Stockholm School of Economics
Department of Accounting
Master Thesis in Accounting and Financial Management
Fall 2016

Trading on Parsimonious Prediction Models: Simply Genius or Genuinely Too Simplistic?

A 'fundamental analysis'-based test of market efficiency in the U.S.

Frederik Motzet[Ω] **Tilman Schwarzenberg**[Ψ]

Abstract

In the mature field of fundamental analysis (FA) and market efficiency tests, the study of Skogsvik & Skogsvik (2010) stands out in that it generates significant abnormal returns using a parsimonious ROE-based investment strategy in the Swedish market. To test the robustness of their results across countries and time periods, the aim of this thesis is twofold: First, it is investigated whether a simple FA-based trading strategy can generate similar excess returns on a large U.S. manufacturing sample. Second, market efficiency and its time-series behavior in the U.S. market are assessed. This paper replicates Skogsvik & Skogsvik's (2010) investment criteria by employing a prediction model of medium-term ROE changes and an indicator variable revealing mispriced stocks based on residual income valuation. Stock positions are taken in a contrarian fashion and held for 36 months over the period 1979-2014. Despite a strong prediction accuracy of 68% for future medium-term ROE changes, no evidence of significant abnormal hedge returns is found. Additional tests suggest that mitigating critical model weaknesses could improve the results, but also impose non-trivial cost of complexity. Moreover, the theoretical abnormal return potential is found to diminish over time in the U.S. market. While this supports the notion of increasing efficiency, preliminary evidence of 'decoupling' between prices and fundamentals in more recent years might point towards 'crazy' rather than 'efficient' prices.

Keywords: *Market efficiency testing; Market mispricing; ROE prediction; Fundamental valuation; Fundamental analysis; Residual income valuation*

Tutor: Kenth Skogsvik, Professor, Department of Accounting at Stockholm School of Economics

Date: December 7th, 2016

Ω 40712@student.hhs.se
Ψ 40711@student.hhs.se

ACKNOWLEDGEMENTS

We would hereby like to take the opportunity to thank our friends and families for their continuous support and invaluable patience over the course of our writing process. Furthermore, we are very grateful for the guidance in statistical and methodological matters provided by Florian Eugster and Per-Olov Edlund. Finally, we would like to show our deepest appreciation and gratitude for our tutor, Kenth Skogsvik, who shared valuable insights and supported us throughout the progress of our work.

Stockholm, December 2016

TABLE OF CONTENTS

1. INTRODUCTION

The ‘efficient market hypothesis’ has attracted researchers – advocates as well as opponents – since its introduction in the 1960s. Of particular interest in an accounting context are tests of market efficiency that adopt fundamental analysis (FA) of publicly available financial statements to uncover mispriced securities and earn abnormal returns. The association of FA with claims of market inefficiency stems from the historical evolution of capital markets research, whose focus shifted from the ‘information content’ of accounting attributes for observable stock prices (Ball & Brown, 1968) to their predictive power for future earnings and stock price changes (Lee, 1999; Bernard, 1995; Penman, 1992). While this research area has matured considerably since its revival in the 1990s, Richardson, Tuna & Wysocki (2010) emphasize its unabated relevance: “Research into accounting anomalies and fundamental analysis is far from dead. […] Indeed, as a profession, we may have barely touched the surface of [the forecasting role of accounting]” (p. 444).

To explore this relationship, a vast amount of literature focuses on complex multivariate prediction models based on a plethora of accounting ratios (Ou & Penman, 1989; Lev & Thiagarajan, 1993; Abarbanell & Bushee, 1997, 1998). The bulk of these studies has been criticized for an inherent “kitchen sink approach” (Richardson et al., 2010, p. 424) that seeks to identify forecasting attributes from a purely statistical angle. In this complexity-driven research branch, the study by Skogsvik & Skogsvik (2010) stands out in that its investment strategy is based on the simple univariate prediction of medium-term ROE. Focusing on practicability and a sound theoretical motivation, their study not only corroborates the predictive power of ROE for future earnings changes (Freeman, Ohlson & Penman, 1982; Skogsvik, 2008), but also provides evidence of modeling mispricing based on the divergence of market prices from ‘historically justified’ fundamental values. By taking investment positions when the ROE predictions and market expectations differ, their indicator variable strategy generates substantial monthly CAPM excess returns of up to 0.8%.

Given the limited sample of Skogsvik & Skogsvik (2010) in the Swedish stock market, our thesis aims to test the validity of their results across countries and time periods. Specifically, we replicate their strategy and apply it to a large sample of U.S. manufacturing firms in the period 1979-2014. In doing so, we intend to investigate the two main research questions of this thesis: First, whether a parsimonious FA-based trading strategy can generate abnormal returns in the U.S. market similar to those observed in the Swedish environment. And second, what observations can be made regarding market efficiency and its development over time in the U.S. context. The second question is particularly intriguing, as Skogsvik & Skogsvik (2010) noticed diminishing abnormal returns over time. The authors consider this an indication of increasing market efficiency, attributable to rising investor sophistication and information accessibility. Finally, as we acknowledge that "deficient research design choices can create the false appearance of market inefficiency" (Kothari, 2001, p. 208), the consideration of data-fitting pitfalls and risk explanations is a complementary research aim.

While the results under a hypothetical 'perfect foreknowledge' scenario reveal substantial rewards to the forecasting exercise, our simple trading strategies fail to produce significant abnormal hedge returns in the U.S. setting. Further empirical analyses point towards central model deficiencies, whose mitigation offers possible return improvements. Similar to observations in the Swedish market, the abnormal return potential is found to diminish over time, indicating either increasing efficiency or rising speculation in U.S. market prices.

The remainder of this paper is structured as follows: Section 2 presents a review of literature in the field of market efficiency testing with a focus on FA-based tests and the Skogsvik & Skogsvik (2010) study. After putting our contributions into context, the research design and examined data sample are presented (section 3 and 4). Section 5 reports the results of our empirical analysis, followed by section 6, which scrutinizes the underlying model assumptions in greater detail to identify potential causes of the poor model performance. In section 7, the main findings are discussed with regards to our two research questions and possible limitations are summarized. Finally, section 8 offers concluding remarks.

2. LITERATURE REVIEW

In order to provide a common understanding for the discussion of our results in section 5 and to put the contributions of our thesis into the context of existing literature, this section first offers a general introduction to market efficiency testing. Thereafter, the concepts of fundamental analysis (FA) and FA-based market efficiency tests are explained in greater detail, culminating in a comprehensive description of the main components of the Skogsvik & Skogsvik (2010) model, which forms the basis for our research design outlined in section 3.

2.1. INTRODUCTION TO MARKET EFFICIENCY TESTING

2.1.1. EFFICIENT MARKET HYPOTHESIS AND ANOMALIES LITERATURE

In the first half of the 20^{th} century, much of the attention of financial research focused on the time series of stock prices and whether recurring patterns could be detected that give insights into the future stock price development (Roberts, 1959). In response to this prominent research field, Fama (1965) introduced the theory of an efficient market in which "large numbers of rational, profit-maximizers [are] actively competing, with each trying to predict future market values of individual securities, and where important current information is almost freely available to all participants" (p. 56). According to this theory, stock prices are a good estimate of intrinsic value and reflect all available information quickly, correctly and to the full extent

(Fama, 1965, 1970). The uncertainty about future news allows room for disagreement between investors, causing prices to wander randomly about their intrinsic values. In this scenario, uninformed investors would obtain a rate of return as generous as that achieved by experts, since there is no long-time benefit in conducting any form of analysis (Malkiel, 2003). The so-called efficient market hypothesis (EMH) as outlined above implies that investors cannot generate abnormal returns due to i) the unpredictability of prices and ii) the competitive nature of a frictionless market. In this context, mispricing will either not occur or be adjusted immediately by the well-informed and sophisticated investors (Malkiel, 2003).

Ever since the concept of efficient markets was introduced in the 1960s, researchers in the field of the anomalies literature have tried to reject the EMH by uncovering predictable abnormal returns. Furthermore, the strict implications of the hypothesis have been widely criticized for being deeply rooted in inflexible statistical concepts (e.g. Ball, 1994) and for disregarding existing market frictions such as transaction costs or behavioral aspects that most likely affect investors' decisions (De Bondt & Thaler, 1985). Therefore, the EMH in its 'purest' form, i.e. conditioned on unpredictable prices and frictionless markets, can indeed be seen as a very strong null hypothesis that leaves ample room for its rejection and thus suspected anomalies. However, several researchers (e.g. Fama, 1970; Kothari, 2001) emphasize that the strict assumptions of statistically independent prices and a frictionless market can be loosened without rejecting the EMH. Given the ambiguous results of other studies in trying to explain market price phenomena, Fama (1998) praises the EMH for not only offering a hypothesis that can be tested and rejected, but also for providing a simple answer to apparent anomalies: the expected value of abnormal returns is zero, but over- and underreaction by chance generate alleged anomalies that split randomly between the two.

Nevertheless, by loosening the strict conditions of the null hypothesis, the question about the tolerable degree of deviations from 'normal returns' remains unanswered and creates controversy. It is in the context of this strong null hypothesis with divergent interpretations that the anomalies literature comes into play, providing evidence for systematic abnormal return possibilities and thereby suggesting a "mounting evidence of apparent market inefficiency" (Kothari, 2001, p. 107). To understand how market efficiency advocates respond to – and justify – the suspected anomalies, the following section depicts some of the most important limitations associated with previous anomalies studies.

2.1.2. COMMON CRITICISM OF THE ANOMALIES LITERATURE

As outlined before, abnormal returns proclaimed in the anomalies literature cannot be lightly accepted as proof of market inefficiency. To draw conclusions for our own research design and further analyses, three of the most common objections are summarized in this section.

Data problems

One of the concerns most widely expressed with regards to anomalies studies is the problem of poor data quality in the underlying sample and false inferences drawn from specific study results. There is a vast amount of literature claiming that researchers are inclined to find unusual results that reject the common notion of market efficiency instead of contributing to the 'boring' alternative by confirming the hypothesis (Fama, 1998; Schwert, 2003; Malkiel, 2003). This focus on surprising results leads to two types of problems that relate to narrow data sets and limited out-of-sample validity.

First, researchers may intentionally fit the data and manipulate their sample until they have found a predictable pattern or anomaly (Malkiel, 2003). One way of doing so is by limiting the sample to stocks with extreme characteristics and thus actively adjusting data sets to achieve desired research results (Kothari, 2001). Second, data biases can occur implicitly through the research design. This so-called 'data snooping' is prevalent when researchers examine positively correlated data of anomalies studies to confirm the previous results (Schwert, 2003). By limiting the data set to the same sample used to derive a certain model, statistical test results may become invalid (Lo & MacKinlay, 1990).

These two types of data biases should be seen as a crucial reminder to treat potential evidence of anomalies with caution. Thus, our research design and sample selection incorporates measures to avoid data problems and acknowledges the limitations regarding inferences.

Risk adjustment in asset-pricing models

Another concern regarding the findings of the anomalies literature is its reliance on asset-pricing models to distinguish expected from abnormal returns (Ball, 1994; Schwert, 2003). Consequently, tests of market efficiency are often claimed to be simultaneous tests of asset-pricing models, since anomalies can either stem from inefficient prices or flaws in the underlying pricing models (Fama, 1970; Ball, 1978; Schwert, 2003).

These model flaws can emerge from i) inaccurately measuring risk or ii) omitting risk factors (Ball, 1978; Kothari, 2001). Risk measurement errors may occur due to the fact that risk proxies (e.g. systematic market beta) and their impact on expected returns need to be estimated for unobservable risk factors (Ball, 1994). The second reason, the omission of risk factors, is a criticism frequently directed towards the widely applied CAPM, which was introduced by Sharpe (1964) and Lintner (1965). By neglecting potential risk factors besides the systematic risk and market volatility, expected returns might be misstated (Kothari, 2001). Thus, apparent abnormal returns might in fact merely represent a fair compensation for risks not taken into account by the CAPM (Fama & French, 1993).

An asset-pricing model that is widely considered to be a superior risk metric for anomalies testing is the 'three-factor model' brought forward by Fama & French in 1993 (Malkiel, 2003; Schwert, 2003; Stambaugh, Yu & Yuan, 2011). Studies show that after incorporating size and value risk proxies of the 'three-factor model', several suspected anomalies were found to be in line with market efficiency (Fama, 1998; Malkiel, 2003; Schwert, 2003). In order to avoid the pitfalls of incomplete risk estimations, our research design therefore extends beyond the CAPM by including the 'three-factor model' as an empirically highly robust return metric.

Trading costs

The third common explanation of abnormal returns by market efficiency advocates refers to the anomalies literature's neglect of costs related to trading. If taken into account, apparent abnormal returns are deemed to be insignificant, as they are eliminated by incurred costs (Schwert, 2003). Consequently, as investors do not see any value in exploiting arbitrage opportunities, the anomalies continue to exist (Beaver, McNichols & Price, 2016).

To understand which costs can decrease abnormal returns to an expected level, Ball (1994) presents two categories: first, the cost of producing information, and second, the cost of acting upon this information – the 'transaction cost'. There is agreement among researchers that the former has been reduced by an increase in information accessibility (e.g. Chordia, Roll & Subrahmanyam, 2011) and that the latter should be taken into account when interpreting returns (e.g. Jensen, 1978). However, measurement problems prevail, as there is no agreed-upon quantification standard due to divergent costs for different groups of investors.

While the inclusion of trading costs in return calculations thus remains subject to debate, related concerns on the practicability of trading strategies deserve a separate discussion. Most of the anomalies studies use hedge returns as the measure of abnormal return. However, institutional and regulatory constraints commonly restrict short-trading in various markets. These restrictions can be associated with high transaction cost, which increase due to low liquidity or additional capital requirements at a later stage (Stambaugh et al., 2011). Beaver et al. (2016) criticize the 'zero-cost assumption' of long-short investment strategies brought forward by the anomalies literature, which unrealistically implies the ability of investors to borrow unlimited amounts at zero cost, despite being prohibited by some regulatory bodies.

Building on the outlined criticism regarding trading costs and regulatory constraints, we are mindful of these issues in the interpretation of our results. As a sophisticated quantification is non-trivial, we refrain from using average values in our research design. However, by scrutinizing the returns of the long and short positions separately, we are able to analyze the results with respect to potential constraints on the short position.

2.2. FUNDAMENTAL ANALYSIS

Following the general introduction to the EMH and the anomalies literature, this section elaborates on the concept of fundamental analysis and associated tests of market efficiency, which form the basis for the model design and trading strategies investigated in this thesis.

According to Penman (2012, p. 84), "fundamental analysis is the method of analyzing information, forecasting payoffs from that information, and arriving at a valuation based on those forecasts". FA is based on the notion that historical financial statement information can be profitably employed to arrive at an intrinsic value that either confirms or routinely deviates from observable market prices, with the latter indicating market inefficiency. While Penman (2012) emphasizes its useful role in conceptualizing relevant value drivers, FA is primarily aimed at identifying mispriced securities and earning excess returns, which increase in the difference between a firm's price and intrinsic value (Kothari, 2001).

As forecasting lies at the heart of FA, fundamental valuation models form a logical starting point by providing the theoretical foundation and informing the selection of value-relevant accounting attributes to be analyzed and predicted (Richardson et al., 2010).

2.2.1. FUNDAMENTAL VALUATION

The fundamental investor critically relies on valuation models that establish robust theoretical and empirical linkages between accounting numbers and firm values. While the theoretical underpinnings of the 'dividend discount model' (DDM) are widely accepted in accounting research, its empirical and practical limitations (Lee, Myers & Swaminathan, 1999; Francis, Olsson & Oswald, 2000) have given rise to several accounting-based alternatives. Among these, the residual income valuation (RIV) model has gained particular prominence in FA research (Lee, 1999; Penman & Sougiannis, 1998; Frankel & Lee, 1998). While its origins can be traced back to Preinreich (1938) and Edwards & Bell (1961), it was not until Ohlson (1995) and Feltham & Ohlson (1995) that the model pioneered the 'measurement perspective' in capital markets research (Bernard, 1995). Following their reasoning, the RIV model is a direct transformation of the DDM and defines firm value as the sum of the book value of owners' equity and the present value of future residual income. Firm value can hence be expressed as a function of future residual earnings ad infinitum, required returns and the present book value of owners' equity:

$$V_0 = BV_0 + \sum_{t=1}^{\infty} \frac{RI_t}{(1 + r_E)^t} \tag{1}$$

where

V_0 = intrinsic value of owners' equity at time t=0,

BV_0 = book value of owners' equity at time t=0,

RI_t = residual income at time t,

r_E = required rate of return on owners' equity.

The usefulness of the Ohlson (1995) RIV model is related to its three underlying assumptions: i) the value of owners' equity is equal to the present value of future expected dividends, ii) the clean surplus relation[1] applies and iii) residual income depicts 'linear information dynamics'. While i) and ii) ensure that the RIV model is equivalent to the DDM, assumption iii) defines income as an autoregressive process, which presumes abnormal earnings to be competed away in the long run. The advantage of this assumption is that it allows investors to linearly derive future residual income from current earnings, reducing the forecasting complexity (Lee, 1999). However, empirical tests of the 'linear information dynamics' property have yielded mixed results (e.g. Myers, 1999; Dechow, Hutton & Sloan, 1999).

Various applications of the RIV model have confirmed its superior relative performance in explaining and predicting cross-sectional variations in stock prices (e.g. Penman & Sougiannis, 1998; Frankel & Lee, 1998; Jorgensen, Lee & Yoo, 2011). However, Skogsvik & Skogsvik (2010) rightfully object that many studies incorporate analyst earnings forecasts, which potentially entail information other than what is implied by firms' fundamentals. Accordingly, one cannot exclude the possibility that the strong empirical performance is at least partly due to the market's overreliance on analyst data, rather than information inherent in current earnings and book values. This is corroborated by Dechow et al. (1999) who conclude that a simple earnings capitalization model using analyst earnings forecasts explains observable stock prices better than the RIV model. Nevertheless, the theoretical underpinnings, the completeness of the value estimate and the practical applicability (Lee et al., 1999) provide strong arguments to adopt the RIV model in the context of FA research.

2.2.2. FUNDAMENTAL RATIOS AND EARNINGS PREDICTION

As implied by the RIV model in Eq. (1), forecasting future (residual) earnings is essential to the assessment of a firm's value and the potential identification of market mispricing. Hence, Feltham & Ohlson's (1995) work fueled an already growing research interest in accounting-based predictions of future firm performance (Lee, 1999). The underlying assumption of this research stream is that accounting ratios derived from a firm's financial statements can serve as reliable predictors to i) forecast future earnings (changes) and ii) effectively identify

[1] The clean surplus relation implies that all changes in the book value of owners' equity can be explained by net income positions, dividends and capital contributions.

mispriced securities (Kothari, 2001). Traditionally, price-earnings (P/E) and price-to-book (P/B) ratios have enjoyed widespread adoption, since price is assumed to reflect information on future earnings (e.g. Penman, 1996, 1998; Fama & French, 2006). However, to be useful in the context of FA research, our focus is on purely fundamentals-based ratios.

In terms of univariate ratios, book return on owners' equity (ROE) is considered the "primary financial ratio" (Penman, 1992, p. 480) and principal component of the DuPont system, which summarizes profitability, operating and financing ratios. The empirical properties of ROE have been explored extensively, with the most intriguing results being: i) current levels of ROE are indicative of future ROE levels, ii) ROE is subject to a mean reversion process, and iii) historical ROE tends to predict future earnings changes (Penman, 1991; Freeman, Ohlson & Penman, 1982; Beaver, 1970). Characteristic ii) is of particular importance in facilitating earnings forecasts, as it presumes (extreme) levels of ROE to revert to an industry-wide average in the long run. However, Penman (1991) concludes that ROE in itself is not sufficient to reliably predict future profitability. He suggests that additional information is required to separate persistent from transitory ROE, and that this information is highly correlated with the firm's P/B ratio, which indicates the speed of mean reversion. Nevertheless, several studies found other accounting ratios to add little incremental value to the forecasting power of ROE (Fairfield, Sweeney & Yohn, 1996; Dechow et al., 1999) or to even deteriorate the prediction results (Skogsvik, 2008).

Furthermore, a large research strand has examined whether multivariate sets of financial ratios have significant predictive ability for future earnings and stock prices. Ou & Penman (1989) pioneered the multivariate analysis by applying rigorous statistical tests to 68 accounting ratios across a large sample of U.S. firms. They arrived at a subset of ratios with the highest forecasting power for future earnings and relatively strong prediction accuracy for stock returns. While several studies replicated this approach (e.g. Holthausen & Larcker, 1992; Setiono & Strong, 1998), it has been repeatedly criticized for the lack of theoretical ex-ante motivation. As a response, Lev & Thiagarajan (1993) conceptually derived 12 fundamental signals of earnings quality from expert judgments and economic theory and found them to be highly correlated with future earnings growth. Their results were further substantiated by Abarbanell & Bushee (1997, 1998) who confirmed the predictive power of fundamentals such as inventory changes, using a subset of Lev & Thiagarajan's (1993) ratios.

Despite growing efforts to enforce a coherent structure upon the financial ratio analysis and determine the incremental value of disaggregated ratios (e.g. Fairfield et al., 1996; Nissim & Penman, 2001; Penman & Zhang, 2006), this area of research remains largely inconclusive regarding the choice of the 'right' financial ratios, i.e. fundamental ratios with the highest forecasting power for future earnings. While combinations of multivariate ratios are often likely to yield marginal improvements, the robust time-series properties and conceptual strength make simple earnings measures such as ROE particularly appealing to FA research.

2.2.3. FUNDAMENTAL TESTS OF MARKET EFFICIENCY

In response to the categorical claims of the EMH (see 2.1), capital markets research in accounting has produced a large body of studies investigating the informational efficiency of stock prices. The literature can be broadly classified into i) long- and short-horizon event studies and ii) cross-sectional tests of return predictability (Kothari, 2001). While the former stream mostly deals with the post-earnings-announcement drift (e.g. Bernard & Thomas, 1990; Foster, Olsen & Shevlin, 1984), the latter constitutes the aforementioned anomalies literature, which tests whether investment strategies based on indicators of market mispricing can systematically earn abnormal returns. FA has become a popular tool to extract such indicators of mispricing using historical accounting ratios and fundamental values. Following Skogsvik & Skogsvik (2010), the relevant literature is discussed with a separate lense on forecasting mispricing and modeling mispricing.

Tests of forecasting mispricing

Forecasting mispricing implies that stock prices do not fully capture the predictive power of fundamentals included in public financial statements (Skogsvik & Skogsvik, 2010). The corresponding tests use ratio-based FA to predict future earnings changes or stock returns and form long-short hedge portfolios accordingly to profitably trade on the accounting signals.

Initial research on fundamental trading strategies produced ambiguous results. Ou & Penman (1989) first introduced a probabilistic prediction model using logit analysis across a large set of U.S. firms' accounting ratios to estimate the likelihood of one-year ahead earnings increases (composite measure 'Pr') and take stock positions thereafter. While two-year excess returns of 12.6% provided strong support for inefficient prices, subsequent studies were quick to contest the robustness of these results (Greig, 1992; Stober, 1992). Holthausen & Larcker (1992) find no evidence for the validity of Ou & Penman's (1989) model in subsequent periods. Instead, they report that an alternative trading strategy based on direct forecasts of stock returns yielded significant abnormal returns up to the year 1988. Conversely, Setiono & Strong's (1998) evidence on U.K. data lends support to Ou & Penman's (1989) model, but finds no evidence for Holthausen & Larcker's (1992) prediction strategy. In line with prior research, they conclude that "using the direct approach […] adds noise to the relation between financial statement information and stock returns" (Setiono & Strong, 1998, p. 655-656).

More recently, the limited effectiveness of applying complex statistical procedures to vast data samples led researchers to focus their efforts on subsets of stocks with the highest mispricing potential. 'Value stocks', i.e. stocks with high book-to-market (B/M) ratios, can be argued to be particularly rewarding for FA (Lakonishok, Shleifer & Vishny, 1994). Piotroski (2000) followed this idea by using a simple summary score ('F-SCORE') of well-established profitability and 'financial health' indicators and taking long (short) positions in high-B/M stocks with the highest (lowest) 'F-SCORE'. The reported one-year average return spread of more than 20% between high and low 'F-SCORE' stocks makes a strong case for this parsimonious strategy. However, Piotroski (2000) observes that the bulk of the gains accrues to firms of small-to-medium size, with low analyst following and share turnover. This raises doubts on the practicability of the strategy, given the implications of market frictions (see 2.1.2). Nevertheless, Piotroski's (2000) findings prompted several subsequent applications of 'contextual fundamental analysis', such as trading strategies on 'growth stocks' (Mohanram, 2005) as well as on extremely performing stocks (Beneish, Lee & Tarpley, 2001).

Tests of modeling mispricing

Modeling mispricing arises when stock prices fail to impound the full valuation implications of predicted fundamentals (Skogsvik & Skogsvik, 2010). To test this, a number of studies have employed fundamental value strategies that use the RIV model to calculate stocks' intrinsic values and trade on the difference to observed market prices.

As discussed earlier, the RIV model has been suggested to yield strong predictive power for cross-sectional stock returns. Both Frankel & Lee (1998) and Lee et al. (1999) show that an intrinsic-value-to-price ratio (V/P) outperforms standard prediction ratios such as the book-to-price ratio (B/P) and generates long-horizon abnormal returns. The results led the authors to conclude that the "price convergence to value is a much slower process than prior evidence suggests" (Frankel & Lee, 1998, p. 315). However, the reliance on analyst earnings forecasts gives reason to question the information content impounded in the intrinsic value measure, given the ample evidence of analysts' over-optimism and their use of non-accounting data (e.g. Frankel & Lee, 1998; Abarbanell & Bernard, 2000; Dechow et al., 1999).

2.3. THE SKOGSVIK & SKOGSVIK (2010) MODEL

Concluding from the previous section, FA-based tests of market efficiency have provided vast evidence of abnormal returns, indicating that market prices process information inefficiently. Nevertheless, Skogsvik (2008) and Skogsvik & Skogsvik (2010) identified critical shortcomings in previous empirical studies, which can be roughly summarized as follows:

1) The complexity of multivariate prediction models is often not justified by their accuracy.
2) There is a lack of studies examining forecasting and modeling mispricing simultaneously.
3) Trading strategies that test market efficiency must be implementable to be relevant.

To address these shortcomings, Skogsvik & Skogsvik (2010) developed an indicator variable strategy based on a ROE prediction model and applied it to Swedish manufacturing firms in the period 1983-2003. Our research interest in validating their model on the U.S. market rests on three pivotal findings: i) the investment strategies produced significant abnormal returns, ii) the results revealed intriguing time-series patterns of market efficiency, and iii) the successful model application substantiated the benefits of parsimonious as opposed to complex multivariate models. In the following, the key characteristics of the Skogsvik & Skogsvik (2010) model are briefly outlined, which forms the basis for our research design.

ROE prediction model

Based on Freeman et al.'s (1982) logit approach, the prediction model uses historical medium-term ROE as a univariate ratio to generate probabilities of future increases in medium-term ROE (Skogsvik, 2008). The high prediction accuracy of 74% is partly due to frequent model estimations, the selection of a homogenous sample of Swedish manufacturing firms as well as the mean reversion properties of ROE discussed earlier (see 2.2.2).

Indicator variable

The key innovation of Skogsvik & Skogsvik (2010) is the introduction of an indicator variable, which compares a firm's market price to its intrinsic value as justified by historical medium-term ROE. With this 'mechanical' approach, they circumvent the need for analyst earnings forecasts, and thus assess modeling mispricing on the grounds of purely accounting-based forecasts. This novelty is particularly appealing, as it provides information on the 'cheapness' of stocks in relation to their projected ROE performance (Penman, 1991).

Investment strategies

Given the joint analysis of forecasting and modeling mispricing, "investment positions are taken when the accounting-based predictions and the market expectations differ" (Skogsvik & Skogsvik, 2010, p. 388). For instance, long positions are taken when the indicator variable is negative (i.e. price below intrinsic value) and the probability of future medium-term ROE increases is at least 50%. This trading rule is expected to maximize abnormal returns, as long (short) positions are only taken in stocks that grant a sufficient discount on purchase (premium on sale). A base case strategy is applied to differentiate forecasting and modeling mispricing.

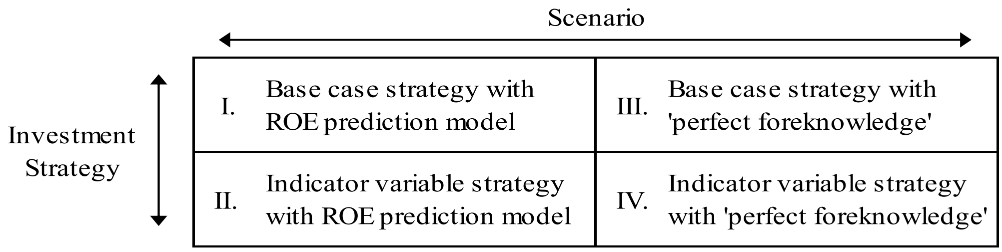
Using a market-adjusted return metric, the indicator variable strategy yielded excess returns of 42% for a three-year holding period, with contributions of both forecasting and modeling mispricing. However, Skogsvik & Skogsvik (2010) point towards several limitations. First, they find abnormal returns to accrue almost exclusively to the long positions and partly attribute this to a 'positive sentiment bias', which might undermine the out-of-sample validity. Second, the mispricing disappears by the mid-1990s, leading the authors to hypothesize that market learning contributed to an increase in market efficiency over time. The application of the Skogsvik & Skogsvik (2010) model in the U.S. market allows us to both draw conclusions regarding market efficiency over time and assess its key limitations.

3. RESEARCH DESIGN

To address the research questions regarding market efficiency and the performance of parsimonious FA-strategies in the U.S. market, the operationalization of the investment strategies as well as the return calculations are central parts of the methodology to be outlined in this section. As our research question is closely linked to the findings of Skogsvik & Skogsvik (2010) in the Swedish market, the research design deliberately aims to replicate core parts of the model in a suitable manner to allow for a comparative analysis of the results.

FIGURE 1

Matrix – Investment strategies and scenarios



In the spirit of the aforementioned study, two main trading strategies are evaluated – a base case and an indicator variable strategy. While the former aims to uncover mispricing in the market solely on the grounds of an accounting-based ROE prediction model, the latter extends this strategy by estimating an indicator variable for potential over- and underpricing in addition to the predicted ROE changes. Each of the two strategies is analyzed for two different scenarios as outlined in Figure 1. The first one is based on the forecasts made by our ROE prediction model, while the second one operates under the assumption that future ROE changes are always predicted correctly. In line with Ball & Brown's (1968) early study, this 'perfect foreknowledge' scenario serves as a benchmark for our results. By defining 18 investment points in time – divided into six subperiods and ranging from 1979 up to 2011 – our research design further allows for an analysis of the returns over time (see Table 1).

TABLE 1

Overview of estimation and investment periods

Subperiods	*Estimation periods*	*Investment periods*
I	1970 - 1975	1979 - 1981
II	1976 - 1981	1985 - 1987
III	1982 - 1987	1991 - 1993
IV	1988 - 1993	1997 - 1999
V	1994 - 1999	2003 - 2005
VI	2000 - 2005	2009 - 2011

Notes: Table 1 depicts the time periods underlying the estimation and investment periods. Estimation periods entail 6 years of calculated medium-term ROE changes based on 12 consecutive years of data, respectively (e.g. 1967-1978 in subperiod I), and are defined to be non-overlapping. Each investment period consists of 3 years with a 36-month holding period, respectively.

3.1. BASE CASE STRATEGY

As pointed out in section 2, ROE is considered the primary value driver in the RIV model and frequently used in determining future firm profitability. Thus, in the base case strategy, investment positions are taken solely on the basis of predicted future ROE changes. In line with our aim of a parsimonious investment strategy, the prediction of future ROE changes is based on historical ROE data only. Similar to Skogsvik & Skogsvik's (2010) original model, changes in ROE are evaluated on a medium-term basis of a three-year arithmetic average. As opposed to a one-year change, this procedure reduces the impact of transitory items with less relevance in the context of valuation. The historical medium-term ROE, $\overline{ROE}^h$, and future medium-term ROE, $\overline{ROE}^f$, for firm *i* at time *t* are operationalized as follows[2]:

$$\overline{ROE}_{i,t}^{h} = \frac{ROE_{i,t-2} + ROE_{i,t-1} + ROE_{i,t}}{3} \tag{2}$$

and

$$\overline{ROE}_{i,t}^{f} = \frac{ROE_{i,t+1} + ROE_{i,t+2} + ROE_{i,t+3}}{3} \tag{3}$$

where

$$ROE_{i,t} = \frac{NI_{i,t}}{BV_{i,t-1}}$$

with net income *($NI_{i,t}$)* and book value of owners' equity ($BV_{i,t}$) for firm *i* at time *t*.

3.1.1. ESTIMATION OF ROE PREDICTION MODEL

In line with previous research in the field, such as Freeman et al. (1982), Ou & Penman (1989) or Skogsvik (2008), a logit analysis is used to estimate the prediction model for the change in medium-term ROE. The latter is defined as follows:

$$\Delta(\overline{ROE}_{i,t}) = \overline{ROE}_{i,t}^{f} - \overline{ROE}_{i,t}^{h} \tag{4}$$

With $\overline{ROE}^h$ as the single independent variable of the logistic regression, the variable for $\Delta\,(\overline{ROE})$ as depicted in Eq. (4) is specified as a binary variable with realizations 1 and 0, indicating an increase or a decrease in $\overline{ROE}$ respectively. Although this procedure does not allow for an analysis of the relative magnitude of the $\overline{ROE}$ changes, it offers three benefits. First, it corresponds to the method used by our reference model in Skogsvik & Skogvik (2010), and therefore facilitates comparisons. Second, it is in line with the aim of a parsimonious investment strategy that determines the probability for an increase in $\overline{ROE}$. And third, it enables a direct comparison between the results of the trading strategy based on the ROE prediction model and the benchmark scenario of 'perfect foreknowledge'. A separate prediction model is

[2] To facilitate the flow of reading, the subscripts *i* and *t* are omitted in in-text notations whenever possible.

estimated for each of the six investment periods to account for time-specific characteristics such as changing levels of ROE. Pooled data over firms and time is used in estimating prediction models for the six subperiods in order to guarantee a sufficient amount of data for the logistic regressions. In contrast to Skogsvik & Skogsvik's (2010) approach, non-overlapping estimation periods are chosen in order to avoid potential problems of statistical overfitting and to make our results more robust. The prediction models are estimated for each period based on the changes in $\overline{ROE}$ in six years prior to the investment period. As $\overline{ROE}^f$ requires accounting data three years into the future, there is an apparent gap of four years between the last year of the estimation period and the first year of the investment period. For the six years included in each estimation period, $\overline{ROE}^h$ is compared to $\overline{ROE}^f$ and a binary variable is assigned to an increase or decrease of $\overline{ROE}$ in accordance with the specifications outlined above.

Based on the assigned binary variables, prediction models for an increase in $\overline{ROE}$ are estimated with the single independent variable $\overline{ROE}^h$. The results of the logit analysis in Table 2 depict negative and highly significant coefficients for $\overline{ROE}^h$. These findings are consistent with observations made by Skogsvik & Skogsvik (2010) and the underlying assumption of mean-reverting ROEs. The results of the model chi-squared tests (*p-value: 0.000*) shown in Table 2 further underline the statistical significance of the estimated models in terms of goodness-of-fit. The underlying sample data is further discussed in section 4.

TABLE 2

Estimated prediction models for the six subperiods

Estimation periods	Subperiod I ***1970-1975***	Subperiod II ***1976-1981***	Subperiod III ***1982-1987***	Subperiod IV ***1988-1993***	Subperiod V ***1994-1999***	Subperiod VI ***2000-2005***
Constant	1.482	0.954	0.741	0.605	0.721	0.581
(p-value)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
$\overline{ROE}^h$	-10.972	-8.871	-7.272	-8.814	-7.815	-5.172
(p-value)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
χ^2	302.77	420.02	307.37	454.94	324.58	209.44
(p-value)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
Increases in $\overline{ROE}$ (%)	57.34%	42.11%	49.00%	43.85%	42.66%	52.59%

Notes: Table 2 shows the intercepts, coefficients and χ^2 test results for the estimated prediction models (based on the six logistic regressions). The sample distribution is indicated by the percentage of increases in the medium-term ROE ($\overline{ROE}$).

3.1.2. INVESTMENT DECISION PROCESS

The coefficients of the logit analyses portrayed in Table 2 allow for the calculation of the probability of an increase in $\overline{ROE}$ at each investment point in time as the only needed additional input variable is the observable $\overline{ROE}^h$. Thus, $\overline{ROE}^h$ is calculated according to Eq. (2) for each year in the six investment periods (see Table 1). The derived $\overline{ROE}^h$ of a company is then entered into the following function to calculate the probability for an increase in $\overline{ROE}$:

$$\hat{p}\left(\Delta\left(\overline{ROE}_{i,t}\right) \geq 0\right) = \frac{1}{\left[1 + e^{-(\phi_1^p + \phi_2^p \times \overline{ROE}_{i,t}^h)}\right]} \tag{5}$$

where

ϕ_1^p= Constant derived from logit analysis for estimation subperiod *p*,

ϕ_2^p= Coefficient for $\overline{ROE}_{i,t}^h$ derived from logit analysis for estimation subperiod *p*.

Following Skogsvik & Skogsvik (2010), the probabilities derived from the prediction models are adjusted according to Skogsvik's (2005) calibration formula (see Eq. (6)) since the proportions of increases in $\overline{ROE}$ in the estimation period samples differ from the a-priori probability of 0.5. This adjustment aims at mitigating the risk of misallocations in the investment periods by counterbalancing the effects of estimation periods with extraordinary proportions of increases.

$$\hat{p}\left(\Delta\left(\overline{ROE}_{i,t}\right) \geq 0\right)^{\star} =$$

$$= \hat{p}\left(\Delta\left(\overline{ROE}_{i,t}\right) \geq 0\right) \cdot \left[\frac{\pi \cdot (1 - prop)}{prop \cdot (1 - \pi) + \hat{p}\left(\Delta\left(\overline{ROE}_{i,t}\right) \geq 0\right) \cdot (\pi - prop)}\right] \tag{6}$$

where

π		= a priori probability of increases in $\overline{ROE}$ (=0.5),
prop		= actual proportion of increases in $\overline{ROE}$ in estimation sample,
$\hat{p}\left(\Delta\left(\overline{ROE}_{i,t}\right) \geq 0\right)$		= model-based probability for increase in $\overline{ROE}$ for company *i*.

Investment logic

In the base case strategy, investment positions are taken three months after the fiscal year-end[3] based on these adjusted probabilities alone. According to the principle of mean reversion in ROEs, a company that has been underperforming in terms of historical medium-term ROE, is more likely to see an increase in future medium-term ROE and vice versa. Given the assumed relationship between stock returns and ROE performance of a company, this strategy invests in previous underperformers and short-sells historical outperformers. In line with Skogsvik & Skogsvik (2010), $\hat{p}(\Delta(\overline{ROE}) \geq 0)^{\star} = 0.5$ is used as the cut-off value for the investment strategy. Consequently, stocks are assigned to investment positions as follows:

- $\hat{p}\left(\Delta\left(\overline{ROE}_{i,0}\right) \geq 0\right)^{\star} > 0.5$: Long portfolio
- $\hat{p}\left(\Delta\left(\overline{ROE}_{i,0}\right) \geq 0\right)^{\star} < 0.5$: Short portfolio

[3] At this point in time firm accounting data of the previous period is expected to be publically available.

3.2. INDICATOR VARIABLE STRATEGY

The ROE prediction model outlined in the previous section is not only relevant for the base case strategy, but forms the foundation for the indicator variable strategy. While the ROE prediction model investigates the issue of forecasting mispricing, the indicator variable sets out to reveal the implicit market expectation of future ROE changes. This section starts out with a brief conceptual introduction to the indicator variable, followed by the underlying RIV model specification, and concludes with the investment decision logic.

Model notations

$BV_{i,t}$	=	ex-dividend book value of owners' equity for firm *i* at time *t*
$Div_{i,t}$	=	common dividends for firm *i* at time *t*
$DS_{i,t}$	=	dividend payout share for firm *i* at time *t*
$[E(r_M) - r_f]$	=	market risk premium
$IND_{i,t}$	=	indicator variable for firm *i* at time *t*
$P_{i,t}$	=	market value of owners' equity for firm *i* at time *t*
$q(BG)_{i,t}$	=	business goodwill of owners' equity for firm *i* at time *t*
$q(CMB)_i$	=	relative cost matching bias of owners' equity for firm *i*
$q(TOT)_{i,t}$	=	relative valuation bias of owners' equity for firm *i* at time *t*
$ROE_{i,t}$	=	return on owners' equity for firm *i* at time *t*
$\overline{ROE}_{i,t}^{h}$	=	historical average medium-term return on equity for firm *i* at time *t*
$r_{E,i}$	=	required rate of return on owners' equity for firm *i*
r_f	=	risk-free rate
T	=	horizon point in time
$V_{i,t}^{h}$	=	'historically justified' intrinsic equity value for firm *i* at time *t*
β_i	=	market risk beta of firm *i*

The indicator variable is defined as the difference between the market value of owners' equity and a 'historically justified' intrinsic value estimate at the valuation date *t* = 0:

$$IND_{i,0} = P_{i,0} - V_{i,0}^{h} \tag{7}$$

Interpretation of $\boldsymbol{IND_{i,0}}$

- If $IND_{i,0} > 0$: market expects higher future ROE than justified by historical realizations, i.e. market-based probability of future medium-term ROE increase > 0.5.
- If $IND_{i,0} = 0$: market expectation of future ROE is on par with historical realizations, i.e. market-based probability of future medium-term ROE increase = 0.5.
- If $IND_{i,0} < 0$: market expects lower future ROE than justified by historical realizations, i.e. market-based probability of future medium-term ROE increase < 0.5.

While $P_{i,0}$ is simply the observable market capitalization of stock i at each investment date (i.e. three months after the fiscal year-end), $V_{i,0}^{h}$ is estimated using a parsimonious RIV model application. In the following, the suggested RIV model design is presented in more detail.

3.2.1. RIV MODEL SPECIFICATION

To estimate V^{h}, critical model design choices need to be made, of which i) the explicit forecast horizon and ii) the terminal value estimate warrant particular consideration (Penman & Sougiannis, 1998; Courteau, Kao & Richardson, 2001). In line with our examination of a parsimonious trading strategy, a sufficient level of parsimony is pivotal for our RIV model specification. Previous literature has produced a large account of the merits and limitations of parsimonious valuation models, with simple RIV applications being widely acknowledged for their superior relative performance (e.g. Francis et al., 2000; Jorgensen et al., 2011; Anesten, Möller & Skogsvik, 2015). Employing these findings, the main RIV model design choices are briefly motivated hereafter.

Explicit forecast horizon

Skogsvik (2002) notes that finite horizon RIV models require careful estimation of the time period preceding the 'steady state', at which earnings growth is often assumed to be constant. Several researchers have approximated the explicit forecast period with estimates of the durations of ROE mean reversion, ranging from 5 to 15 years (Penman, 1991). Nevertheless, the RIV model has been found to be relatively insensitive to extensions of the forecast horizon (Bernard, 1995; Jorgensen et al., 2011), owing to the high proportion of value that is anticipated in current book values of owners' equity (Penman, 2005, 2012). Thus, even one-year horizon RIV applications have yielded comparatively strong valuation results (Norrman & Rahmn, 2016) and can be deemed appropriate for the purpose of large-scale equity valuations. In contrast to Skogsvik & Skogsvik's (2010) three-year horizon model, a simple one-year horizon RIV model is therefore suggested for this thesis.

Terminal value estimate

The terminal value equals the present value of infinite residual earnings at the horizon point in time and is identical to the firm's total goodwill at *t*=T. To circumvent the empirically problematic extrapolation of expected future residual earnings at truncation, Skogsvik (1998) offers a useful restatement of the terminal value based on a 'goodwill to book ratio':

$$\frac{(V_{i,T} - BV_{i,T})}{(1 + r_{E,i})^{T}} = \frac{BV_{i,T} \cdot q(TOT)_{i,T}}{(1 + r_{E,i})^{T}} \tag{8}$$

According to Skogsvik (1998), $q(TOT)$ consists of both business goodwill $q(BG)$ and the cost matching bias $q(CMB)$ proceeding from conservative accounting rules. While the former is transitory and presumably disappears with long-run competitive forces, $q(CMB)$ is expected to be fairly persistent and similar in size for companies of the same industry. Skogsvik & Skogsvik (2010) utilize this empirical property by incorporating industry estimates of $q(CMB)$ that were reported by Runsten (1998) for the Swedish market. However, two caveats undermine the utility of $q(CMB)$ for our model application: First, given our single-period horizon model, the assumption of a 'steady state' or even competitive equilibrium is unreasonable. Thus, $q(TOT)$ is likely to contain a substantial amount of business goodwill, which disqualifies $q(CMB)$ as a measure of the total valuation bias. Second, no empirical estimates of industry $q(CMB)$ have been reported for the U.S. market to date. Considering the potentially significant impact of differing accounting regimes on the cost matching bias, an application of Runsten's (1998) Swedish estimates for U.S. companies is considered inappropriate.

A fruitful alternative is presented by Norrman & Rahmn (2016). The authors empirically tested the valuation accuracy of the RIV model based on reverse-engineered $q(TOT)$ values for a similar one-year horizon model. Drawing from their 'Method I' model, an implicit firm-specific estimate of 'last year's q(TOT)' is obtained from historical market values using the RIV model for the period preceding the valuation date. A necessary assumption is that the previous year's $q(TOT)$ is a valid representation of the current year's estimate. Norrman & Rahmn (2016) find this simplified method to yield stable or even enhanced valuation accuracies as compared to using the three-year historical average of $q(TOT)$. Thus, we adopt their parsimonious approach and thoroughly scrutinize the resultant valuation accuracy.

Final model specification

V^h is hence calculated based on the following RIV model specification:

$$V_{i,0}^h = BV_{i,0} + \sum_{t=1}^{T} \frac{\left(\overline{ROE}_{i,t}^h - r_{E,i}\right) \cdot BV_{i,t-1}}{(1 + r_{E,i})^t} + \frac{BV_{i,T} \cdot q(TOT)_{i,T}}{(1 + r_{E,i})^T} \tag{9}$$

where T=1.

In line with Skogsvik & Skogsvik (2010), $\overline{ROE}_{i,t}^h$ is selected as the earnings measure for the computation of 'historically justified' intrinsic values to avoid that $V_{i,0}^h$ estimates are contaminated by non-fundamental information (e.g. analyst earnings forecasts). To estimate $q(TOT)_{i,T}$, the specification in Eq. (9) is applied analogously to the period preceding each investment date, with the difference that $V_{i,0}^h$ is replaced by the historical market price of each firm at the valuation date, $P_{i,-1}$.[4]

[4] $P_{i,-1}$ is equal to the product of the number of common shares outstanding and the closing share price at valuation. To ensure ex-dividend prices, the valuation date coincides with the dividend payment date. If the latter is not reported, the valuation date is set to the first trading day of June of the respective year, at which prices can be assumed to be ex-dividend.

Solving the adjusted equation (9) for $q(TOT)_{i,T}$ yields Eq. (10), which describes the applied reverse-engineering procedure:

$$q(TOT)_{i,T} =$$

$$q(TOT)_{i,T-1} = \frac{\left[P_{i,-1} - BV_{i,-1} - \left(\frac{\left(\overline{ROE}_i^h - r_{E,i}\right) \cdot BV_{i,-1}}{\left(1 + r_{E,i}\right)}\right)\right] \cdot \left(1 + r_{E,i}\right)}{BV_{i,0}} \tag{10}$$

For detailed explanations regarding the operationalization and the underlying assumptions of the remaining RIV model input parameters, the interested reader is referred to Appendix A.

3.2.2. INVESTMENT DECISION PROCESS

The indicator variable strategy refines the base case strategy by introducing a contrarian trading logic. Thus, investment positions are only taken, if the accounting-based probabilistic ROE prediction differs from the market-based probability of future ROE increases (Skogsvik & Skogsvik, 2010). While the former is measured by the ROE prediction model, the latter is captured by the firm-specific indicator variable as defined in section 3.2.

Investment logic

In line with the base case strategy, investment portfolios are formed at the end of the third month following the fiscal year-end date. The long and short portfolios are held for 36 consecutive months, and the hedge position is assessed ex-post (Skogsvik & Skogsvik, 2010).

Long investment position:

Stocks are allocated to the long portfolio, if either of the following holds true:

- $IND_{i,0} < 0$ and $\hat{p}\left(\Delta\left(\overline{ROE}_{i,0}\right) \geq 0\right)^{\star} \geq 0.5$
- $IND_{i,0} = 0$ and $\hat{p}\left(\Delta\left(\overline{ROE}_{i,0}\right) \geq 0\right)^{\star} > 0.5$

Short investment position:

Stocks are allocated to the short portfolio, if either of the following holds true:

- $IND_{i,0} > 0$ and $\hat{p}\left(\Delta\left(\overline{ROE}_{i,0}\right) \geq 0\right)^{\star} \leq 0.5$
- $IND_{i,0} = 0$ and $\hat{p}\left(\Delta\left(\overline{ROE}_{i,0}\right) \geq 0\right)^{\star} < 0.5$

To account for potential measurement errors related to the estimation of the indicator variable, the following specification is made in accordance with Skogsvik & Skogsvik (2010):

$$IND_{i,0} = 0, if\ \frac{IND_{i,0}}{BV_{i,0}} \in [-0.1, 0.1]$$

3.3. EXCESS RETURN CALCULATION

To distinguish between 'normal' rates of return and predictable abnormal returns, an examination of potential risk factors is essential. The CAPM is chosen as the standard return metric to allow for comparisons with Skogsvik & Skogsvik (2010). Given the deficiencies of the market risk proxy[5], we further benchmark the returns to the Fama & French (1992, 1993) 'three-factor model' estimates. In doing so, we extend the sensitivity tests applied by Skogsvik & Skogsvik (2010) with a robust risk-adjusted metric.

3.3.1. CAPM EXCESS RETURNS

We adopt the monthly return regression procedure proposed by Greig (1992) to arrive at estimates of 'Jensen's Alpha' as the abnormal return measure (Jensen, 1968). In a first step, the monthly returns to the short and long portfolios are calculated as the value-weighted average of the corresponding returns on the individual stocks.[6] In addition, the monthly hedge returns are computed as the difference between the average long and short portfolio returns:

$$\bar{r}_{L,x} = \frac{1}{\sum_{i=1}^{n_L} S_{i,x-1}} \cdot \sum_{i=1}^{n_L} r_{i,x} \cdot S_{i,x-1} \tag{11}$$

$$\bar{r}_{S,x} = \frac{1}{\sum_{i=1}^{n_S} S_{i,x-1}} \cdot \sum_{i=1}^{n_S} r_{i,x} \cdot S_{i,x-1} \tag{12}$$

$$\bar{r}_{H,x} = \bar{r}_{L,x} - \bar{r}_{S,x} \tag{13}$$

where

- $\bar{r}_{L,x}$ = average weighted return on the long position for month *x*,
- $\bar{r}_{S,x}$ = average weighted return on the short position for month *x*,
- $\bar{r}_{H,x}$ = average weighted return on the hedge position for month *x*,
- $S_{i,x}$ = closing price of stock i at the end of month *x*,
- $r_{i,x}$ = return on stock i for month *x*,
- $n_{(.)}$ = number of stocks in the long (L) or short (S) position.

Subsequently, the monthly average long and short returns are reduced by the monthly risk-free rate to yield excess returns. Finally, the monthly long and short excess returns as well as the monthly hedge returns are regressed over the corresponding market excess returns[7]:

[5] See 2.1.2 for a summary of deficiencies with respect to return metrics.
[6] In line with Skogsvik & Skogsvik (2010), receipts from companies delisted during the investment period are reinvested in the market index and the returns are included in subsequent value-weighted portfolio returns.
[7] The market index is the S&P 500 Composite Index. The risk-free rate is the one-month U.S. treasury bill rate.

$$\bar{r}_{L,x} - r_{f,x} = \alpha_L + \beta_L^M \cdot (r_{M,x} - r_{f,x}) + \varepsilon_{L,x} \tag{14}$$

$$\bar{r}_{S,x} - r_{f,x} = \alpha_S + \beta_S^M \cdot (r_{M,x} - r_{f,x}) + \varepsilon_{S,x} \tag{15}$$

$$\bar{r}_{H,x} = \alpha_H + \beta_H^M \cdot (r_{M,x} - r_{f,x}) + \varepsilon_{H,x} \tag{16}$$

where

$r_{f,x}$ = risk-free rate for month *x*,
$r_{M,x}$ = market index return for month *x*,
$\alpha_{(.)}$ = abnormal return to the long (L), short (S) or hedge (H) position,
$\beta_{(.)}^M$ = market risk beta of the long (L), short (S) or hedge (H) position,
$\varepsilon_{(.),x}$ = error term of the long (L), short (S) or hedge (H) position for month x.

3.3.2. THREE-FACTOR EXCESS RETURNS

The procedure applied is analogous to 3.3.1. However, instead of estimating abnormal returns through linear regressions over a single explanatory variable $(r_{M,x} - r_{f,x})$, multiple regressions over market, size and book-to-market premia are used for 'three-factor' returns [8]:

$$\bar{r}_{L,x} - r_{f,x} = \alpha_L + \beta_L^M \cdot (r_{M,x} - r_{f,x}) + \beta_L^{SMB} \cdot SMB_x + \beta_L^{HML} \cdot HML_x + \varepsilon_{L,x} \tag{17}$$

$$\bar{r}_{S,x} - r_{f,x} = \alpha_S + \beta_S^M \cdot (r_{M,x} - r_{f,x}) + \beta_S^{SMB} \cdot SMB_x + \beta_S^{HML} \cdot HML_x + \varepsilon_{S,x} \tag{18}$$

$$\bar{r}_{H,x} = \alpha_H + \beta_H^M \cdot (r_{M,x} - r_{f,x}) + \beta_H^{SMB} \cdot SMB_x + \beta_H^{HML} \cdot HML_x + \varepsilon_{H,x} \tag{19}$$

where

SMB_x = size (Small Minus Big) mimicking portfolio return for month *x*,
HML_x = book-to-market (High Minus Low) mimicking portfolio return for month *x*,
$\beta_{(.)}^{SMB}$ = size risk beta of the long (L), short (S) or hedge (H) position,
$\beta_{(.)}^{HML}$ = book-to-market risk beta of the long (L), short (S) or hedge (H) position.

[8] Monthly size (SMB) and book-to-market (HML) portfolio returns are retrieved from Kenneth R. French's (2016) continuously updated online database. The factors are constructed based on six value-weighted mimicking portfolios formed on size (market value of equity) and book-to-market values. SMB corresponds to the difference between the average monthly returns on the three small portfolios and the three large portfolios. HML corresponds to the average monthly returns on the two high book-to-market portfolios and the two low book-to-market portfolios. For detailed explanations, see Fama & French (1993). Database link: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html.

4. DATA SAMPLE

In the following, the time period and industry focus of the investigated sample are outlined, before more specific selection criteria for the periods and scenarios are introduced and motivated. This section concludes with a description of the final sample for each period.

4.1. TIME PERIOD AND INDUSTRY FOCUS

To analyze the evolution of market efficiency in the U.S. setting and the potential to generate abnormal returns over time, this study investigates the time period 1979-2014 for U.S. manufacturing companies. The year 2014 is chosen as the final year, since accounting and stock price information for this year is assumed to be fully available. Following the logic of an investment strategy with 36-months holding periods, the last investments take place in 2011. The first year of investment in 1979 is decided on by counting backwards from the year 2014, while accounting for non-overlapping estimation periods and a total of six investment periods, consisting of three investment years each (see Table 1).

The restriction to the manufacturing industry is motivated by three decisive factors: First, it is in line with the criterion outlined by Skogsvik & Skogsvik (2010). Second, the methodology of the ROE prediction model specified in the previous section depends on a certain degree of homogeneity within the sample with regards to levels of and consistency in ROE. Including multiple industries with a wide range of business models and potentially industry-specific accounting techniques could weaken the predictive ability of the proposed model. Lastly, given the vast amount of data for the entire U.S. market, the focus on one substantial sector allows for an investigation over a long time period with conventional data-processing tools.

All company and stock data was retrieved from the Compustat and CRSP databases. The SIC-code industry classification by French (2016) was used to follow a consistent sample selection process throughout the estimation and investment periods. Within his specification of '10 Industry Portfolios', the manufacturing category was chosen, which includes machinery, trucks, planes, chemicals, office furniture, paper and commercial printing firms. The use of the 957 SIC codes included in this industry classification throughout the entire time period under investigation guarantees consistency, while reducing the risk of 'data snooping' biases.

4.2. SELECTION CRITERIA

Apart from the industry classification outlined above, several other selection criteria are applied to the sample. Before pointing out specific criteria for the estimation and investment periods as well as the 'perfect foreknowledge' scenario, there are two universal criteria:

1. Firm-year observations with $\overline{ROE}^h$ > +100% or $\overline{ROE}^h$ < -100% are excluded.
2. Firm-year observations with reported negative equity are excluded.

The first criterion is helpful in eliminating extreme values from our sample that would impair our regression results and thereby our prediction model. In our estimation samples, the firm-year observations affected by this elimination account for only 0.2% to 1.2%. By investigating this low figure of eliminations, we found data recording errors to be the primary cause.

The reason for excluding firm-year observations with negative equity is twofold: on the one hand, negative equity could signal data errors or financial distress. And on a more practical note, the inclusion of negative equity observations would affect the automated calculation of ROE, as negative net income and negative equity would mistakenly result in a positive ROE.

In addition to the aforementioned criteria, a necessary condition for the estimation periods is that required firm data is available for the 12 consecutive years needed to calculate the six historical and six future medium-term ROEs. To take subperiod I as an example, data for the years 1967 to 1978 needs to be available to calculate the changes in medium-term ROEs. Similarly, data needs to be available for four consecutive years for the investment periods in order to calculate the three historical medium-term ROEs necessary to make investment decisions. Another condition for the investment samples is the coincidence of the fiscal year-end with the calendar year-end. This criterion is essential to ensure return comparability across all firms, as investment positions are assumed to be taken three months after the fiscal year-end. Finally, companies are excluded from the sample in a given investment year if the stock price is not available or zero in the database to guarantee correct return calculations.

The benchmark 'perfect foreknowledge' scenario does not rely on an estimation model and therefore only has one additional criterion regarding the investment periods. For this scenario, seven consecutive years are needed, starting with the year prior to the first year of each investment period, as historical and future medium-term ROEs need to be calculated in order to make the investment decision. Compared to the ROE prediction model scenario, the sample size is therefore reduced by the number of firms delisted over the investment period.

4.3. FINAL SAMPLE

Estimation periods

Appendix B gives an overview of the sample selection procedure for the six estimation periods (see Table 9). The final number of firms used in estimating the prediction model ranged from 324 in the last subperiod (1997-2008) to 512 in the second subperiod (1973-1984).

Investment periods

A similar sample selection procedure as for the estimation periods was completed for the investment periods and the results are summarized in Appendix B (see Table 10). In the base case strategy, between 275 and 465 firms were included in each year's investment portfolio, while a smaller number of firms remained for the indicator variable strategy due to the additional investment criteria applied. In both strategies, the sample size diminishes over time, which could point towards a growing consolidation of the manufacturing industry or a shift in the classification from traditional manufacturing to more service-oriented business models.

5. RESULTS AND PRELIMINARY ANALYSIS

Prior to the presentation of the obtained investment returns, this section discloses important findings from the conducted analyses that are necessary in order to implement the proposed investment strategies. Therefore, in a first step, the input variables of the ROE prediction model as well as the indicator variable are assessed and the general model fit is evaluated with the help of prediction and valuation accuracy measures. Thereafter, descriptive statistics regarding the investment periods are shown before the monthly excess returns of the base case and indicator variable strategies are presented for both scenarios.

5.1. BASE CASE – MODEL FIT

5.1.1. INPUT VARIABLES AND BACKGROUND STATISTICS

The thorough assessment of the two input variables for the ROE prediction model estimation – $\overline{ROE}^h$ and $\overline{ROE}^f$ – aims at verifying the validity of our sample set. As can be seen in Table 3, the level of $\overline{ROE}^h$ and $\overline{ROE}^f$ has fluctuated significantly over time. While the lowest mean $\overline{ROE}^h$ of 8.6% in the 2000-2005 period is trailing the highest figure of 15.0% in the 1976-1981 period by 6.4%, a similar discrepancy can be observed in the mean $\overline{ROE}^f$ (i.e. a difference of 7.3% between the high in subperiod II and the low in subperiod IV).

TABLE 3

Descriptive statistics of input variables – ROE prediction model estimation

Subperiod	*Estimation period*	*Number of firms*	$\overline{ROE}^h$		$\overline{ROE}^f$		$(\overline{ROE}^f - \overline{ROE}^h)$	
			Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*
I	1970-1975	386	0.103	0.106	0.125	0.121	0.022	0.016
II	1976-1981	512	0.150	0.149	0.136	0.138	-0.014	-0.012
III	1982-1987	391	0.106	0.120	0.110	0.121	0.004	0.001
IV	1988-1993	390	0.097	0.110	0.063	0.102	-0.034	-0.008
V	1994-1999	343	0.136	0.138	0.127	0.125	-0.009	-0.012
VI	2000-2005	324	0.086	0.098	0.133	0.110	0.048	0.013
I-VI	All years		0.115	0.122	0.116	0.121	0.001	-0.001

Notes: Table 3 shows arithmetic mean and median values of the input variables used for the six estimated prediction models. $\overline{ROE}^f - \overline{ROE}^h$ denotes the spread of historical ($\overline{ROE}^h$) and future ($\overline{ROE}^f$) medium-term ROE as defined in Eq. (4).

In addition to the analysis of the absolute levels of the input variables, the spread between and $\overline{ROE}^f$ and $\overline{ROE}^h$ is assessed. Although the binary variables of the prediction model are assigned on the basis of the sign of change rather than its magnitude, the values reported in Table 3 regarding the spread provide interesting insights into time-specific patterns. The signs of the spread alternate almost perfectly between positive and negative and thus hint at cyclical movements. Interestingly, despite the aforementioned inter-period discrepancies in the ROE levels as well as the varying magnitudes of the spreads over the years, mean and median $\overline{ROE}^h$ and $\overline{ROE}^f$ are almost identical when calculated over the 13,763 firm-year observations throughout all estimation periods. This not only corroborates the finding of cyclicality, but could furthermore be interpreted as evidence of ROE following a mean-reversion process. However, it has to be noted that problems of statistical overfitting could be a contributing factor to this finding, as firm-year observations are overlapping in the calculation of medium-term ROEs of subsequent estimation periods.

5.1.2. PREDICTION ACCURACY

The measure of prediction accuracy reveals the ability of the ROE prediction model to correctly forecast the direction of changes in future medium-term ROE. Defined in line with Skogsvik & Skogsvik (2010), this measure not only provides a goodness-of-fit indication over time, but also allows for a comparison between the performance of our model and the one of the original study in the Swedish market. It is calculated by comparing the correctly predicted binary variables of change to the actual changes in future medium-term ROE. Table 4 provides an overview of the measure and its development over time.

TABLE 4

Prediction accuracy for the six investment subperiods

Investment periods	Sub-period I ***1979-81***	Sub-period II ***1985-87***	Sub-period III ***1991-93***	Sub-period IV ***1997-99***	Sub-period V ***2003-05***	Sub-period VI ***2009-11***	All years
Number of observations	1187	846	907	901	822	724	5387
Increases	391	446	398	303	532	373	2443
Decreases	796	400	509	598	290	351	2944
Correctly predicted (%)	68.6%	62.5%	69.3%	68.5%	68.9%	72.1%	68.3%
Increases	41.4%	78.5%	75.6%	47.2%	82.1%	65.7%	67.0%
Decreases	81.9%	44.8%	64.4%	79.3%	44.5%	78.9%	69.3%

Notes: Table 4 presents the prediction accuracy and corresponding number of observations for medium-term ROE ($\overline{ROE}$) increases, decreases and overall changes for the six investment periods. The prediction accuracy is measured as the respective proportion of predictions of $\overline{ROE}$ changes that match the observed ex-post realizations.

68.3% of $\overline{ROE}$ changes were correctly predicted over the entire period 1979-2011, with the negative outlier of 62.5% in subperiod II and the positive extreme of 72.1% in subperiod VI and otherwise relatively stable figures of approximately 69%. To put the figures into context, a comparison to Skogsvik & Skogsvik's (2010) findings seems appropriate: despite the fact that the prediction accuracy in one of their investigated periods reaches an astonishing 81.2%, the prediction accuracy over all periods amounts to 73.7%, thereby indicating a slightly better, but overall quite similar performance compared to our model.

Interestingly, similar to the findings of 5.1.1. regarding the observed cyclicality in ROE, the inter-period prediction accuracy for increases and decreases is subject to strong fluctuations, yet relatively stable when computed over the whole period. Despite the heavily skewed performance towards decreases in subperiods I and IV and towards increases in subperiods II and V, the overall prediction accuracy for increases and decreases differs within a range of only 2.3%. This even overall prediction performance for increases and decreases is interesting, as it contrasts with Skogsvik & Skogsvik's (2010) finding of a prediction accuracy of 84.7% for decreases and only 61.1% for increases, with a consistently skewed performance towards decreases. While it is impossible to further investigate these differences due to a lack of information regarding the underlying data in Skogsvik & Skogsvik's (2010) study, the patterns observed in our sample set leave room for further testing in section 6.

5.2. INDICATOR VARIABLE – MODEL FIT

5.2.1. INPUT VARIABLES AND BACKGROUND STATISTICS

To assess the 'quality' of the intrinsic equity value estimate – and thus ensure the robustness of the indicator variable – the key RIV model input parameters have been thoroughly examined. Table 5 presents the mean and median values of $q(TOT)$, $\overline{ROE}^h$, r_E and DS for each investment year as well as the corresponding number of firm-year observations.

Overall, the mean and median $q(TOT)$ are consistent with estimates reported by Norrman & Rahmn (2016) for the Scandinavian markets. Further, the median of 0.50 is close to the 0.49 $q(CMB)$ estimate used by Skogsvik & Skogsvik (2010) based on Runsten's (1998) findings. Examining the time-series properties, mean and median $q(TOT)$ exhibit a significant upward trend prior to the year 2000, ranging from a median low of -0.11 in 1979 to a median peak of 1.25 in 1999. Subsequently, $q(TOT)$ varied substantially, reaching another high in 2005. The observed volatility stands in stark contrast to the empirical persistence of the cost-matching bias (Skogsvik, 1998). Thus, fluctuations seem to be largely driven by extreme market expectations of future business goodwill (badwill) proceeding from the short-horizon RIV model. Accordingly, the low median $q(TOT)$ values of the initial periods may be considered leading indicators of the imminent U.S. recession in the early 1980s. Conversely, the peak $q(TOT)$

levels of the late 1990s and mid-2000s most likely reflect the speculative goodwill anticipation in the run-up to the 'Dot-com crash' (1999-2001) and the global financial crisis (2007-2008).

TABLE 5

Descriptive statistics of input variables – Intrinsic value estimation

Year	*Number of firms*	$q(TOT)$		$\overline{ROE}^h$		r_E		DS	
		Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*	*Mean*	*Median*
1979	296	0.015	-0.105	0.139	0.145	0.144	0.142	0.041	0.042
1980	300	0.152	-0.020	0.157	0.156	0.168	0.165	0.048	0.046
1981	264	0.207	-0.072	0.158	0.164	0.192	0.190	0.060	0.042
1985	305	0.594	0.303	0.090	0.101	0.167	0.163	0.037	0.034
1986	313	0.666	0.402	0.081	0.113	0.141	0.138	0.039	0.035
1987	269	1.028	0.580	0.074	0.107	0.124	0.123	0.036	0.033
1991	249	1.339	0.470	0.088	0.117	0.130	0.132	0.046	0.032
1992	264	1.000	0.344	0.068	0.097	0.118	0.120	0.035	0.025
1993	277	1.655	0.624	0.045	0.060	0.111	0.111	0.037	0.021
1997	280	1.783	1.171	0.149	0.150	0.104	0.106	0.038	0.023
1998	298	1.649	1.094	0.141	0.155	0.093	0.097	0.031	0.016
1999	300	1.823	1.253	0.132	0.145	0.084	0.082	0.029	0.009
2003	281	1.375	0.650	0.048	0.064	0.081	0.079	0.030	0.000
2004	275	1.365	0.439	0.024	0.046	0.084	0.079	0.033	0.000
2005	276	2.432	1.150	0.062	0.070	0.090	0.085	0.030	0.003
2009	247	1.589	1.138	0.125	0.145	0.096	0.089	0.032	0.016
2010	246	0.750	0.222	0.087	0.103	0.115	0.117	0.037	0.010
2011	244	1.202	0.848	0.078	0.084	0.111	0.107	0.030	0.011
All years		1.137	0.502	0.098	0.115	0.120	0.120	0.037	0.026

Notes: Table 5 exhibits yearly arithmetic means and medians of the primary input variables used in the intrinsic value calculation according to Eq. (9). $q(TOT)$ denotes the total valuation bias, r_E denotes the cost of equity capital, DS denotes the dividend payout ratio and $\overline{ROE}^h$ represents historical medium-term ROE.

In terms of $\overline{ROE}^h$, the mean and median values are roughly in line with the 10.05% U.S. industry average reported by Damodaran (2016) for our sample industries. Despite being smoothed over three historical years, $\overline{ROE}^h$ shows significant time variation and seems to closely track the business cycle movements with an inherent lag (NBER, 2010). Surprisingly, $\overline{ROE}^h$ is lower than r_E for most of the periods, with the latter averaging 12.00% over the period 1979-2011. Taken at face value, this would suggest that, on average, the sample firms failed to earn their cost of equity and destroyed rather than created shareholder value. Correspondingly, the median residual income is slightly negative. Further analyses revealed that – up to the early 1990s – the high levels of r_E were predominantly driven by inflated risk-free rates.[9] In contrast,

[9] Approximated by the U.S. 10-year treasury bond rates.

a steep post-crisis increase in security price volatilities[10] resulted in the relatively high r_E values in 2010 and 2011, despite historically low risk-free rates. Given that residual income on average accounts for less than 2% of the total intrinsic value estimate[11], the aforementioned caveats are deemed to be acceptable within the scope of our modeling.

Lastly, the mean DS indicates that dividend policies have been relatively stable, which corroborates previous research (Lintner, 1956). However, the median reveals a marked downward trend over the sample period, culminating in a payout halt in the post-'Dot-com' years. Excluding the crisis years, the lower average DS is mostly attributable to surging book values of owners' equity, while aggregate dividend levels have grown somewhat less rapidly.

5.2.2. VALUATION ACCURACY

The purpose of assessing the valuation performance of our RIV model specification in Eq. (9) is twofold: First and foremost, an empirically sound estimate of V^h is a prerequisite to render the indicator variable strategy effective. Second, the valuation accuracy of the parsimonious RIV model is in itself a matter of utmost interest to academics and practitioners alike. Our analysis thus intends to validate recent research efforts investigating the relative RIV model performance and the importance of complexity adjustments (e.g. Anesten et al., 2015). The performance of the RIV model specification is assessed by comparing the obtained intrinsic value estimate V^h to the prevailing market price P at the investment date. Additionally, four valuation accuracy measures – commonly used in previous literature – are examined.

The four valuation accuracy measures are briefly defined in the following. For all of these indicators, values closer to zero (ceteris paribus) indicate a higher valuation accuracy.

i) The mean absolute prediction error (MAPE) serves as the primary indicator of accuracy (e.g. Francis et al., 2000) and is calculated as the arithmetic sample average of the absolute difference between V^h and P, relative to P.
ii) The signed prediction error (SPE) distinguishes between under- and overvaluations and is calculated as the difference between V^h and P, relative to P.
iii) The interquartile range (IQRPE) is a measure of the spread between the third and the first quartile of the SPE (Liu, Nissim & Thomas, 2002).
iv) The 15% APE is equal to the proportion of all V^h estimates that exhibit an absolute prediction error of more than 15% (Kim & Ritter, 1999).

Table 6 presents the key results of our examination. Overall, the precision and spread measures exhibit a relatively strong valuation performance that broadly confirms previous empirical

[10] Measured as median market betas of 1.57 in 2010 and 1.50 in 2011.
[11] On average, BV_0 accounts for ca. 50% and the discounted terminal value accounts for ca. 48% of V_0^h.

findings (Jorgensen et al., 2011; Courteau et al., 2001; Anesten et al., 2015). However, the time series shows a significant deterioration across the performance indicators in the late-1990s.

Starting with MAPE as the principal accuracy measure, the mean of 0.43 is somewhat higher than the values reported by Norrman & Rahmn (2016), whose RIV model specification is consistent with ours. Prior to 1999, our model performs on par or outperforms their results, with MAPE values below 0.30 in the 1980s. Compared to more sophisticated RIV model applications on similar U.S. data, the obtained MAPE falls within the range estimated for various long-horizon models (Jorgensen et al., 2011). Hence, our findings support previous notions of the limited marginal utility of RIV complexity adjustments (Anesten et al., 2015).

TABLE 6

Valuation accuracy indicators for intrinsic value estimates

Year	*Number of firms*	*MAPE*	*Median V_0/P_0*	*Mean SPE*	*Median SPE*	*Std dev SPE*	*15% APE*	*IQRPE*
1979	296	0.199	0.955	-0.030	-0.045	0.277	0.493	0.277
1980	300	0.291	1.183	0.186	0.183	0.335	0.703	0.365
1981	264	0.271	0.801	-0.176	-0.199	0.404	0.670	0.253
1985	305	0.258	0.947	0.016	-0.053	1.004	0.551	0.281
1986	313	0.275	0.832	-0.129	-0.168	0.381	0.649	0.289
1987	269	0.314	0.918	0.006	-0.082	0.847	0.561	0.315
1991	249	0.421	1.054	0.215	0.054	1.672	0.639	0.466
1992	264	0.385	0.858	-0.027	-0.142	1.172	0.636	0.358
1993	277	0.298	0.870	-0.094	-0.130	0.416	0.664	0.404
1997	280	0.370	1.041	0.212	0.041	0.712	0.568	0.422
1998	298	0.309	0.857	-0.047	-0.143	0.479	0.681	0.340
1999	300	0.739	1.535	0.656	0.535	0.810	0.873	0.765
2003	281	0.554	1.232	0.365	0.232	0.913	0.733	0.609
2004	275	0.368	0.708	-0.260	-0.292	0.419	0.745	0.411
2005	276	0.517	0.976	0.211	-0.024	2.343	0.620	0.425
2009	247	1.460	1.864	1.369	0.864	1.671	0.915	1.663
2010	246	0.431	0.619	-0.342	-0.381	0.388	0.805	0.447
2011	244	0.328	0.782	-0.215	-0.218	0.424	0.709	0.329
All years		0.425	0.947	0.104	-0.053	1.042	0.676	0.478

Notes: Table 6 shows seven valuation performance indicators, which assess the goodness-of-fit of the intrinsic value estimates as per RIV model specification in Eq. (9). MAPE is the principal accuracy measure, while median V_0/P_0 as well as the mean and median SPE indicate the valuation bias. Finally, the standard deviation of SPE, 15% APE and IQRPE indicate the spread.

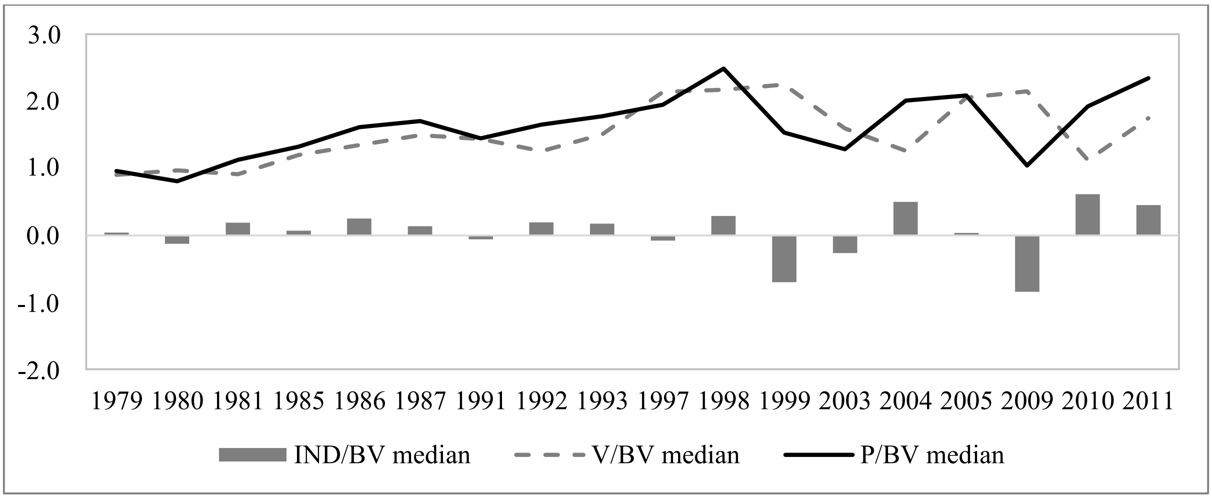
In terms of bias, some intriguing observations can be made. Over the period 1979-2011, both the median V_0/P_0 and the median SPE signal a lower bias towards understated value estimates than the one reported by previous studies (e.g. Penman, 2005; Courteau et al., 2001; Norrman & Rahmn, 2016). This suggests that incorporating the market's goodwill expectation at least partially offsets the RIV model's inherent tendency towards undervaluations (Dechow et al., 1999). Nevertheless, one can observe a significant rise in the skewedness and volatility of the model estimates following the year 1998. While the average V^h was largely understated in 2004 and 2010, extreme overvaluations were found in 1999 and 2009.

The analysis of the spread measures produces somewhat mixed results. While the relatively high standard deviation in SPE underscores the aforementioned volatility, IQRPE and 15% APE are in line with or more favorable than previous studies' findings (Jorgensen et al., 2011; Anesten et al., 2015). Overall, the measures of spread further support the notion of a fairly well-performing RIV model specification with regard to the entire sample period.

To further examine the time-series patterns of our intrinsic value estimates, Figure 2 presents the evolution of the median V and P, scaled by BV. In addition, median values of the scaled indicator variables (IND/BV) are shown. Appendix C reports both mean and median values for each investment year. Over the entire sample period, the median IND/BV is slightly positive, yet lower than the equivalent stated by Skogsvik & Skogsvik (2010). In line with our valuation analysis, the IND estimates can hence be considered an unbiased metric of over-/ underpricing. Further, the median V/BV and P/BV are similar in magnitude to the ones found by Skogsvik & Skogsvik (2010), which indicates sample consistency. Notably, while P/BV and V/BV seem to covary in a rather stable fashion prior to 1998, both ratios become much more volatile thereafter. Further, V/BV appears to trail the spikes and declines in P/BV with a lag of one period. This explains the phenomenon of extreme over- and undervaluations in several of the more recent investment years, which further materializes in highly skewed indicator variables. A more thorough discussion of these findings is reserved to section 7.

FIGURE 2

Time-series of normalized median price, intrinsic value and indicator variable



Notes: Figure 2 illustrates the time-series evolution in median values of price (P), intrinsic value (V) and the indicator variable (IND) across all investment years, scaled by the book value of owners' equity (BV). P is the ex-dividend market price of equity, V denotes the ex-dividend intrinsic value and IND is the difference between P and V at the investment date t=0.

5.3. INVESTMENT PERIOD SUMMARY STATISTICS

Table 7 provides an overview of final descriptive statistics for all investment years and both strategies. For the base case sample companies, $\overline{ROE}^h$ fluctuates significantly between subperiods, which is in line with the observed patterns in 5.1.1 regarding the estimation periods. A peak in the mean value of 15.3% in 1981 is accompanied by a low of 1.4% in 2004. Consistent with the logic of the prediction model, these fluctuations in turn impact the computed probabilities for future increases of $\overline{ROE}$: years with peaks in $\overline{ROE}^h$ (e.g. 1981) coincide with probabilities smaller than the cut-off value of 0.5, while troughs in $\overline{ROE}^h$ (e.g. 2004) typically observe high probabilities for future increases of $\overline{ROE}$.

TABLE 7

Summary statistics for investment portfolio samples

Year	*Number of firms (BC)*	$\overline{ROE}^h$		$\hat{p}(\Delta(\overline{ROE}) \geq 0)^*$		*Number of firms (IND)*	IND_0/BV_0		$\hat{p}(\Delta(\overline{ROE}) \geq 0)^*$	
		Mean	*Median*	*Mean*	*Median*		*Mean*	*Median*	*Mean*	*Median*
1979	465	0.138	0.142	0.426	0.409	205	0.113	0.053	0.432	0.418
1980	444	0.150	0.152	0.402	0.382	160	0.049	-0.026	0.460	0.461
1981	434	0.153	0.155	0.398	0.376	199	0.360	0.186	0.349	0.344
1985	396	0.078	0.097	0.608	0.602	194	0.010	0.014	0.595	0.587
1986	377	0.081	0.110	0.598	0.573	162	0.227	0.077	0.539	0.488
1987	358	0.075	0.105	0.602	0.584	154	-0.104	0.013	0.594	0.574
1991	344	0.094	0.120	0.499	0.477	157	0.025	-0.039	0.504	0.508
1992	343	0.074	0.096	0.539	0.521	151	0.192	0.083	0.508	0.480
1993	348	0.050	0.061	0.573	0.584	129	0.167	0.050	0.530	0.525
1997	398	0.139	0.145	0.426	0.396	154	-0.075	0.022	0.440	0.426
1998	408	0.132	0.149	0.430	0.388	185	0.597	0.414	0.368	0.348
1999	410	0.124	0.135	0.444	0.417	111	-0.321	-0.359	0.587	0.590
2003	352	0.039	0.062	0.626	0.629	186	-0.452	-0.293	0.650	0.650
2004	358	0.014	0.038	0.657	0.672	85	0.172	0.079	0.548	0.517
2005	345	0.059	0.072	0.600	0.612	132	-0.254	-0.095	0.619	0.629
2009	282	0.124	0.142	0.460	0.436	100	-0.453	-0.484	0.589	0.591
2010	275	0.092	0.102	0.497	0.488	123	0.816	0.579	0.417	0.417
2011	278	0.084	0.087	0.507	0.507	112	0.705	0.362	0.434	0.431
All years		0.097	0.114	0.511	0.501		0.102	0.038	0.503	0.486

Notes: Table 7 presents yearly arithmetic means and medians of the main decision variables, upon which stock investments are made. For the base case strategy (BC) on the left-hand side, historical medium-term ROE ($\overline{ROE}^h$) and the adjusted probability of $\overline{ROE}$ increases ($\hat{p}(\Delta(\overline{ROE}) \geq 0)^*$) are reported, while the indicator variable (IND_0/BV_0) is additionally shown for the indicator variable strategy (IND) on the right-hand side. The number of portfolio firms is presented for both strategies.

The statistics for the indicator variable strategy investment sample in Table 7 allow for a comparison with the total sample described in 5.1.2. Overall, IND_0/BV_0 in the limited sample is slightly lower in mean and median than the equivalent values of the total sample, pointing towards an even lower level of skewedness.[12] Compared to the base case statistics, a slightly lower overall mean and median probability in addition to a minor positive IND_0/BV_0 bias throughout the investment years leads to a marginally higher number of short allocations. In other words, the combination of a probability below the cut-off value of 0.5 and a positive indicator variable is on average more frequent than the criteria necessary for a long allocation. However, no systematic bias towards short positions can be found.

Additionally, it is worth noting that the distribution between short and long allocations is roughly even over the entire time period (see Appendix D). This outcome is particularly interesting with respect to the imbalanced allocations reported in Skogsvik & Skogsvik (2010), where 63.3% and 76.9% of all investments were allocated to the short position for the base case and indicator variable strategy, respectively.

5.4. MONTHLY CAPM EXCESS RETURNS

5.4.1. PERFECT FOREKNOWLEDGE

Table 8 (Panel A) reports estimates of 'Jensen's Alpha' and the corresponding beta coefficients for the long, short and hedge position, as obtained from the regressions described in 3.3.1.

For the base case strategy, the hedge position produces significant monthly abnormal returns of 1.2% (*p-value: 0.000*) over the entire investment period. Interestingly, the hedge return is almost identical to the equivalent reported by Skogsvik (2008) for the 1983-1991 investment period on the Swedish market. The sign of the obtained α estimates is consistent with the tested alternative hypotheses, i.e. positive for the long and negative for the short position. Notably, the short position contributes about 75% to the total hedge return and is highly significant (*p-value: 0.000*), while the abnormal return on the long position is non-significant at the 5%-level (*p-value: 0.097*). These findings stand in stark contrast to Skogsvik (2008), who observed the reverse asymmetrical relation between the long and short positions.

The analysis of the two time periods 1979-1993 and 1997-2011 yields further valuable insights. First, α values diminish significantly over the two time periods, with monthly excess return to the hedge position declining from 1.5% (*p-value: 0.001*) to 0.9% (*p-value: 0.000*). Second, the asymmetry of the long-short return contribution increases markedly over time, which is evident in clearly non-significant long returns of only 0.1% (*p-value: 0.332*) over subperiod II. In effect, the hedge return is, on average, almost entirely attributable to the statistically significant short position during the more recent investment periods.

[12] A mean (median) of 0.1 (0.04) compared to 0.22 (0.07) in the total sample.

TABLE 8

Monthly CAPM excess returns

Panel A: 'Perfect foreknowledge' scenario

Time period	*Number of obs.*	*Position*	*Base case*		*Indicator variable*	
			α	β^M	α	β^M
1979 - 2011	648	Long	0.003	1.080	0.007	1.032
		(p-value)	*(0.097)*	*(0.000)*	*(0.000)*	*(0.000)*
		Short	-0.009	0.936	-0.009	0.940
		(p-value)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		Hedge	0.012	0.144	0.016	0.092
		(p-value)	*(0.000)*	*(0.013)*	*(0.000)*	*(0.080)*
1979 - 1993	324	Long	0.005	1.032	0.009	1.054
		(p-value)	*(0.123)*	*(0.000)*	*(0.000)*	*(0.000)*
		Short	-0.010	0.940	-0.010	1.034
		(p-value)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		Hedge	0.015	0.092	0.019	0.020
		(p-value)	*(0.001)*	*(0.367)*	*(0.000)*	*(0.770)*
1997 - 2011	324	Long	0.001	1.142	0.004	1.013
		(p-value)	*(0.332)*	*(0.000)*	*(0.052)*	*(0.000)*
		Short	-0.008	0.928	-0.008	0.828
		(p-value)	*(0.001)*	*(0.000)*	*(0.004)*	*(0.000)*
		Hedge	0.009	0.214	0.012	0.186
		(p-value)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.019)*

Panel B: ROE prediction model scenario

Time period	*Number of obs.*	*Position*	*Base case*		*Indicator variable*	
			α	β^M	α	β^M
1979 - 2011	648	Long	-0.003	0.996	-0.001	0.944
		(p-value)	*(−)*	*(0.000)*	*(−)*	*(0.000)*
		Short	-0.004	1.018	-0.003	0.995
		(p-value)	*(0.002)*	*(0.000)*	*(0.009)*	*(0.000)*
		Hedge	0.001	-0.021	0.002	-0.051
		(p-value)	*(0.199)*	*(0.568)*	*(0.143)*	*(0.328)*
1979 - 1993	324	Long	-0.002	0.940	-0.002	0.965
		(p-value)	*(−)*	*(0.000)*	*(−)*	*(0.000)*
		Short	-0.005	1.058	-0.004	1.060
		(p-value)	*(0.011)*	*(0.000)*	*(0.014)*	*(0.000)*
		Hedge	0.003	-0.119	0.002	-0.095
		(p-value)	*(0.138)*	*(0.044)*	*(0.289)*	*(0.221)*
1997 - 2011	324	Long	-0.004	1.065	0.001	0.916
		(p-value)	*(−)*	*(0.000)*	*(0.426)*	*(0.000)*
		Short	-0.003	0.967	-0.002	0.915
		(p-value)	*(0.058)*	*(0.000)*	*(0.165)*	*(0.000)*
		Hedge	-0.001	0.097	0.003	0.000
		(p-value)	*(−)*	*(0.030)*	*(0.182)*	*(0.995)*

Notes: Table 8 shows the monthly CAPM excess returns (α) and corresponding beta coefficients (β^M) obtained from the regression procedure outlined in section 3.3 for both the base case and the indicator variable strategy. Panel A reports the returns given 'perfect foreknowledge', while Panel B presents the returns to investments based on the ROE prediction model. *P-values* are reported in parenthesis, unless the sign of α is inconsistent with the alternative hypothesis. For long and hedge positions, the null hypothesis of non-positive α is tested against the alternative of a positive α. For short positions, the null hypothesis of a non-negative α is tested against the alternative of a negative α. Tests of β-values are two-tailed.

When applying the indicator variable strategy under the 'perfect foreknowledge' scenario, several noteworthy observations can be made. Overall, the monthly hedge excess returns increase by about 0.4% points to 1.6% (*p-value: 0.000*) compared to the base case. Different from the previous findings, this increase is entirely due to the long position, which generates highly significant abnormal returns of 0.7% (*p-value: 0.000*). Thus, the long position accounts for roughly half of the total hedge returns. At the same time, the short position effectively remains constant. Therefore, the application of the indicator variable appears to have unilaterally benefitted the long position, while the impact on the short position is negligible.

The time-series analysis across the two subperiods largely confirms the findings of the base case strategy. Specifically, the abnormal hedge returns in the 1979-1993 period (α: 1.9%; *p-value: 0.000*) are considerably higher than those in the 1997-2011 period (α: 1.2%; *p-value: 0.000*). Moreover, while the significance of the long excess returns improves across all time periods, they are still non-significant at the 5%-level in period 1997-2011 (α: 0.4%; *p-value: 0.052*) and contribute considerably less to the total hedge return compared to the first period.

Concluding, the returns to the 'perfect foreknowledge' scenario underscore the high value-relevance of historical medium-term ROE and corroborate its usefulness for the prediction of $\overline{ROE}$ changes and future stock returns. Further, the indicator variable strategy proves to be highly effective in enhancing the abnormal return yield, with the benefits concentrated on the long position. These findings serve as a benchmark for the ROE prediction model results.

5.4.2. ROE PREDICTION MODEL

Table 8 (Panel B) reports α and β estimates for the long, short and hedge position based on the ROE prediction model. Base case and indicator variable strategy results are reported separately.

Overall, the returns to both investment strategies reveal a strikingly poor performance of the ROE prediction model as compared to the 'perfect foreknowledge' benchmark. Starting with the base case strategy, the monthly abnormal hedge returns amount to only 0.1% and are statistically non-significant (*p-value: 0.199*). The lack of statistical significance of the hedge returns applies to all investment periods and both strategies investigated. Further, the long position produces consistently negative and non-significant α estimates, which undermines the tested hypothesis. As a result of the negative long returns, the hedge returns turn negative in period 1997-2011. The monthly short returns amount to -0.4% (*p-value: 0.002*) overall and -0.5% (*p-value: 0.011*) over the period 1979-1993, while being non-significant thereafter. Hence, the short position fails to render the hedge returns significant.

In addition, no significant improvement is obtained by applying the indicator variable strategy. On the contrary, the significance of the α values to the long and short positions deteriorates when compared to the base case results. Albeit being non-significant, the long excess returns slightly increase under the indicator variable and turn positive from 1997 to 2011, resulting in hedge returns of 0.3% (*p-value: 0.182*) over this period. This substantiates our aforementioned findings in 5.4.1. Moreover, the short excess returns slightly decline.

Our results sharply contrast with those of Skogsvik & Skogsvik (2010) in the Swedish market. The authors report significant monthly α estimates to the hedge position of 0.4% for the base case strategy and up to 0.8% for the indicator variable strategy. Further, their positive hedge returns are almost entirely due to the long position – in line with Skogsvik (2008).

5.5. MONTHLY THREE-FACTOR EXCESS RETURNS

The most common objection directed towards anomalies studies concerns the omission of potential risk factors that – if carefully accounted for – would render abnormal returns non-significant (see 2.1.2). To ensure the robustness of our results, the Fama & French (1992, 1993) 'three-factor model' is applied according to the procedure outlined in 3.3.2.

Appendix E presents the intercepts and beta coefficients obtained from the regression procedure. Overall, the 'SMB' and 'HML' factors add little explanatory power to the monthly excess returns. This is both evident in negligible increases in R^2 and mostly non-significant coefficients for the size (β^{SMB}) and B/M (β^{HML}) factors. Using Fama & French's (1993, 1996) portfolio estimates as a benchmark, realizations of β^{SMB} and β^{HML} are below the median, suggesting sample firms to be relatively large and feature relatively low B/M values.

Perfect foreknowledge

In terms of the 'perfect foreknowledge' scenario, the abnormal hedge returns remain unchanged in all periods when applying the 'three-factor' regressions. In addition, the long excess returns diminish by 0.1% points, while the short excess returns 'appreciate' to the same extent (i.e. -0.1%). Interestingly, these observations hold for both the base case and the indicator variable strategy. This phenomenon seems to be partially driven by β^{HML}, which is found to be considerably higher and significant for the long and short positions in period 1997-2011. Thus, the sample composition might have shifted to higher B/M firms over time.

ROE prediction model

The returns to the ROE prediction model reveal somewhat different patterns. The abnormal hedge returns diminish by 0.1% points for the base case and 0.2% points for the indicator variable strategy in both time periods, but remain non-significant. The decrease is mainly driven by equivalent declines in the long returns, which become even more negative relative to the CAPM metric, and turn significant overall for the base case. Compared to the 'perfect foreknowledge' scenario, the β^{HML} estimates are higher and asymmetrically distributed between the long and short position. However, as β^{HML} is found to have no significant explanatory power for the observed excess returns, no in-depth analysis is conducted.

Concluding this section, our findings suggest that applying the ROE prediction model proves ineffective in generating abnormal returns for both return metrics, given the time period and firm sample examined in this thesis. At the same time, the results of the 'three-factor' model further emphasize the robustness of the abnormal return potential as reflected by the 'perfect foreknowledge' scenario. Taken together, these observations are indicative of several important research conclusions that require further scrutiny in the following section.

6. ADDITIONAL TESTS

The results presented in the previous section differ markedly from the findings in Skogsvik & Skogvik's (2010) study with regards to the abnormal returns generated by both investment strategies using the ROE prediction model. To systematically investigate the underlying root causes of the poor model performance, the following five additional tests are conducted:

i) **Calibration of model (6.1):** Test to examine whether results are affected by methodological deficiencies or false technical applications of the ROE prediction model, the investment strategies and the return calculation.

ii) **Mean reversion assumption (6.2):** Test to examine the empirical validity of the critical three-year mean reversion assumption underlying the ROE prediction model.

iii) **Impact of time-varying ROE levels (6.3):** Test to examine the sensitivity of the ROE prediction model to $\overline{ROE}^h$ fluctuations between the estimation and investment periods.

iv) **Misallocation impact (6.4):** Test to examine the return impact of alternative prediction cut-off values and to identify characteristics of systematic misallocation drivers.

v) **Sentiment bias (6.5):** Test to examine whether results are adversely affected by a systematic market sentiment bias, as evidenced by Skogsvik & Skogsvik (2010).

While test i) aims to confirm the proper technical operationalization, tests ii) and iii) problematize the impact of two pivotal ROE prediction model assumptions. Lastly, tests iv) and v) seek to detect specific sample characteristics with a particularly adverse return impact.

6.1. CALIBRATION OF MODEL

One of the main challenges in replicating a model and applying it to a new market context is the lack of benchmarks for the obtained results. While the 'perfect foreknowledge' scenario aims to fill this void, it is not able to remove uncertainties about the correct calibration of the ROE prediction model. Hence, to further test our methodology, we apply it to the same period and a similar sample as Ou & Penman (1989). Their study serves as a calibration mechanism for our approach, given the substantial abnormal returns they found in the U.S. market with a similar fundamental investment strategy. Appendix F presents the methodology, the sample characteristics as well as the achieved returns and the associated prediction accuracy.

To begin with, the prediction accuracy amounts to 63.0% and is therefore slightly weaker than in our original sample (68.3%). The proportion of actual increases and decreases during the entire period is 49.9% and 51.1%, respectively. While the model predicts decreases correctly in 75.9% of the cases, only 50.0% of the increases are correctly predicted. In contrast to the balanced prediction accuracy reported in 5.1.2, this performance bias towards decreases is similar to the findings of Skogsvik & Skogsvik (2010).

Panel B in Table 14 (Appendix F) reports the estimates of 'Jensen's Alpha' and the corresponding beta coefficients for the long, short and hedge position under the 'perfect foreknowledge' scenario as well as the ROE prediction model. The 'perfect foreknowledge' scenario generates monthly abnormal returns of 1.9% that are highly significant (*p-value: 0.000*) and somewhat higher than observed in our original sample. Contrary to our results in the base case 'perfect foreknowledge' scenario, the short position is barely significant at the 5%-level (*p-value: 0.049*) and contributes marginally to the abnormal hedge returns, while the long position is highly significant (*p-value: 0.000*), contributing 1.4% to the monthly hedge returns. In terms of the ROE prediction model scenario, it can be observed that positive and significant (*p-value: 0.015*) monthly abnormal hedge returns of 0.9% are generated. Once again, the long position is the driving force behind this hedge return with a 0.7% (*p-value: 0.032*) return contribution, while the short position remains non-significant (*p-value: 0.181*).

To put the hedge returns presented in Appendix F into context, Ou & Penman's (1989) results provide a rough benchmark. Their 'perfect foreknowledge' scenario generates cumulative market-adjusted monthly hedge returns of 28.41% over a 36-month period, with long and short positions contributing evenly. The investment strategy based on 'Pr', which serves as a benchmark for our prediction model strategy, yields cumulative market-adjusted monthly hedge returns of 20.83%. Although our method and return calculations differ and thus make one-to-one comparisons with Ou & Penman's (1989) results neither possible nor appropriate, the positive and significant hedge returns as well as the similarity in magnitude corroborate our confidence in the 'correct' application of the ROE prediction model.

In addition, the test results mitigate concerns regarding a systematic bias in the methodology underlying the returns to the long position. While the returns in our original sample show non-significant and weak (or negative) long excess returns in all years, the calibration in the period 1973-1983 proved that significant returns to the long position can indeed be obtained.

6.2. MEAN REVERSION ASSUMPTION

One of the crucial assumptions of the investment strategy is the concept of mean-reverting ROEs. While the intuition for the latter has been outlined earlier, it is worth testing to what extent the theory holds true in the investigated sample. Apart from the general properties of time-series changes in ROE, the duration of the mean-reversion process is of particular importance in the context of the Skogsvik & Skogsvik's (2010) model. By defining $\overline{ROE}$ as a three-year average and by taking contrarian positions depending on $\overline{ROE}^h$, the three-year horizon is assumed to be sufficient to reflect the mean-reverting process of ROE. This assumption seems reasonable given Penman's (1991) observation of a three-year period capturing most of the ROE portfolio changes in the U.S. market. However, due to differing samples and time periods, further scrutiny is warranted.

Appendix G provides an overview of the average changes in $\overline{ROE}$ as the difference between $\overline{ROE}^f$ and $\overline{ROE}^h$ for the quartiles based on mean $\overline{ROE}^h$ performance. In line with the general notion of mean-reverting ROEs, the changes in $\overline{ROE}$ for the first quartile, i.e. the strongest past performers, are negative while the changes for the fourth quartile, i.e. the weakest past performers, are positive for all time spans. Within the second and third quartile, this logic holds true for the time periods following 1979 and most of 2003 (with the exception of 2003 to 2005). Therefore, the condition of mean-reverting ROEs is substantiated by the sample set.

However, the observed magnitude of the $\overline{ROE}$ changes across different time spans casts doubts on the assumption that a three-year period captures a sufficiently large fraction of the mean-reversion drift. While the changes in the first three-year period in Panel A range from -4.87% to +5.95% and from -0.12% to +15.52% in Panel B, a look at the subsequent $\overline{ROE}$ development reveals a significantly longer reversion trend. Particularly, the firms in the first quartile show considerably higher changes in absolute values six or nine years after the investment point in time when compared to the first three-year period.

These findings point towards two central limitations. First, assuming a correlation between stock prices and ROE development, this test calls into question whether stock prices fully incorporate the effects of mean reversion within a three-year time window. Considering that ROE reversion trends extend significantly beyond the investment period, one may assume stock prices to react more strongly at a later stage, i.e. six or nine years after the investment date. Second, limiting the definition of medium-term to a three-year time period could adversely affect the prediction accuracy, since the accordingly defined $\overline{ROE}^h$ might not be a good indicator of the current state of a firm's relative profitability trend. However, while these results suggest extensions of the mean-reversion period as a potential model improvement, caution is warranted due to the difficult isolation of the mean reversion impact from other industry- or firm-specific factors. Further, the relationship between stock price and ROE development in a given market context needs to be scrutinized more thoroughly to understand at which point in time the market reacts most strongly to signals of $\overline{ROE}$ changes.

6.3. IMPACT OF TIME-VARYING ROE LEVELS

Another inherent weakness of the ROE prediction model relates to its inability to accommodate time-varying levels of ROE between the estimation and investment periods. As discussed in section 5, the $\overline{ROE}^h$ levels are found to be subject to strong cyclical fluctuations over the sample period. However, the prediction model relies on the assumption that ROE levels measured over the 12-year estimation periods are representative of subsequent levels in the investment periods. This discrepancy has significant implications for the bias in the prediction accuracy between $\overline{ROE}$ increases and decreases.

Appendix H depicts the association of the change in mean $\overline{ROE}^h$ levels between each investment year and the corresponding estimation period, with the prediction accuracy for $\overline{ROE}$ increases, decreases and overall changes in the respective investment year. The time-series analysis reveals the following systematic pattern: Positive (negative) changes in the mean $\overline{ROE}^h$ level from the estimation to the investment period result in strong (weak) predictions of future $\overline{ROE}$ decreases and weak (strong) predictions of future $\overline{ROE}$ increases. For instance, the higher mean $\overline{ROE}^h$ level in 1980 (+4.9%) leads to 81.5% correct predictions of decreases and only 41.0% correct predictions of $\overline{ROE}$ increases. Conversely, the markedly lower $\overline{ROE}^h$ level in 2004 (-12.2%) involves 84.6% correct predictions of $\overline{ROE}$ increases, but only 42.3% correct predictions of $\overline{ROE}$ decreases. Importantly, the distribution of the actual $\overline{ROE}$ increases (decreases) in the investment periods does not explain the skewness in the prediction accuracy. However, it affects the strength of the overall prediction accuracy.

The observed patterns result from a simple deficiency: the coefficients of the estimation models are based on a specific historical ROE distribution, where the 50% probability of an $\overline{ROE}$ increase is roughly attributed to the average historical ROE as the threshold. If the ROE level increases in the investment period, a larger proportion of firms will surpass this estimation threshold and be projected to decrease in the future, thereby boosting the prediction accuracy for the short position. The reverse relation holds for ROE-level decreases.

Assuming this pattern to be representative of the ROE prediction model, one can conclude that Skogsvik & Skogsvik's (2010) consistently stronger prediction of $\overline{ROE}$ decreases may have been due to a persistent upward trend in historical ROE over time. A strong case can thus be made for model amendments that account for time-varying levels of ROE, thereby enhancing the model robustness. One particularly fruitful approach would be to use an indexed measure of historical ROE instead of the absolute level. Further, it may be beneficial to incorporate business cycle indicators that help predict changing industry levels of ROE.

6.4. MISALLOCATION IMPACT

Another critical assumption in the ROE prediction model is the choice of the probability cut-off value of 0.5. A more stringent cut-off criterion could potentially improve the prediction accuracy and in turn the returns. Accordingly, Freeman et al. (1982) attribute the weak performance of their prediction model to the fact that two-thirds of the observations ranged between 0.4 and 0.6. They conclude that the predictive ability of ROE is limited to deviations from the mean larger than 0.1. Following this logic and Ou & Penman (1989), cut-off values of 0.4 and 0.6 are chosen for the short and long positions respectively to test the impact of the 'borderline cases'. However, the prediction accuracy only exhibits marginal improvements and the return performance of both investment strategies shows no significant changes.

Thus, the majority of the misallocations with severe return impact seems to be robust to changes in probability cut-off values. Based on the indicator variable strategy, we further investigate this issue for the long and short positions separately and focus solely on misallocations, i.e. actual increases in $\overline{ROE}$ that were allocated to the short position and vice versa. By splitting them into quartiles according to the weighted return impact, typical characteristics of the largest negative return contributors can be assessed (see Table 16 in Appendix I). As a first result, the firms in the first quartile, i.e. firms with the most adverse impact on the respective position, are found to have mean and median probabilities of below 0.4 for the short, and above 0.6 for the long position for most of the investigated years. This observation largely explains why the change in cut-off values does not lead to significant improvements.

Another intriguing observation can be made regarding the portfolio composition. For each investment year, between 17% and 59% of the misallocations originate from firms that were delisted in the subsequent 36-month holding period. However, the conventional prediction accuracy measure compares predicted ROE changes to the actual ones, without accounting for delistings. Given the considerable proportion of delistings, the denominator of the prediction accuracy measure is reduced significantly in several years. When incorporating delistings into the denominator, the overall prediction accuracy is reduced by 9.6% and moves below 50% (see Table 17 in Appendix I). Thus, the conventional prediction accuracy measure can be misleading in terms of the predictive performance of the model when delistings are frequent. The relevance of delisted firms is substantiated by the return impact of the respective misallocations. In line with Skogsvik & Skogsvik (2010), proceeds from delisted stocks are reinvested in the market index for the remainder of the investment period. With this approach, the impact on the returns is entirely contingent on the overall market development, which may explain why the impact of misallocated delistings shows the same sign for the short and long positions in four of the six investigated periods (see Table 18 in Appendix I). If delisted firms were to be excluded, the adverse return impact of the misallocations would be significantly alleviated.

The insights from this preliminary analysis cast some doubts on the feasibility of the Skogsvik & Skogsvik (2010) method for larger samples. By investigating a small sample of established firms in the Swedish market, Skogsvik & Skogsvik's (2010) results were potentially less sensitive to delistings. However, given the significant impact on our sample, the reinvestment of delisted proceeds in the market index may be inadequate. With bankruptcy being a primary reason for delistings, the adoption of a bankruptcy prediction model (e.g. Skogsvik, 1990) may be useful as an additional investment criterion. However, further investigations would need to assess to what extent bankruptcy contributes to delistings and whether the added complexity (in addition to ROE prediction model and indicator variable) would undermine the parsimony and viability of the investment strategy.

6.5. SENTIMENT BIAS

An important caveat regarding the out-of-sample validity of the trading strategies was brought forward by Skogsvik & Skogsvik (2010), who identified a 'positive sentiment bias' in the Swedish sample data. By testing the market reaction to earnings surprises under the indicator variable strategy, it was concluded that 'positive news' ($\overline{ROE}$ increases after negative market expectations) on average produced much stronger price reactions than 'negative news' ($\overline{ROE}$ decreases after positive market expectations). This observation raises the question whether the significant abnormal returns to the long position are replicable in other market contexts.

To compare the price reaction to positive and negative news, average 36-month market-adjusted returns are assessed based on the indicator variable strategy under the 'perfect foreknowledge' scenario. As the ROE prediction model produces non-significant returns for our sample data, the 'perfect foreknowledge' scenario is argued to be the more accurate measure of relative market response. By design, it rules out the possibility of an asymmetric prediction accuracy for $\overline{ROE}$ increases and decreases, which might otherwise explain an observed return bias. Appendix J presents the methodology for the return metric, the relevant sample statistics and the 36-month long, short and hedge returns for each investment year.

Of the investigated 5,453 firm-year observations, 45.1% show an increase and 54.9% a decrease in $\overline{ROE}$, which is similar to Skogsvik & Skogsvik's (2010) finding. Further, the indicator variable is positive for 47.1%, negative for 34.7% and zero for 18.2% of the observations. Accordingly, 1,205 short positions and 811 long positions are taken over the entire period 1979-2011. Interestingly, no indication of a systematic bias in market prices can be found. For 'positive earnings surprises' (i.e. long positions), the average market-adjusted 36-month return over the entire investment period amounts to +36.4%. On the other hand, the equivalent market-adjusted return for 'negative earnings surprises' (i.e. short positions) is -39.7%. While the market-adjusted returns vary considerably on a period-by-period basis, the overall results corroborate the absence of a systematic sentiment bias in market prices. This contrasts sharply with Skogsvik & Skogsvik's (2010) finding of market-adjusted long returns of +144.7% and equivalent short returns of -23.1%. While the latter results were considered evidence of over-optimistic market expectations with regard to future medium-term ROE, our findings suggest a rather symmetric distribution of market expectations in the U.S. sample.

These results are in line with our assessment of monthly excess returns. An intriguing question is to which extent the significant abnormal long returns reported by Skogsvik & Skogsvik (2010) were driven by 'positive sentiment' as a form of mispricing. However, the assessment of the relative impact of the 'sentiment bias' is left to future research.

7. DISCUSSION

By replicating and applying the Skogsvik & Skogsvik (2010) model to the U.S. market, this thesis set out to answer two central questions. First, whether a parsimonious fundamental trading strategy can earn abnormal returns similar to those observed in the Swedish market. And second, what observations can be made regarding market efficiency and its development in the U.S. market over the time period 1979-2014. Subsequent to the previous – rather technical – discussion of potential explanations for the weak model performance, this section aims to take a step back and pinpoint the main findings of our research with regards to the aforementioned research questions. Lastly, some methodological limitations are presented.

7.1. PERFORMANCE OF THE PARSIMONIOUS MODEL

The positive and significant abnormal hedge returns to both strategies in the 'perfect foreknowledge' scenario prove ROE to be an effective fundamental predictor of future earnings changes. However, despite the relatively high prediction accuracy, which is broadly in line with Skogsvik & Skogsvik (2010), the investment strategies based on the estimated ROE prediction models do not yield significant abnormal returns in any examined period.

The results indicate that the returns to the prediction-based trading strategies are profoundly influenced by the investigated time period and data sample. However, a fundamental investment strategy can only be labelled effective, if it demonstrates the ability to generate systematic abnormal returns across different time periods and market contexts. Otherwise, limited data sets could lead to false inferences being drawn from a specific time period (Lo & MacKinlay, 1990). Our results with positive and significant hedge returns in the calibration period are a good example of this 'danger', as they illustrate how a model can work for a certain sample in a specific period, but fail to deliver positive results in subsequent periods.

As the discussion of model flaws and possible improvements in the previous section revealed, there is the potential to at least partially mitigate the weaknesses of the examined investment strategies. An extension of the time period implied by medium-term ROE could contribute to a more accurate indication of a firm's state within the mean reversion process. Furthermore, the use of an indexed measure of historical medium-term ROE could make the model more robust to changes in ROE levels between estimation and investment periods. And the addition of a bankruptcy prediction model is likely to impose an effective complementary hurdle to reduce the risk of investing in stocks that are delisted throughout the holding period.

However, there are two pivotal caveats concerning further amendments to the model. First, in line with the nature of research, improvements can only be made ex-post, meaning based on historical data and circumstances. The risk of (implicitly) constructing a 'manipulated' data sample for which a specific model amendment proves to be an enhancement is large (Schwert, 2003). And second, any amendment to the Skogsvik & Skogsvik (2010) model would further increase the already substantial complexity of the forecasting procedure and thereby undermine the goal of parsimony. As it is, the model poses a considerable amount of data processing and analysis work to the investor – a fact that Skogsvik & Skogsvik (2010) acknowledge in their original study. The ROE prediction model may be considered simplistic in its underlying assumptions, yet it requires an extensive set of company data and statistical expertise in order to run and analyze the regressions. The indicator variable strategy adds to the difficulty of the data processing workload by requiring a thorough valuation process with sound decision-making. Thus, while straightforward in its logic, the model in its current form imposes quite high requirements on the user, while achieving non-significant returns.

These results indicate that the so-called parsimonious prediction model based on ROE is not sufficient to predict and trade on future mispricing in the U.S. market. These findings could point towards higher investor sophistication in the U.S. compared to the Swedish market, which requires more elaborate multivariate models to reveal mispricing. In line with Penman's (1991) finding, the inclusion of earnings persistence indicators – such as P/B ratios – could be a fruitful enhancement. However, given the current level of complexity, any amendments that would simultaneously increase returns and model sophistication are likely to reflect a 'fair' compensation for the information cost incurred by the investor, rather than abnormal returns.

7.2. MARKET EFFICIENCY ASSESSMENT IN THE U.S. MARKET

Central to achieving our second research aim is the question whether the Skogsvik & Skogsvik (2010) model can reveal mispricing in the U.S. stock market and whether the evidence on market efficiency differs from or complies with findings in the Swedish market. To this end, applying an accounting-based prediction model and trading strategy to a vast U.S. firm sample and time period offers an intriguing opportunity for market efficiency testing.

Our findings generally support the notion that the U.S. market was efficient in the semi-strong form over the period 1979-2014. This conclusion draws from three important observations: First and foremost, as noted in 7.1, the ROE-based trading strategies fail to produce significant abnormal hedge returns throughout all investigated periods, despite a favorable prediction performance. The absence of apparent mispricing suggests that both the predictive power of historical medium-term ROE as well as its valuation implications were by and large impounded in market prices. Accordingly, neither forecasting mispricing nor modeling mispricing are supported. Second, even when applying the 'perfect foreknowledge' scenario, which indicates the hypothetical abnormal return potential, no significant returns to the long position are evident in the base case. One interpretation of this would be that investor anticipation is particularly strong for ROE increases, while decreases are either more difficult to predict or less likely to result in trading due to potential short-selling cost. Third, no evidence of a systematic sentiment bias is found in U.S. market prices, which corroborates the out-of-sample validity of our findings. This means that the conclusions drawn with respect to mispricing in the U.S. market can be considered an unbiased representation of other time periods and data samples and do not seem to be compromised by market distortions such as systematic over- or underpricing.

Our market efficiency analysis is supported by the consideration of three major pitfalls in anomalies research, i.e. data manipulation, risk adjustment and trading costs (see 2.1.2). While the latter has been considered essential in explaining the short returns, both data-fitting and risk concerns were preemptively addressed in our research design. Reassuring of our hypotheses, the results are robust to non-overlapping investment periods (see Appendix K) and common risk factors including market, size and book-to-market value (Fama & French, 1993). Thus, our empirical research is broadly affirmative of the efficient market hypothesis.

The results differ markedly from Skogsvik & Skogsvik's (2010) observation of significant and large abnormal returns over the period 1983-2003 in the Swedish market. Given the comparable sample characteristics, one might infer that the U.S. market was relatively more efficient within the overlapping time frame. However, as it is unclear to which extent the Swedish return observations were due to a 'positive sentiment bias' rather than actual mispricing (Skogsvik & Skogsvik, 2010), the comparison remains somewhat inconclusive.

Besides the apparent differences in the degree of market efficiency, interesting commonalities are revealed in the time series. In line with Skogsvik & Skogsvik (2010), the abnormal return potential is found to diminish substantially over time (based on the 'perfect foreknowledge' scenario). The notion of increasing market efficiency is corroborated by the disproportionately higher return contribution of the short position after 1997, as abnormal short returns can be considered less 'reliable' due to potential short-selling constraints. Considering the likely trading cost implied by short investments, it can be argued that – even with 'perfect foresight' – excess returns could have hardly been earned in recent years. The findings point towards the role of market learning, whereby facilitated data processing and greater investor sophistication have reduced the information cost over time. This implies that the higher return potential observed in the 1980s most likely reflects the prohibitively costly data analysis. Given the advances in knowledge and computing power, our observations support the notion of Richardson et al. (2010) that "in some sense, it is relatively 'easy' to find an 'anomaly' in the 1960s, 1970s, and even in the 1980s. It is much harder to do so in the last ten years" (p. 450).

A puzzling empirical finding, however, is missing in the aforementioned discussion, which may weaken our reasoning on market efficiency. To begin with, the indicator variable has received little attention in our assessment, as it merely improves the returns in the 'perfect foreknowledge' scenario. While this shows its general usefulness as a measure of stocks' 'cheapness' by fundamental standards, it is found to be highly contingent on a well-performing base case. Nevertheless, our analysis of the valuation implied by the indicator variable shed light on an intriguing phenomenon. As evident from our findings in 5.2.2, an abrupt decline in the valuation accuracy of the RIV model is identified after 1998, accompanied by soaring volatility in various ratios. Further investigations (see Appendix L) provide preliminary evidence of a structural break between the periods 1979-1993 and 1997-2011. The once closely aligned relationship between intrinsic values and prices appears to dissolve after 1998, as visible in unprecedented spikes and declines in the V/P ratio. Interestingly, the apparent decoupling cannot be solely explained by the change in the market's implicit goodwill expectation, which admittedly increases the volatility of the intrinsic value estimates.[13] Even measures unaffected by our modeling, such as the P/B ratio, cease to follow their almost linear trend lines with the rise of the 'Dot-com crash' and fluctuate wildly thereafter. Taken together, our findings are at least indicative of what Curtis (2012) phrased a "significant change in the time-series dynamics of price" in the mid-1990s, which led to the "lack of cointegration between price and accounting fundamentals" (p. 143).

Given that further statistical tests and continuous time-series analyses are outside the scope of our thesis, this rather naïve analysis is by no means definite. Nevertheless, our observations coupled with previous research (Curtis, 2012) could point towards increasing speculation in

[13] Based on the 'last year's q(TOT)' approach (Norrman & Rahmn, 2016), strong market price fluctuations are impounded in the subsequent period's intrinsic value estimates via the reverse-engineered goodwill expectation. See 7.3 for further remarks on the limitations of the 'last year's q(TOT)' approach.

U.S. market prices. This would not only substantially lower the value of fundamental analysis and valuation, but at worst render such assessments of market efficiency impossible.

7.3. LIMITATIONS OF THE RESEARCH DESIGN

Throughout this thesis, we aspired to be mindful of potential limitations regarding the methodology used and critically reflect upon the central assumptions that were made. However, despite the consideration of the pitfalls outlined in 2.1.2 in our research design, the choice of a parsimonious model inevitably requires a trade-off between accuracy and utility. This section aims to give a brief and structured summary of the most important confinements.

Data and sample selection

In an attempt to process large amounts of data, our work relies on accurate figures in electronic databases. Besides using the well-reputed Compustat and CRSP databases, we defined strict selection criteria in order to mitigate the risk of including falsely reported data points (e.g. extreme values). While the definition of specific selection and exclusion criteria is necessary to establish a consistent data set, it leaves room for a considerable amount of decisions to be made by the researcher. In this context we acknowledge that although it was our explicit objective to avoid data biases as outlined in 2.1.2, some decisions regarding the sample selection might have influenced our results. The definition of cut-off points for extreme values, the use of the industry classification by French, as well as the handling of delisted companies (see 6.4) are prominent examples. However, we believe to have cautiously considered all options and that our careful interpretation mitigates the risks involved.

ROE prediction model

The main limitation of the ROE prediction model is the underlying assumption regarding the mean reversion duration in the investigated sample (see 6.2). Furthermore, by requiring full data availability for several consecutive years in the past, the prediction model is subject to an implicit survivorship bias. Nevertheless, given our aim to replicate Skogsvik & Skogsvik's (2010) model, these limitations are negligible from a comparative analysis point of view.

Indicator variable

To calculate the indicator variable, we applied the reverse-engineering method by Norrman & Rahmn (2016) to arrive at estimates of the market's goodwill expectation at the horizon point in time. This approach rests on two critical assumptions: First, it implicitly assumes prices to be efficient. The method is, however, deemed appropriate in a context of relaxed informational efficiency. Thus, even prices that (temporarily) deviate from intrinsic values are assumed to entail value-relevant goodwill information that significantly enhances our RIV model estimates. Second, using 'last year's q(TOT)' is presumed to be more accurate than a multi-year average. Norrman & Rahmn (2016) indicated the superior valuation accuracy of the most recent market information. A drawback of this method is its susceptibility to price volatility, which enters the intrinsic value estimate with a one-period lag through the inclusion of the market's implicit goodwill estimate. This exacerbates the valuation discrepancy in volatile market contexts, as can be seen in the markedly deteriorating valuation accuracy measures after the year 1998. However, the 'cost' of this approach are argued to be outweighed by the strong overall valuation performance and the parsimonious model benefits.

8. CONCLUDING REMARKS

The aim of this thesis was to find answers to the two principal research questions regarding the abnormal return potential of a parsimonious FA-based trading strategy and the development of market efficiency in the U.S market. In line with Skogsvik & Skogsvik (2010), two investment strategies were investigated. The base case strategy subjects investment decisions solely to the prediction of medium-term ROE changes. The indicator variable strategy adds another investment criterion by comparing stock prices to fundamental values based on the residual income valuation to reveal mispricing in the market.

The results of the empirical analyses suggest that both investment strategies based on the outlined parsimonious prediction model are not able to achieve significant abnormal hedge returns in the investigated setting, despite a relatively strong prediction accuracy of 68% for future medium-term ROE changes. In order to validate whether investors could benefit from an improved forecasting performance, we further investigated a scenario in which predictions were made under the assumption of 'perfect foreknowledge'. In this scenario, the base case strategy generates monthly CAPM excess returns of 1.2% for the hedge position, with 75% of the returns proceeding from short sales. The indicator variable strategy improves these monthly returns by another 0.4%.[14] Interestingly, the indicator variable appears to have exclusively benefitted the long position, while the short position's performance remains virtually unchanged. The significant returns of the 'perfect foreknowledge' scenario not only underline the predictive power of ROE for future earnings changes, they also substantiate the existence of an abnormal return potential and thus motivated a series of additional tests conducted in order to uncover flaws of the model logic. While findings regarding ROE cyclicality, the mean reversion assumption and the risk of delistings offer room for improvements of the prediction model, the cost of additional complexity may undermine the approach of parsimony.

The theoretical abnormal return potential furthermore allows for intriguing insights into market efficiency. First, the reliance of the base case returns on the short position stands in stark contrast to the strong performance of the long position reported by Skogsvik & Skogsvik (2010). Given the legal barriers and high trading costs associated with short sales, it has to be

[14] All of the returns are robust to the 'three-factor' risk metric introduced by Fama & French (1993).

questioned whether the apparent abnormal returns are in fact realizable and can thus be seen as a sign of market inefficiency. Second, in line with Skogsvik & Skogsvik (2010), the abnormal return potential decreases significantly over time – by 0.6% (0.7%) in the base case (indicator variable) strategy. Thus, it can be argued that the U.S. market has become more efficient over time, potentially due to market learning and decreasing information cost.

However, while improved market efficiency is one possible explanation for diminishing returns, the findings of our thorough valuation assessment point towards another likely cause: a 'decoupling' of prices from fundamental values in recent years. From the late-1990s onwards, we not only observe a significant deterioration in several valuation performance indicators of our RIV model, but also a soaring volatility in key valuation ratios. If understood as a sign of rising speculation (Curtis, 2012), this could in turn indicate 'crazy' rather than 'efficient' prices that undermine the utility of fundamental analysis and related trading strategies. We leave it to future research to explore whether more recent U.S. stock prices are best explained by the long-standing 'efficient market hypothesis' or a yet-to-be-defined 'crazy market hypothesis'.

9. REFERENCES

Abarbanell, J. S. & Bushee, B. J. (1997). Fundamental Analysis, Future Earnings, and Stock Prices. *Journal of Accounting Research, 35*, 1-24.

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

Abarbanell, J. & Bernard, V. (2000). Is the U.S. Stock Market Myopic? *Journal of Accounting Research, 38*, 221-242.

Anesten, S., Möller, N. & Skogsvik, K. (2015). The Accuracy of Parsimonious Equity Valuation Models - Empirical Tests of the Dividend Discount, Residual Income and Abnormal Earnings Growth Models. *SSE Working Paper Series in Business Administration, 2015*, 1-39.

Ball, R. (1978). Anomalies in Relationships between Securities' Yields and Yield-Surrogates. *Journal of Financial Economics, 6*, 103-126.

Ball, R. (1994). The Development, Accomplishments and Limitations of the Theory of Stock Market Efficiency. *Managerial Finance, 20*, 3-48.

Ball, R. & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research, 6*, 159-178.

Beaver, W. H. (1970). The Time Series Behavior of Earnings. *Journal of Accounting Research, 8*, 62-99.

Beaver, W., McNichols, M. & Price, R. (2016). The Cost and Benefits of Long-Short Investing: A Perspective on the Market Efficiency Literature. *Journal of Accounting Literature, In Press, Accepted Manuscript*.

Beneish, M. D., Lee, C. M. & Tarpley, R. L. (2001). Contextual Fundamental Analysis Through the Prediction of Extreme Returns. *Review of Accounting Studies, 6*, 165-189.

Bernard, V. L. (1995). The Feltham-Ohlson Framework: Implications for Empiricists. *Contemporary Accounting Research, 11*, 733-747.

Bernard, V. L. & Thomas, J. K. (1990). Evidence that Stock Prices do not Fully Reflect the Implications of Current Earnings for Future Earnings. *Journal of Accounting and Economics, 13*, 305-340.

Chordia, T., Roll, R. & Subrahmanyam, A. (2011). Recent Trends in Trading Activity and Market Quality. *Journal of Financial Economics, 101*, 243-263.

Courteau, L., Kao, J. L. & Richardson, G. D. (2001). Equity Valuation Employing the Ideal versus Ad Hoc Terminal Value Expressions. *Contemporary Accounting Research, 18*, 625-661.

Curtis, A. (2012). A Fundamental-Analysis-Based Test for Speculative Prices. *The Accounting Review, 87*, 121-148.

Damodaran, A. (2016). *Return on Equity decomposition by Industry sector: US [Data file]*. Retrieved October 20, 2016, from Damodaran Online: http://people.stern.nyu.edu/adamodar/New_Home_Page/datacurrent.html#growth

DeBondt, W. F. & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance, 40*, 793-805.

Dechow, P. M., Hutton, A. P. & Sloan, R. G. (1999). An Empirical Assessment of the Residual Income Valuation Model. *Journal of Accounting and Economics, 26*, 1-34.

Dimson, E., Marsh, P. & Staunton, M. (2016). In *Credit Suisse Global Investment Returns Yearbook 2016*. Credit Suisse.

Edwards, E. O. & Bell, P. W. (1961). *The Theory and Measurement of Business Income*. London: Cambridge University Press.

Fairfield, P. M., Sweeney, R. J. & Yohn, T. L. (1996). Accounting Classification and the Predictive Content of Earnings. *The Accounting Review, 71*, 337-355.

Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial Analyst Journal*, 55-59.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 383-417.

Fama, E. F. (1998). Market Efficiency, Long-Term Returns, and Behavioral Finance. *Journal of Financial Economics, 49*, 283-306.

Fama, E. F. & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance, 47*, 427-465.

Fama, E. F. & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics, 33*, 3-56.

Fama, E. F. & French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance, 51*, 55-84.

Fama, E. F. & French, K. R. (2006). Profitability, Investment and Average Returns. *Journal of Financial Economics, 82*, 491-518.

Feltham, G. A. & Ohlson, J. A. (1995). Valuation and Clean Surplus Accounting for Operating and Financial Activities. *Contemporary Accounting Research, 11*, 689-731.

Fernandez, P., Ortiz, A. & Acín, I. (2016). Market Risk Premium Used in 71 Countries in 2016: a Survey with 6,932 Answers. *Navarra: IESE Business School*.

Foster, G., Olsen, C. & Shevlin, T. (1984). Earnings Releases, Anomalies, and the Behavior of Security Returns. *The Accounting Review, 59*, 574-603.

Francis, J., Olsson, P. & Oswald, D. R. (2000). Comparing the Accuracy and Explainability of Dividend, Free Cash Flow, and Abnormal Earnings Equity Value Estimates. *Journal of Accounting Research, 38*, 45-70.

Frankel, R. & Lee, C. M. (1998). Accounting Valuation, Market Expectation, and Cross-sectional Stock Returns. *Journal of Accounting and Economics, 25*, 283-319.

Freeman, R. N. Ohlson, J. A., & Penman, S. H. (1982). Book Rate-of-Return and Prediction of Earnings Changes: An Empirical Investigation. *Journal of Accounting Research, 20*, 639-653.

French, K. R. (2016). *Industry definitions [Data file]*. Retrieved August 25, 2016, from Kenneth R. French's Data Library: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/changes_ind.html

Greig, A. C. (1992). Fundamental Analysis and Subsequent Stock Returns. *Journal of Accounting and Economics, 15*, 413-442.

Holthausen, R. W. & Larcker, D. F. (1992). The Prediction of Stock Returns Using Financial Statement Information. *Journal of Accounting and Economics, 15*, 373-411.

Jensen, M. C. (1968). The Performance of Mutual Funds in the Period 1945-1964. *The Journal of Finance, 23*, 389-416.

Jensen, M. C. (1978). Some Anomalous Evidence Regarding Market Efficiency. *Journal of Financial Economics, 6*, 95-101.

Jorgensen, B. N., Lee, Y. G. & Yoo, Y. K. (2011). The Valuation Accuracy of Equity Value Estimates Inferred from Conventional Empirical Implementations of the Abnormal Earnings Growth Model: US Evidence. *Journal of Business Finance & Accounting, 38*, 446-471.

Kim, M. & Ritter, J. R. (1999). Valuing IPOs. *Journal of Financial Economics, 53*, 409-437.

Koller, T., Goedhart, M. & Wessels, D. (2010). *Valuation. Measuring and Managing the Value of Companies (5th ed.)*. Hoboken: John Wiley and Sons.

Kothari, S. P. (2001). Captial Market Research in Accounting. *Journal of Accounting and Economics, 31*, 105-231.

Lakonishok, J., Shleifer, A. & Vishny, R. W. (1994). Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance, 49*, 1541-1578.

Lee, C. M. (1999). Accounting-Based Valuation: Impact on Business Practices and Research. *Accounting Horizons, 13*, 413-425.

Lee, C. M., Myers, J. & Swaminathan, B. (1999). What is the Intrinsic Value of the Dow? *The Journal of Finance, 54*, 1693-1741.

Lev, B. & Thiagarajan, S. R. (1993). Fundamental Information Analysis. *Journal of Accounting Research, 31*, 190-215.

Lintner, J. (1956). Distribution of Incomes of Corporations among Dividends, Retained Earnings, and Taxes. *The American Economic Review, 46*, 97-113.

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains from Diversification. *The Journal of Finance, 20*, 587-615.

Liu, J., Nissim, D. & Thomas, J. (2002). Equity Valuation Using Mulitples. *Journal of Accounting Research, 40*, 135-172.

Lo, A. W. & MacKinlay, A. C. (1990). Data-Snooping Biases in Tests of Financial Asset Pricing Models. *The Review of Financial Studies, 3*, 431-467.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives, 17*, 59-82.

Mohanram, P. S. (2005). Separating Winners from Losers Among Low Book-to-Market Stocks Using Financial Statement Analysis. *Review of Accounting Studies, 10*, 133-170.

Myers, J. N. (1999). Implementing Residual Income Valuation with Linear Information Dynamics. *The Accounting Review, 74*, 1-28.

NBER. (2010). *US Business Cycle Expansions and Contractions*. Retrieved November 1, 2016, from The National Bureau of Economic Research: http://www.nber.org/cycles.html

Nissim, D. & Penman, S. H. (2001). Ratio Analysis and Equity Valuation: From Research to Practice. *Review of Accounting Studies, 6*, 109-154.

Norrman, A. & Rahmn, J. (2016). *Valuation Accuracy of the RIV Model*. Stockholm School of Economics, Stockholm.

Ohlson, J. A. (1995). Earnings, Book Values, and Dividends in Equity Valuation. *Contemporary Accounting Research, 11*, 661-687.

Ou, J. A. & Penman, S. H. (1989). Financial Statement Analysis and the Prediction of Stock Returns. *Journal of Accounting and Economics, 11*, 295-329.

Penman, S. H. (1991). An Evaluation of Accounting Rate-of-Return. *Journal of Accounting, Auditing & Finance, 6*, 233-255.

Penman, S. H. (1992). Return to Fundamentals. *Journal of Accounting, Auditing & Finance, 7*, 465-483.

Penman, S. H. (1996). The Articulation of Price-Earnings Ratios and Market-to-Book Ratios and the Evaluation of Growth. *Journal of Accounting Research, 34*, 235-259.

Penman, S. H. (1998). A Synthesis of Equity Valuation Techniques and the Terminal Value Calculation for the Dividend Discount Model. *Review of Accounting Studies, 2*, 303-323.

Penman, S. H. (2005). Discussion of ''On Accounting-Based Valuation Formulae'' and ''Expected EPS and EPS Growth as Determinants of Value''. *Review of Accounting Studies, 10*, 367-378.

Penman, S. H. (2012). *Financial Statement Analysis and Security Valuation*. McGraw-Hill Education.

Penman, S. H. & Sougiannis, T. (1998). A Comparison of Dividend, Cash Flow, and Earnings Approaches to Equity Valuation. *Contemporary Accounting Research, 15*, 343-383.

Penman, S. H. & Zhang, X.-J. (2006). Modeling Sustainable Earnings and P/E Ratios Using Financial Statement Information. *Working Paper*.

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.

Preinreich, G. A. (1938). Annual Survey of Economic Theory: The Theory of Depreciation. *Econometrica, 6*, 219-241.

Richardson, S., Tuna, I. & Wysocki, P. (2010). Accounting Anomalies and Fundamental Analysis: A Review of Recent Research Advances. *Journal of Accounting and Economics, 50*, 410-454.

Roberts, H. V. (1959). Stock-Market "Patterns" And Financial Analysis: Methodological Suggestions. *The Journal of Finance, 14*, 1-10.

Runsten, M. (1998). The Association between Accounting Information and Stock Prices. *Ph. D. Stockholm School of Economics*.

Schwert, G. W. (2003). Anomalies and Market Efficiency. In *Handbook of the Economics of Finance* (Vol. 1, pp. 939-974).

Setiono, B. & Strong, N. (1998). Predicting Stock Returns Using Financial Statement Information. *Journal of Business Finance & Accounting , 25*, 631-657.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance, 19*, 425-442.

Skogsvik, K. (1990). Current Cost Accounting Ratios as Predictors of Business Failure: The Swedish Case. *Journal of Business, Finance & Accounting, 17*, 137-160.

Skogsvik, K. (1998). Conservative Accounting Principles, Equity Valuation and the Importance of Voluntary Disclosures. *The British Accounting Review, 30*, 361-381.

Skogsvik, K. (2002). A Tutorial on Residual Income Valuation and Value Added Valuation. *SSE/EFI Working Paper Series in Business Administration (SWOBA)*.

Skogsvik, K. (2005). On the Choice-Based Sample Bias in Probabilistic Business Failure Prediction. *SSE/EFI Working Paper Series in Business Administradion (SWOBA)*.

Skogsvik, S. (2008). Financial Statement Information, the Prediction of Book Return on Owners' Equity and Market Efficiency: The Swedish Case. *Journal of Business Finance & Accounting, 35*, 795-817.

Skogsvik, S. & Skogsvik, K. (2010). Accounting-Based Probabilisitc Prediction of ROE, the Residual Income Valuation Model and the Assessment of Mispricing in the Swedish Market. *ABACUS, 46*, 387-418.

Stambaugh, R. F. Yu, J. & Yuan, Y. (2012). The Short of it: Investor Sentiment and Anomalies. *Journal of Financial Economics, 104*, 288-302.

Stober, T. L. (1992). Summary Financial Statement Measures and Analysts' Forecasts of Earnings. *Journal of Accounting and Economics, 15*, 347-372.

10. APPENDIX

APPENDIX A

RIV model input parameters and underlying assumptions

1) *Required rate of return on owners' equity* $(\boldsymbol{r_{E,i}})$

The CAPM is applied to determine the firm-specific cost of owners' equity as follows:

$$r_{E,i} = r_f + [E(r_M) - r_f] \cdot \beta_i \tag{A.1}$$

The risk-free rate is approximated by the 10-year U.S. treasury bond rate at each valuation date, i.e. the prior fiscal year-end date (Koller, Goedhart & Wessels, 2010). Estimates for the market risk premium are derived from studies of long-term annualized U.S. equity excess returns over the period 1900-2015 (Dimson, Marsh & Staunton, 2016) and recent survey data provided by Fernandez, Ortiz & Acín (2016). Accordingly, $[E(r_M) - r_f]$ is set to 5%. In addition, β_i is determined for each sample firm by regressing 48 months of historical adjusted stock return data over the corresponding market index returns (Skogsvik & Skogsvik, 2010). The market index used is the S&P 500 Composite Index. Stock returns are adjusted for distributions and stock splits to reflect actual ex-post realized returns.

2) *Dividend payout share* $(\boldsymbol{DS_{i,t}})$

Our definition of the dividend payout share draws from Penman's (2012) observation of dividends being paid out of financial assets and ensures a smooth distribution measure:

$$DS_{i,t} = \frac{Div_{i,t}}{BV_{i,t-1}} \tag{A.2}$$

Instead of a three-year historical average of $DS_{i,t}$ (Skogsvik & Skogsvik, 2010), the most recent period's dividend share serves as a proxy for the forecast period. This is motivated by the robust empirical 'stickiness' of dividends (Lintner, 1956) and data availability constraints.

3) *Book value of owners' equity* $(\boldsymbol{BV_{i,t}})$

To satisfy the CSR condition (Ohlson, 1995), the book value of owners' equity is defined as comprehensive stockholders' equity ('SEQ' in Compustat) including ordinary and preferred stock capital. Historical $BV_{i,t}$ correspond to SEQ values at the previous period's fiscal year-end date. To derive the future value of $BV_{i,T}$, the following CSR transformation is applied:

$$BV_{i,t+1} = BV_{i,t} + \left(\overline{ROE}^{h}_{i,t+1} \cdot BV_{i,t}\right) - (DS_{i,t+1} \cdot BV_{i,t}) \tag{A.3}$$

APPENDIX B

TABLE 9: Data sample reduction and selection criteria for the six estimation subperiods

Estimation periods	Subperiod I ***1970-1975***	Subperiod II ***1976-1981***	Subperiod III ***1982-1987***	Subperiod IV ***1988-1993***	Subperiod V ***1994-1999***	Subperiod VI ***2000-2005***
Number of firm-year observations available	9429	11110	10541	10580	10449	8618
To be excluded						
Negative equity	0	235	242	376	405	340
Data unavailable for 12 years	4785	4731	5607	5512	5928	4390
Number of firm-year observations incl. in calculation of ROE changes	4644	6144	4692	4692	4116	3888
Number of firm-year observations with calculated ROE changes	2287	3024	2301	2292	2021	1904
To be excluded						
Extreme medium-term ROE values (>+100% or <-100%)	4	6	11	9	24	12
Number of observations used in regression	2283	3018	2290	2283	1997	1892
(Number of firms)	*(386)*	*(512)*	*(391)*	*(390)*	*(343)*	*(324)*

Notes: Table 9 presents the sample selection process and underlying criteria applied to the total industry firm-year population for the logistic regressions conducted over the six estimation periods.

APPENDIX B (continued)

TABLE 10: Data sample reduction and selection criteria for the 18 investment years

Investment periods	Subperiod I			Subperiod II			Subperiod III			Subperiod IV			Subperiod V			Subperiod VI		
	1979	***1980***	***1981***	***1985***	***1986***	***1987***	***1991***	***1992***	***1993***	***1997***	***1998***	***1999***	***2003***	***2004***	***2005***	***2009***	***2010***	***2011***
Number of firm-year observations available for defined SIC codes	3809	3723	3533	3567	3544	3416	3425	3361	3202	3772	3871	3759	3064	2875	2650	2289	2204	2069
To be excluded																		
Fiscal year-end ≠ calendar year-end	1640	1626	1492	1634	1639	1527	1598	1555	1382	1531	1538	1396	1066	974	840	722	691	611
Negative equity	36	28	28	16	21	27	73	78	85	101	111	112	100	99	93	81	75	65
Data unavailable for 4 years	229	257	257	305	360	394	362	320	307	508	534	567	470	346	297	330	286	241
Number of firm-year observations incl. in calculation of historical ROE	1904	1812	1756	1612	1524	1468	1392	1408	1428	1632	1688	1684	1428	1456	1420	1156	1152	1152
Number of firm-year observations with calculated historical ROE	476	453	439	403	381	367	348	352	357	408	422	421	357	364	355	289	288	288
To be excluded																		
Extreme medium-term ROE values (>+100% or <-100%)	2	3	2	4	2	6	4	9	7	9	13	10	3	2	7	6	11	10
Stock price not available or equal to zero at investment point in time	9	6	3	3	2	3	0	0	2	1	1	1	2	4	3	1	2	0
Number of firms allocated in base-case strategy	465	444	434	396	377	358	344	343	348	398	408	410	352	358	345	282	275	278
Number of firms allocated in indicator variable strategy	205	160	199	194	162	154	157	151	129	154	185	111	186	85	132	100	123	112

Notes: Table 10 presents the sample selection process and underlying criteria applied to the total industry firm-year population for the 18 investment years with the resultant allocations.

APPENDIX C

TABLE 11: Normalized prices, intrinsic values and indicator variables (annual median)

Year	*Number of firms*	V_0/BV_0		P_0/BV_0		IND_0/BV_0	
		Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*
1979	296	1.019	0.904	1.093	0.959	0.074	0.039
1980	300	1.145	0.969	1.031	0.809	-0.114	-0.128
1981	264	0.975	0.910	1.554	1.127	0.579	0.189
1985	305	1.495	1.197	1.594	1.324	0.099	0.071
1986	313	1.560	1.345	1.939	1.612	0.379	0.250
1987	269	1.865	1.495	2.183	1.703	0.318	0.138
1991	249	1.908	1.439	2.184	1.446	0.276	-0.058
1992	264	1.915	1.256	2.374	1.651	0.459	0.193
1993	277	2.066	1.503	4.001	1.776	1.936	0.172
1997	280	2.840	2.140	2.520	1.950	-0.320	-0.080
1998	298	2.709	2.173	3.225	2.487	0.516	0.289
1999	300	2.959	2.248	2.063	1.534	-0.896	-0.700
2003	281	2.181	1.595	1.766	1.285	-0.415	-0.267
2004	275	1.991	1.259	2.842	2.010	0.851	0.500
2005	276	3.208	2.052	2.959	2.087	-0.248	0.038
2009	247	2.531	2.147	1.543	1.041	-0.988	-0.840
2010	246	1.541	1.121	2.356	1.922	0.815	0.615
2011	244	2.149	1.747	2.950	2.347	0.801	0.451
All years		2.000	1.438	2.220	1.549	0.220	0.072

Notes: Table 11 shows arithmetic mean and median values of price (P_0), intrinsic value (V_0) and the indicator variable (IND_0) for each investment year, scaled by the book value of owners' equity (BV_0). P_0 is the ex-dividend market price of equity, V_0 denotes the ex-dividend intrinsic value estimate and IND_0 is the difference between P_0 and V_0 at the investment date t = 0.

APPENDIX D

TABLE 12: Stock allocations to investment positions

Year	*Number of investment positions (BC)*			*Number of investment positions (IND)*		
	Total	*Long*	*Short*	*Total*	*Long*	*Short*
1979	465	140	325	205	57	148
1980	444	122	322	160	67	93
1981	434	119	315	199	30	169
1985	396	285	111	194	127	67
1986	377	250	127	162	74	88
1987	358	244	114	154	96	58
1991	344	159	185	157	80	77
1992	343	193	150	151	66	85
1993	348	218	130	129	68	61
1997	398	132	266	154	56	98
1998	408	135	273	185	35	150
1999	410	157	253	111	86	25
2003	352	264	88	186	164	22
2004	358	284	74	85	44	41
2005	345	238	107	132	96	36
2009	282	106	176	100	79	21
2010	275	129	146	123	18	105
2011	278	145	133	112	21	91
All years	6615	3320	3295	2699	1264	1435

Notes: Table 12 shows the number of stocks allocated to the long and short positions respectively and overall for each investment year. Numbers for the base case strategy (BC) and the indicator variable strategy (IND) are reported separately.

APPENDIX E

TABLE 13: Monthly three-factor excess returns

Panel A: 'Perfect foreknowledge' scenario

Time period	*Number of obs.*	*Position*	*Base case* α	β^M	β^{SMB}	β^{HML}	*Indicator variable* α	β^M	β^{SMB}	β^{HML}
1979-2011	648	Long	0.002	1.080	0.077	0.131	0.006	1.035	0.033	0.136
		(p-value)	*(0.171)*	*(0.000)*	*(0.299)*	*(0.109)*	*(0.001)*	*(0.000)*	*(0.513)*	*(0.016)*
		Short	-0.010	0.938	0.038	0.136	-0.010	0.942	0.047	0.142
		(p-value)	*(0.000)*	*(0.000)*	*(0.441)*	*(0.014)*	*(0.000)*	*(0.000)*	*(0.439)*	*(0.034)*
		Hedge	0.012	0.142	0.038	-0.004	0.016	0.092	-0.013	-0.006
		(p-value)	*(0.000)*	*(0.015)*	*(0.607)*	*(0.958)*	*(0.000)*	*(0.080)*	*(0.845)*	*(0.941)*
1979-1993	324	Long	0.004	0.997	0.336	0.085	0.009	1.058	-0.043	-0.031
		(p-value)	*(0.175)*	*(0.000)*	*(0.096)*	*(0.624)*	*(0.000)*	*(0.000)*	*(0.687)*	*(0.737)*
		Short	-0.011	0.938	0.058	0.088	-0.011	1.010	0.233	0.052
		(p-value)	*(0.000)*	*(0.000)*	*(0.588)*	*(0.345)*	*(0.000)*	*(0.000)*	*(0.049)*	*(0.609)*
		Hedge	0.015	0.058	0.277	-0.003	0.020	0.049	-0.276	-0.083
		(p-value)	*(0.001)*	*(0.579)*	*(0.182)*	*(0.988)*	*(0.000)*	*(0.496)*	*(0.051)*	*(0.495)*
1997-2011	324	Long	0.000	1.141	0.005	0.184	0.003	1.014	0.044	0.232
		(p-value)	*(0.981)*	*(0.000)*	*(0.918)*	*(0.003)*	*(0.136)*	*(0.000)*	*(0.450)*	*(0.001)*
		Short	- 0.009	0.928	0.028	0.167	-0.009	0.827	-0.029	0.226
		(p-value)	*(0.000)*	*(0.000)*	*(0.615)*	*(0.016)*	*(0.002)*	*(0.000)*	*(0.693)*	*(0.012)*
		Hedge	0.009	0.214	-0.023	0.017	0.012	0.186	0.073	0.006
		(p-value)	*(0.000)*	*(0.000)*	*(0.625)*	*(0.768)*	*(0.000)*	*(0.019)*	*(0.353)*	*(0.950)*

Panel B: ROE prediction model scenario

Time period	*Number of obs.*	*Position*	*Base case* α	β^M	β^{SMB}	β^{HML}	*Indicator variable* α	β^M	β^{SMB}	β^{HML}
1979-2011	648	Long	-0.004	0.999	0.060	0.200	-0.003	0.947	0.122	0.303
		(p-value)	*(–)*	*(0.000)*	*(0.242)*	*(0.000)*	*(–)*	*(0.000)*	*(0.073)*	*(0.000)*
		Short	-0.004	1.019	0.003	0.057	-0.004	1.001	-0.066	0.063
		(p-value)	*(0.003)*	*(0.000)*	*(0.941)*	*(0.230)*	*(0.007)*	*(0.000)*	*(0.117)*	*(0.176)*
		Hedge	0.001	-0.020	0.057	0.144	0.001	-0.054	0.188	0.239
		(p-value)	*(0.756)*	*(0.591)*	*(0.238)*	*(0.007)*	*(0.364)*	*(0.292)*	*(0.005)*	*(0.001)*
1979-1993	324	Long	-0.003	0.944	0.033	0.134	-0.004	0.969	0.112	0.284
		(p-value)	*(–)*	*(0.000)*	*(0.761)*	*(0.155)*	*(–)*	*(0.000)*	*(0.455)*	*(0.029)*
		Short	-0.005	1.048	0.055	-0.055	-0.004	1.053	0.029	-0.064
		(p-value)	*(0.032)*	*(0.000)*	*(0.566)*	*(0.506)*	*(0.023)*	*(0.000)*	*(0.751)*	*(0.409)*
		Hedge	0.002	-0.104	-0.022	0.189	0.000	-0.083	0.083	0.348
		(p-value)	*(0.236)*	*(0.085)*	*(0.856)*	*(0.066)*	*(0.497)*	*(0.293)*	*(0.594)*	*(0.010)*
1997-2011	324	Long	-0.005	1.066	0.073	0.232	-0.001	0.917	0.119	0.315
		(p-value)	*(–)*	*(0.000)*	*(0.216)*	*(0.001)*	*(–)*	*(0.000)*	*(0.112)*	*(0.001)*
		Short	-0.004	0.967	-0.025	0.134	-0.003	0.914	-0.113	0.159
		(p-value)	*(0.029)*	*(0.000)*	*(0.582)*	*(0.018)*	*(0.107)*	*(0.000)*	*(0.018)*	*(0.007)*
		Hedge	-0.002	0.098	0.098	0.098	0.001	0.003	0.233	0.156
		(p-value)	*(–)*	*(0.027)*	*(0.027)*	*(0.071)*	*(0.340)*	*(0.964)*	*(0.001)*	*(0.055)*

Notes: Table 13 presents the monthly three-factor excess returns (α) and the corresponding coefficients for the market (β^M), size (β^{SMB}) and book-to-market (β^{HML}) factors. The results proceed from the regressions specified in 3.3.2 and are reported for both the base case and the indicator variable strategy under the 'perfect foreknowledge' (Panel A) and ROE prediction model (Panel B) scenario. *P-values* are reported in parenthesis, unless the sign of α is inconsistent with the alternative hypothesis. For long and hedge positions, the null hypothesis of non-positive α is tested against the alternative of a positive α. For short positions, the null hypothesis of a non-negative α is tested against the alternative of a negative α. Tests of β-values are two-tailed.

APPENDIX F

TABLE 14: Ou & Penman (1989) calibration test results

Panel A: Prediction accuracy assessment

Investment periods	Subperiod I ***1973-1977***	Subperiod II ***1978-1983***	All years ***1973-1983***
Number of observations	6667	10257	16924
Increases	3869	4584	8453
Decreases	2798	5673	8471
Correctly predicted (%)	62.56%	63.22%	62.96%
Increases	57.84%	43.39%	50.01%
Decreases	69.09%	79.23%	75.88%

Panel B: Monthly CAPM excess returns

Time period	*Number of obs.*	*Position*	*ROE Prediction Model*		*Perfect Foreknowledge*	
			α	β^M	α	β^M
1973 - 1983	396	Long	0.007	1.270	0.014	1.269
		(p-value)	*(0.032)*	*(0.000)*	*(0.000)*	*(0.000)*
		Short	-0.002	1.115	-0.005	1.153
		(p-value)	*(0.181)*	*(0.000)*	*(0.049)*	*(0.000)*
		Hedge	0.009	0.155	0.019	0.116
		(p-value)	*(0.015)*	*(0.101)*	*(0.000)*	*(0.195)*

Notes: Table 14 presents the prediction accuracy results (Panel A) and the monthly CAPM excess returns and beta coefficients (Panel B) based on the investment period 1973-1983. In line with the Ou & Penman (1989) sample, all listed U.S. companies available in the Compustat and CRSP databases are included in the estimation and investment periods, regardless of their industry classification. In Panel A, the accuracy of two ROE prediction models is assessed. The first (second) is estimated based on years 1964 to 1972 (1969 to 1977) for the investment period 1973-1977 (1978-1983). For the results in Panel B, investment decisions are made solely on the basis of the ROE prediction models and historical medium-term ROE. This procedure follows the base case strategy specified in 3.1 and differs from Ou & Penman's (1989) one-year forecasting procedure based on the summary measure 'Pr'. The cut-off value for the probability of an increase in medium-term ROE remains at 0.5, which differs from Ou & Penman's (1989) 0.4 and 0.6 cut-off values. Estimates of α and β^M are reported in Panel B. For long and hedge positions, the null hypothesis of non-positive α is tested against the alternative of a positive α. For short positions, the null hypothesis of a non-negative α is tested against the alternative of a negative α. Tests of β-values are two-tailed.

APPENDIX G

TABLE 15: ROE mean reversion test results

Panel A: Mean reversion period 1979-1990

Mean historical medium-term ROE performance (Quartiles)	*Q(I)*	*Q(II)*	*Q(III)*	*Q(IV)*
$\overline{ROE}_{1976-1978}$	0.258	0.170	0.131	0.057
$\overline{ROE}_{1979-1981} - \overline{ROE}_{1976-1978}$	-0.049	-0.006	0.013	0.060
$\overline{ROE}_{1982-1984} - \overline{ROE}_{1976-1978}$	-0.098	-0.064	-0.051	0.059
$\overline{ROE}_{1985-1987} - \overline{ROE}_{1976-1978}$	-0.150	-0.060	-0.047	0.009
$\overline{ROE}_{1988-1990} - \overline{ROE}_{1976-1978}$	-0.093	-0.027	-0.013	0.026

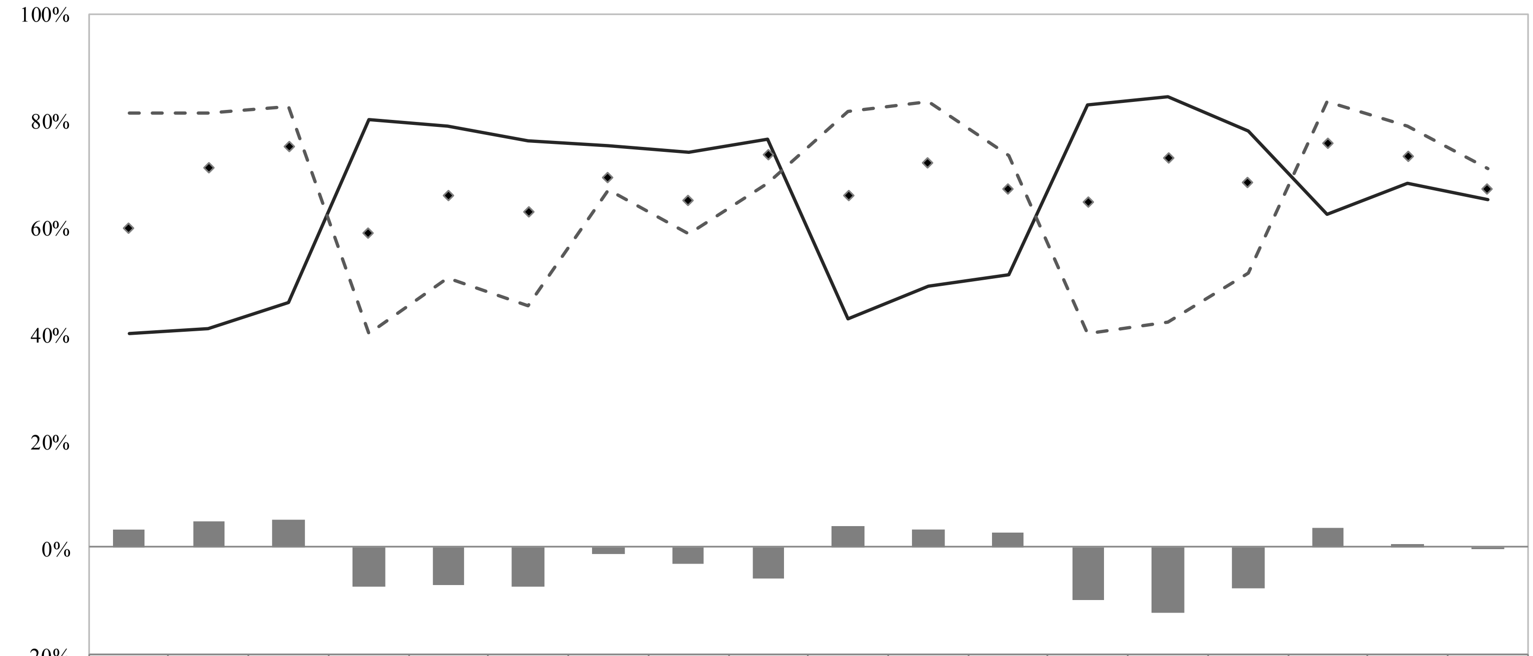
Panel B: Mean reversion period 2003-2014

Mean historical medium-term ROE performance (Quartiles)	*Q(I)*	*Q(II)*	*Q(III)*	*Q(IV)*
$\overline{ROE}_{2000-2002}$	0.257	0.117	0.059	-0.066
$\overline{ROE}_{2003-2005} - \overline{ROE}_{2000-2002}$	-0.001	0.067	0.037	0.155
$\overline{ROE}_{2006-2008} - \overline{ROE}_{2000-2002}$	-0.032	0.041	0.093	0.216
$\overline{ROE}_{2009-2011} - \overline{ROE}_{2000-2002}$	-0.055	-0.032	0.114	0.156
$\overline{ROE}_{2012-2014} - \overline{ROE}_{2000-2002}$	-0.035	0.058	0.186	0.181

Notes: Table 15 shows the changes in medium-term ROE ($\overline{ROE}$) as the difference between future medium-term ROE and historical medium-term ROE. Tests regarding the actual change in $\overline{ROE}$ are conducted for the first investment year (Panel A) and at the end of the investment period (Panel B). Firms included in the tests correspond to our sample definition. The firms are divided into quartiles based on their historical medium-term ROE at the investment point in time. For each firm, four future medium-term ROEs are calculated based on data retrieved from Compustat for the years 1975 to 1990 and 1999 to 2014. The first ROE figure calculated is the future medium-term ROE as defined by the model logic – with the investment point in time as the starting point for the three-year range. Additional future medium-term ROEs are calculated three, six, and nine years after the initial investment year. Given the need for twelve consecutive data points, the years 1979 and 2003 are selected.

APPENDIX H

FIGURE 3: Impact of changing historical medium-term ROE levels (mean) on the prediction performance



	1979	1980	1981	1985	1986	1987	1991	1992	1993	1997	1998	1999	2003	2004	2005	2009	2010	2011
Δ Hist. ROE (mean)	3.4%	4.9%	5.1%	-7.2%	-7.1%	-7.5%	-1.2%	-3.2%	-5.9%	4.1%	3.5%	2.7%	-9.9%	-12.2%	-7.8%	3.7%	0.7%	-0.2%
PA - Overall	59.9%	71.2%	75.3%	58.9%	66.1%	62.9%	69.3%	65.1%	73.6%	66.1%	72.1%	67.2%	64.9%	73.1%	68.5%	75.7%	73.3%	67.3%
PA - Increases	40.0%	41.0%	46.1%	80.3%	78.9%	76.3%	75.3%	74.2%	76.7%	43.0%	49.0%	51.2%	83.0%	84.6%	78.2%	62.6%	68.5%	65.1%
PA - Decreases	81.4%	81.5%	82.6%	40.0%	50.4%	45.2%	67.1%	58.8%	68.2%	81.7%	83.5%	73.4%	40.2%	42.3%	51.6%	83.8%	79.1%	71.0%

Notes: Figure 3 illustrates the relationship between the changes in the arithmetic mean historical medium-term ROE (between each investment year and the corresponding estimation period) and the subsequent prediction model performance in terms of the prediction accuracy for increases (PA – Increases) and decreases (PA – Decreases) in medium-term ROE and overall (PA – Overall).

APPENDIX I

TABLE 16: Misallocation impact test results

Investm. Year	Weighted return impact (Quartiles)	Long misallocations			Short misallocations		
		Weighted return impact	$\hat{\rho}(\Delta(\overline{ROE}) \geq 0)^*$		Weighted return impact	$\hat{\rho}(\Delta(\overline{ROE}) \geq 0)^*$	
			Mean	Median		Mean	Median
1979	Q(I)	-0.099	0.658	0.609	0.226	0.335	0.328
	Q(II)	-0.015	0.642	0.538	0.044	0.362	0.380
	Q(III)	-0.002	0.611	0.605	0.011	0.359	0.369
	Q(IV)	0.081	0.611	0.593	-0.021	0.389	0.410
1985	Q(I)	-0.341	0.753	0.746	0.244	0.407	0.401
	Q(II)	-0.005	0.713	0.703	0.087	0.402	0.430
	Q(III)	0.010	0.639	0.609	0.045	0.407	0.414
	Q(IV)	0.068	0.683	0.627	0.015	0.402	0.438
1991	Q(I)	-0.241	0.696	0.658	0.025	0.387	0.378
	Q(II)	-0.003	0.691	0.644	0.008	0.362	0.336
	Q(III)	0.019	0.618	0.573	-0.001	0.443	0.441
	Q(IV)	0.126	0.564	0.533	-0.015	0.255	0.224
1997	Q(I)	-0.198	0.618	0.567	0.079	0.353	0.377
	Q(II)	-0.011	0.688	0.704	0.022	0.326	0.287
	Q(III)	0.038	0.663	0.609	0.003	0.376	0.414
	Q(IV)	0.189	0.606	0.595	-0.065	0.380	0.385
2003	Q(I)	-0.012	0.712	0.640	0.333	0.388	0.430
	Q(II)	0.031	0.685	0.703	0.145	0.400	0.402
	Q(III)	0.090	0.664	0.601	0.056	0.392	0.438
	Q(IV)	0.387	0.629	0.624	0.002	0.359	0.438
2009	Q(I)	-0.013	0.733	0.767	0.191	0.271	0.257
	Q(II)	0.034	0.679	0.643	0.075	0.276	0.276
	Q(III)	0.084	0.601	0.591	0.058	0.324	0.324
	Q(IV)	0.261	0.617	0.613	0.010	0.431	0.431

Notes: Table 16 illustrates the impact of short and long misallocations on the generated returns for 6 of the 18 investment years. Long (short) misallocations are defined as allocations in which the future medium-term ROE decreases (increases) while an increase (decrease) was predicted by the model. The firms that were falsely allocated are divided into quartiles based on the magnitude of their adverse impact on the respective position, i.e. firms with positive (negative) return impact in the short (long) position Additionally, the table presents arithmetic averages and medians of the adjusted probability of medium-term ROE increases ($\hat{\rho}(\Delta(\overline{ROE}) \geq 0)^*$) for each quartile. The weighted return impact is measured as the sum of the value-weighted 36-month returns of the misallocated stocks in the respective quartile, for the long and short position respectively:

$$\bar{r}_{MS,t}^{WRI} = \frac{1}{\sum_{i=1}^{n_{TS}} S_{i,0}} \cdot \sum_{i=1}^{n_{MS}} S_{i,0} \cdot \left[\prod_{x=1}^{36} (1 + r_{i,x})\right] \tag{I.1}$$

where $\bar{r}_{MS,t}^{WRI}$ = weighted return impact (WRI) of misallocated stocks (MS) for year t

$S_{i,x}$ = adjusted closing price of stock i at the end of month x

$r_{i,x}$ = adjusted return on stock i for month x

$n_{(.)}$ = number of total stocks (TS) or misallocated stocks (MS) in the respective position

APPENDIX I (continued)

TABLE 17: Impact of delisted firms on prediction accuracy

Panel A: Prediction accuracy including delisted firms

Investment years	*1979*	*1985*	*1991*	*1997*	*2003*	*2009*	*Total*
Number of observations	205	194	157	154	186	100	996
Increases	57	127	80	56	164	79	563
Decreases	148	67	77	98	22	21	433
Correctly predicted (%)	41.46%	44.33%	57.32%	43.51%	48.92%	57.00%	47.79%
Increases	49.12%	40.94%	32.50%	33.93%	50.00%	56.96%	44.76%
Decreases	38.51%	50.75%	83.12%	48.98%	40.91%	57.14%	51.73%

Panel B: Prediction accuracy excluding delisted firms

Investment years	*1979*	*1985*	*1991*	*1997*	*2003*	*2009*	*Total*
Number of observations	177	159	143	116	152	83	830
Increases	48	102	71	34	133	64	452
Decreases	129	57	72	82	19	19	378
Correctly predicted (%)	48.02%	54.09%	62.94%	57.76%	59.87%	68.67%	57.35%
Increases	58.33%	50.98%	36.62%	55.88%	61.65%	70.31%	55.75%
Decreases	44.19%	59.65%	88.89%	58.54%	47.37%	63.16%	59.26%

Notes: Table 17 exhibits the prediction accuracy and corresponding number of observations for the first year of each of the six investment periods respectively. In Panel A, the prediction accuracy is defined as the proportion of correctly predicted medium-term ROE ($\overline{ROE}$) increases, decreases and overall changes based on ex-post realizations including delisted firms. Panel B presents the equivalent results under the assumption, that delisted firms are excluded from the prediction accuracy assessment.

TABLE 18: Impact of delisted firms on weighted return

Inv. Year	*Long misallocations*				*Short misallocations*			
	Misallocation Cause		*Weighted return impact*		*Misallocation Cause*		*Weighted return impact*	
	Prediction	*Delisting*	*Prediction*	*Delisting*	*Prediction*	*Delisting*	*Prediction*	*Delisting*
1979	68.97%	31.03%	-0.091	0.056	79.12%	20.88%	0.171	0.089
1985	66.67%	33.33%	-0.034	-0.233	69.70%	30.30%	0.182	0.208
1991	83.33%	16.67%	-0.099	-0.001	61.54%	38.46%	0.001	0.017
1997	40.54%	59.46%	-0.115	0.133	68.00%	32.00%	0.006	0.033
2003	62.20%	37.80%	0.350	0.147	76.92%	23.08%	0.448	0.088
2009	55.88%	44.12%	0.167	0.199	77.78%	22.22%	0.284	0.050

Notes: Table 18 compares the share of misallocations due to false predictions to the share of misallocations due to delistings over the holding period. Additionally, the weighted return impact for each of the two misallocation causes is depicted for the long and short position separately. The weighted return impact is measured as the sum of the value-weighted 36-month returns of the misallocated stocks in the respective quartile, for the long and short position respectively (see Eq. (I.1) in Table 16).

APPENDIX J

TABLE 19: Sentiment bias test – Sample statistics and 36-month market-adjusted returns

Year	*Number of firms*	$\hat{\rho}\left(\Delta(\overline{ROE}) \geq 0\right)^*$		IND_0/BV_0			*Investment positions*		*Market-adjusted 36-month returns*		
		> 0.5	*< 0.5*	*> 0*	*= 0*	*< 0*	*Long*	*Short*	*Long*	*Short*	*Hedge*
1979	416	215	201	95	104	61	71	76	0.011	-0.263	0.274
1980	395	101	294	45	70	149	54	72	0.680	-0.536	1.216
1981	382	77	305	155	57	22	14	157	0.835	-0.412	1.247
1985	305	143	162	108	73	61	58	89	0.500	-0.518	1.019
1986	278	152	126	151	44	40	33	83	0.395	-0.367	0.762
1987	271	152	119	104	48	55	52	53	0.380	-0.525	0.905
1991	306	82	224	73	55	96	32	89	0.386	-0.287	0.673
1992	306	126	180	141	50	43	33	110	0.864	-0.424	1.288
1993	305	193	112	142	37	58	45	47	0.105	-0.414	0.519
1997	306	124	182	74	40	98	52	56	-0.977	-1.462	0.484
1998	307	101	206	133	29	53	19	100	0.026	-0.569	0.595
1999	304	85	219	30	10	188	43	19	-0.126	-0.139	0.014
2003	278	159	119	43	40	143	99	26	1.207	0.011	1.196
2004	288	208	80	179	19	28	33	39	0.423	-0.115	0.538
2005	264	165	99	101	27	88	65	35	0.452	-0.404	0.857
2009	244	92	152	20	10	183	69	15	1.263	-0.120	1.384
2010	249	132	117	172	15	28	17	78	0.272	-0.278	0.550
2011	249	153	96	169	18	33	22	61	-0.153	-0.325	0.172
All years		2460	2993	1935	746	1427	811	1205	0.364	-0.397	0.761

Notes: Table 19 shows the firm distribution across the two primary investment criteria as well as the number of investment positions taken and the market-adjusted returns for each investment year. The underlying sample corresponds to the one specified under the indicator variable strategy in the 'perfect foreknowledge' scenario. The firm distribution is shown for the two dimensions of the probability of medium-term ROE increases ($\hat{\rho}(\Delta(\overline{ROE}) \geq 0)^*$), i.e. probabilities above and below 0.5, as well as the three dimensions for the scaled indicator variable (IND_0/BV_0), i.e. values above, below and equal to zero. The market-adjusted returns are calculated over the respective 36-month holding period for each portfolio stock and investment year. Long and short returns are computed as the value-weighted average of the individual stock returns in each position:

$$\bar{r}_{L,t}^{MJBH} = \frac{1}{\sum_{i=1}^{n_L} S_{i,0}} \cdot \sum_{i=1}^{n_L} S_{i,0} \cdot \left[\prod_{x=1}^{36}(1 + r_{i,x}) - \prod_{x=1}^{36}(1 + r_{m,x})\right] \quad \text{(K.1)}$$

$$\bar{r}_{S,t}^{MJBH} = \frac{1}{\sum_{i=1}^{n_S} S_{i,0}} \cdot \sum_{i=1}^{n_S} S_{i,0} \cdot \left[\prod_{x=1}^{36}(1 + r_{i,x}) - \prod_{x=1}^{36}(1 + r_{m,x})\right] \quad \text{(K.2)}$$

where $\bar{r}_{L,t}^{MJBH}$ = average market-adjusted return to the long position for year t
$\bar{r}_{S,t}^{MJBH}$ = average market-adjusted return to the short position for year t
$S_{i,x}$ = adjusted closing price of stock i at the end of month x
$r_{i,x}$ = adjusted return on stock i for month x
$r_{m,x}$ = return on the market index (S&P 500 Composite) for month x
$n_{(.)}$ = number of stocks in the long (L) or short (S) position

APPENDIX K

TABLE 20: Monthly CAPM excess returns (non-overlapping investment periods)

Panel A: 'Perfect foreknowledge' scenario

Time period	*Number of obs.*	*Position*	*Base case*		*Indicator variable*	
			α	β^M	α	β^M
1979 - 2009	216	Long	0.005	1.062	0.005	1.198
		(p-value)	*(0.198)*	*(0.000)*	*(0.057)*	*(0.000)*
		Short	-0.008	0.980	-0.009	0.962
		(p-value)	*(0.004)*	*(0.000)*	*(0.007)*	*(0.000)*
		Hedge	0.013	0.082	0.014	0.236
		(p-value)	*(0.007)*	*(0.464)*	*(0.001)*	*(0.008)*
1980 - 2010	216	Long	0.002	1.160	0.009	0.969
		(p-value)	*(0.314)*	*(0.000)*	*(0.001)*	*(0.000)*
		Short	-0.009	0.871	-0.009	0.811
		(p-value)	*(0.001)*	*(0.000)*	*(0.006)*	*(0.000)*
		Hedge	0.012	0.288	0.018	0.158
		(p-value)	*(0.006)*	*(0.007)*	*(0.000)*	*(0.070)*
1981 - 2011	216	Long	0.003	1.004	0.006	0.895
		(p-value)	*(0.136)*	*(0.000)*	*(0.041)*	*(0.000)*
		Short	-0.010	0.950	-0.009	1.066
		(p-value)	*(0.000)*	*(0.000)*	*(0.001)*	*(0.000)*
		Hedge	0.013	0.054	0.015	-0.171
		(p-value)	*(0.000)*	*(0.468)*	*(0.000)*	*(0.079)*

Panel B: ROE prediction model scenario

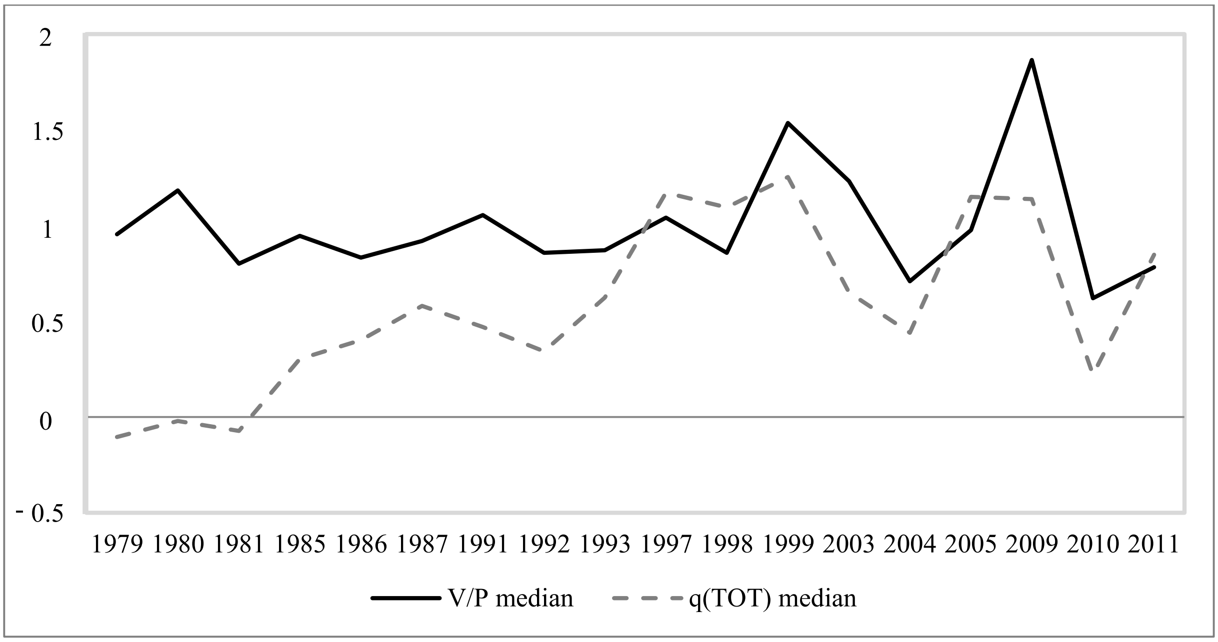
Time period	*Number of obs.*	*Position*	*Base case*		*Indicator variable*	
			α	β^M	α	β^M
1979 - 2009	216	Long	-0.004	1.055	-0.003	1.187
		(p-value)	*(−)*	*(0.000)*	*(−)*	*(0.000)*
		Short	-0.002	1.014	-0.002	0.920
		(p-value)	*(0.164)*	*(0.000)*	*(0.155)*	*(0.000)*
		Hedge	-0.002	0.042	-0.001	0.268
		(p-value)	*(−)*	*(0.491)*	*(−)*	*(0.328)*
1980 - 2010	216	Long	-0.003	0.933	-0.001	0.694
		(p-value)	*(−)*	*(0.000)*	*(−)*	*(0.000)*
		Short	- 0.005	1.012	-0.005	1.012
		(p-value)	*(0.031)*	*(0.000)*	*(0.054)*	*(0.000)*
		Hedge	0.002	-0.079	0.004	-0.318
		(p-value)	*(0.295)*	*(0.239)*	*(0.242)*	*(0.007)*
1981 - 2011	216	Long	-0.001	1.035	0.001	0.933
		(p-value)	*(−)*	*(0.000)*	*(0.405)*	*(0.000)*
		Short	-0.005	1.040	-0.003	1.070
		(p-value)	*(0.074)*	*(0.000)*	*(0.101)*	*(0.000)*
		Hedge	0.003	-0.005	0.003	-0.138
		(p-value)	*(0.131)*	*(0.951)*	*(0.129)*	*(0.063)*

Notes: Table 20 shows the monthly CAPM excess returns (α) and beta coefficients (β^M) obtained from the regression procedure for both strategies. The investment returns are assessed over non-overlapping periods based on the first (1979-2009), the second (1980-2010) and the third (1981-2011) year of each investment period respectively. Panel A reports the returns in the 'perfect foreknowledge' scenario, while Panel B presents the returns to the investment strategies based on the ROE prediction model. *P-values* are reported in parenthesis, unless the sign of α is inconsistent with the alternative hypothesis. For long and hedge positions, the null hypothesis of non-positive α is tested against the alternative of a positive α. For short positions, the null hypothesis of a non-negative α is tested against the alternative of a negative α. β-value tests are two-tailed.

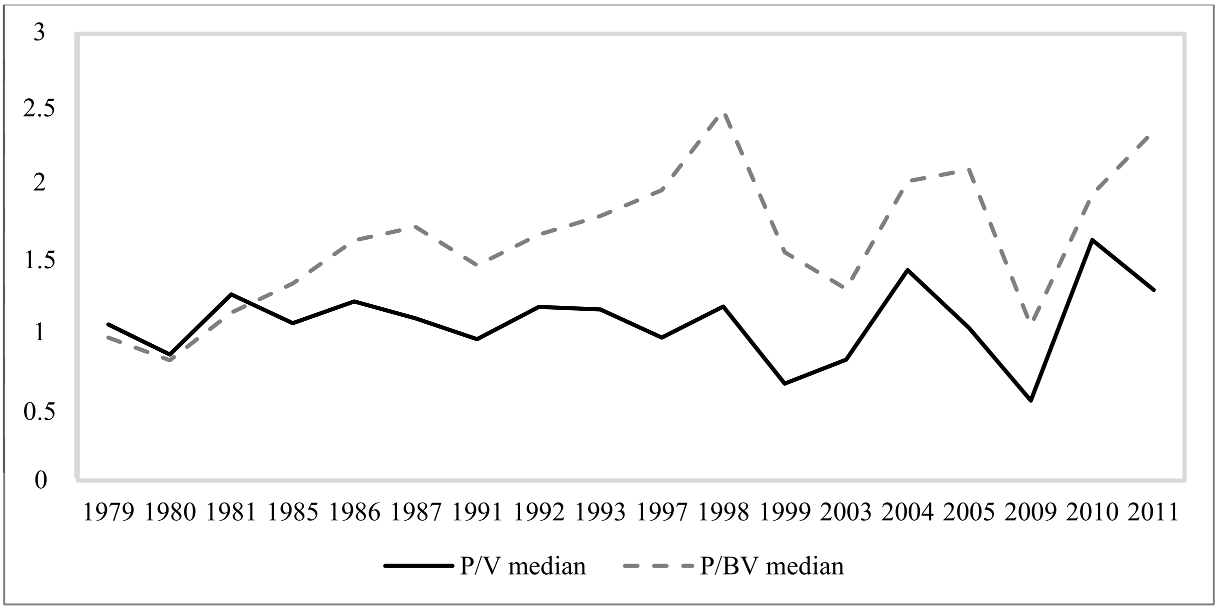
APPENDIX L

FIGURE 4: Time-series variation in price and fundamental value

Panel A: Median intrinsic-value-to-price ratio and total valuation bias



Panel B: Median price-to-intrinsic-value and price-to-book ratios



Notes: Figure 4 illustrates the time-series variation of primary price and fundamental ratios across the entire investment period 1979-2011. Panel A shows the median intrinsic-value-to-price ratio (V/P) and the total valuation bias ($q(TOT)$). Panel B depicts the median price-to-intrinsic-value ratio (P/V) in comparison to the median price-to-book ratio (P/BV).