outliers and missing data in the same manner as for the financial statement data employed in the PCA. Namely, we standardize the data and apply Shapiro-Wilk test at significance level α=0.001%. Again, if the null hypothesis is rejected, the data point with highest absolute value is declared an outlier and omitted from further analysis. The remaining data is again standardized. The procedure is then iterated until hypothesis cannot be rejected. We find this approach an effective way to successfully detect outliers and on average exclude 3% of the dataset and a maximum of 7%.

After the missing data and outliers are removed from the sample, we estimate the model betas estimated using the OLS regression. Based on obtained regression equation we are able to calculate implied price multiple for each company. The latter is presumed to reflect the company's intrinsic value based on company's fundamental data as presented in the financial statements. In that manner, we are able to determine over and undervalued companies. The companies with market price multiple greater than implied price multiple are determined as overvalued, whereas companies with market price multiple less than implied price multiple are determined as undervalued. By the EMH, we anticipate market prices will eventually revert to its true, intrinsic value.

Here we implicitly assume all companies are subject to the same systemic factors such as macroeconomic environment, monetary policy, and political environment. This is reasonable given all companies in our sample are leading representatives of U.S. market (albeit with a share of international operations). The relative differences in value are thus solely a reflection of idiosyncratic factors, which are projected through each company's financial statements.

4.2.3 Investment strategy

In year *t* we perform a regression analysis of price multiple using calculated principal components and derive implied price multiples. Next, the companies are ranked according to the ratio of implied price multiple to market price multiple. We then form an investment strategy, where we take long positions in undervalued stocks. We construct an equally weighted portfolio by selecting 20 most undervalued stocks and invest funds for a period of one year. Here we additionally require the minimum 10% difference between implied price multiple and market price multiple. This can be seen as a prerequisite for an active investment. If less than 20 stocks meet this criteria, the corresponding share of funds is invested at a risk-free rate represented by the 1-year U.S. Treasury note. After one year, the positions are liquidated using current market prices, yielding one-year portfolio return. If the company is acquired or delisted in the interim period, the corresponding position is evaluated using the last known market price (using one-week price periodicity).

In the year *t*+1, we update accounting ratios using refreshed financial statements, which are then used to repeat the PCA and recalculate the principal components (As described in the following sections, the repetition of PCA is redundant and omitted when Kalman filter is used). The components are then again used as independent variables in regression analysis to determine the updated model parameters. From here on, we follow the layout of the stock picking process as described above.

By calculating regression parameters at each one-year point, we correspondingly adjust the estimations to the most recent financial statement data. However, this may introduce volatility and a high degree of uncertainty into the model, resulting in unstable model parameters and subsequently implied price multiples. By extending the time series of financial statement data to, say, three years, more historical information can be utilized in the regression model. This would also reduce model sensitivity to aggravated market swings. Simultaneously, a longer history of market data could increase model awareness of broad market peaks and troughs.

However, it is challenging to determine the optimal length of historical data series. A too short window will result in a more volatile parameter estimates and higher sensitivity to market swing, while too long window may be insufficiently sensitive to changes in the environment and altered relations between the factors and market prices.

To that end, we apply Kalman filter algorithm. First, we obtain initial beta estimates using multiple regression as described above. In each subsequent year, we adjust the betas in accordance with the Bayes' theorem. In the notation of Kalman filter, the initial betas represent the state of system **B**, principal components (i.e. factors) represent the observation model x_t, and y_t represent the vector of price multiples. The transition model is given by the identity matrix **I**. We assume measurement error **R** and process noise **Q** are equal to 10%.

4.3 Performance measurement indicators

We compare the applied investment strategy using a number of risk-adjusted indicators. A simple buy-and-hold S&P 100 Index portfolio serves as the benchmark and 1-year U.S. Treasury note is a proxy for a risk-free investment. In the subsequent analysis we use the following indicators:

- Tracking error: The tracking error (hereinafter: TE) expresses the deviation of active portfolio return (R_p) from the benchmark return (R_b) and is defined as standard deviation of the difference in returns

$$TE = \sigma\left(R_p - R_b\right) \tag{17}$$

  A lower value of TE indicates risk of active investment strategy is close to the benchmark. Such deviations from the benchmark, however, incur additional risk, which should be sufficiently compensated by additional return (i.e. alpha).

- Information ratio: The information ratio (hereinafter: IR) evaluates the performance of the active portfolio versus the benchmark. The IR is a generalization of Sharpe ratio where risk-free return is replaced by a benchmark portfolio and is thus defined as

$$IR = \frac{R_p - R_b}{\sigma\left(R_p - R_b\right)} \tag{18}$$

  The IR compares the residual return of the portfolio to its residual risk. The residual return is represented by the return not explained by the benchmark and residual risk as the deviation of the difference in returns (i.e. TE). The IR reflects manager's information and skill compared to public information available. As such, it expresses if deviations from the benchmark yield sufficient return.

- Sharpe ratio: The Sharpe ratio is one of the most common measures of performance. It measures the risk premium of a portfolio compared to its total risk and is defined as

$$Sharpe\ ratio = \frac{R_p - R_F}{\sigma\left(R_p\right)} \tag{19}$$

where R_F is the return on a risk-free asset. The Sharpe ratio enables relative comparison of portfolios where higher ratio indicates better performance.

- M-squared ratio: M-squared ratio (hereinafter: M^2) is derived from the Sharpe ratio and evaluates the portfolio's excess return to the benchmark and risk-free asset. The ratio adjusts the excess return of the active portfolio with respect to its riskiness compared to the benchmark and is defined as

$$M^2 = \left(R_p - R_b\right)\frac{\sigma\left(R_m\right)}{\sigma\left(R_p\right)} - \left(R_b - R_F\right) \tag{20}$$

The M^2 results in an identical ranking of portfolios as the Sharpe ratio. However, the M^2 values are meaningful since they express risk-adjusted return in percentage points. The portfolio with M^2 zero has the same performance as the benchmark on a risk-adjusted basis, whereas the portfolio with a positive M^2 outperforms the benchmark on a risk-adjusted basis.

- Jensen's alpha: The Jensen's alpha is defined as the difference between the active portfolio return and the return explained by the market-based CAPM model and is given by

$$\alpha_p = R_p - \left[R_F + \beta_p\left(R_b - R_F\right)\right] \tag{21}$$

Here we use benchmark index as a proxy for the market. The Jensen's alpha is also expressed in percentage term and expresses a superior return that is due to the manager's skill.

- Maximum drawdown: The maximum drawdown (hereinafter: MDD) is a measure of downside risk and expresses a maximum decline from a peak to a trough of a portfolio. The MDD is computed as

$$MDD = \frac{Trough\,Value - Peak\,Value}{Peak\,Value} \tag{22}$$

5 FINANCIAL STATEMENT FACTORS

5.1 Determination of financial statement factors

Financial statement data were selected to reflect different aspects of company's operations and financial strength. Since the data is drawn from the interconnected financial statements, multiple accounting ratios represent similar concepts in term of profitability, financial strength, and growth prospects.

We refer to those concepts as factors, which influence company's value. We apply the PCA to restate the original financial statement data and determine financial statement factors. Hereafter we use a three-year history of financial statement data obtained from the in-sample testing period, March 2001 to March 2003. We use three consecutive years to increase the number of data in the sample and thus derive a more robust analysis of principal components. The obtained financial data are restated in terms of normalized values, and apparent outliers are removed as described in Section 4.2.1.

The pairwise correlation coefficients show interesting details regarding interconnectedness of financial statement data. E.g., operating income growth (OPER_INC_GROWTH) contributes to higher earnings, which in turn translates into higher EBIT growth (EBIT_YR_GROWTH). The corresponding correlation coefficient is thus high, as expected (0,92). Some measures such as NET_DEBT_TO_CASHFLOW (ratio of a company's total debt to trailing 12-month cash flow from operations) are negatively correlated with almost all remaining accounting ratios, with exception of indebtedness measures (i.e., Total Debt to EBITDA, Total Debt to Total Assets, Total Debts to Total Capital, and Total Debts to Total Equity). The latter indebtedness measures are also highly cross-correlated. These measures reflect higher financing with debt, which generally has a lower required rate of return compared to equity. In turn, higher debt financing directly contributes to the lower weighted cost of capital (WACC). Noticeable is a strong correlation among stock volatility measures, i.e. risk proxies, corresponding to different periods (VOLATILITY_260D, VOLATILITY_180D, and VOLATILITY_90D). The correlation matrix is presented in the Appendix.

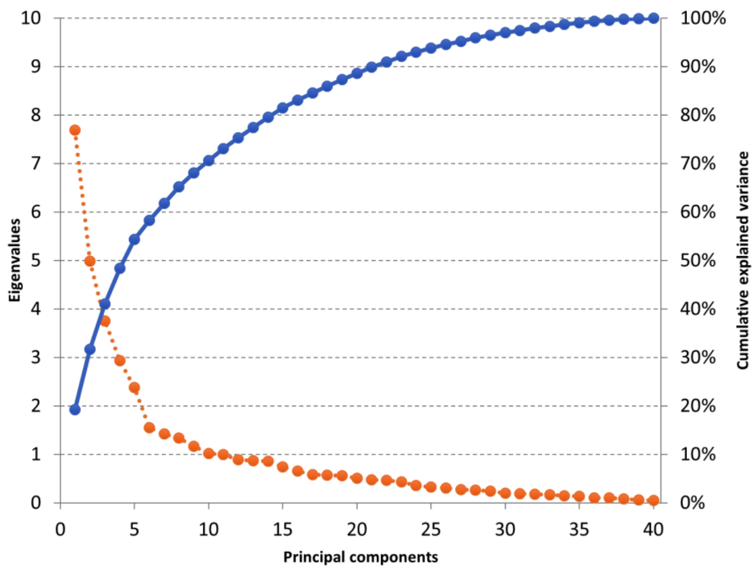
Next, the validity of the PCA is confirmed by a high value of Kaiser-Meyer-Olkin (KMO) test for sampling adequacy, KMO=0.79. Bartlett's test of sphericity also demonstrates a significant divergence of observed correlation matrix from the identity matrix with a p-value less than 0.0001. This confirms a certain redundancy between the variables exists, and we can summarize the data with a smaller number of factors. Results of KMO and Barlett's test are summarized in Table 2.

Table 2. Results of KMO test and Bartlett's test (2001-2003)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.791
Bartlett's Test of Sphericity	Approx. Chi-Square	8,222.375
	df	780
	Sig.	.000

Next, we determine a suitable number of principal components *k*. The corresponding Scree plot is shown in Figure 2. Applying the Kaiser's rule, only the principal components with eigenvalues greater than one should be retained. This can be interpreted as the upper limit for a suitable number of principal components, which in our case yields *k* = 10. The Scree plot criteria puts a suitable number of principal components at around six. We additionally take into account the tendency to retain a larger share of total variance while also obtain components with a meaningful economic interpretation. Additionally, we also examined Scree plots and principal components obtained for a single year of financial statement data (discussed below). Taking into account the above, we decide to hereafter use *k* = 6 principal components as a significant number of financial statement factors. The selected principal components jointly account for 58% of the total variance.

Figure 2. Scree plot and cumulative share of explained variance (2001 - 2003)



The selected initial six principal components are then rotated using the varimax rotation. The rotation results in each accounting ratio being strongly associated with only one of principal components. This allows us to attribute a distinct economic concept to each principal component and refer to them as financial statement factors. The rotated component matrix is presented in Table 3.

Table 3. Rotated component matrix (2001 - 2003)

	Accounting ratio	**PC 1**	**PC 2**	**PC 3**	**PC 4**	**PC 5**	**PC 6**
Margins	NORMALIZED_PROFIT_MARGIN	**0.40**	0.00	0.02	-0.01	0.00	-0.05
	FREE_CASH_FLOW_MARGIN	**0.26**	-0.04	-0.03	0.20	-0.08	0.10
	EBIT_MARGIN	**0.35**	-0.01	0.01	-0.04	0.07	0.04
	EBITDA_MARGIN	**0.37**	0.02	0.02	0.01	0.03	-0.09
	PROF_MARGIN	**0.34**	-0.03	0.00	-0.02	0.02	0.03
	OPER_MARGIN	**0.42**	0.03	0.01	0.00	0.04	-0.05
	WORK_CAP_GROWTH	0.10	-0.01	0.00	0.08	-0.06	0.08
	GROSS_MARGIN	**0.32**	-0.03	-0.01	-0.03	-0.08	0.01
Indebtedness	WACC	-0.05	**-0.31**	0.01	0.15	-0.10	0.08
	TOT_DEBT_TO_TOT_EQY	0.01	**0.41**	-0.01	0.00	-0.04	0.05
	NET_DEBT_TO_CASHFLOW	-0.13	**0.26**	0.04	-0.05	0.01	-0.06
	TOT_DEBT_TO_TOT_ASSET	0.00	**0.41**	-0.02	0.03	-0.01	0.06
	TOT_DEBT_TO_EBITDA	-0.09	**0.34**	0.01	0.02	0.00	-0.09
	COM_EQY_TO_TOT_ASSET	-0.06	**-0.36**	0.00	-0.06	0.04	-0.04
	TOT_DEBT_TO_TOT_CAP	0.02	**0.44**	0.00	0.01	-0.02	0.05
Operating growth	OPER_INC_GROWTH	-0.02	-0.01	**0.38**	-0.01	-0.03	0.02
	NET_INC_GROWTH	0.03	0.02	**0.34**	0.00	0.04	0.00
	EBITDA_GROWTH	0.04	-0.02	**0.28**	0.05	0.04	0.02
	CONT_INC_GROWTH	0.00	0.00	**0.37**	-0.02	0.00	-0.03
	EPS_GROWTH	0.01	0.02	**0.40**	-0.04	-0.01	0.00
	INC_TAX_EXP_YR_GROWTH	-0.01	0.01	**0.31**	-0.02	-0.02	-0.03
	EBIT_YR_GROWTH	-0.01	0.01	**0.38**	0.02	-0.03	0.01
	REINVEST_EARN_YR_GROWTH	0.00	-0.02	**0.31**	0.03	0.00	0.02
Volatility	BETA_RAW_OVERRIDABLE	0.01	-0.05	0.01	**0.37**	0.01	-0.12
	CASH_FLOW_GROWTH	0.03	0.02	0.01	0.09	0.00	0.07
	VOLATILITY_260D	-0.02	0.01	-0.02	**0.51**	0.02	0.00
	VOLATILITY_180D	-0.06	0.00	0.01	**0.49**	0.00	-0.01
	VOLATILITY_90D	-0.04	0.03	0.01	**0.46**	0.07	-0.01
Company growth	EMPL_GROWTH	0.04	-0.03	-0.04	0.02	0.17	-0.13
	ASSET_GROWTH	-0.03	0.01	0.01	-0.02	**0.49**	-0.03
	SALES_GROWTH	-0.07	-0.02	0.11	0.08	0.20	0.19
	NET_WORTH_GROWTH	0.02	-0.01	0.03	0.02	**0.43**	0.05
	GROWTH_IN_CAP	-0.01	0.00	-0.04	-0.06	**0.51**	-0.05
	BVPS_GROWTH	0.02	-0.01	0.00	0.06	**0.40**	0.08
Profitability	NORMALIZED_ROE	0.11	0.09	-0.04	0.01	0.02	**0.42**
	RETURN_ON_INV_CAPITAL	-0.08	-0.05	0.00	0.00	-0.07	**0.45**
	RETURN_ON_ASSET	-0.13	-0.02	0.03	-0.02	0.00	**0.41**
	CAP_EXPEND_RATIO	0.14	0.02	-0.01	0.15	-0.13	**0.22**
	RETURN_COM_EQY	0.06	0.07	-0.01	-0.02	0.11	**0.39**
	RETURN_ON_CAP	-0.02	-0.14	-0.03	-0.14	0.03	**0.32**

Note. Coefficients with an absolute value greater than 0.25 are bolded.

The first factor is referred to as margins factor. The factor accounts for 19% of total variance, and groups together different margin ratios. As such, it represents company's ability to generate profit, and indirectly reflects company's competitive position and quality of the management. Margin is a key component of company's long-term success and highly important measure for equity analysts. The factor does not reveal sources of earnings or other structural information and merely sees the margin ratios as highly correlated, which allows us to represent them jointly in a single, margins factor.

The second factor combines accounting ratios that refer to company's indebtedness and represents an additional 12% of total variance. The factor combines debt ratios (debt to balance sheet items), which reflects company's financial structure, and coverage ratios (debt to income statement items), which show the adequacy of company's cash flow and earnings to service debt obligations. As such, these ratios reflect company's long-term financial health, i.e. solvency. Ratio Common equity to total assets (COM_EQY_TO_TOT_ASSET) is similar to debt ratios and conversely impacted by the changes in indebtedness, as demonstrated with a negative sign. The weighted average cost of capital (WACC) also displays negative relation to company's indebtedness. Namely, as debt is generally a cheaper source of financing compared to equity, rising share of debt in capital structure lowers average cost of financing.

The third ratio reflects company's growth in operating activities and profitability. This factor thus represents a trend in company's operations and relates to improvements in its operations or expansion of activities. Such growth in operating performance is a positive indicator in company valuation. This factor accounts for an additional 9% of the total variance.

The fourth factor reflects the volatility of share price. The factor combines annualized standard deviation of historical price changes for a different number of recent trading days. Similarly, the beta also measures the volatility of the stock price relative to the volatility of the market index. Thus, this factor can be seen as a proxy for risk and accounts for 7% of total variance.

The fifth factor combines accounting ratios related to company's growth of its overall business. The factor is predominantly determined by the accounting ratios referring to the changes in the balance sheet, e.g. asset growth and BVPS growth. The income statement items, such as sales growth, also point to the expanding business but with a smaller impact. Similarly, an employee growth is also related to overall growth of the business, but the relation is somewhat weaker and only partially correlates with the growth of balance sheet items. The factor accounts for 6% of the total variance.

The sixth factor represents profitability and projects company's ability to effectively use available resources in terms of equity and assets. This factor is closely related to the margins factor but expresses company's operational performance in terms of balance sheet items. The ratios such as ROE and ROA are widely used by analysts as a benchmark for company's success. These ratios express company's ability to deliver a return to all its stakeholders. As such, they are a key determinant of value for investors. The factor accounts for 4% of the total variance.

5.2 Analysis of stability of financial statement factors

We examine the stability of the PCA in different periods. First, we use financial statement data from 2001 only and demonstrate the validity of the PCA as denoted by a high value of the Kaiser-Meyer-Olkin (KMO) test, KMO=0.72. Likewise, Bartlett's test of sphericity results in a p-value less than 0.0001, thus demonstrating a significant divergence of observed correlation matrix from the identity matrix. Results of KMO and Barlett's test are summarized in Table 4.

The values of the KMO test and Bartlett's test are slightly lower compared to the data from the 2001-2003 period. We attribute that to the smaller data sample, where hidden structural relations in terms of factor structure are more subjected to noise and incidental correlations.

Table 4. Results of KMO test and Barlett's test (2001)

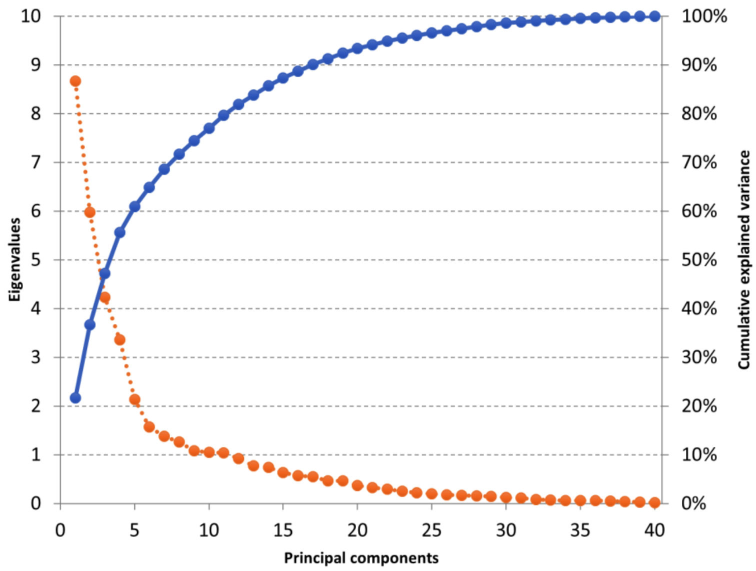
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.717
Bartlett's Test of Sphericity	Approx. Chi-Square	3,575.174
	df	780
	Sig.	.000

Figure 3 shows a Scree plot using 2001 financial statement data. Notice the similarities in qualitative and quantitative characteristics with a Scree plot obtained using three years of financial statement data (2001-2003) in Figure 2. Here, the first six principal components account for 65% of total variance. The result is somewhat expected, given that data corresponds to a single cut-off date rather than a longer period, which contains a time variable and hence a potentially larger dispersion of the data.

Again we apply varimax rotation using the selected six initial principal components. The rotated component matrix is presented in Table 5. We analyze the accounting ratios associated with each principal component and compare them to principal components obtained by analyzing three years of data (Table 3.).

Overall, we obtain a similar factor structure as in Table 3. Namely, accounting ratios are similarly grouped and associated with a distinct principal component. Only a few accounting ratios are associated with a different factor as compared to the 2001-2003 dataset (see Table 3). Such migrations between factors are typically encountered for accounting ratios with small absolute weights. E.g., net debt to cash flow ratio has the largest weight placed on the margins factor (-0.22), as opposed to indebtedness factor in the 2001-2003 dataset. However, the net debt to cash flow ratio's weight on indebtedness factor is comparable, 0.17. This value suggests that changes in corresponding weights could be attributed to the noise due to the small sample size.

Figure 3. Scree plot and cumulative share of explained variance (2001)



The explained share of total variance by each factor and its corresponding eigenvalue are different compared to the 2001-2003 dataset. E.g., volatility factor accounts for 15% of total variance whereas in the 2001-2003 dataset its corresponding share is only 7%. Again, the margins factor explains the largest share of total variance, 22%.

The two examples of rotated component matrix demonstrate that applied PCA using six rotated principal components yields consistent results in term of financial statement factors. Differences in the variability of variables in a given year change factor eigenvalues, which results in rearrangement of factor sequence. However, the total share of explained variance by six principal components is roughly similar.

Figure 4 present the PCA results using annual data at corresponding cut-off dates for the period March 2001 to March 2016. We focus on the validity of the PCA using the KMO test and explained share of total variance by each factor, without further examining the structure of each principal component.

Table 5. Rotated component matrix (2001)

	Accounting ratio	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Margins	NORMALIZED_PROFIT_MARGIN	**0.33**	0.05	0.03	0.05	0.05	-0.10
	FREE_CASH_FLOW_MARGIN	0.24	0.15	0.05	0.13	0.00	-0.08
	EBIT_MARGIN	**0.38**	-0.04	-0.01	-0.04	0.02	0.02
	EBITDA_MARGIN	**0.36**	0.09	-0.03	0.11	0.07	0.11
	PROF_MARGIN	**0.30**	-0.04	0.02	-0.03	-0.02	-0.05
	OPER_MARGIN	**0.33**	0.06	0.09	0.12	-0.03	-0.06
	NET_DEBT_TO_CASHFLOW	-0.22	-0.03	0.04	0.17	0.04	0.01
	WORK_CAP_GROWTH	0.18	-0.10	-0.03	-0.11	-0.10	0.11
	GROSS_MARGIN	**0.33**	-0.05	-0.07	-0.04	0.02	0.04
Volatility	BETA_RAW_OVERRIDABLE	-0.01	**0.40**	-0.05	0.01	0.10	0.01
	VOLATILITY_260D	-0.01	**0.45**	-0.02	-0.05	-0.05	0.04
	VOLATILITY_180D	0.06	**0.44**	-0.02	-0.06	-0.01	0.08
	VOLATILITY_90D	0.03	**0.46**	0.00	-0.04	-0.06	0.00
Operating growth	OPER_INC_GROWTH	-0.08	0.02	**0.37**	-0.01	0.05	-0.10
	NET_INC_GROWTH	0.01	-0.05	**0.36**	-0.05	0.02	0.04
	EBITDA_GROWTH	-0.03	0.01	**0.30**	0.00	0.11	0.11
	CONT_INC_GROWTH	0.03	-0.02	**0.37**	0.02	-0.03	-0.02
	EPS_GROWTH	0.04	-0.05	**0.38**	0.02	-0.04	-0.03
	INC_TAX_EXP_YR_GROWTH	-0.01	-0.01	**0.26**	-0.05	0.00	0.03
	EBIT_YR_GROWTH	-0.02	0.07	**0.40**	0.02	-0.08	-0.06
	REINVEST_EARN_YR_GROWTH	0.14	-0.13	0.21	-0.04	-0.06	0.18
Indebtedness	WACC	-0.09	0.20	0.01	**-0.30**	0.02	-0.07
	TOT_DEBT_TO_TOT_EQY	0.07	0.05	-0.04	**0.41**	-0.04	-0.01
	TOT_DEBT_TO_TOT_ASSET	0.05	-0.01	-0.04	**0.35**	-0.03	0.07
	TOT_DEBT_TO_EBITDA	-0.11	0.00	-0.01	**0.26**	0.00	0.13
	COM_EQY_TO_TOT_ASSET	0.07	0.08	0.00	**-0.37**	-0.01	0.14
	TOT_DEBT_TO_TOT_CAP	0.01	-0.04	-0.03	**0.40**	0.00	-0.05
Company growth	ASSET_GROWTH	0.01	0.07	0.02	-0.01	**0.41**	0.10
	SALES_GROWTH	0.07	0.04	0.17	-0.02	0.20	0.15
	NET_WORTH_GROWTH	-0.04	-0.01	0.00	0.06	**0.48**	-0.17
	GROWTH_IN_CAP	-0.04	-0.02	-0.03	0.00	**0.51**	-0.04
	BVPS_GROWTH	0.06	-0.07	-0.03	-0.09	**0.40**	0.02
Profitability	EMPL_GROWTH	0.11	-0.12	-0.04	-0.03	0.21	**0.37**
	NORMALIZED_ROE	0.10	-0.10	-0.06	0.01	0.02	**-0.32**
	RETURN_ON_INV_CAPITAL	0.06	-0.10	-0.07	-0.16	0.03	**-0.31**
	RETURN_ON_ASSET	0.18	-0.05	-0.01	-0.17	-0.06	-0.19
	CAP_EXPEND_RATIO	0.03	0.17	0.07	0.16	0.03	**-0.36**
	RETURN_COM_EQY	0.15	-0.07	-0.01	-0.01	-0.01	**-0.28**
	CASH_FLOW_GROWTH	-0.08	0.13	0.09	0.12	0.13	**-0.26**
	RETURN_ON_CAP	-0.09	-0.09	-0.01	-0.19	0.00	**-0.36**

Note. Coefficients with an absolute value greater than 0.25 are bolded.

Overall, the PCA is relatively stable through the entire period. The values of the KMO test are sufficiently high, and ranging between 0.63 and 0.73. The first six principal components account for 64% to 72% of the total variance. The relative shares of explained variance by each principal component are relatively constant. The exception is a significant increase in explained variance by the first principal component in from 2007 (19%) to 2009 (29%). The increase is a result of a significant increase in volatility in that period. Namely, the volatility factor accounted for 8% of the variance in 2007 and rose to 29% of explained variance in 2009. The rotated component matrices for years 2007 and 2009 are presented in the Appendix.

Figure 5 presents results for a multiple year PCA using a three-year rolling window. I.e., the PCA in 2012 is calculated using financial statement data from years 2010, 2011, and 2012 where respective financial statement data are collected annually on the corresponding cut-off dates.

Compared to a single year PCA, the value of the KMO test and share of the explained variance is less volatile form year to year. The KMO values are high and range between 0.76 and 0.82. The first six principal components account for 54% to 61% of the total variance. The explained variance by each principal component is relatively constant, and only a slight increase in the explained variance by the first principal component is present from 2007 to 2010.

Table 6 represents the rotated component matrix of initial six principal components using the end of sample data period from 2014 to 2016. The table further demonstrates the stability of the PCA factor structure. The accounting ratios are grouped in the same manner as using the data from 2001 to 2003. However, in period 2014-2016 business growth is the dominant factor, accounting for 19% of the total variance. Margins factor is significantly less prominent compared to the period 2001-2003, accounting for only 7% of total variance as opposed to 19% in 2001-2003

Figure 4. Principal components and KMO test (annual data; 2001-2016)

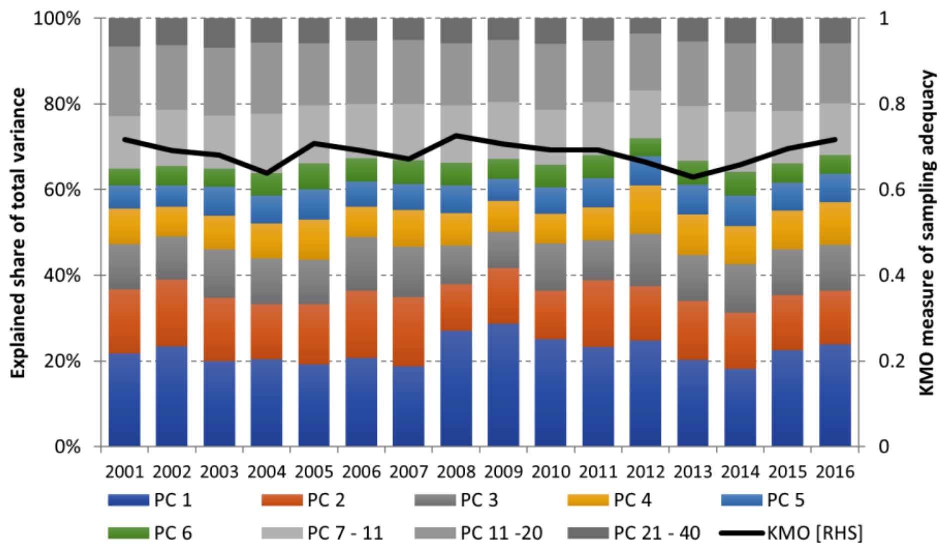


Figure 5. Principal components and KMO test (three-year rolling window; 2003-2016)

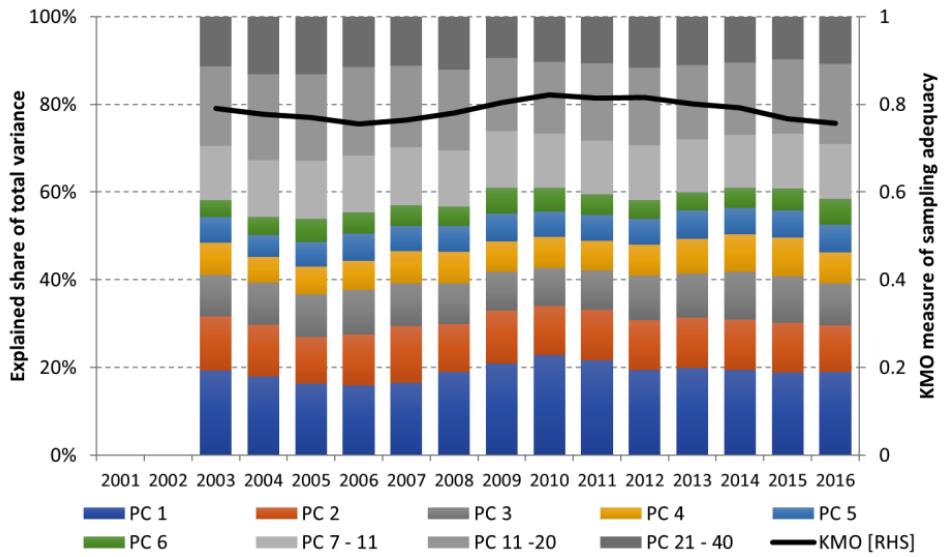


Table 6. Rotated component matrix (2014 - 2016)

	Accounting ratio	**PC 1**	**PC 2**	**PC 3**	**PC 4**	**PC 5**	**PC 6**
Operating growth	OPER_INC_GROWTH	**0.36**	0.00	0.00	0.02	0.03	0.03
	NET_INC_GROWTH	**0.38**	-0.03	-0.02	0.01	-0.04	-0.06
	EBITDA_GROWTH	**0.32**	0.03	0.01	0.02	0.05	0.08
	CONT_INC_GROWTH	**0.40**	-0.06	-0.02	0.00	-0.05	-0.06
	EPS_GROWTH	**0.41**	-0.04	0.00	0.01	-0.04	-0.07
	INC_TAX_EXP_YR_GROWTH	0.21	0.07	0.03	0.00	-0.02	0.04
	EBIT_YR_GROWTH	**0.35**	0.01	-0.03	0.02	0.05	0.06
	REINVEST_EARN_YR_GROWTH	**0.30**	0.00	-0.01	-0.05	0.01	-0.01
Indebtedness	WACC	0.00	**-0.27**	0.14	-0.01	0.15	-0.02
	TOT_DEBT_TO_TOT_EQY	0.02	**0.42**	0.04	0.04	0.04	-0.01
	NET_DEBT_TO_CASHFLOW	-0.03	**0.26**	-0.07	0.02	-0.06	-0.06
	TOT_DEBT_TO_TOT_ASSET	0.00	**0.42**	0.10	0.04	0.04	0.02
	TOT_DEBT_TO_EBITDA	-0.02	**0.35**	-0.09	-0.02	0.03	0.03
	COM_EQY_TO_TOT_ASSET	-0.03	**-0.35**	0.00	0.08	0.02	0.03
	TOT_DEBT_TO_TOT_CAP	0.02	**0.44**	0.04	0.00	0.02	0.00
Profitability	NORMALIZED_ROE	0.00	0.10	**0.43**	-0.01	0.05	0.02
	RETURN_ON_INV_CAPITAL	0.01	-0.02	**0.42**	-0.02	-0.05	0.00
	RETURN_ON_ASSET	-0.03	-0.09	**0.39**	0.08	-0.03	-0.01
	CAP_EXPEND_RATIO	-0.01	-0.05	**0.09**	0.06	-0.04	-0.01
	RETURN_COM_EQY	0.03	0.07	**0.43**	-0.01	0.00	-0.01
	RETURN_ON_CAP	0.01	-0.10	**0.40**	-0.05	-0.04	0.01
Margins	NORMALIZED_PROFIT_MARGIN	-0.01	-0.01	-0.03	**0.42**	0.00	-0.05
	FREE_CASH_FLOW_MARGIN	0.04	0.00	-0.03	**0.31**	0.06	0.04
	EBIT_MARGIN	-0.02	0.04	0.05	**0.43**	-0.01	0.01
	EBITDA_MARGIN	-0.02	-0.07	-0.11	**0.34**	-0.05	-0.03
	PROF_MARGIN	0.02	-0.01	0.06	**0.41**	0.02	0.01
	OPER_MARGIN	0.02	0.04	0.06	**0.43**	0.01	0.05
	GROSS_MARGIN	-0.04	0.01	0.11	0.13	0.06	-0.10
Volatility	BETA_RAW_OVERRIDABLE	0.05	-0.04	-0.14	0.04	**0.32**	-0.03
	VOLATILITY_260D	-0.05	-0.01	-0.05	0.02	**0.50**	0.04
	VOLATILITY_180D	-0.02	0.00	0.02	-0.01	**0.51**	-0.01
	VOLATILITY_90D	0.01	0.01	0.02	0.00	**0.53**	-0.03
Company growth	EMPL_GROWTH	0.01	-0.06	0.06	-0.07	0.10	0.23
	ASSET_GROWTH	-0.01	0.00	0.05	-0.03	0.06	**0.41**
	SALES_GROWTH	0.15	0.06	0.09	-0.07	0.11	**0.25**
	NET_WORTH_GROWTH	-0.02	-0.03	-0.10	0.02	-0.09	**0.45**
	GROWTH_IN_CAP	-0.04	0.00	0.06	-0.01	0.01	**0.45**
	BVPS_GROWTH	-0.02	-0.02	-0.07	0.02	-0.07	**0.44**
	WORK_CAP_GROWTH	-0.03	0.01	-0.03	0.13	-0.08	0.15
	CASH_FLOW_GROWTH	0.05	0.02	-0.03	0.06	-0.03	0.22

Note. Coefficients with an absolute value greater than 0.25 are bolded.

5.3 Discussion

The PCA is a suitable tool to group accounting ratios and represent the original data using new independent variables. By applying rotation, we were able to determine six financial statement factors, each with a distinct economic meaning. We demonstrated that derived factors have stable structure across the entire analyzed period. Moreover, weights for each accounting ratio do not change significantly and have a meaningful argument signs.

Admittedly, a few accounting ratios tend to shift year to year and are associated with different factors (e.g. employee growth). However, they tend to be only weakly associated with the corresponding factor and have similarly small weights for two or more factors.

This shifting of accounting ratios is much less pronounced when three years of data is used in the PCA compared to one year only. This demonstrates that changes in the principal component structure are small and could be attributed to small sample size effect and the associated noise. Using multiple years of data thus induces stability in the PCA and results in more stable factor structure. However, this is achieved at the expense of a poorer sensitivity to possible structural changes in the dataset from year to year. Nevertheless, the chosen three-year period seems reasonable to mitigate noisy data issues while maintaining sufficient sensitivity to changes in the data series.

Although the factor structure is highly coherent, their corresponding eigenvalues vary considerably. In other words, the explained share of total variance by each factor depends on the relative variability of the corresponding accounting ratios. E.g., this is evident when comparing rotated component matrix for years 2007 and 2009. As a result of financial turmoil in 2008 and beginning of 2009, market volatility rose sharply. The volatility factor thus accounted for 8% and 29% in 2007 and 2009, respectively. Again, a three-year PCA produces more stable results across the entire sample period.

6 APPLICATION OF INVESTMENT STRATEGY

We use the derived financial statement factors to determine the company's intrinsic value. In year *t* we collect financial statement data across all companies currently representing the S&P Index and calculate six financial statement factors using the means of the PCA as described above. In the following regression analysis, the factors are then used as explanatory variables and selected price multiple as the independent variable. In this manner, we obtain regression betas and calculate implied price multiple. The latter is presumed to reflect the intrinsic value based on company's financial statements data (see Section 4.2.2 for more details).

We then apply the investment strategy as outlined in Section 4.2.3. We construct an equally weighted portfolio by selecting 20 relatively most undervalued stocks. The constructed active portfolio is held unchanged for a period of one year. In the year $t+1$, the value of the positions in the portfolio is calculated using most recent market prices, yielding one-year active portfolio return. Next, we update the list of S&P Index constituents and repeat the procedure outlined above. Namely, we collect updated financial statement data, perform the calculation of implied price multiples, and construct a revised portfolio of relatively most undervalued stocks.

First, we focus on the price-to-earnings (P/E) multiple as arguably the most common and widely used measure of relative value. We test three different investment strategies, which are all based on financial statement factors determined using the PCA but differ in length of the historical moving window and use of Kalman filter algorithm. In the next step, we use the most perspective investment strategy and apply it to alternative price multiples, i.e. price-to-sales (P/S), price-to-book (P/B), price-to-EBITDA (P/EBITDA), and price-to-free cash flow (P/FCF).

We compare the investment strategies using a number of risk-adjusted indicators. The S&P 100 Index portfolio serves as the benchmark and the 1-year U.S. Treasury note is a proxy for a risk-free investment. All investment strategies are evaluated in period March 2003 to March 2017.

6.1 Investment strategy using P/E ratio

6.1.1 Investment strategy using single year PCA

In the first investment strategy (*Strategy I*) we use only the most recent data corresponding to the cut-off date in year t. This imposes that the model and investment strategy are highly responsive to the changes in the environment and interdependency of the data. As such, it represents a very short-term view, which is particularly susceptible to market sentiment and perceptions.

In year $t = 2001$, we derive the principal components using the rotated component matrix in Table 5. As discussed above, we refer to the obtained principal components as factors representing profitability, volatility, etc. Next, we relate the factors to P/E ratios obtained on the cut-off date in year $t = 2001$ using the OLS regression.

The overall model fit is presented in Table 7. The model is statistically significant at $p = 0.001$ (F-statistics). The total explained share of the variance of the P/E ratio explained by financial statement factors is 22.8%. The R-squared value suggests idiosyncratic effects importantly contribute to the variability of P/E ratio. However, the value is in line with similar studies of cross-sectional asset returns. The Q-Q plot and associated distribution of the residuals are presented in the Appendix. The two figures demonstrate that residuals are well approximated by a normal distribution.

The model parameter estimates are listed in Table 8. We observe three financial statement factors that are statistically significant at $p \leq 0.05$, i.e. Volatility, Operating growth, and Indebtedness. The three factors have a similar effect on the P/E ratio, denoted by the standardized beta values. We notice that volatility factor is positively related to the P/E ratio, thus increasing the value of the company. Operating growth reflects improvements in operating activities and growth in margins and is a positive indicator of company's value. Conversely, higher debt levels are diminishing the P/E ratio. The remaining factors have an order of magnitude smaller effect on the P/E ratio and are statistically insignificant. Their respective betas are thus suppressed in the calculation of implied P/E ratio.

Table 7. Regression statistics (Strategy I; 2001)

	R Squared	Adjusted R Squared	F-value	Significance
Strategy I	0.228	0.173	4.17	0.0010

Table 8. Regression parameters (Strategy I; 2001)

		Beta	Standardized Beta	Std. Error	t-value	p-value
Margins	PC1	0.006	0.026	0.028	0.233	0.817
Volatility	PC2	0.068*	0.221	0.033	2.061	0.042
Operating growth	PC3	0.078**	0.295	0.027	2.879	0.005
Indebtedness	PC4	-0.079**	-0.318	0.026	-3.031	0.003
Company growth	PC5	-0.014	-0.044	0.034	-0.430	0.668
Profitability	PC6	0.033	0.089	0.042	0.787	0.434

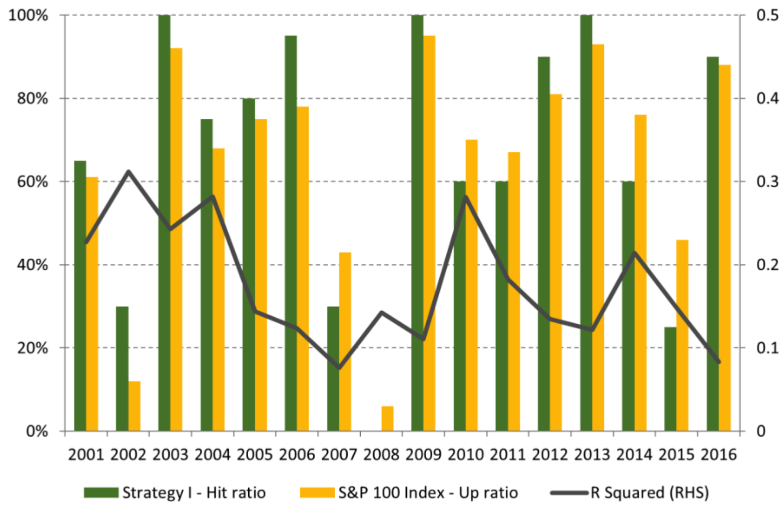
Note. Parameter are statistically significant at p = 0.05 (*), p = 0.01 (**), and p = 0.001 (***).

We then construct an equally weighted portfolio by selecting 20 relatively most undervalued stocks and evaluate the performance after one year, $t+1 = 2002$. The investment strategy procedure is then repeated each consecutive year. However, a detailed and systematic analysis of betas for each year is cumbersome. Namely, although the overall factor structure is constant, the ranking of factors changes. Additionally, the argument signs of factor constituents can change, thus introducing an additional degree of complexity. The analysis shows that statistical significance of factors varies year to year, however, the number of statistically significant factors does not fluctuate considerably.

Figure 6 presents annual returns for *Strategy I* compared to returns of the S&P 100 Index and 1-year U.S. Treasury note. The successfulness of the investment strategy to select undervalued stock over a one-year horizon is shown in Figure 7. The Hit ratio reflects the percentage of selected stocks that increased in value, while the comparable Up ratio for the S&P 100 Index reflects the percentage of stocks in the index that increased in value.

Figure 6. Annual returns (Strategy I; 2001-2016)

Figure 7. Hit ratio and R-squared (Strategy I; 2001-2016)



In the period 2004-2017 *Strategy I* on average achieved annual return of 14.7%. This is considerably higher than average return of the S&P 100 Index of 7.8%. However, the volatility of returns in *Strategy I* (28.9%) is substantially higher than the volatility of the S&P 100 Index (16.8%). The *Strategy I* Hit ratio and the S&P 100 Up ratio are similar and equal 69% and 70%, respectively. This suggests the performance of *Strategy I* is strongly correlated with the S&P 100 Index. The percentage of selected stock with a positive return is thus on par with the overall performance of the S&P 100 Index, however, *Strategy I* on average selects stocks, which yield a higher return than the S&P 100 Index. The paired samples t-test confirms the superior performance of *Strategy I* at p=0.05 significance level. The t-test summary is presented in Table 9.

Table 9. Paired samples test (Strategy I; 2001-2016)

	Paired Differences					**t**	**df**	**Sig. (1-tailed)**
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Strategy I - S&P 100	0.07189	0.12832	0.03208	0.00351	0.14027	2.241	15	0.021

6.1.2 Investment strategy using multiple year PCA

In the second investment strategy (*Strategy II*), we use a three-year rolling window to determine financial statement factors and calculate the regression betas. As such, the model incorporates a more long-term view compared to *Strategy I*. In turn, the model is less responsive to market swings and induces greater stability throughout the model. At the same time, the moving window approach allows the model to adapt to the changes in the changes in underlying data characteristics.

Starting in year t = 2003, we derived the financial statement factors as presented Table 3. We recall that factors are determined using financial statement data from years 2001, 2002, and 2003. Next, we perform the OLS regression where factors are used as independent variables and the P/E ratios from the same historical window as the dependent variable.

The overall model fit is presented in Table 10. The model is statistically highly significant at p = 0.00001 (F-statistics). The explained share of the variance of the P/E ratio is again somewhat low (R^2=18.1%), albeit similar as in *Strategy I* despite longer historical window. The Q-Q plot and associated distribution of the residuals are presented in the Appendix. The two figures demonstrate a near-normal distribution of the data and residuals, albeit slight deviations are present in the tails of the distribution.

The model parameter estimates are listed in Table 11. Compared to *Strategy I*, factor betas exhibit higher statistical significance. We attribute that to a more stable factor structure and a larger dataset. We find three financial statement factors are statistically significant at p <= 0.05. The indebtedness factor exhibits the greatest effect on the P/E ratio, denoted by the standardized beta values, followed by profitability factor and company growth factor. Indebtedness increases the riskiness of the company and, in turn, implies a lower P/E ratio. Also, profitability factor and company growth factor exhibit negative dependency with the P/E ratio. Higher profitability implies higher earnings, which increase the denominator of the P/E ratio. The negative dependency suggests that on average the change in price does not sufficiently compensate earnings increase. Company growth factor relates to the growth in the balance sheet items. However, negative dependency suggests that such growth does not, on average, translate into higher P/E ratio, but rather diminishes it. The remaining factors have an order of magnitude smaller effect on the P/E ratio and are statistically insignificant. Their respective betas are thus suppressed in the calculation of implied P/E ratio.

Table 10. Regression statistics (Strategy II; 2003)

	R Squared	**Adjusted R Squared**	**F-value**	**Significance**
Strategy II	0.181	0.162	9.72	0.0000

Table 11. Regression parameters (Strategy II; 2003)

		Beta	**Standardized Beta**	**Std. Error**	**t-value**	**p-value**
Margins	PC1	0.025	0.044	0.036	0.701	0.484
Indebtedness	PC2	-0.211***	-0.389	0.033	-6.323	0.000
Operating growth	PC3	0.040	0.068	0.034	1.166	0.245
Volatility	PC4	0.039	0.058	0.041	0.947	0.345
Company growth	PC5	-0.096*	-0.136	0.041	-2.345	0.020
Profitability	PC6	-0.164**	-0.234	0.047	-3.514	0.001

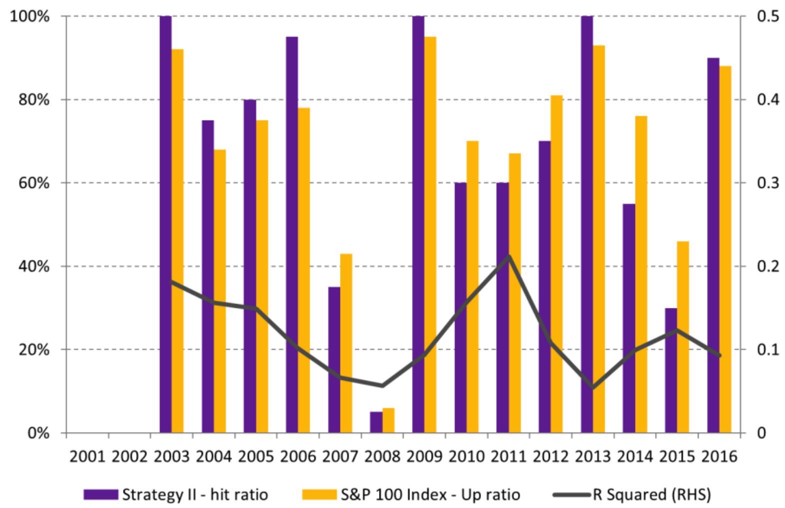
Note. Parameter are statistically significant at p = 0.05 (*), p = 0.01 (**), and p = 0.001 (***).

We then construct an equally weighted portfolio by selecting 20 relatively most undervalued stocks. However, only companies representing the S&P 100 Index in current year t represent the eligible investment universe. We evaluate the performance after one year, t+1= 2004. The historical data series is then adjusted to include most recent year t+1 while dropping the most distant year (i.e. t-2). The investment strategy procedure is then repeated as described above. The variations in factor ranking and argument signs of factor loadings complicate the systematic analysis of factor betas on a yearly basis. Similarly to *Strategy I*, the statistical significance of factors varies, however, the number of statistically significant factors does not fluctuate considerably.

Figure 8 presents annual returns for *Strategy II* compared to returns of the S&P 100 Index and 1-year U.S. Treasury note. The successfulness of investment strategy to select undervalued stock over a one-year horizon is shown in Figure 9. The Hit ratio reflects the proportion of selected stocks that increased in value, while the comparable Up ratio for the S&P 100 Index reflects the ratio of stocks in the index that increased in value.

Figure 8. Annual returns (Strategy II; 2003-2016)

Figure 9. Hit ratio and R-squared (Strategy II; 2003-2016)



In the period 2004-2017 *Strategy II* on average achieved annual return of 14.1%. This is considerably higher that average return of the S&P 100 Index of 7.8%. Again, the volatility of active *Strategy II* (27.4%) is significantly higher than the volatility of the S&P 100 Index (16.8%). The average *Strategy II* Hit ratio and the S&P 100 Up ratio are 68% and 70%, respectively. This suggests the performance of *Strategy II* is highly correlated with the S&P 100 Index and percentage of selected stock with a positive return is on par with the overall performance of the S&P 100 Index. However, *Strategy II* on average selects stocks, which yield considerably higher return than the S&P 100. The paired samples t-test confirms that performance of *Strategy II* is superior to the performance of S&P 100 at significance level p=0.05. The t-test summary is presented in Table 12.

Table 12. Paired samples test (Strategy II; 2003-2016)

	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Strategy II - S&P 100	0.06218	0.1269	0.03392	-0.01109	0.13545	1.833	13	0.045

6.1.3 Investment strategy using PCA and KF

The third investment strategy (*Strategy III*) also applies PCA and additionally introduces Kalman filtering. We demonstrated in Section 5.2 that a three-year historical window yields a factor structure, which is consistent in different historical periods and only varies in the respective factor ranking. To that end, we adopt the factor structure (i.e. loadings) as determined based on period 2001-2003 and is presented in Table 3. Hereafter we assume that determined factor structure remains valid throughout the analyzed period 2001-2016.

We proceed by obtaining the initial regression coefficients. We apply the OLS regression for the year 2001, where factors are used as independent variables and P/E ratios as the dependent variable. The overall model fit is presented in Table 13. The model is statistically highly significant at $p = 0.001$ (F-statistics). We note that the explained share of the variance of P/E ratio is higher than for *Strategy I* and *Strategy II* (R^2=25.3%). The Q-Q plot and distribution of the residuals exhibit a near-normal distribution and are presented in the Appendix.

The model parameter estimates are listed in Table 14. Operating growth factor and volatility are statistically significant at $p <= 0.05$. The former has three times larger effect on the P/E ratio as the latter, as denoted by the standardized beta values. As noted previously, operating growth reflects improvements in operating activities and growth in profitability, and can thus be understood as a positive indicator of company's value. We again find volatility factor as a positive determinant of equity value.

Table 13. Regression statistics (Strategy III; 2001)

	R Squared	Adjusted R Squared	F-value	Significance
Strategy II	0.253	0.2	4.8	0.0003

Table 14. Regression parameters (Strategy III; 2001)

		Beta	Standardized Beta	Std. Error	t-value	p-value
Margins	PC1	0.039	0.146	0.055	-0.074	0.941
Indebtedness	PC2	-0.115	-0.440	0.031	1.285	0.202
Operating growth	PC3	0.066***	0.244	0.029	-4.005	0.000
Volatility	PC4	0.025*	0.077	0.029	2.294	0.024
Company growth	PC5	-0.034	-0.101	0.035	0.705	0.483
Profitability	PC6	-0.084	-0.257	0.035	-0.984	0.328

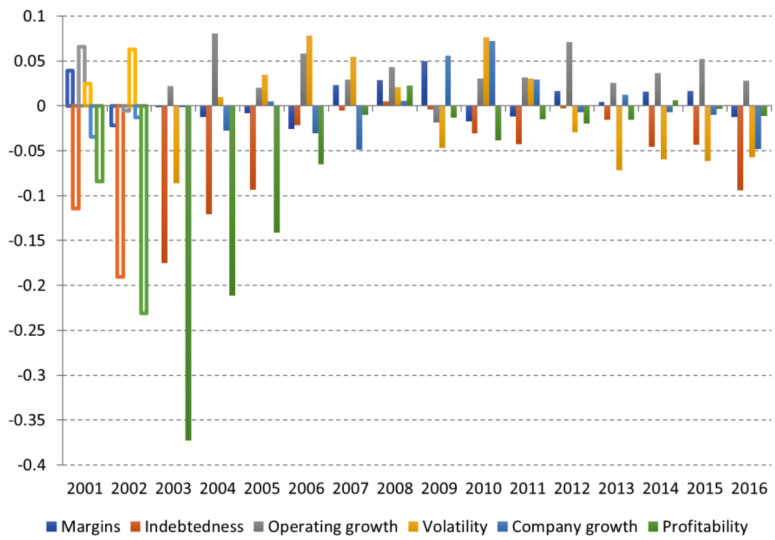
Note. Parameter are statistically significant at p = 0.05 (*), p = 0.01 (**), and p = 0.001 (***).

The calculated factor betas are used as an initial starting point for Kalman filter algorithm. In the subsequent years, we retain the initial factor loadings and only calculate financial statement factors using the most recent data. The Kalman filter algorithm is then utilized to amend factor betas. Based on the updated regression equation we calculate implied P/E ratio and construct an equally weighted portfolio consisting of 20 relatively most undervalued stocks. The additional two years (2001-2003) are used as an adjustment period for Kalman filter, and the performance of *Strategy III* is evaluated in the period 2003-2016.

Using the described approach, factors betas dynamically adapt to the changes in the market environment and underlying data characteristics. This eliminates the need to select the moving window length beforehand, while the model implicitly utilizes the entire data history. As such, it combines responsiveness to abrupt market swings and stability of parameters over a longer term. Moreover, since factor structure and ranking is fixed, the evolution of corresponding factor betas can be performed.

Figure 10 presents the evolution of factor betas. The figure gives an interesting insight into how the market prices financial statement factors in different periods. E.g., at the beginning of the sample period, 2003-2005, profitability and indebtedness were the main discriminating factors of company value. In the remaining years, their effect subsided while remaining mostly negative. We also notice indebtedness beta is increasing in absolute terms from 2012 onward. The negative relation suggests that companies with higher debt have P/E ratio below average. Higher profitability implies higher earnings, which increase the denominator of the P/E ratio. The negative beta of the profitability factor suggests that on average the change in stock price does not sufficiently compensate earnings increase generated by a higher profitability. Operating growth is, with the exception of 2009, mainly considered as a positive factor for the P/E ratio. Volatility, margins, and company growth exhibit shifting relations with the P/E ratio.

Figure 10. Evolution of factor betas (Strategy III; 2001-2016)



Note. Open bars denote the in-sample testing period, while solid bars correspond to the out-of-sample period.

Figure 11 presents factor betas for the subset of the sample period, 2006-2010. We notice significant changes in factor betas during the period of global financial crisis, as many reverse their dynamics in the years 2008 and 2009. E.g., company growth factor beta is significantly negative before 2008 but becomes a significant positive determinant of the P/E ratio afterward. Profitability, which is mainly negatively related to the P/E ratio, appears highly sought during the peak period of the global financial crisis in 2008. Volatility, which is mainly a positive determinant for the P/E ratio, is deeply negative in 2009, which suggests market's high aversion to uncertainty. Likewise, margins factor is seen by the market as highly valuable in the aftermath of the global financial crisis in 2009. While indebtedness is not a significant discriminating factor before and during the financial crisis (2007-2009), the market seems to develop a stronger aversion to higher indebtedness in the following year 2010. Admittedly, some share of beta volatility can be attributed to the measurement noise and Kalman filtering procedure. Nevertheless, the stable response of Kalman filter in the initial three-year training period suggests that changes in the financial statement data are a primary driver of changes of beta.

Figure 11. Evolution of factor betas – a subset (Strategy III; 2006-2010)

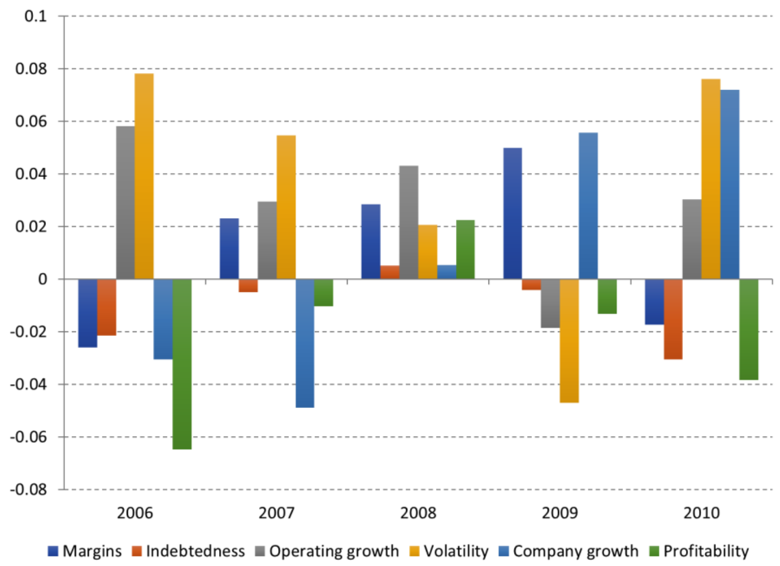


Figure 12 presents annual returns for *Strategy III* compared to the return of the S&P 100 Index and 1-year U.S. Treasury note. The ability of investment strategy to select undervalued stocks over a one-year horizon is in shown in Figure 13. The Hit ratio reflects the proportion of selected stocks that increased in value, while the comparable Up ratio for S&P 100 Index reflects the ratio of stocks in the index that increased in value.

Figure 12. Annual returns (Strategy III; 2001-2016)

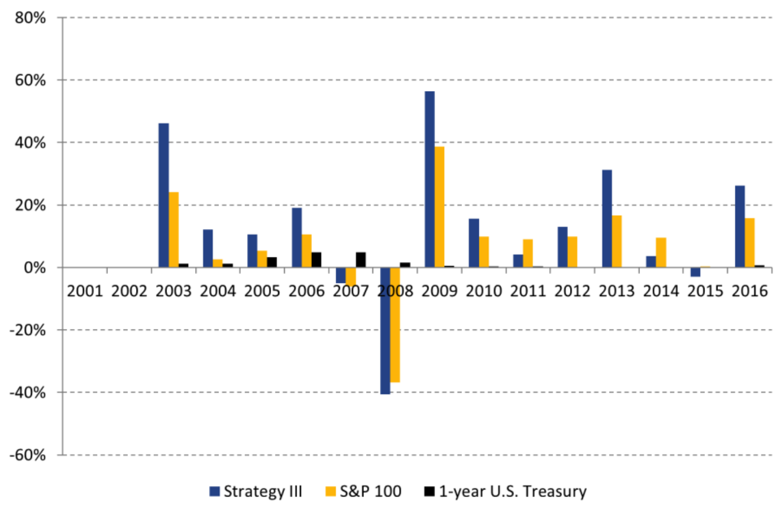
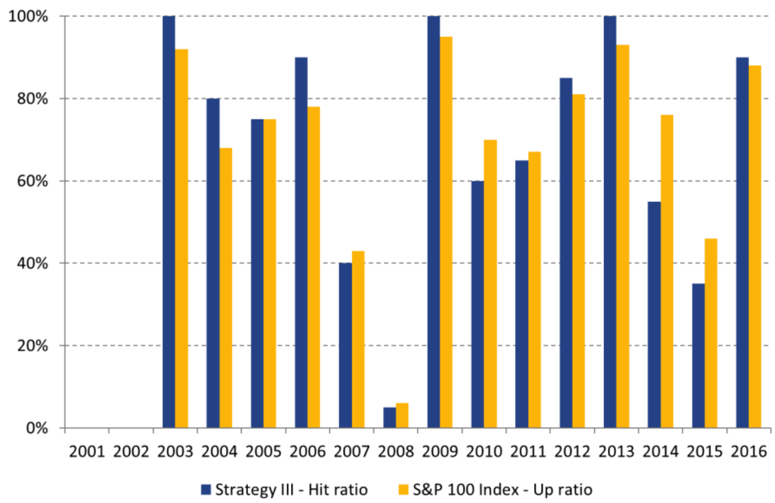


Figure 13. Hit ratio and R-squared (Strategy III; 2003-2016)



In the period 2004-2017 *Strategy III* on average achieved annual return of 13.5%. This is again higher than the average return of the S&P 100 Index of 7.8%, but lower than the return of *Strategy I* and *Strategy II*. However, the volatility of *Strategy III* (23.5%) is lower compared to *Strategy I* and *Strategy II*, albeit still considerably higher than the volatility of the S&P 100 Index (16.8%). The average *Strategy III* Hit ratio and the S&P 100 Up ratio are equal to 70%. Again, the performance of *Strategy III* is strongly correlated with the S&P 100 Index, and percentage of selected stock with a positive return is on par with the overall performance of the S&P 100 Index. However, *Strategy III* on average selects stocks, which yield considerably higher return than the S&P 100. The paired samples t-test confirms that performance of *Strategy III* is statistically significantly higher than the performance of the S&P 100 Index at significance level p=0.05. The t-test summary is presented in Table 15.

Table 15. Paired samples test (Strategy III; 2003-2016)

	Paired Differences					**t**	**df**	**Sig. (1-tailed)**
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Strategy III - S&P 100	0.05686	0.08674	0.02318	0.00678	0.10695	2.453	13	0.015

6.1.4 Discussion

We compare the three proposed investment strategies using a set of risk-adjusted performance measurement indicators. We use the S&P 100 Index portfolio as the market benchmark, and 1-year U.S. Treasury notes serve as a proxy for a risk-free investment.

The analyzed period spans 14 years from 2003 to 2016. The performance indicators are presented in Table 16.

All three strategies result in statistically significant higher average return (see Table 9, Table 12, Table 15) compared to the S&P 100 Index. *Strategy I* yields highest average return among active investment strategies. The strategy incorporates only one year of historical data and is thus most responsive to changes in market conditions and hence also most volatile, as denoted by high standard deviation of the returns.

On the contrary, *Strategy II* incorporates a longer historical window. The associated factor structure is thus more stable compared to *Strategy I*, the overall model also has higher statistical significance, and factor betas are determined at a lower significance level. The overall performance of *Strategy II* is comparable to *Strategy I*. However since the former is less responsive to the market sentiment, volatility, average return, and total return are slightly lower. When return is adjusted for the risk incurred, the performance of S*trategy I* and *Strategy II* are almost equivalent. The maximum drawdown, which reflects downside risk, favors a more stable *Strategy II*. However, the measures comparing portfolio's excess return to the benchmark and risk-free investment (M^2, Jensen' alpha) are marginally higher for *Strategy I*.

Strategy III utilizes the Kalman filter algorithm to derive factor betas. This eliminates the selection of the historical window and indirectly uses the entire available data history. Additionally, it allows the model to optimally adapt betas when new information is available. The approach also enables a straightforward tracking of factor betas and offers a transparent insight into how market prices financial statement factors in different periods.

Strategy III is also a superior active strategy based on all risk-adjusted indicators. Although the average return of *Strategy III* is slightly lower than for other two active portfolios, the total return and CAGR are highest for *Strategy III*. *Strategy III* also has the smallest volatility and tracking error, thus indicating the smallest excess deviations from a broad market S&P 100 Index.

We note that Information ratio is positive for all active strategies. This indicates that active strategies yield sufficient return for deviations from the benchmark. Similarly, Sharpe ratio of all active strategies exceeds benchmark's Sharpe ratio. This shows that active strategies yield sufficient risk premium compared to their total risk. The M^2 values demonstrate, in percentage terms, that active portfolios' outperformance is significant on a risk-adjusted basis. The outperformance of *Strategy III* is approximately 70% higher compared to *Strategy I* and *Strategy II* and equals 2.23% on a risk-adjusted basis. Similarly, Jensen's alpha is positive for all active strategies and highest for *Strategy III*, suggesting highest superior return due to skill in stock selection. Since active strategies are more volatile, the downside risk (i.e. MDD) exceeds the benchmark. However, the MDD for *Strategy III* is only slightly higher than the benchmark's MDD.

Table 16. Trading strategy performance (2003-2016)

	Strategy I	***Strategy II***	***Strategy III***	**S&P 100 Index**	**1-year U.S. Treasury**
Average return (in %)	**14.75**	14.05	13.52	7.83	1.40
Std. of return (in %)	28.90	27.44	**23.47**	16.78	1.70
Total return (in %)	330.04	317.58	**330.18**	138.86	21.23
CAGR (in %)	10.98	10.75	**10.98**	6.42	1.38
Sharpe Ratio	0.48	0.48	**0.54**	0.40	0
M^2 (in %)	1.31	1.30	**2.23**	0	-
Jensen's alpha (in %)	2.73	2.67	**3.47**	0	-
IR	0.52	0.51	**0.68**	-	-
TE	0.13	0.12	**0.08**	-	-
MDD (in %)	52.46	48.54	**43.68**	40.69	0

Note. Bolded values denote a superior strategy in each category.

In summary, *Strategy III* is a superior active strategy in terms of performance measurement indicators. In addition, we consider *Strategy III* as favorable due to its methodological advantages associated with the use of Kalman filter. To that end, we hereafter focus on *Strategy III* as our preferred investment strategy.

We look more closely at the return generation in *Strategy III*. Figure 14 presents the distribution of invested funds by sector. The investment strategy actively adapts allocation in response to market conditions and correspondingly over- and underweights specific sectors. Overall, the funds are well distributed across different sectors. The highest and most stable percentage of funds is allocated to Investment Technology sector, which on average account for 20%. Allocation to Financials sector on average accounts for 19% and exhibits high volatility. E.g., following the peak of financial crisis its share in 2009 dropped to 5%, but increased steadily afterward and reached 40% in 2016. Healthcare and Energy are also more prominent sectors, which on average account for 12% and 14%, respectively.

In Figure 15 we present contribution of sectors to the annual return. On average, Information Technology sector contributes the highest share of total return, 3.85%, followed by Financials sector, which on average contributes 2.98%. Financials sector also has the highest volatility-adjusted return and is also the only sector, which had a positive return contribution in 2008, 0.44%. Relatively strongest performance is attributed to Utilities sector, which on average contributes 1.08% of total return where its share of allocated funds is only 4% on average. Materials and Real Estate sector are weakest performers in term of absolute and relative return contribution.

Figure 14. Distribution of invested funds by sector (Strategy III; 2003-2016)

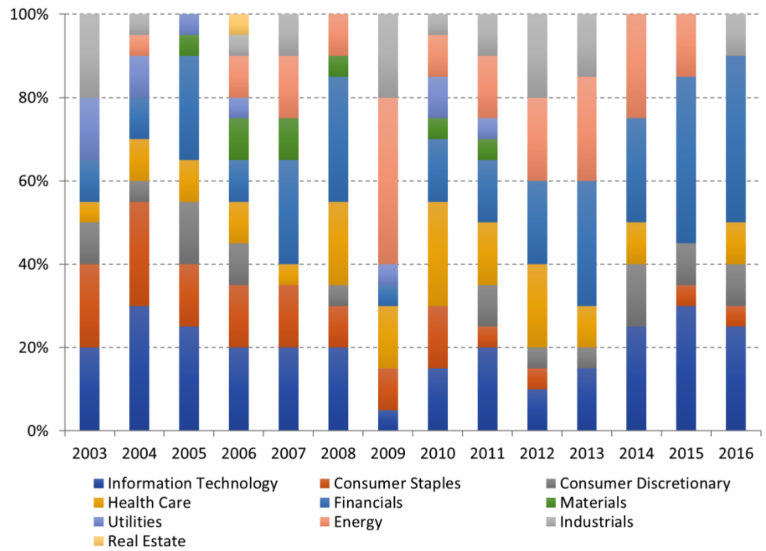
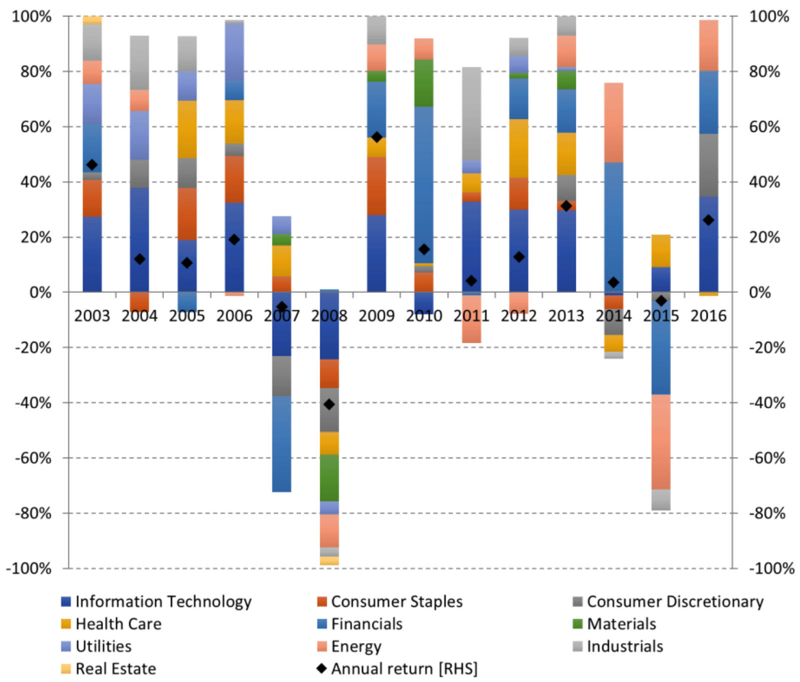


Figure 15. Contribution of sectors to annual return (Strategy III; 2003-2016)



6.2 Investment strategy using other valuation ratios

In Chapter 6.1 we discussed three proposed investment strategies using the P/E ratio. We compared the strategies based on their narrative, statistical properties of the model, and performance using a number of risk-adjusted measures. We find that proposed *Strategy III*, which is based on a three-year PCA and utilizes Kalman filter, gives superior results and allows a meaningful analysis of the relation between price ratio and financial statement factors.

Here we apply *Strategy III* using other commonly used price ratios, i.e. price-to-sales (P/S), price-to-book (P/B), price-to-EBITDA (P/EBITDA), and price-to-free cash flow (P/FCF). To that end, we follow the same methodological procedure as presented for the P/E ratio in Chapter 6.1.3. The only distinction is the choice of the independent variable in the linear regression model. The overall model fit is presented in Table 17 and model parameter estimates are listed in Table 18.

Among the presented models, only price-to-free cash flow (P/FCF) ratio yields statistically insignificant results. This indicates the P/FCF ratio is not appropriate for equity value estimation using the proposed financial statement factor analysis. A possible reason might be that the P/FCF ratio focuses on realized cash flows but neglects the other contributions to a broader economic value such as, e.g., deferred revenue. The other models are statistically highly significant, and their explained shares of the total variance (R^2) are also higher than for the P/E ratio model. The same findings were also observed in the literature (Damodaran, 2012).

Table 17. Regression statistics of different price ratios (Strategy III; 2001)

	R Squared	**Adjusted R Squared**	**F-value**	**Significance**
P/E	0.253	0.200	4.80	0.0003
P/S	0.550	0.519	18.10	0.0000
P/FCF	0.091	0.010	1.12	0.3600
P/B	0.389	0.348	9.45	0.0000
P/EBITDA	0.490	0.448	11.70	0.0000

Table 18. Regression parameters of different price ratios (Strategy III; 2001)

		P/E	**P/S**	**P/FCF**	**P/B**	**P/EBITDA**
Margins	PC1	0.039	0.266***	-0.039	0.109**	0.022
Indebtedness	PC2	-0.115	-0.170***	-0.148*	-0.006	-0.238***
Operating growth	PC3	0.066***	0.034	0.024	0.038	0.073*
Volatility	PC4	0.025*	0.025	0.002	0.023	0.003
Company growth	PC5	-0.034	-0.006	-0.036	-0.057	-0.063
Profitability	PC6	-0.084	-0.082	-0.097	0.182***	-0.003

Note. Parameter are statistically significant at p = 0.05 (*), p = 0.01 (**), and p = 0.001 (***).

We notice each price ratio is uniquely related to financial statement factors. E.g., the main determinant of the P/S ratios is the margins factor. We recall that the P/S ratio is internally inconsistent. Namely, the market value of equity is compared to revenue, which is distributed among all stakeholders. The financial leverage significantly impacts the P/S ratio, which is reflected in statistically significant and negative indebtedness beta. For the P/B ratio, the profitability factor and margins factor are the key determinants. Both factors reflect how skillfully the company is using the available resources to generate a return on invested equity. The P/EBITDA ratio is closely related to the P/E ratio but is calculated before adjustment for capital expenditures and debt servicing obligations. Thus, the financial leverage will significantly impact the valuation, which is denoted by a negative indebtedness beta. Operating growth is a positive indicator of company's value as it reflects improvements in operating activities and growth in profitability.

We compare the gross performance of active trading strategies based on different price ratios using a set of performance measurement indicators. The S&P 100 Index portfolio is used as a market benchmark, and 1-year U.S. Treasury notes serve as a proxy for the risk-free investment. The results are presented in Table 19. The P/FCF ratio strategy is omitted since it is not statistically significant.

The P/S strategy on average yields the highest return of 17.31%. However, the associated volatility and downside risk (MDD) is high. Although the P/S strategy does outperform the benchmark index on a risk-adjusted basis (Sharpe ratio, M^2), its performance is inferior to the P/E strategy.

The P/EBITDA strategy yields the lowest return among active trading strategies, but still considerably exceeds the benchmark. The strategy also has the lowest volatility, which is also only slightly higher than the benchmark volatility. Importantly, the strategy has the lowest downside risk, and the maximum drawdown (MDD) is lower than the benchmark MDD. Nonetheless, the P/EBITDA strategy performance exceeds the P/S strategy and the P/E strategy on a risk-adjusted basis.

The P/B strategy yields high average return and has superior CAGR. The volatility is high but similar to the P/E strategy. The deviations from the benchmark, denoted by the tracking error, yield the highest excess return, which is reflected in high information ratio. On a risk-adjusted basis, the P/B strategy also shows superior performance where all indicators (Sharpe Ratio, Jensen's alpha, and M^2) demonstrate a high excess return.

Among the presented active strategies, the P/B strategy demonstrates a superior return, albeit return volatility is high. A more risk-averse investor might thus prefer the P/EBITDA strategy, which yields lower return but offers protection against high volatility and excess downside risk. The P/E strategy represents a balanced approach compared to the P/B strategy and the P/EBITDA strategy. The performance of the P/S strategy is inferior to all other active strategies.

Table 19: Algorithm gross performance of different price ratios (Strategy III; 2003-2016)

	P/E	P/S	P/B	P/EBITDA	S&P 100 Index	1-year U.S. Treasury
Average return (in %)	13.52	**17.31**	16.75	11.70	7.83	1.40
Std. of return (in %)	23.47	33.02	25.57	**19.13**	16.78	1.70
Total return (in %)	330.18	448.71	**538.62**	278.24	138.86	21.23
CAGR (in %)	10.98	12.93	**14.16**	9.97	6.42	1.38
Sharpe Ratio	0.54	0.50	**0.62**	0.56	0.40	0
M^2 (in %)	2.23	1.6	**3.63**	2.60	0	-
Jensen's alpha (in %)	3.5	4.2	**6.2**	3.8	0	-
IR	0.68	0.53	**0.80**	0.45	-	-
TE	**0.08**	0.18	0.11	0.09	-	-
MDD (in %)	43.7	52.0	39.0	**36.2**	40.69	0

Note. Bolded values denote a superior strategy in each category.

6.2.1 Inclusion of transaction costs

Up to this point, we have implicitly assumed frictionless markets and did not account for any additional costs incurred. Such costs could include information-gathering costs, transaction costs (e.g. bid/ask spread, dividend reinvestment fees, brokerage fees), taxes, liquidity constraints, and any other cost related to the trading account. This would lower realized return and affect the performance of active trading strategies.

Here we assume that all buy and sell transactions are subjected to 1% brokerage fee. We also assume that a 25% tax is levied on all capital gains. Since the positions in active trading strategies are set up each year and then fully liquidated at year's end, they are annually subject to brokerage fees and potential tax on capital gains. However, the S&P 100 Index portfolio is only subject to the initial 1% and final 1% brokerage fee as well as a 25% tax on a capital gain at the end of the period, when we assume the position is liquidated. Since we consider a buy-and-hold strategy, we presume no costs are incurred in the interim period.

The net performance of active trading strategies is presented in Table 20. The S&P 100 Index portfolio is used as a market benchmark, and 1-year U.S. Treasury notes serve as a proxy for the risk-free investment. The P/FCF ratio is omitted since the corresponding strategy is not statistically significant. The total cumulative wealth for each price ratio and the benchmark portfolio are given in Figure 16.

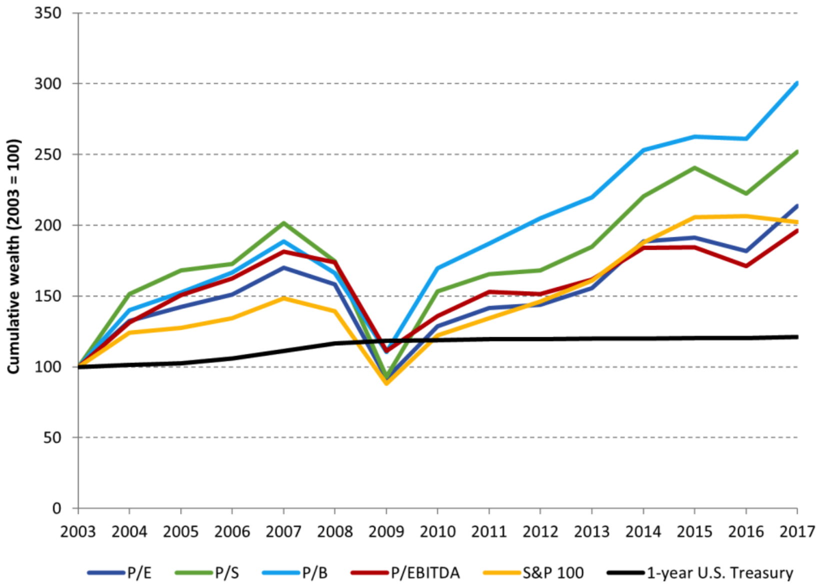
The inclusion of brokerage fees and taxes significantly lowers return of active trading strategies. However, the comparative performance of different price ratios remains unchanged compared to gross performance in Table 19. While the P/S strategy yields the highest average return, the P/B strategy results in highest total return and demonstrates a superior performance on a risk-adjusted basis. More importantly, the P/B strategy is also the only strategy that outperforms the benchmark S&P 100 Index on a risk-adjusted basis. The risk-adjusted indicators (Sharpe Ratio, Jensen's alpha, and M^2) demonstrate a positive, albeit low, excess net return.

Table 20. Algorithm net performance of different price ratios (Strategy III; 2003-2016)

	P/E	P/S	P/B	P/EBITDA	S&P 100 Index	1-year U.S. Treasury
Average return (in %)	7.49	**10.02**	10.00	6.24	7.83	1.40
Std. of return (in %)	19.38	26.42	20.35	**16.01**	16.78	1.70
Total return (in %)	113.81	152.12	**200.37**	96.19	102.07	21.23
CAGR (in %)	5.58	6.83	**8.17**	4.93	5.15	1.38
Sharpe Ratio	0.33	0.34	**0.44**	0.31	0.40	0
M^2 (in %)	-1.17	-0.96	**0.65**	-1.36	0	-
Jensen's alpha (in %)	-1.16	-1.02	**1.13**	-0.68	0	-
IR	-0.07	0.19	**0.34**	-0.22	-	-
TE	**0.05**	0.11	0.06	0.07	-	-
MDD (in %)	45.9	53.9	41.4	**38.7**	40.69	0

Note. Bolded values denote a superior strategy in each category.

Figure 16. Cumulative wealth (Strategy III; 2003-2016)



CONCLUSIONS

Information processing and price discovery are highly important for markets to function efficiently. In equity valuation, financial statement data provide a basis for all valuation approaches and contain comprehensive information about the state of the company. Previous studies focused on the formulation of summary metrics to aggregate financial statement data in order to systematically compare companies. However, the selection and aggregation of financial ratios are challenging and subjective. We designed an approach to systematically and efficiently utilize financial statement data for the purpose of stock valuation and on that basis proposed an active investment strategy.

The financial statement data are highly interconnected. We used principal component analysis (PCA) to aggregate financial statement data and formed principal components. We demonstrated that six principal components are a suitable representation of the original data and represent financial statement factors with a meaningful economic interpretation. The derived financial statement factors have highly coherent structure, and its constituents have a meaningful argument signs. Additionally, we find that combining three years of data provides a more stable factor structure compared to only one year of data.

We used regression analysis to determine the implied P/E ratio where financial statement factors serve as explanatory variables. We estimated the regression betas using the OLS regression and constructed an active investment portfolio based on the difference between market price and implied price. We demonstrate that such active investment strategy results in statistically significant outperformance of benchmark S&P 100 Index portfolio. We find that using a longer historical window in the analysis improves the stability of the model while preserving the outperformance.

We additionally utilized the Kalman filter algorithm. This allows us to eliminate the selection of historical window and indirectly use the entire available data history, while it enables the model to optimally adapt factor betas to new information. Additionally, this approach offers a direct insight into how market prices financial statement factors in different periods. Performance measurement indicators show that strategy with Kalman filtering yields superior results on a risk-adjusted basis.

We applied the proposed methodology also using other commonly reported price multiples, i.e. price-to-sales (P/S), price-to-book (P/B), price-to-EBITDA (P/EBITDA), and price-to-free cash flow (P/FCF). While the P/FCF ratio yields statistically insignificant results, the models using the remaining price ratios are statistically highly significant, and their explained share of the total variance (R^2) is even higher than for the P/E ratio strategy. We find that the P/B strategy yields superior return, but is highly volatile. In contrast, the P/EBITDA strategy has very low volatility while still considerably exceeds the benchmark. The P/E strategy represents a mixture of the P/B strategy and the P/EBITDA strategy, whereas the P/S strategy yields an inferior performance.

In summary, we successfully demonstrated a coherent approach to systematically and efficiently aggregate financial statement data. We showed that obtained financial statement factors form a suitable basis to derive implied stock prices and determine undervalued companies. We demonstrated that proposed investment strategy yields superior risk-adjusted performance compared to the benchmark S&P 100 Index.

However, our results indicate that presence of transaction costs significantly lowers the performance of active trading strategy, and excess return nearly disappears when performance is adjusted for expenses and transaction costs. Additionally, we acknowledge that the analysis includes some idealistic assumptions (e.g., fractional investment) and does not necessarily reflect all potential costs (e.g., information-gathering costs, transaction costs, taxes, liquidity constraints, etc.).

The presented results are in line with previous studies, which demonstrate that stock picking strategies can successfully determine outperforming stocks, but the excess portfolio return nearly disappears when adjusted for expenses and transaction costs. E.g., Cohen et. al. (2010) noted that active managers can pick outperforming stocks and yield 6% excess return. Similarly, Wermers (2000) reported that stocks of mutual funds outperform the market by 1.3 percent per year, but funds' performance is comparable to the market portfolio when expenses and transaction costs are accounted for.

This work presents a valuable contribution to the topic of financial statement analysis for a structured cross-sectional comparison of companies. It additionally demonstrates a novel approach to determine significant financial statement factors for stock valuation. Future work could extend the analysis to a broader market index or focus on a specific sector, where sector-specific ratios could provide a valuable enhancement of the existing methodology. Additionally, an investment portfolio could be constructed according to the financial statement risk factors to further enhance risk-adjusted performance.

REFERENCE LIST

1. Abarbanell, J. S., & Bushee, B. J. (1997). Fundamental Analysis, Future Earnings, and Stock Prices. *Journal of Accounting Research*, *35*(1), 1.
2. Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459.
3. Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18.
4. Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models: Realized Volatility. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 253–280.
5. Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Chichester, New York: Wiley.
6. Basu, S. (1977). Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *Journal of Finance*, *32*(3), 663–82.
7. Bessembinder, H., & Chan, K. (1998). Market Efficiency and the Returns to Technical Analysis. *Financial Management*, *27*(2), 5–17.
8. Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, *104*(3), 535–559.
9. Bloomberg. (2017). *Bloomberg Terminal*. Retrieved April 4, 2017, from https://www.bloomberg.com/professional/solution/bloomberg-terminal/
10. Brealey, R. A. (1986). *An Introduction to Risk and Return from Common Stocks* (2nd ed.). Cambridge: The MIT Press.
11. Capaul, C., Rowley, I., & Sharpe, W. F. (1993). International Value and Growth Stock Returns. *Financial Analysts Journal*, *49*(1), 27–36.
12. Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276.
13. Cerny, B. A., & Kaiser, H. F. (1977). A Study Of A Measure Of Sampling Adequacy For Factor-Analytic Correlation Matrices. *Multivariate Behavioral Research*, *12*(1), 43–47.
14. Chan, E. P. (2013). Algorithmic trading: winning strategies and their rationale. Hoboken, New Jersey: Wiley.
15. Chan, L. K. C., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and Stock Returns in Japan. *The Journal of Finance*, *46*(5), 1739–1764.
16. Chen, K.-J., & Li, X.-M. (2006). Is Technical Analysis Useful for Stock Traders in China? Evidence from the Szse Component a-Share Index. *Pacific Economic Review*, *11*(4), 477–488.
17. Chordia, T., Roll, R., & Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, *76*(2), 271–292.

18. Choudhry, T., & Wu, H. (2008). Forecasting ability of GARCH vs Kalman filter method: evidence from daily UK time-varying beta. *Journal of Forecasting*, *27*(8), 670–689.
19. Cohen, R. B., Polk, C., & Silli, B. (2010). Best Ideas. *SSRN Electronic Journal*. Retrieved April 12, 2017, from http://www.ssrn.com/abstract=1364827
20. Conrad, J., & Kaul, G. (1988). Time-Variation in Expected Returns. *The Journal of Business*, *61*(4), 409–25.
21. Cortazar, G., & Naranjo, L. (2006). An N-Factor Gaussian Model of Oil Futures Prices. *Journal of Futures Markets*, *26*(3), 243–268.
22. Cox, N. J. (2011). Stata tip 96: Cube roots. *The Stata Journal*, *11*(1), 149–154.
23. Cragg, J. G., & Malkiel, B. G. (1968). The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings. *The Journal of Finance*, *23*(1), 67–84.
24. Damodaran, A. (2002). Investment valuation: tools and techniques for determining the value of any asset (2nd ed.). New York: Wiley.
25. Damodaran, A. (2012). Investment Valuation: Tools and Techniques for Determining the Value of Any Asset (3rd ed.). New Jersey: Wiley.
26. Driessen, J., Nijman, T., & Melenberg, B. (2000). Common Factors in International Bond Returns. *Journal of International Money and Finance*, *22*(5), 629–656.
27. Ebner, M., & Neumann, T. (2005). Time-Varying Betas of German Stock Returns. *Financial Markets and Portfolio Management*, *19*(1), 29–46.
28. Emir, Ş., Dinçer, H., & Timor, M. (2012). A Stock Selection Model Based on Fundamental and Technical Analysis Variables by Using Artificial Neural Networks and Support Vector Machines. *Review of Economics & Finance*, *2*, 106–122.
29. Faff, R. W., & Brooks, R. D. (1998). Time-varying Beta Risk for Australian Industry Portfolios: An Exploratory Analysis. *Journal of Business Finance*, *25*(5&6), 721–745.
30. Faff, R. W., Hillier, D., & Hillier, J. (2000). Time Varying Beta Risk: An Analysis of Alternative Modelling Techniques. *Journal of Business Finance*, *27*(5&6), 523–554.
31. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383.
32. Fama, E. F. (1976). Foundations of finance: portfolio decisions and securities prices. New York: Basic Books.
33. Fama, E. F., & French, K. R. (1995). Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance*, *50*(1), 131.
34. Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.
35. Fenn, D. J., Porter, M. A., Williams, S., McDonald, M., Johnson, N. F., & Jones, N. S. (2011). Temporal Evolution of Financial Market Correlations. *Physical Review E*, *84*(2).
36. Fifield, S. G. M., Power, D. M., & Sinclair, C. D. (2005). An analysis of trading strategies in eleven European stock markets. *The European journal of finance*, *11*(6), 531–548.

37. Financial accounting standards board. (2017). 2017 US GAAP FINANCIAL REPORTING TAXONOMY. Retrieved April 4, 2017, from http://www.fasb.org/cs/ContentServer?c=Page&pagename=FASB%2FPage%2FSectionPage&cid=1176168807223
38. Fung, W., & Hsieh, D. A. (1997). Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds. *The Review of Financial Studies*, *10*(2), 275–302.
39. Fung, W., & Hsieh, D. A. (2002). Asset-Based Style Factors for Hedge Funds. *Financial Analysts Journal*, *58*(5), 16–27.
40. Gan, C., Lee, M., Hwa, A. Y. H., & Zhang, J. (2005). Revisiting Share Market Efficiency: Evidence from the New Zealand Australia, US and Japan Stock Indices. *American Journal of Applied Sciences*, *2*(5), 996–1002.
41. Gibbons, M. R., & Hess, P. (1981). Day of the Week Effects and Asset Returns. *The Journal of Business*, *54*(4), 579–96.
42. Gomez Biscarri, J., & Lopez Espinosa, G. (2008). The influence of differences in accounting standards on empirical pricing models: An application to the Fama–French model. *Journal of Multinational Financial Management*, *18*(4), 369–388.
43. Gordon, M. J. (1962). The investment, financing, and valuation of the corporation. R.D. Irwin.
44. Greene, W. H. (2012). *Econometric analysis*. Pearson series in economics (7th ed.). Boston, Mass.: Pearson.
45. Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, *70*(3), 393–408.
46. Gultekin, M. N., & Gultekin, N. B. (1983). Stock market seasonality: International Evidence. *Journal of Financial Economics*, *12*(4), 469–481.
47. Hamilton, J. D. (1994). *Time series analysis*. Princeton, N.J: Princeton University Press.
48. Harvey, A. C. (1991). *Forecasting, Structural Time Series Models and the Kalman Filter* (Reprint ed.). Cambridge; New York: Cambridge University Press.
49. Haugen, R. A., & Lakonishok, J. (1988). *The incredible January effect: the stock market's unsolved mystery*. Homewood Ill: Dow Jones-Irwin.
50. Hirshleifer, D., Hou, K., & Teoh, S. H. (2009). Accruals, cash flows, and aggregate stock returns. *Journal of Financial Economics*, *91*(3), 389–406.
51. Jacobs, P. D., & Levy, K. N. (1987). Calendar Anomalies: Abnormal Returns at Calendar Turning Points. *Financial Analysts Journal*, *44*(6), 28–39.
52. Jaffe, J. F. (1974). Special Information and Insider Trading. *The Journal of Business*, *47*(3), 410–28.
53. Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York: Springer.
54. Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200.
55. Kaiser, H. F. (1961). A Note on Guttman's Lower Bound for the Number of Common Factors1. *British Journal of Statistical Psychology*, *14*(1), 1–2.

56. Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35.
57. Keim, D. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics*, *12*(1), 13–32.
58. Kim, D. (2006). On the Information Uncertainty Risk and the January Effect. *The Journal of Business*, *79*(4), 2127–2162.
59. Kim, D.-H., & Jeong, H. (2005). Systematic analysis of group identification in stock markets. *Physical Review E*, *72*(4), 46133.
60. Kritzman, M., Li, Y., Page, S., & Rigobon, R. (2010). Principal Components as a Measure of Systemic Risk. *SSRN Electronic Journal*.
61. Lallemand, J., & Strauss, J. (2016). Can We Count on Accounting Fundamentals for Industry Portfolio Allocation? *The Journal of Portfolio Management*, *42*(4), 70–87.
62. Lev, B., & Thiagarajan, S. R. (1993). Fundamental Information Analysis. *Journal of Accounting Research*, *31*(2), 190.
63. Lie, F., Brooks, R., & Faff, R. (2000). Modelling the Equity Beta Risk of Australian Financial Sector Companies. *Australian Economic Papers*, *39*(3), 301–311.
64. Manoliu, M., & Tompaidis, S. (2002). Energy futures prices: term structure models with Kalman filter estimation. *Applied Mathematical Finance*, *9*(1), 21–43.
65. Meucci, A. (2010). Managing Diversification. *Risk*, *22*(5), 74–79.
66. Miles, B. W. J., Stokes, C. R., Vieli, A., & Cox, N. J. (2013). Rapid, climate-driven changes in outlet glaciers on the Pacific coast of East Antarctica. *Nature*, *500*(7464), 563–566.
67. Mobarek, A., Mollah, A. S., & Bhuyan, R. (2008). Market Efficiency in Emerging Stock Market: Evidence from Bangladesh. *Journal of Emerging Market Finance*, *7*(1), 17–41.
68. Mohanram, P. S. (2005). Separating winners from losers among lowbook-to-market stocks using financial statement analysis. *Review of Accounting Studies*, *10*(2–3), 133–170.
69. Olson, D., & Mossman, C. (2003). Neural network forecasts of Canadian stock returns using accounting ratios. *International Journal of Forecasting*, *19*(3), 453–465.
70. Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, *11*(4), 295–329.
71. Partovi, M. H., & Caputo, M. (2004). Principal Portfolios: Recasting the Efficient Frontier. *Economics Bulletin*, *7*(3), 1–10.
72. Pastor, L., & Stambaugh, R. F. (2009). Predictive Systems: Living with Imperfect Predictors. *The Journal of Finance*, *64*(4), 1583–1628.
73. Pennacchi, G. G. (1991). Identifying the Dynamics of Real Interest Rates and Inflation: Evidence Using Survey Data. *The Review of Financial Studies*, *4*(1), 53–86.
74. Perignon, C., Smith, D., & Villa, C. (2007). Why common factors in international bond returns are not so common. *Journal of International Money and Finance*, *26*(2), 284–304.

75. Pinto, J. E., Henry, E., Robinson, T. R., Stowe, J., & Wilcox, S. E. (Eds.). (2015). *Equity asset valuation* (Third edition.). Hoboken, New Jersey: Wiley.
76. Piotroski, J. D. (2000). Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research*, *38*, 1–41.
77. Racicot, F.-É., & Théoret, R. (2010). Forecasting stochastic volatility using the kalman filter: An application to canadian interest rates and price-earnings ratio. *Aestimatio: The IEB International Journal of Finance*, *1*, 28–47.
78. Rajamani, M. R., & Rawlings, J. B. (2009). Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. *Automatica*, *45*(1), 142–148.
79. Razali, N. M., & Wah, Y. B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.
80. Roll, R. (2010). Vas Ist Das? *The Journal of Portfolio Management*, *9*(2), 18–28.
81. Rosenbaum, J., & Pearl, J. (2013). Investment banking: valuation, leveraged buyouts, and mergers & acquisitions (2. ed.). Hoboken, NJ: Wiley.
82. Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, *11*(3), 9–16.
83. Rozeff, M. S., & Zaman, M. A. (1988). Market Efficiency and Insider Trading: New Evidence. *The Journal of Business*, *61*(1), 25–44.
84. Rudin, A. M., & Morgan, J. S. (2006). A portfolio diversification index. *The Journal of Portfolio Management*, *32*(2), 81–89.
85. Rytchkov, O. (2012). Filtering Out Expected Dividends and Expected Returns. *Quarterly Journal of Finance*, *2*(3), 1250012.
86. Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611.
87. Sørensen, C. (2002). Modeling seasonality in agricultural commodity futures: Modeling Seasonality. *Journal of Futures Markets*, *22*(5), 393–426.
88. S&P Dow Jones Indices. (2017a). S&P U.S. Indices Methodology. S&P Global. Retrieved from https://us.spindices.com/documents/methodologies/methodology-sp-us-indices.pdf?force_download=true
89. S&P Dow Jones Indices. (2017b). S&P 100. *S&P 100*. Retrieved April 4, 2017, from https://us.spindices.com/indices/equity/sp-100
90. Wells, C. (1995). *The Kalman Filter in Finance* (1996 ed.). Boston, Mass.: Springer.
91. Wermers, R. (2000). Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. *The Journal of Finance*, *55*(4), 1655–1695.
92. Zahedi, J., & Rounaghi, M. M. (2015). Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange. *Physica A: Statistical Mechanics and its Applications*, *438*, 178–187.

93. Zhang, Y., & Choudhry, T. (2016). Forecasting the Daily Time-Varying Beta of European Banks During the Crisis Period: Comparison Between GARCH Models and the Kalman Filter: Forecasting the Daily Time-Varying Beta of European Banks. *Journal of Forecasting*.
94. Zheng, Z., Podobnik, B., Feng, L., & Li, B. (2012). Changes in cross-correlations as an indicator for systemic risk. *Scientific Reports*, *2*, 888.

APPENDIXES

TABLE OF APPENDIXES

Appendix A: List of Abbreviations

Table 1. List of Abbreviations

Abbreviation	Definition
CAGR	Compound annual growth rate
DCF	Discounted cash flow
EMH	Efficient market hypothesis
IR	Information ratio
KF	Kalman filter
KMO test	Kaiser-Meyer-Olkin test
M^2 ratio	M-squared ratio
MDD	Maximum drawdown
OLS regression	Ordinary least square regression
P/B ratio	Price-to-Book ratio
P/E ratio	Price-to-Earnings ratio
P/EBITDA ratio	Price-to-EBITDA ratio
P/FCF ratio	Price-to-Free cash flow ratio
P/S ratio	Price-to-Sales ratio
PCA	Principal component analysis
RV	Relative valuation
SVD	Singular value decomposition
TE	Tracking error

Appendix B: S&P 100 Index constituents from March 2001 to March 2016

Table 2. S&P 100 Index constituents from March 2001 to March 2016

	Company name	Equity Ticker	Sector
1	3M CO	MMM US Equity	Industrials
2	ABBOTT LABORATORIES	ABT US Equity	Health Care
3	ABBVIE INC	ABBV US Equity	Health Care
4	ACCENTURE PLC-CL A	ACN US Equity	Information Technology
5	AES CORP	AES US Equity	Utilities
6	AIG LIFE HOLDINGS INC	950967Q US Equity	Financials
7	ALCATEL-LUCENT USA INC	LU US Equity	Information Technology
8	ALCOA CORP	AA US Equity	Materials
9	ALLEGHENY TECHNOLOGIES INC	ATI US Equity	Materials
10	ALLERGAN PLC	AGN US Equity	Health Care
11	ALLSTATE CORP	ALL US Equity	Financials
12	ALPHABET INC-CL C	GOOG US Equity	Consumer Discretionary
13	ALTRIA GROUP INC	MO US Equity	Consumer Staples
14	AMAZON.COM INC	AMZN US Equity	Consumer Discretionary
15	AMERICAN ELECTRIC POWER	AEP US Equity	Utilities
16	AMERICAN EXPRESS CO	AXP US Equity	Financials
17	AMERICAN INTERNATIONAL GROUP	AIG US Equity	Financials
18	AMGEN INC	AMGN US Equity	Health Care
19	ANADARKO PETROLEUM CORP	APC US Equity	Energy
20	ANHEUSER-BUSCH COS INC	3393199Q US Equity	Consumer Staples
21	APACHE CORP	APA US Equity	Energy
22	APPLE INC	AAPL US Equity	Information Technology
23	AT&T INC	T US Equity	Information Technology
24	AVON PRODUCTS INC	AVP US Equity	Consumer Staples
25	BAKER HUGHES INC	BHI US Equity	Energy
26	BANK OF AMERICA CORP	BAC US Equity	Financials
27	BANK OF NEW YORK MELLON CORP	BK US Equity	Financials
28	BANK ONE CORP	3621240Q US Equity	Financials
29	BAXTER INTERNATIONAL INC	BAX US Equity	Health Care
30	BERKSHIRE HATHAWAY INC	BRK/B US Equity	Financials
31	BIOGEN INC	BIIB US Equity	Health Care
32	BLACK & DECKER CORP/THE	BDK US Equity	Industrials
33	BLACKROCK INC	BLK US Equity	Financials
34	BOEING CO/THE	BA US Equity	Industrials
35	BRISTOL-MYERS SQUIBB CO	BMY US Equity	Health Care
36	BURLINGTON NORTHERN SANTA FE	BNI US Equity	Industrials
37	CAESARS ENTERTAINMENT CORP	HET US Equity	Consumer Discretionary
38	CAMPBELL SOUP CO	CPB US Equity	Consumer Staples
39	CAPITAL ONE FINANCIAL CORP	COF US Equity	Financials
40	CATERPILLAR INC	CAT US Equity	Industrials
41	CBS CORP-CLASS	CBS US Equity	Consumer Discretionary
42	CELGENE CORP	CELG US Equity	Health Care
43	CHEVRON CORP	CVX US Equity	Energy
44	CIGNA CORP	CI US Equity	Health Care
45	CISCO SYSTEMS INC	CSCO US Equity	Information Technology

(Table continues)

(Continued)

46	CITIGROUP INC	C US Equity	Financials
47	COCA-COLA CO/THE	KO US Equity	Consumer Staples
48	COLGATE-PALMOLIVE CO	CL US Equity	Consumer Staples
49	COMCAST CORP	CMCSA US Equity	Consumer Discretionary
50	COMPUTER SCIENCES CORP	CSC US Equity	Information Technology
51	CONOCOPHILLIPS	COP US Equity	Energy
52	COSTCO WHOLESALE CORP	COST US Equity	Consumer Staples
53	COVIDIEN LTD	COV US Equity	Health Care
54	CVS HEALTH CORP	CVS US Equity	Consumer Staples
55	DANAHER CORP	DHR US Equity	Health Care
56	DELL INC	DELL US Equity	Information Technology
57	DELTA AIR LINES INC	DALRQ US Equity	Consumer Discretionary
58	DEVON ENERGY CORP	DVN US Equity	Energy
59	DOW CHEMICAL CO/THE	DOW US Equity	Materials
60	DU PONT (E.I.) DE NEMOURS	DD US Equity	Materials
61	DUKE ENERGY CORP	DUK US Equity	Utilities
62	EASTMAN KODAK CO	EKDKQ US Equity	Information Technology
63	EBAY INC	EBAY US Equity	Consumer Discretionary
64	EL PASO LLC	EP US Equity	Energy
65	ELI LILLY & CO	LLY US Equity	Health Care
66	EMC CORP/MA	EMC US Equity	Information Technology
67	EMERSON ELECTRIC CO	EMR US Equity	Industrials
68	ENRON CREDITORS RECOVERY COR	ENRNQ US Equity	Energy
69	ENTERGY CORP	ETR US Equity	Utilities
70	EXELON CORP	EXC US Equity	Utilities
71	EXXON MOBIL CORP	XOM US Equity	Energy
72	FACEBOOK INC	FB US Equity	Consumer Discretionary
73	FEDEX CORP	FDX US Equity	Industrials
74	FORD MOTOR CO	F US Equity	Consumer Discretionary
75	FREEPORT-MCMORAN INC	FCX US Equity	Materials
76	GENERAL DYNAMICS CORP	GD US Equity	Industrials
77	GENERAL ELECTRIC CO	GE US Equity	Industrials
78	GENERAL MOTORS CO	GM US Equity	Consumer Discretionary
79	GILEAD SCIENCES INC	GILD US Equity	Health Care
80	GILLETTE COMPANY	1028411Q US Equity	Consumer Staples
81	GLOBAL CROSSING LTD	GBLXQ US Equity	Information Technology
82	GOLDMAN SACHS GROUP INC	GS US Equity	Financials
83	HALLIBURTON CO	HAL US Equity	Energy
84	HARTFORD FINANCIAL SVCS GRP	HIG US Equity	Financials
85	HCA INC	3605818Q US Equity	Health Care
86	HILLSHIRE BRANDS CO/THE	HSH US Equity	Consumer Staples
87	HOME DEPOT INC	HD US Equity	Consumer Staples
88	HONEYWELL INTERNATIONAL INC	HON US Equity	Industrials
89	HP INC	HPQ US Equity	Information Technology
90	IHEARTCOMMUNICATIONS INC	2968900Q US Equity	Consumer Discretionary
91	INTEL CORP	INTC US Equity	Information Technology
92	INTERNATIONAL PAPER CO	IP US Equity	Materials
93	INTL BUSINESS MACHINES CORP	IBM US Equity	Information Technology
94	JOHNSON & JOHNSON	JNJ US Equity	Health Care

(Table continues)

(Continued)

95	JPMORGAN CHASE & CO	JPM US Equity	Financials
96	KINDER MORGAN INC	KMI US Equity	Energy
97	KRAFT HEINZ CO/THE	KHC US Equity	Consumer Staples
98	L BRANDS INC	LB US Equity	Consumer Staples
99	LEHMAN BROTHERS HOLDINGS INC	LEHMQ US Equity	Financials
100	LOCKHEED MARTIN CORP	LMT US Equity	Industrials
101	LOWE'S COS INC	LOW US Equity	Consumer Staples
102	MASTERCARD INC – A	MA US Equity	Financials
103	MAY DEPARTMENT STORES CO	987200Q US Equity	Consumer Staples
104	MCDONALD'S CORP	MCD US Equity	Consumer Staples
105	MEDIMMUNE LLC	MEDI US Equity	Health Care
106	MEDTRONIC PLC	MDT US Equity	Health Care
107	MERCK & CO. INC.	MRK US Equity	Health Care
108	MERRILL LYNCH & CO INC	MER US Equity	Financials
109	METLIFE INC	MET US Equity	Financials
110	MICROSOFT CORP	MSFT US Equity	Information Technology
111	MONDELEZ INTERNATIONAL INC	MDLZ US Equity	Consumer Staples
112	MONSANTO CO	MON US Equity	Materials
113	MORGAN STANLEY	MS US Equity	Financials
114	MOTORS LIQUIDATION CO	MTLQQ US Equity	Consumer Discretionary
115	NATIONAL OILWELL VARCO INC	NOV US Equity	Energy
116	NATIONAL SEMICONDUCTOR CORP	0203524D US Equity	Information Technology
117	NESTLE PURINA PETCARE CO	RAL US Equity	Consumer Staples
118	NEXTEL COMMUNICATIONS INC	NXTL US Equity	Information Technology
119	NEXTERA ENERGY INC	NEE US Equity	Utilities
120	NIKE INC	NKE US Equity	Consumer Discretionary
121	NORFOLK SOUTHERN CORP	NSC US Equity	Industrials
122	NORTEL NETWORKS CORP	NRTLQ US Equity	Information Technology
123	NYSE EURONEXT	NYX US Equity	Financials
124	OCCIDENTAL PETROLEUM CORP	OXY US Equity	Energy
125	OFFICEMAX INC	OMX US Equity	Consumer Staples
126	ORACLE CORP	ORCL US Equity	Information Technology
127	PAYPAL HOLDINGS INC	PYPL US Equity	Information Technology
128	PEPSICO INC	PEP US Equity	Consumer Staples
129	PFIZER INC	PFE US Equity	Health Care
130	PHARMACIA LLC	748957Q US Equity	Health Care
131	PHILIP MORRIS INTERNATIONAL	PM US Equity	Consumer Staples
132	PRICELINE GROUP INC/THE	PCLN US Equity	Consumer Discretionary
133	PROCTER & GAMBLE CO/THE	PG US Equity	Consumer Staples
134	QUALCOMM INC	QCOM US Equity	Information Technology
135	RAYTHEON COMPANY	RTN US Equity	Industrials
136	REGIONS FINANCIAL CORP	RF US Equity	Financials
137	ROCKWELL AUTOMATION INC	ROK US Equity	Industrials
138	RS LEGACY CORP	RSHCQ US Equity	Consumer Staples
139	SCHERING-PLOUGH CORP/PRE-MER	SGP US Equity	Health Care
140	SCHLUMBERGER LTD	SLB US Equity	Energy
141	SEARS ROEBUCK & CO	605555Q US Equity	Consumer Staples
142	SIMON PROPERTY GROUP INC	SPG US Equity	Real Estate
143	SOUTHERN CO/THE	SO US Equity	Utilities

(Table continues)

(Continued)

144	SPRINT COMMUNICATIONS INC	0848680D US Equity	Information Technology
145	STARBUCKS CORP	SBUX US Equity	Consumer Staples
146	TARGET CORP	TGT US Equity	Consumer Staples
147	TEXAS INSTRUMENTS INC	TXN US Equity	Information Technology
148	TIME WARNER INC	TWX US Equity	Consumer Discretionary
149	TOYS R US INC	TOY US Equity	Consumer Staples
150	TWENTY-FIRST CENTURY FOX	FOXA US Equity	Consumer Discretionary
151	TYCO INTERNATIONAL PLC	TYC US Equity	Industrials
152	UNION PACIFIC CORP	UNP US Equity	Industrials
153	UNISYS CORP	UIS US Equity	Information Technology
154	UNITED PARCEL SERVICE	UPS US Equity	Industrials
155	UNITED TECHNOLOGIES CORP	UTX US Equity	Industrials
156	UNITEDHEALTH GROUP INC	UNH US Equity	Health Care
157	US BANCORP	USB US Equity	Financials
158	VERIZON COMMUNICATIONS INC	VZ US Equity	Information Technology
159	VISA INC-CLASS A SHARES	V US Equity	Financials
160	WACHOVIA CORP	1255173D US Equity	Financials
161	WALGREENS BOOTS ALLIANCE INC	WBA US Equity	Consumer Staples
162	WAL-MART STORES INC	WMT US Equity	Consumer Staples
163	WALT DISNEY CO/THE	DIS US Equity	Consumer Discretionary
164	WELLS FARGO & CO	WFC US Equity	Financials
165	WEYERHAEUSER CO	WY US Equity	Real Estate
166	WILLIAMS COS INC	WMB US Equity	Energy
167	WYETH LLC	WYE US Equity	Health Care
168	XEROX CORP	XRX US Equity	Information Technology

Note. The annual cut-off date is fourth Friday in March. Companies are listed in alphabetic order. Sectors are determined based on GICS (as determined on April 4th, 2017).

Source: Bloomberg, *Bloomberg Terminal*, 2017.

Appendix C: List of financial statement data with definitions

Table 3. List of financial statement data with definitions

	Accounting ratio	Description
1	ASSET_GROWTH	A percentage increase or decrease of total assets by comparing current period with same period prior year. Calculated as: (Total Assets – Total Assets Same Period Prior Year) * 100 / Total Assets from Same Period Prior Year Where: Total Assets is BS035, BS_TOT_ASSET
2	BETA_RAW_OVERRIDABLE	Raw (historical) beta measures the volatility of the stock price relative to the volatility in the market index. Beta is the percent change in the price of the stock given a 1% change in the market index. This is calculated from the overrides in fields Beta Start Date Override (RK390, BETA_OVERRIDE_START_DT), Beta End Date Override (RK391, BETA_OVERRIDE_END_DT), Beta Relative Index Override (RK392, BETA_OVERRIDE_REL_INDEX) and Beta Periodicity Override (RK393, BETA_OVERRIDE_PERIOD). This field requires a minimum of 3 data points for it to work.
3	BVPS_GROWTH	Percentage increase or decrease of book value per share by comparing current period with same period prior year. Calculated as: (Book Value per Share - Book Value per Share same period prior year) * 100 / Book Value per Share from same period prior year Where: Book Value per Share s is RR020, BOOK_VAL_PER_SH
4	CAP_EXPEND_RATIO	Measures how much of the cash generated from operations will be left after payment of capital expenditures to service the company's debt. Unit: Actual. INDUSTRIALS, BANKS, FINANCIALS, INSURANCE, UTILITIES & REITS Calculated as: Cash From Operations / Capital Expenditures Where: Cash From Operations is CF015, CF_CASH_FROM_OPER Capital Expenditures is RR014, CAPITAL_EXPEND
5	CASH_FLOW_GROWTH	One year growth measure of the company cash flow. Calculated as: [(Cash Flow from Operations Current Period / Cash Flow from Operations Same Period Prior Year) – 1] * 100 Where: Cash From Operations is CF015, CF_CASH_FROM_OPER Note: For interim periods, the comparative period is the same interim period of the preceding year.
6	COM_EQY_TO_TOT_ASSET	One of many financial ratios (in Percentage) used to determine the financial health and long-term profitability of a corporation. Calculated as: Common Equity * 100 / Total Assets Where: Common Equity is RR010, TOT_COMMON_EQY Total Assets is BS035, BS_TOT_ASSET
7	CONT_INC_GROWTH	A percentage increase or decrease of income before extraordinary items by comparing current period with same period prior year. Calculated as: (Income Before XO Items from Current Period – Income Before XO Items from Same Period Prior Year) * 100 / Income Before XO Items from Same Period Prior Year Where: Income Before Extraordinary Items is RR092, INC_BEF_XO_LESS_MIN_INT_PREF_DVD Income Before Extraordinary Items Growth is not computed if it changes signs from prior year to current period.
8	EBIT_MARGIN	INDUSTRIALS, UTILITIES & REITS Ratio which measures the company's profitability. Unit: Actual. Calculated as: (Trailing 12M Operating Inc (Loss) / Trailing 12M Net Sales)*100. Where: Trailing 12M Operating Inc (Loss) is RR803, TRAIL_12M_OPER_INC Trailing 12M Net Sales is RR800, TRAIL_12M_NET_SALES

(Table continues)

(Continued)

9	EBIT_YR_GROWTH	Percentage change in earnings before interest and taxes (EBIT) from last year to the current year. For interim periods, the comparative period is the same interim period of the preceding year. Unit: Actual. INDUSTRIALS, UTILITIES, & REITS Calculated as: Growth 1 Year (Interest Expense) Where: Earnings before Interest and Taxes is RR002, EBIT
10	EBITDA_GROWTH	Percentage change in Earnings Before Interest Taxes Depreciation Amortization (RR009, EBITDA) from last year to the current year. For interim periods, the comparative period is the same interim period of the preceding year. Unit: Actual. INDUSTRIALS, FINANCIALS, UTILITIES, REITS, MUNICIPAL REVENUE Calculated as: Growth 1 Year (RR009) Where: EBITDA is RR009, EBITDA Please reference EBITDA Growth Adjusted Year over Year (F1151, EBITDA_GROWTH_ADJUSTED_YOY) for the adjusted value that excludes the impact of abnormal items.
11	EBITDA_MARGIN	Percentage margin of trailing 12 month Earnings Before Interest Taxes Depreciation and Amortization (EBITDA) divided by the trailing 12 month Sales. Unit: Actual. INDUSTRIALS, FINANCIALS, UTILITIES & REITS Calculated as: (Trailing 12 month EBITDA / Trailing 12 month Sales) * 100 Where: Trailing 12 Month EBITDA is RR841, TRAIL_12M_EBITDA Trailing 12 Month Sales is RR800, TRAIL_12M_NET_SALES MUNICIPAL REVENUE Calculated as: (EBITDA / Sales) * 100 Where: EBITDA is RR009, EBITDA Sales is IS010, SALES_REV_TURN
12	EMPL_GROWTH	Percentage increase or decrease of employee number by comparing the current period with the same period in the prior year. Calculated as: (Number of Employees - Number of Employees Same Period Prior Year)*100/Number of Employees from Same Period Prior Year Where: Number of Employees is RR121, NUM_OF_EMPLOYEES
13	EPS_GROWTH	Percentage increase or decrease of earning before extraordinary items by comparing current period with same period prior year. Calculated as: (EPS before XO Items - EPS before XO Items same period prior year) * 100 / EPS before XO Items from same period prior year Where: EPS before XO Items is IS064, IS_EARN_BEF_XO_ITEMS_PER_SH
14	FREE_CASH_FLOW_MARGIN	Free Cash Flow as a percentage of Revenue. Unit: Actual. Calculated as: INDUSTRIALS, INSURANCE, UTILITIES, & REITS (Free Cash Flow / Sales/Revenue/Turnover) * 100. Where: Free Cash Flow is RR008, CF_FREE_CASH_FLOW Sales/Revenue/Turnover is IS010, SALES_REV_TURN. BANKS & FINANCIALS (Free Cash Flow / Net Revenue) * 100 Where: Free Cash Flow is RR008, CF_FREE_CASH_FLOW Net Revenue is RR209, NET_REV

(Table continues)

(Continued)

15	GROSS_MARGIN	INDUSTRIALS & UTILITIES Gross margin represents the percent of total sales revenue that the company retains after incurring the direct costs associated with producing the goods and services sold by a company. Calculated as: (Net Sales - Cost of Goods Sold) * 100 / Net Sales Where: Net Sales is IS010, SALES_REV_TURN Cost of Goods Sold is IS021, IS_COGS_TO_FE_AND_PP_AND_G Please reference Gross Margin Adjusted (F1172, GROSS_MARGIN_ADJUSTED) for the adjusted value that excludes the impact of abnormal items.
16	GROWTH_IN_CAP	Percentage increase or decrease of total capital by comparing current period with same period prior year. Calculated as: (Total Capital - Total Capital same period prior year) * 100 / Total Capital from same period prior year Where: Total Capital is RR006, BS_TOT_CAP
17	INC_TAX_EXP_YR_GROWTH	Percentage change in income tax expense from last year to the current year. For interim periods, the comparative period is the same interim period of the preceding year. Unit: Actual. INDUSTRIALS, BANKS, FINANCIALS, INSURANCE, & UTILITIES Calculated as: Growth 1 Year (Income Tax Expense) Where: Income Tax Expense is IS038, IS_INC_TAX_EXP
18	NET_DEBT_TO_CASHFLOW	Ratio of a company's total debt to trailing 12-month cash flow from operations . Unit: Actual. Calculated as: Net Debt / Cash Flow from Operations Where: Net Debt is RR208, NET_DEBT Trailing 12M Cash From Operations is RR824, TRAIL_12M_CASH_FROM_OPER
19	NET_INC_GROWTH	A percentage increase or decrease of net income by comparing current period with same period prior year. Calculated as: (Net Income from Current Period - Net Income from Same Period Prior Year) * 100 / Net Income from Same Period Prior Year Where: Net Income is IS050, NET_INCOME Net Income is IS534, IS_CHANGE_IN_NET_ASSETS for Municipal Revenue Net Income Growth is not computed if Net Income changes signs from prior year to current period.
20	NET_WORTH_GROWTH	Percentage increase or decrease of net worth (also named as common equity) by comparing current period with same period prior year. Calculated as: (Total Common Equity - Total Common Equity same period prior year) * 100 / Total Common Equity from same period prior year Where: Total Common Equity is RR010, TOT_COMMON_EQY
21	NORMALIZED_PROFIT_MARGIN	Profitability ratio that shows how much of revenue contributes to net income before extraordinary items, one time charges, minus preferred dividends, minority interest and other adjustments, divided by sales, figure is in percentage. Calculated as: INDUSTRIAL, BANKS, FINANCIAL, INSURANCE, UTILITY, & REITS (Normalized Income / Sales) * 100 Where: Normalized Income is RX062, NORMALIZED_INCOME Sales is IS010, SALES_REV_TURN MUNICIPAL REVENUE ((Normalized Income / Sales) * 100 Where: Normalized Income is F0222, TOTAL_NORMALIZED_PROFIT Sales is IS894, IS_TOTAL_REVENUES
22	NORMALIZED_ROE	Returns on Common Equity based on net income excluding one-time charges. Calculated as: [Trailing 12 Month Normalized Income / Average of Current and Prior Period (Common Equity)] *100 Where: T12M Normalized Income is RX114, T12_NORMALIZED_INCOME Total Common Equity is RR010, TOT_COMMON_EQY

(Table continues)

(Continued)

23	OPER_INC_GROWTH	A percentage increase or decrease of operating income by comparing current period with same period prior year. EBIT (earnings before interest and taxes) is also commonly known as Operating Income. Calculated as: (Operating Income from Current Period - Operating Income from Same Period Prior Year) * 100 / Operating Income from Same Period Prior Year Where: Operating Income is IS033, IS_OPER_INC Operating Income is RR002, EBIT for REITs format. Operating Income Growth is not computed if Operating Income changes signs from prior year to current period. Please reference EBIT Growth Adjusted Year over Year (F1145, EBIT_GROWTH_ADJUSTED_YOY) for the adjusted value that excludes the impact of abnormal items.
24	OPER_MARGIN	Ratio used to measure a company's pricing strategy and operating efficiency, in percentage. INDUSTRIALS, INSURANCE, UTILITIES, & MUNICIPAL REVENUE Calculated as: Operating Income (Losses) / Total Revenue * 100 Where: Operating Income is IS033, IS_OPER_INC Total Revenue is IS010, SALES_REV_TURN BANKS & FINANCIALS Calculated as: Operating Income (Losses) / Net Income * 100 Where: Operating Income is IS033, IS_OPER_INC Net Revenue is RR209, NET_REV REITS Calculated as: EBIT / Total Revenue * 100 Where: EBIT is RR002, EBIT Total Revenue is IS010, SALES_REV_TURN
25	PROF_MARGIN	Measuring the company's profitability, this ratio is the comparison of how much of the revenue incurred during the period was retained in income. Calculated as: INDUSTRIALS, FINANCIAL, INSURANCE, UTILITIES, & REITS (Net Income / Revenue) * 100 Where: Net Income is IS050, NET_INCOME Revenue is IS010, SALES_REV_TURN BANKS (Net Income / Net Revenue) *100 Where: Net Income is IS050, NET_INCOME Net Revenue is RR209, NET_REV MUNICIPAL REVENUE (Change in Net Assets / Revenue) *100 Where: Change in Net Assets is IS534, IS_CHANGE_IN_NET_ASSETS Revenue is IS010, SALES_REV_TURN Municipal Revenue is for Municipal issues under the Equity key only.
26	REINVEST_EARN_YR_GROWTH	* Reinvested Earnings year change is calculated as follows: ((Reinvested earnings in Current period - Reinvested earnings in previous period)/(Reinvested earnings in previous period)) * 100 Reinvested earnings = Net income(losses) - Total preferred dividends - Total common dividends. This ratio is not computed if the sign changes from year to year in value.

(Table continues)

(Continued)

27	RETURN_COM_EQY	Measure of a corporation's profitability by revealing how much profit a company generates with the money shareholders have invested, in percentage. Calculated as: (T12 Net Income Available for Common Shareholders / Average Total Common Equity) * 100 Where: T12 Net Income Available for Common Shareholders is T0089, TRAIL_12M_NET_INC_AVAI_COM_SHARE Average Total Common Equity is the average of the beginning balance and ending balance of RR010, TOT_COMMON_EQY If either the beginning or ending total common equity is negative, Return on Equity will not be calculated. Please reference Return on Common Equity Adjusted (F1169, RETURN_ON_COMMON_EQUITY_ADJUSTED) for the adjusted value that excludes the impact of abnormal items.
28	RETURN_ON_ASSET	Indicator of how profitable a company is relative to its total assets, in percentage. Return on assets gives an idea as to how efficient management is at using its assets to generate earnings. INDUSTRIALS, BANKS, FINANCIALS, UTILITIES, & REITS Calculated as: (Trailing 12M Net Income / Average Total Assets) * 100 Where: Trailing 12M Net Income is RR813, TRAIL_12M_NET_INC Average Total Assets is the average of the beginning balance and ending balance of BS035, BS_TOT_ASSET INSURANCE ((Trailing 12M Net Income + Trailing 12M Policyholders' Surplus) / Average Total Assets) * 100 Where: Trailing 12M Net Income is RR813, TRAIL_12M_NET_INC Trailing 12M Policyholders' Surplus is RR713, TRAIL_12M_POLICY_HOLDER_SURPLUS Average Total Assets is the average of the beginning balance and ending balance of BS035, BS_TOT_ASSET Please reference Return on Assets Adjusted (F1170, RETURN_ON_ASSETS_ADJUSTED) for the adjusted value that excludes the impact of abnormal items.
29	RETURN_ON_INV_CAPITAL	Indicates how effectively a company uses the sources of capital (equity and debt) invested in its operations. Average Invested Capital is the average of the beginning and ending balance of Total Invested Capital (RX215, TOTAL_INVESTED_CAPITAL). It is computed as: 100 x (T12M Net operating profit after tax / Average invested capital) Where: T12M Net operating profit after tax is RX216, TRAIL_12M_NET_OP_PROF_AFTER_TAX Invested Capital is RX215, TOTAL_INVESTED_CAPITAL Average of Invested Capital is calculated based on the average of invested capital for current period and invested capital for the same period a year ago. ROIC will not compute if: 1. The year-over-year average of invested capital is negative. 2. Effective Tax Rate (RR037, EFF_TAX_RATE) or Trailing 12 Months Effective Tax Rate (RR712, TRAIL_12M_EFF_TAX_RT) is not available. The calculation of this field includes only the basic adjustments listed under RX214, NET_OPER_PROFIT_AFTER_TAX and RX215, TOTAL_INVESTED_CAPITAL and therefore may not show the same value as WACC Return on Invested Capital (VM013, WACC_RETURN_ON_INV_CAPITAL).

(Table continues)

(Continued)

30	RETURN_ON_CAP	Metric that measures the return that an investment generates for capital contributors, in percentage. It indicates how effective a company is turning capital into profits. INDUSTRIALS & UTILITIES Calculated as: ((T12 Net Income (Losses) + T12 Minority Interest + T12 Interest Expense * (1 – (T12 Effective Tax Rate / 100))) / Average of Total Capital) * 100 Where: Trailing 12 M Net Income is RR813, TRAIL_12M_NET_INC Trailing 12M Minority Interest is RR812, TRAIL_12M_MINORITY_INT Trailing 12M Interest Expense is RR804, TRAIL_12M_INT_EXP Trailing 12M Effective Tax Rate is RR712, TRAIL_12M_EFF_TAX_RT Total Capital is RR006, BS_TOT_CAP Average is the average of the beginning and ending balances. Trailing 12 month values use the latest 4 quarters, 2 semi annuals or annual. Return on Capital is not computed if the Effective Tax Rate is negative or Interest Expense (IS034, IS_INT_EXPENSE) is not available. BANKS, FINANCIALS & REITS Calculated as: ((T12 Net Income (Losses) + T12 Minority Interest) / Average of Total Capital) * 100 Where: Trailing 12 M Net Income is RR813, TRAIL_12M_NET_INC Trailing 12M Minority Interest is RR812, TRAIL_12M_MINORITY_INT Trailing 12M Effective Tax Rate is RR712, TRAIL_12M_EFF_TAX_RTTotal Capital is RR006, BS_TOT_CAP Average is the average of the beginning and ending balances. Trailing 12 month values use the latest 4 quarters, 2 semi annuals or annual. Return on Capital is not computed if the Effective Tax Rate is negative INSURANCE Calculated as: ((T12 Policyholders' Surplus + T12 Net Income (Losses) + T12 Minority Interest + T12 Interest Expense * (1 – (T12 Effective Tax Rate / 100))) / Average of Total Capital) * 100 Where: Trailing 12 M Net Income is RR813, TRAIL_12M_NET_INC Trailing 12M Minority Interest is RR812, TRAIL_12M_MINORITY_INT Trailing 12M Interest Expense is RR804, TRAIL_12M_INT_EXP Trailing 12M Effective Tax Rate is RR712, TRAIL_12M_EFF_TAX_RT Total Capital is RR006, BS_TOT_CAP Trailing 12M Policyholders' Surplus is RR713, TRAIL_12M_POLICY_HOLDER_SURPLUS Average is the average of the beginning and ending balances. Trailing 12 month values use the latest 4 quarters, 2 semi annuals or annual. Return on Capital is not computed if the Effective Tax Rate is negative or Interest Expense (IS034, IS_INT_EXPENSE) is not available. Please reference Return on Capital Adjusted (F1171, RETURN_ON_CAPITAL_ADJUSTED) for the adjusted value that excludes the impact of abnormal items.
31	SALES_GROWTH	A percentage increase or decrease of sales revenue by comparing current period with same period prior year. Calculated as: (Revenue from Current Period – Revenue from Same Period Prior Year) * 100 / Revenue from Same Period Prior Year Where: Revenue is IS010, SALES_REV_TURN Revenue Growth is not computed if Revenue changes signs from prior year to current period. Please reference Revenue Growth Adjusted Year over Year (F1139, REVENUE_GROWTH_ADJUSTED_YOY) for the adjusted value that excludes the impact of abnormal items.

(Table continues)

(Continued)

32	TOT_DEBT_TO_EBITDA	INDUSTRIALS, FINANCIALS, UTILITIES, & REITS Measure of a company's ability to pay off its incurred debt. This ratio gives the investor the approximate amount of time that would be needed to pay off all debt, ignoring the factors of interest, taxes, depreciation and amortization. Unit: Actual. Calculated as: Total Debt / Trailing 12 Month EBITDA Where: Total Debt is RR251, SHORT_AND_LONG_TERM_DEBT Trailing 12 Month EBITDA is RR841, TRAIL_12M_EBITDA
33	TOT_DEBT_TO_TOT_ASSET	Leverage ratio in percentage that defines the total amount of debt relative to assets. This enables comparisons of leverage to be made across different companies. Calculated as: Total Debt *100 / Total Assets Where: Total Debt is RR251, SHORT_AND_LONG_TERM_DEBT Total Assets is BS035, BS_TOT_ASSET
34	TOT_DEBT_TO_TOT_CAP	INDUSTRIALS, BANKS, FINANCIALS, INSURANCE, UTILITIES & REITS Measure of a company's financial leverage that presents its total debt as a percentage of total capital. Calculated as: Total Debt x 100 / Total Capital Where: Total Debt is RR251, SHORT_AND_LONG_TERM_DEBT Total Capital is RR006, BS_TOT_CAP
35	TOT_DEBT_TO_TOT_EQY	INDUSTRIALS, BANKS, FINANCIALS, INSURANCE, UTILITIES & REITS Total debt divided by total shareholders' equity. Calculated as: Short and Long Term Debt / Shareholders' Equity * 100 Where: Short and Long Term Debt is RR251, SHORT_AND_LONG_TERM_DEBT Shareholders' Equity is RR007, TOTAL_EQUITY
36	VOLATILITY_180D	A measure of the risk of price moves for a security calculated from the standard deviation of day to day logarithmic historical price changes. The 180-day price volatility equals the annualized standard deviation of the relative price change for the 180 most recent trading days closing price, expressed as a percentage.When looking at current value, the last price point will be the most recently traded price of the security. Currencies: Non-CDS (Credit Default Swap) currency securities are not supported for historical data.
37	VOLATILITY_260D	A measure of the risk of price moves for a security calculated from the standard deviation of day to day logarithmic historical price changes. The 260-day price volatility equals the annualized standard deviation of the relative price change for the 260 most recent trading days closing price, expressed as a percentage. When looking at current value, the last price point will be the most recently traded price of the security. Portfolio: Standard deviation of daily total returns as computed in the Portfolio & Risk Analytics function.
38	VOLATILITY_90D	Measure of the risk of price moves for a security calculated from the standard deviation of day to day logarithmic historical price changes. The 90-day price volatility equals the annualized standard deviation of the relative price change for the 90 most recent trading days closing price, expressed as a percentage. Currencies: Non-CDS (Credit Default Swap) currency securities are not supported for historical data.
39	WACC	* WACC = [KD x (TD/V)] + [KP x (P/V)] + [KE x (E/V)] * KD = Cost of Debt, TD = Total Debt, V = Total Capital * KP = Cost of Preferred, P = Preferred Equity, KE = Cost of Equity, E = Equity Capital * Total Capital = Total Debt + Preferred Equity + Equity Capital. Figures are drawn from the company's most recent report, annual or interim.
40	WORK_CAP_GROWTH	INDUSTRIALS Measure of one year working capital growth. Unit in Percentage. Calculated as: [(Working Capital in current period - Working Capital in previous period) / Working Capital in previous period * 100] Where: Working Capital is RR150, WORKING_CAPITAL

Note. All definitions, abbreviations, notations are listed as reported by Bloomberg (2017).

Source: Bloomberg, *Bloomberg Terminal*, 2017.

Appendix D: Correlation matrix of financial statement data (2001-2003)

Table 4. Correlation matrix of financial statement data (2001-2003)

	ASSET_GROWTH	BETA_RAW_OVERRIDABLE	BVPS_GROWTH	CAP_EXPEND_RATIO	CASH_FLOW_GROWTH	COM_EQY_TO_TOT_ASSET	CONT_INC_GROWTH	EBIT_MARGIN	EBIT_YR_GROWTH	EBITDA_GROWTH	EBITDA_MARGIN	EMPL_GROWTH	EPS_GROWTH	FREE_CASH_FLOW_MARGIN	GROSS_MARGIN	GROWTH_IN_CAP	INC_TAX_EXP_YR_GROWTH	NET_DEBT_TO_CASHFLOW	NET_INC_GROWTH	NET_WORTH_GROWTH
ASSET_GROWTH	1	0.01	0.33	-0.01	0.06	0.05	0.13	0.21	0.07	0.1	0.12	0.2	0.09	0.1	-0.03	0.77	0.06	-0.13	0.14	0.4
BETA_RAW_OVERRIDABLE		1	0.06	-0.01	0.02	0	-0.1	-0.26	0.02	-0.07	-0.08	0.1	-0.13	0.07	-0.1	-0.04	-0.02	-0.11	-0.07	-0.01
BVPS_GROWTH			1	0.08	0.02	0.12	0.1	0.24	0.03	0.16	0.19	0.09	0.14	0.12	0.17	0.39	0.06	-0.14	0.14	0.79
CAP_EXPEND_RATIO				1	0.08	-0.03	0.09	0.14	0	0.07	0.09	-0.05	0.08	0.6	0.19	-0.02	0.06	-0.25	0.07	0.1
CASH_FLOW_GROWTH					1	0	0.01	0.06	0.07	0.07	0.11	0.07	-0.01	0.19	0.05	-0.02	-0.12	-0.01	0	0.05
COM_EQY_TO_TOT_ASSET						1	0.02	0.07	0.06	0.12	0.03	-0.02	-0.02	0.08	0	0.03	-0.01	-0.36	-0.02	0.09
CONT_INC_GROWTH							1	0.16	0.46	0.32	0.15	0.05	0.85	0.07	0.1	0.01	0.54	-0.06	0.6	0.13
EBIT_MARGIN								1	0.22	0.22	0.79	0.05	0.22	0.31	0.52	0.24	0.12	-0.27	0.18	0.29
EBIT_YR_GROWTH									1	0.54	0.18	-0.09	0.51	0.01	0.12	0.03	0.41	0	0.42	0.09
EBITDA_GROWTH										1	0.21	0.01	0.43	0.13	0.14	0.09	0.26	-0.02	0.42	0.17
EBITDA_MARGIN											1	0.02	0.2	0.25	0.52	0.16	0.07	-0.15	0.2	0.18
EMPL_GROWTH												1	-0.01	0.04	0.04	0.2	-0.01	0.01	-0.08	0
EPS_GROWTH													1	0.05	0.13	0	0.55	-0.06	0.67	0.16
FREE_CASH_FLOW_MARGIN														1	0.3	0.04	0.05	-0.32	0.07	0.13
GROSS_MARGIN															1	0.06	0.11	-0.28	0.12	0.11
GROWTH_IN_CAP																1	-0.01	-0.07	0.08	0.47
INC_TAX_EXP_YR_GROWTH																	1	0	0.35	0.11
NET_DEBT_TO_CASHFLOW																		1	-0.07	-0.13
NET_INC_GROWTH																			1	0.22
NET_WORTH_GROWTH																				1

(Table continues)

(Continued)

	NORMALIZED_PROFIT_MARGIN	NORMALIZED_ROE	OPER_INC_GROWTH	OPER_MARGIN	PROF_MARGIN	REINVEST_EARN_YR_GROWTH	RETURN_COM_EQY	RETURN_ON_ASSET	RETURN_ON_CAP	RETURN_ON_INV_CAPITAL	SALES_GROWTH	TOT_DEBT_TO_EBITDA	TOT_DEBT_TO_TOT_ASSET	TOT_DEBT_TO_TOT_CAP	TOT_DEBT_TO_TOT_EQY	VOLATILITY_180D	VOLATILITY_260D	VOLATILITY_90D	WACC	WORK_CAP_GROWTH
ASSET_GROWTH	0.15	0.16	0.08	0.17	0.1	0.07	0.21	0.05	0.12	-0.02	0.34	-0.1	-0.03	-0.04	-0.07	-0.03	-0.02	0	-0.08	0.01
BETA_RAW_OVERRIDABLE	-0.09	-0.32	-0.01	-0.16	-0.12	-0.01	-0.29	-0.24	-0.32	-0.21	-0.07	0	-0.12	-0.13	-0.16	0.59	0.56	0.5	0.42	-0.01
BVPS_GROWTH	0.21	0.18	0.04	0.23	0.28	0.13	0.33	0.16	0.16	0.12	0.23	-0.18	-0.15	-0.14	-0.15	-0.06	0.01	0.05	0.03	0.05
CAP_EXPEND_RATIO	0.22	0.22	0	0.22	0.24	0.09	0.17	0.05	0.17	0.18	0.14	-0.21	-0.12	-0.07	-0.07	-0.09	-0.03	-0.11	0.11	0.08
CASH_FLOW_GROWTH	0.02	0.03	0.07	0.06	0.03	0.09	0.04	-0.01	0.05	0.04	0.1	0.02	-0.03	-0.04	-0.03	0.02	-0.03	0.03	0.06	-0.03
COM_EQY_TO_TOT_ASSET	0.02	0.03	0.08	-0.05	0.02	0.05	0.04	0.22	0.26	0.14	0.17	-0.46	-0.56	-0.79	-0.65	0.09	0.1	0.06	0.49	0.09
CONT_INC_GROWTH	0.13	0.04	0.46	0.15	0.17	0.39	0.11	0.11	0.05	0.03	0.13	-0.08	-0.08	-0.03	-0.05	-0.09	-0.1	-0.07	-0.03	0.05
EBIT_MARGIN	0.62	0.51	0.19	0.71	0.58	0.16	0.46	0.16	0.34	0.19	0.21	-0.32	-0.05	-0.08	-0.11	-0.26	-0.19	-0.19	-0.04	0.17
EBIT_YR_GROWTH	0.19	0.09	0.92	0.18	0.09	0.42	0.09	0.07	0.09	0.07	0.22	-0.1	-0.05	-0.05	-0.07	-0.06	-0.12	0.02	0	0.03
EBITDA_GROWTH	0.24	0.11	0.48	0.28	0.18	0.37	0.16	0.08	0.07	0.06	0.36	-0.17	-0.09	-0.11	-0.16	-0.05	-0.02	-0.04	0.08	0.04
EBITDA_MARGIN	0.55	0.29	0.15	0.59	0.46	0.14	0.3	0.04	0.1	0.01	0.09	-0.12	0.03	-0.02	0.01	-0.15	-0.09	-0.09	-0.07	0.11
EMPL_GROWTH	0	-0.11	-0.12	0.08	-0.03	-0.02	-0.1	-0.06	0.05	-0.05	0.06	-0.01	0.01	0	-0.03	0.04	0.1	0.12	0.06	0.05
EPS_GROWTH	0.19	0.1	0.51	0.17	0.19	0.49	0.17	0.13	0.05	0.06	0.13	-0.06	-0.04	0.02	-0.01	-0.16	-0.17	-0.15	-0.05	0.05
FREE_CASH_FLOW_MARGIN	0.35	0.19	0	0.42	0.39	0.07	0.12	0.02	0.13	0.11	0.19	-0.31	-0.19	-0.18	-0.18	0.02	0.07	-0.01	0.2	0.17
GROSS_MARGIN	0.51	0.36	0.13	0.48	0.4	0.1	0.25	0.09	0.27	0.16	0.08	-0.3	-0.05	-0.07	-0.12	-0.22	-0.16	-0.18	0.05	0.07
GROWTH_IN_CAP	0.15	0.16	0.04	0.2	0.17	0.05	0.21	0.04	0.19	0	0.27	-0.1	-0.05	-0.03	-0.05	-0.09	-0.03	-0.02	-0.12	-0.04
INC_TAX_EXP_YR_GROWTH	0.09	0.07	0.42	0.09	0.1	0.3	0.08	0.04	0.06	-0.02	0.08	-0.04	-0.04	0.01	-0.01	-0.07	-0.1	-0.05	0	-0.01
NET_DEBT_TO_CASH_FLOW	-0.27	-0.21	0.02	-0.17	-0.24	-0.07	-0.23	-0.14	-0.35	-0.22	-0.15	0.6	0.52	0.56	0.52	-0.03	-0.07	-0.01	-0.41	-0.19
NET_INC_GROWTH	0.25	0.07	0.42	0.22	0.26	0.58	0.19	0.05	0.07	0.09	0.25	-0.06	-0.09	-0.02	-0.03	-0.11	-0.12	-0.1	0	0.07
NET_WORTH_GROWTH	0.23	0.22	0.09	0.25	0.29	0.16	0.33	0.08	0.13	0.09	0.23	-0.19	-0.17	-0.12	-0.17	-0.09	-0.04	0.02	-0.06	0.03

(Table continues)

(Continued)

	NORMALIZED_PROFIT_MARGIN	NORMALIZED_ROE	OPER_INC_GROWTH	OPER_MARGIN	PROF_MARGIN	REINVEST_EARN_YR_GROWTH	RETURN_COM_EQY	RETURN_ON_ASSET	RETURN_ON_CAP	RETURN_ON_INV_CAPITAL	SALES_GROWTH	TOT_DEBT_TO_EBITDA	TOT_DEBT_TO_TOT_ASSET	TOT_DEBT_TO_TOT_CAP	TOT_DEBT_TO_TOT_EQY	VOLATILITY_180D	VOLATILITY_260D	VOLATILITY_90D	WACC	WORK_CAP_GROWTH
NORMALIZED_PROFIT_MARGIN	1	0.41	0.2	0.8	0.63	0.19	0.35	0.02	0.16	0.13	0.07	-0.25	-0.1	-0.03	-0.05	-0.23	-0.16	-0.18	-0.06	0.14
NORMALIZED_ROE		1	0.1	0.4	0.4	0.03	0.72	0.33	0.49	0.43	0.14	-0.27	0.02	0.01	0.01	-0.27	-0.21	-0.23	-0.06	0.13
OPER_INC_GROWTH			1	0.18	0.15	0.45	0.11	0.11	0.11	0.1	0.24	-0.14	-0.09	-0.07	-0.08	-0.11	-0.17	-0.06	0.03	0.04
OPER_MARGIN				1	0.74	0.16	0.31	0.04	0.18	0.1	0.13	-0.24	-0.02	0.03	-0.02	-0.25	-0.15	-0.16	-0.14	0.19
PROF_MARGIN					1	0.18	0.4	0.15	0.28	0.18	0.12	-0.38	-0.19	-0.1	-0.14	-0.26	-0.17	-0.18	0.02	0.18
REINVEST_EARN_YR_GROWTH						1	0.11	0.06	0.11	0.15	0.19	-0.11	-0.14	-0.08	-0.07	-0.06	-0.06	-0.07	0.09	0.05
RETURN_COM_EQY							1	0.36	0.46	0.39	0.17	-0.28	-0.05	-0.03	-0.03	-0.31	-0.27	-0.21	-0.04	0.1
RETURN_ON_ASSET								1	0.42	0.43	0.19	-0.27	-0.09	-0.17	-0.16	-0.18	-0.14	-0.13	0.15	0.04
RETURN_ON_CAP									1	0.47	0.26	-0.5	-0.29	-0.3	-0.32	-0.34	-0.31	-0.29	0.3	0.11
RETURN_ON_INV_CAPITAL										1	0.21	-0.39	-0.24	-0.21	-0.19	-0.21	-0.19	-0.23	0.26	0.11
SALES_GROWTH											1	-0.16	-0.13	-0.18	-0.19	0	0	-0.02	0.16	0.03
TOT_DEBT_TO_EBITDA												1	0.72	0.71	0.72	0.04	0.01	0.01	-0.51	-0.11
TOT_DEBT_TO_TOT_ASSET													1	0.87	0.74	-0.12	-0.1	-0.05	-0.59	-0.06
TOT_DEBT_TO_TOT_CAP														1	0.85	-0.17	-0.17	-0.12	-0.65	-0.09
TOT_DEBT_TO_TOT_EQY															1	-0.16	-0.16	-0.14	-0.59	-0.05
VOLATILITY_180D																1	0.9	0.75	0.29	0
VOLATILITY_260D																	1	0.75	0.27	0.03
VOLATILITY_90D																		1	0.13	0.03
WACC																			1	0.02
WORK_CAP_GROWTH																				1

Appendix E: Rotated component matrix (2007)

Table 5. Rotated component matrix (2007)

	Accounting ratio	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Indebtedness	WACC	**0.25**	-0.10	0.02	0.21	-0.03	0.12
	TOT_DEBT_TO_TOT_EQY	**-0.38**	-0.03	0.02	-0.01	-0.04	0.03
	NET_DEBT_TO_CASHFLOW	-0.16	-0.07	0.01	-0.14	0.00	-0.13
	CASH_FLOW_GROWTH	0.12	-0.11	0.07	0.06	-0.01	0.03
	TOT_DEBT_TO_TOT_ASSET	**-0.37**	0.04	0.00	0.03	0.01	0.00
	TOT_DEBT_TO_EBITDA	**-0.30**	-0.05	0.00	0.09	0.01	-0.14
	COM_EQY_TO_TOT_ASSET	**0.38**	0.08	0.00	0.05	-0.02	-0.07
	TOT_DEBT_TO_TOT_CAP	**-0.41**	0.01	0.01	0.02	-0.03	0.01
Margins	NORMALIZED_PROFIT_MARGIN	-0.03	**0.40**	0.03	0.00	0.05	0.05
	FREE_CASH_FLOW_MARGIN	0.08	0.24	-0.02	0.03	-0.01	-0.08
	EBIT_MARGIN	0.06	**0.35**	-0.02	-0.04	-0.01	0.09
	EBITDA_MARGIN	-0.14	**0.31**	-0.03	0.08	0.09	-0.04
	PROF_MARGIN	0.03	**0.37**	0.02	-0.03	-0.06	0.07
	OPER_MARGIN	-0.08	**0.39**	0.02	-0.01	0.08	0.01
	CAP_EXPEND_RATIO	0.12	0.16	-0.03	0.04	0.03	-0.08
	GROSS_MARGIN	0.05	**0.36**	0.06	-0.08	-0.12	-0.04
Operating growth	OPER_INC_GROWTH	0.08	-0.08	**0.33**	-0.10	0.23	-0.03
	NET_INC_GROWTH	-0.02	0.00	**0.37**	-0.01	-0.07	0.07
	EBITDA_GROWTH	0.06	0.03	**0.32**	0.01	0.18	-0.12
	CONT_INC_GROWTH	-0.04	-0.04	**0.41**	0.02	-0.08	0.02
	EPS_GROWTH	-0.07	-0.01	**0.37**	0.05	-0.15	0.09
	INC_TAX_EXP_YR_GROWTH	0.02	0.07	**0.25**	0.04	0.00	-0.12
	EBIT_YR_GROWTH	0.02	-0.11	**0.35**	-0.10	0.14	0.10
	REINVEST_EARN_YR_GROWTH	0.01	0.12	**0.29**	0.10	-0.21	-0.09
Volatility	BETA_RAW_OVERRIDABLE	-0.02	-0.06	0.03	**0.40**	0.06	0.02
	VOLATILITY_260D	0.01	-0.01	0.01	**0.48**	-0.03	-0.02
	VOLATILITY_180D	0.01	-0.01	-0.02	**0.47**	0.00	0.00
	VOLATILITY_90D	-0.02	0.04	0.00	**0.47**	0.04	-0.02
Company growth	EMPL_GROWTH	0.05	0.05	-0.06	0.04	**0.33**	-0.04
	ASSET_GROWTH	-0.05	0.01	0.01	0.07	**0.39**	-0.01
	SALES_GROWTH	-0.01	0.10	0.20	-0.01	**0.26**	-0.02
	NET_WORTH_GROWTH	0.03	-0.04	0.00	-0.04	**0.36**	0.03
	GROWTH_IN_CAP	-0.07	0.04	-0.02	0.10	**0.40**	-0.01
	BVPS_GROWTH	0.03	-0.03	-0.07	-0.09	**0.38**	0.10
Profitability	NORMALIZED_ROE	-0.17	0.04	0.03	-0.01	0.00	**0.50**
	RETURN_ON_INV_CAPITAL	0.08	-0.05	0.01	-0.01	0.05	**0.36**
	RETURN_ON_ASSET	0.15	0.09	-0.01	0.04	0.00	**0.31**
	RETURN_COM_EQY	-0.16	0.04	0.00	0.06	-0.02	**0.49**
	WORK_CAP_GROWTH	-0.07	-0.05	-0.05	0.01	0.05	0.09
	RETURN_ON_CAP	0.18	-0.05	0.00	-0.05	-0.02	**0.33**

Note. Coefficients with an absolute value greater than 0.25 are bolded.

Appendix F: Rotated component matrix (2009)

Table 6. Rotated component matrix (2009)

	Accounting ratio	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Volatility	BETA_RAW_OVERRIDABLE	**-0.35**	0.06	-0.02	-0.07	-0.05	-0.04
	EMPL_GROWTH	-0.19	-0.15	0.02	0.00	0.17	-0.07
	VOLATILITY_260D	**-0.37**	0.01	0.03	-0.04	-0.08	-0.08
	VOLATILITY_180D	**-0.43**	0.01	0.01	-0.02	-0.03	-0.05
	VOLATILITY_90D	**-0.36**	0.01	-0.04	-0.01	0.00	-0.01
Indebtedness	WACC	-0.01	**-0.35**	0.03	-0.07	0.07	0.02
	TOT_DEBT_TO_TOT_EQY	-0.07	**0.40**	-0.01	-0.01	0.10	0.03
	NET_DEBT_TO_CASHFLOW	-0.03	0.22	0.03	-0.01	-0.13	0.01
	TOT_DEBT_TO_TOT_ASSET	0.06	**0.41**	0.00	0.04	0.01	-0.05
	TOT_DEBT_TO_EBITDA	0.06	**0.33**	0.03	0.00	-0.13	-0.05
	COM_EQY_TO_TOT_ASSET	0.13	**-0.33**	-0.06	0.10	-0.18	-0.14
	TOT_DEBT_TO_TOT_CAP	-0.06	**0.40**	0.00	-0.04	0.14	0.05
Operating growth	OPER_INC_GROWTH	-0.08	-0.03	**0.41**	0.10	-0.02	0.12
	NET_INC_GROWTH	0.20	0.06	**0.31**	-0.05	-0.09	-0.15
	SALES_GROWTH	-0.07	-0.06	0.24	0.02	0.12	-0.09
	EBITDA_GROWTH	-0.14	-0.05	**0.29**	0.11	0.14	0.09
	CONT_INC_GROWTH	0.15	0.03	**0.36**	-0.09	-0.04	-0.08
	EPS_GROWTH	0.17	0.05	**0.33**	-0.07	-0.04	-0.09
	CASH_FLOW_GROWTH	-0.10	0.05	0.14	-0.05	0.09	0.05
	INC_TAX_EXP_YR_GROWTH	-0.01	0.03	0.17	0.11	0.00	-0.01
	EBIT_YR_GROWTH	-0.18	-0.03	**0.40**	0.12	0.01	0.15
	REINVEST_EARN_YR_GROWTH	0.18	0.02	**0.29**	-0.09	-0.12	-0.06
Margins	NORMALIZED_PROFIT_MARGIN	0.03	-0.02	0.03	**0.34**	0.07	0.01
	FREE_CASH_FLOW_MARGIN	0.07	0.04	-0.14	**0.20**	0.08	-0.17
	EBIT_MARGIN	-0.10	-0.04	-0.02	**0.44**	0.00	0.06
	EBITDA_MARGIN	-0.08	0.04	-0.02	**0.47**	-0.14	0.00
	PROF_MARGIN	0.11	-0.01	0.04	0.23	0.12	-0.01
	OPER_MARGIN	0.06	-0.03	0.08	**0.29**	0.05	0.04
	GROSS_MARGIN	0.09	0.09	-0.01	**0.39**	-0.09	-0.10
Profitability	NORMALIZED_ROE	0.03	0.05	-0.02	-0.02	**0.47**	0.08
	RETURN_ON_INV_CAPITAL	-0.06	-0.10	0.07	0.03	**0.35**	0.07
	RETURN_ON_ASSET	0.12	-0.01	-0.06	0.10	**0.29**	-0.15
	CAP_EXPEND_RATIO	-0.02	0.01	-0.07	0.11	0.21	-0.12
	RETURN_COM_EQY	0.12	0.14	0.01	-0.02	**0.41**	-0.10
	RETURN_ON_CAP	0.18	-0.16	-0.01	-0.10	0.24	0.01
Company growth	ASSET_GROWTH	-0.16	-0.06	0.11	-0.05	0.12	**-0.31**
	NET_WORTH_GROWTH	-0.09	0.01	0.04	0.00	-0.02	**-0.47**
	GROWTH_IN_CAP	-0.04	0.00	0.01	0.00	0.05	**-0.44**
	BVPS_GROWTH	-0.05	-0.02	0.01	0.00	-0.01	**-0.46**
	WORK_CAP_GROWTH	0.09	0.03	-0.06	0.02	-0.13	-0.19

Note. Coefficients with an absolute value greater than 0.25 are bolded.

Appendix G: Distribution of the residuals and Q-Q plot (Strategy I)

Figure 1. Distribution of the residuals (Strategy I)

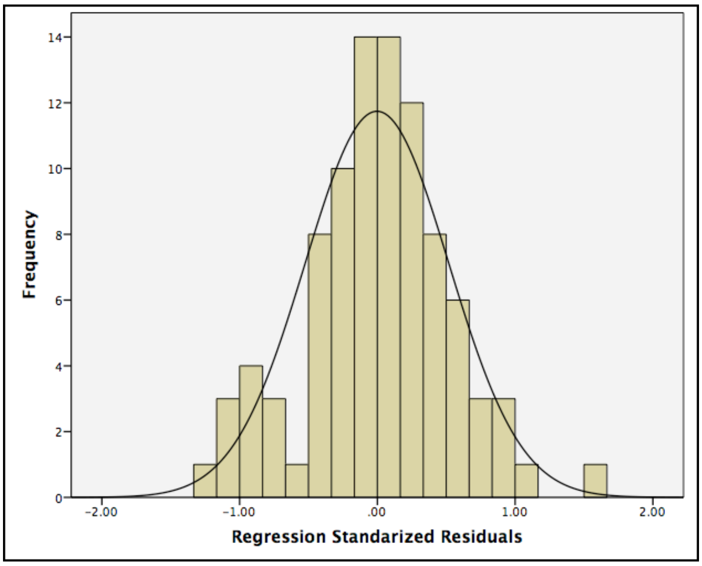
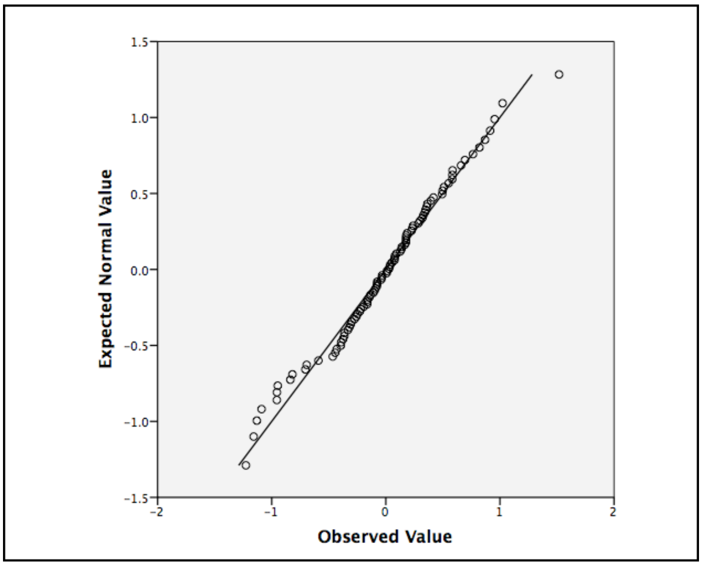


Figure 2. Q-Q plot (Strategy I)



Appendix H: Distribution of the residuals and Q-Q plot (Strategy II)

Figure 3. Distribution of the residuals (Strategy II)

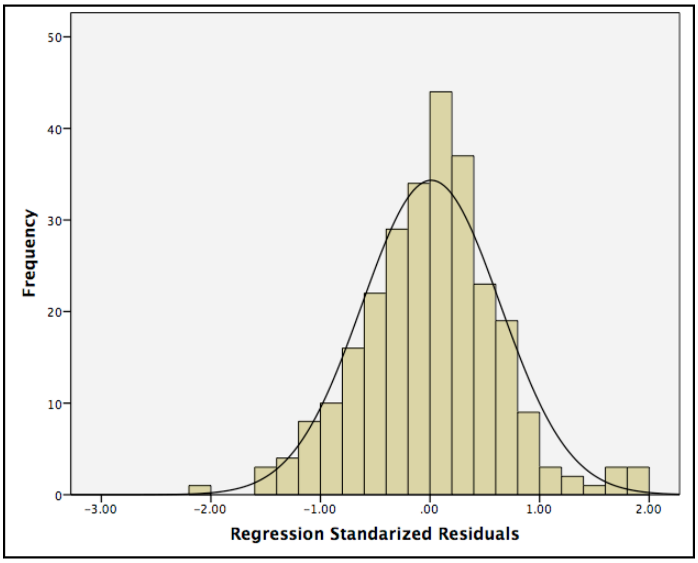
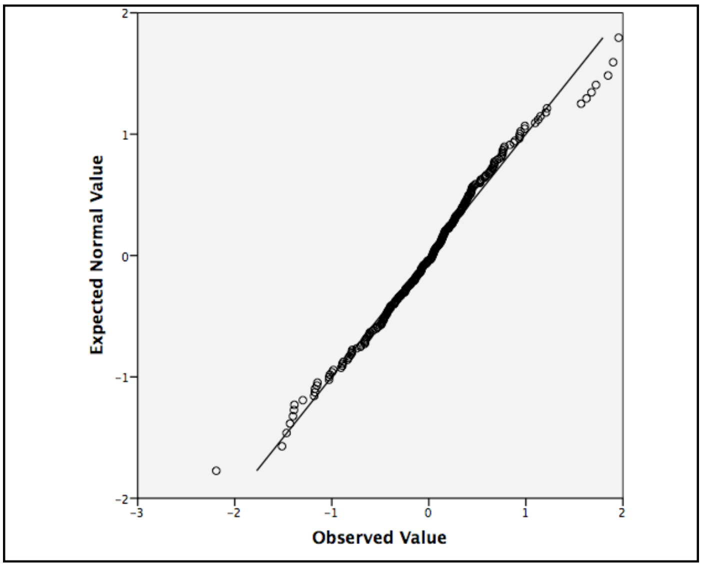


Figure 4. Q-Q plot (Strategy II)



Appendix I: Distribution of the residuals and Q-Q plot (Strategy III)

Figure 5. Distribution of the residuals (Strategy III)

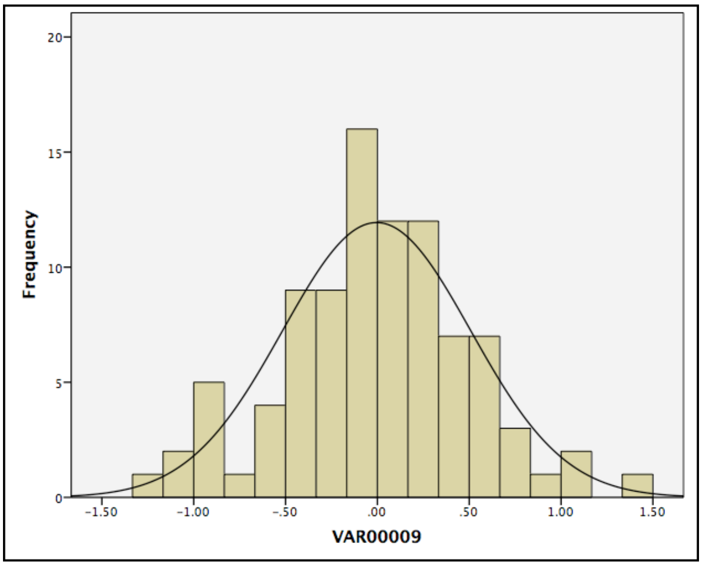


Figure 6. Q-Q plot (Strategy III)