Aalto University
School of Economics

Usefulness of financial statement information in forecasting earnings of high-technology companies

Accounting

Master's thesis

Raakel Leppähaara

2012

Department of Accounting
Aalto University
School of Economics

Aalto University
School of Economics
Master's Thesis
Raakel Leppähaara

Abstract
April 19, 2012

USEFULNESS OF FINANCIAL STATEMENT INFORMATION IN FORECASTING EARNINGS OF HIGH-TECHNOLOGY COMPANIES

PURPOSE OF THE STUDY

The purpose of this study is to examine whether financial statement information is useful in explaining and predicting the future earnings of high-technology companies. In addition, the study further analyzes whether the firm-specific uncertainty factors recognized by the prior literature are associated with the forecasting errors calculated from the hold-out sample.

DATA

The data collection consists of 10 642 active and inactive high-tech companies of which 5 341 were listed during 1990-2010. High-tech industries were identified based on the U.S. Bureau of Labor Statistics research which emphasizes companies' human effort on research and development (Hecker 2005). The earnings forecast model replicates the forecast model of Hou et al. (2010). The study was conducted as a cross-sectional analysis.

RESULTS

The findings of the study indicate that basic financial statement information has explanatory power in association with future earnings based on the IPO sample during 1991-2010. Especially lagged earnings, dividends and accruals were identified to explain future earnings.

When analyzing the forecasting errors in the hold-out sample 2001-2010, the estimated earnings forecast model performed rather well in terms of the precision and bias of the forecasts. Forecasting errors slightly increased with the forecasting horizon. There were no substantial differences documented in the performance of forecasts for IPO and all listed high-tech companies.

The observed forecasting errors could not be explained by the generally recognized uncertainty factors and, therefore, the model should be cautiously applied in high-tech samples.

KEYWORDS

High-technology, earnings forecasting, usefulness of financial statement information, cross-sectional analysis

Aalto-yliopiston kauppakorkeakoulu Tiivistelmä

Pro Gradu - tutkielma 19. huhtikuuta 2012

Raakel Leppähaara

TILINPÄÄTÖSTIETOJEN HYÖDYNNETTÄVYYS HIGH-TECH YRITYSTEN TULOKSEN ENNUSTAMISESSA

TUTKIELMAN TAVOITTEET

Tutkimuksen tavoitteena on selvittää, voidaanko tilinpäätösinformaatiolla selittää ja ennustaa high-tech yritysten tulevaisuuden tulosta. Lisäksi tutkimus analysoi, selittävätkö aikaisemmassa kirjallisuudessa tunnistetut yhtiökohtaiset epävarmuustekijät ennustemallin riippumaton otoksen ennustevirheitä.

LÄHDEAINEISTO

Tutkimuksen aineisto koostuu 10 642 aktiivisesta ja inaktiivisesta high-tech yrityksestä, joista 5 341 listautui vuosien 1990–2010 aikana. High-tech toimialojen määritelmä perustui U.S. Bureau of Labor Statisticsin tutkimukseen, joka korostaa yhtiöiden henkilöstön tutkimuspainotteisuutta (Hecker 2005). Tutkimuksessa käytetty tuloksen ennustemalli jäljittelee Hou ym. (2010) tuloksen ennustemallia. Tutkimus toteutettiin poikkileikkausanalyysina.

TULOKSET

Tutkimuksen tulokset osoittavat, että perinteinen tilinpäätösinformaatio pystyy selittämään tulevaisuuden tulosta 1990–2010 aikana listautuneiden yhtiöiden otoksessa. Etenkin edellisen tilikauden tulos, osingot ja jaksotukset todettiin selittävän tulevaa tulosta.

Estimoitu ennustemalli toimii kohtalaisen hyvin, kun analysoidaan riippumattoman otoksen (2001–2010) ennustevirheiden suuruutta ja suuntaa. Ennustevirheet kasvavat hieman ennusteajan pidetessä. Mallin toimivuudessa ei havaittu merkittäviä eroja verrattaessa ennusteita juuri listautuneiden ja kaikkien listautuneiden high-tech yhtiöiden välillä.

Ennustevirheitä ei kyetty selittämään yleisesti tunnetuilla epävarmuustekijöillä ja tämän vuoksi mallia tulisi käyttää varoen high-tech otoksissa.

AVAINSANAT

High-tech, tuloksen ennustaminen, tilinpäätösinformaation hyödynnettävyys, poikkileikkausanalyysi

Acknowledgements

This year has been a year of change not least because of this master thesis and the upcoming graduation. This year has also taught a lot, taught me to survive the downturns and to enjoy the ray of light. I would not have accomplished these changes without the sincere support of several people. I am deeply grateful to all of those people who contributed in this process.

First my deepest thanks to my supervisors. Harri Seppänen, Assistant Professor at Aalto University School of Economics, encouraged me through the whole process and offered all his expertise when I needed it the most. Juha Kinnunen, Professor of accounting at Aalto University School of Economics, provided the utmost important and supportive feedback. I would also like to thank my fellow student and “partner in crime”, Saara Hietanen, who gave valuable support in the data collection phase.

Thanks to my co-workers for understanding the pressure to accomplish this piece.

Then my friends, first my profound apologies that I have not been there for you during this process, and at the same time, thank you for still being there for me.

But lastly, my dearest thanks to my family. Thank you Mom for nursing me when I was at my weakest. Thank you Dad for setting everything up as you have always done for me. And Laura, thanks for cheering up and giving me a shelter when I really needed it.

Helsinki, April 19, 2012

Raakel Leppähaara

TABLE OF CONTENTS

List of Tables

List of Figures

1 Introduction

1.1 Background

Earnings forecasts are often incorporated into analyses of the future performance and value of companies. Past research has extensively addressed the usefulness of accounting information in valuation and predicting earnings. Financial statement information is useful for investors if they can exploit it to forecast earnings and evaluate the risk associated to future earnings (Richardson et al. 2010). Useful information should be relevant and reliable. Relevant financial information makes a difference in investment decisions and has predictive and confirmatory value (IFRS Framework 2010). “To date, very little research combines multiple accounting attributes to forecast future earnings or returns” (Richardson et al. 2010, 412). On the other hand, there is lack of generally agreed variables which should be included in earnings forecast models (Richardson et al. 2010).

Especially the usefulness of financial statement information in high-tech context has been widely debated (see e.g. Keating 2000, Demers & Lev 2001, Hand 2000, Bowen et al. 2002, Trueman et al. 2000). Keating (2000, 163) captures the essential idea of analyzing the financial performance of high-tech companies: “New economy firms are of particular interest because of concerns about whether the traditional accounting model can be used to value firms in instances where most of that value consists of growth opportunities”. High-tech industries are of particular interest also because high-tech is commonly referred to the economic growth of economies, the source of employment growth, innovation and technological opportunities (see e.g. OECD Science, Technology and Industry Scoreboard 2005, Tong 2005, Kask & Sieber 2002).

High growth potential, but also a high degree of uncertainty, characterizes the technological markets. Volatile earnings, young age with lack of historical data and the intangible intensive business model impede the forecasting of high-tech firms’ performance compared with more established, seasoned firms (Trueman et al. 2000). Uncertain prospects may also result in positive unexpected performance. For instance, analysts tend to underestimate the earnings of Internet firms (Trueman et al. 2001). Moreover, financial press is currently arguing the social media bubble related to the new and possible IPOs of social media firms (e.g. LinkedIn, Facebook, and Twitter, see e.g. Helsingin Sanomat 13 July, 2011).

1.2 Research question

Motivated by the challenges and importance of high-technology industry, the objective of the thesis is to analyze whether financial statement information is useful in explaining and predicting the future earnings of high-tech companies. Briefly, the research question can be stated as:

> *Does financial statement information have explanatory and predictive power with respect to the future earnings of high-technology companies?*

The study is executed as a fundamental analysis of accounting-based metrics which presumably have explanatory and predictive power in association with the future earnings of high-technology companies. The performance of the earnings forecast model is evaluated in terms of the bias and precision of the forecasts. In addition, the study further analyzes whether the uncertainty factors recognized by the prior literature are associated with the forecasting errors of the estimated earnings forecast model.

1.3 Sample and methods

The study is carried out by analyzing the financial statements of global listed high-technology companies from 1990–2010. The data is retrieved mainly from Thomson Reuters database. High-tech industries were identified based on the U.S. Bureau of Labor Statistics research which emphasizes companies' human effort on research and development (Hecker 2005). In total, 10 642 active and inactive companies were identified operating in high-tech industries and of which 5 341 firms were listed during 1991-2010.

Firstly, the explanatory power of accounting metrics was analyzed with the sample of newly listed high-tech firms during 1991-2010. The IPO sample was used because the accounting has been argued to poorly capture the economics of high-tech firms (see e.g. Lev & Zarowin 1999, Francis & Schipper 1999). The earnings forecast model used in this study replicates the forecast model in Hou et al. (2010). Large cross-sectional analyses across industries have proven the applicability of the model. Lee et al. (2011) find that Hou et al.'s (2010) earnings forecasting model produces reliable forecasts in their comparison of the implied cost of capital (ICC) estimates. However, the model has not yet been tested in industry-level context. The independent variables of the model include basic accounting variables, for instance total assets, dividends paid, lagged earnings and accruals, completed with market capitalization. The dependent variable consists of absolute dollar earnings of one, three and five years ahead. The data was first analyzed with multivariate OLS regression. However, the results of the explanatory power are founded on the ordinal logistic regression because of the statistical issues encountered in OLS regression analyses. For the ordinal regression, the dependent variable was divided into 20 equal sized groups based on the size of the future earnings.

Secondly, for the analysis of predictive power, high-tech companies were divided into two samples, IPOs and all listed companies. The split was made to compare whether the forecasts for IPOs are less accurate than the forecasts for seasoned firms. Hou et al.'s (2010) model was estimated based on the cross-sectional analysis of the data from 1990-2000 and tested in a hold-out sample from 1991-2010. The model was estimated with multivariate OLS regression separately for both samples and the forecast horizons of one, three and five years. The predictive performance of the model was analyzed in terms of bias and precision of the forecasts in the hold-out sample. Precision describes the absolute deviation of the expected value from the actual earnings (calculated as abs(actual – expected)). Bias takes into account the direction of the error. Bias is defined as the raw difference of the expected value from the actual earnings (calculated as actual – expected). Both statistics are scaled by market capitalization.

Thirdly, the study examines whether the uncertainty factors recognized by the prior literature are associated with the estimated forecasting errors of the hold-out period 2001-2010 with the samples of IPO and all listed high-tech companies. Uncertainty factors include firm age, sales growth, analyst coverage, research and development expenditures and intangible assets. Both the bias and precision of the forecasts were tested as dependent variables. The association of the uncertainty factors with the forecasting errors was first analyzed with OLS regression but also binomial logistic regressions were employed.

1.4 Results

Regarding the explanatory power of the model, the findings indicate that financial statement variables provide valuable information regarding the future earnings of newly listed high-tech companies. These results are founded on the ordinal regression analysis where the future earnings were divided into 20 equal sized groups based on the size of the future earnings. Ordinal regression was employed because of the statistical problems encountered in OLS regression. Especially lagged earnings, dividends and accruals were identified to explain future earnings whereas the relation of the size variables (market capitalization and total assets) to future earnings was weak. In a conclusion, financial statement information is useful in explaining the future earnings of newly listed high-tech firms.

Usefulness of financial statement information requires that the riskiness of forecasted earnings should be reliably evaluated (Richardson et al. 2010). When analyzing the predictive power of the model, overall, the Hou et al.'s (2010) model performed rather well in terms of the precision and bias of the forecasts in the hold-out sample. Previously, the model has only been tested in large pooled samples across industries (see Lee et al. 2011). Contrary to the findings of Hou et al. (2010), the bias of the forecasts in both samples was close to neutral or slightly positive with an average bias of 0.045 for IPOs and 0.09 for all listed firms over all forecasting horizons. Surprisingly, the earnings forecast model did not perform better with the sample consisting of all listed high-tech companies than with the IPO sample. In fact, the performance of the model was rather similar with both samples in terms of mean and median errors in precision. The forecasting errors tend to increase with the forecast horizon, however, the increase was not substantial. Approximately half of the forecasts in both samples and throughout the forecasting horizons produced forecasting errors within 10% of the market value. In a conclusion, financial statement information is useful in forecasting the earnings of high-tech firms.

The association of the uncertainty factors with the estimated forecasting errors was examined to improve the Hou et al.'s (2010) model. There was no conclusive evidence found that the chosen uncertainty factors explain either the precision or bias of the forecasts in the hold-out sample. These findings impact on the applicability of Hou et al.'s (2010) model in high-tech context. Even though the model seems to perform rather well in terms of the precision and bias of the forecasts, the observed forecasting errors could not be explained by the generally recognized uncertainty factors. In order to improve the forecasting model of Hou et al. (2010), the root causes of the forecasting errors should be identified. Meanwhile, the model should be cautiously applied in high-tech analyses. In a conclusion, the uncertainty factors do not improve the Hou et al.'s (2010) earnings forecast model.

1.5 Implications

The findings of this study have several practical implications. Firstly, the results of the study provide confirmatory evidence that financial statement information is useful in analyzing high-tech firms' future earnings. Basic accounting variables are documented to have explanatory power for the future earnings of IPOs in high-tech context which is generally recognized as hard to predict.

Secondly, the study provides some answers to Richardson et al.'s (2010) question which variables should be included in earnings forecasting models. The findings of the study indicate that lagged earnings, dividends and accruals have explanatory power in relation with future earnings.

Thirdly, the results benefit especially the investors and analysts evaluating the performance of high-tech firms. Based on the findings of the study, Hou et al.'s (2010) cross-sectional multivariate model does fairly good job in forecasting of the earnings of high-tech firms. The model performed rather well in terms of the bias and precision of the forecasts in an out-of-sample test. Also the variance in the earnings with longer forecast horizon was captured by the model. Financial statement information seems to take into account the uncertainty factors recognized by the prior literature because the findings indicate that none of the uncertainty factors explained the forecasting errors. However, the high-tech firms are generally identified to be hard-to-predict and, therefore, it is not surprising that there are also some large forecasting errors detected. Still, approximately half of the forecasts for IPOs and all listed high-tech firms produced forecasting errors within 10% of the market value throughout the forecasting horizons. The forecasts of absolute dollar earnings can also be compared with the analysts' estimates. However, further analysis of the forecast errors produced by the model is left for future research. Factors causing the forecasting errors should be identified in order to ensure the quality and applicability of the earnings forecast model.

1.6 Structure of the study

The study is structured as follows. Chapter 2 provides insight into prior literature on forecasting earnings. The chapter covers the main trends in earnings forecasting and presents recent multivariate earnings forecast models. Chapter 3 continues with outlining the empirical studies related to high-tech firms' future performance. Especially the usefulness of financial statement information in context of high-tech industries is discussed. Chapter 4 describes the research design including the expectations and sample of the study. Results of the empirical analyses are presented in Chapter 5. Lastly, Chapter 6 concludes the study as well as gives proposals for future research.

2 Prior research on earnings forecasting

2.1 Introduction to earnings forecasting

The goal of this chapter is to review the prior earnings forecasting literature. First, the purpose of forecasting earnings and the development of earnings forecasting research are discussed. Then the main trends in prior empirical studies including the mean reversion of earnings and profitability as well as time-series and cross-sectional estimation are reviewed more profoundly. Recent forecasting models using accounting variables as predictors of future performance are introduced first in Chapter 2.3. The DuPont model is briefly discussed in order to understand the properties of earnings. Also the usefulness of forecasting at industry-level is surveyed since this study is focused on a specific industry, high-tech. Lastly, the uncertainty factors commonly related to forecasting errors are discussed in Chapter 2.4 and, in fact, there are several uncertainty characteristics which can be recognized in the firms operating in high-tech industry.

2.1.1 Motivations of earnings forecasting

Forecasting firms' operating performance and earnings can be motivated by several approaches. The Framework of International Financial Reporting Standards emphasizes the qualitative characteristics of useful financial information (IFRS Framework 2010). Useful information should be relevant and reliable. Information which makes a difference in decision making is defined as relevant information. Relevant information should have predictive value and confirmatory value. Richardson et al. (2010) conclude that the financial statement information is useful for investors if they can use the information for forecasting earnings, evaluating the riskiness of these earnings and finally make an analysis of the firm value.

Forecasting companies' performance is important for investors but also for other stakeholders. "Managers need forecasts for planning and to provide performance targets; analysts need forecasts to help communicate their views of the firm's prospects to investors; bankers and debt market participants need assess the likelihood of loan repayment." (Palepu et al. 2007, 260) The key motivation for investors to forecast earnings is the link between companies' expected performance and stock valuation. Valuation requires knowledge of future earnings and cash flows (Kinnunen 1988). Beaver (1998) argues that the current earnings provide information to predict future earnings. Bowen et al. (1986) note that the earnings based on accrual accounting provide valuable information about future cash flows.

Kothari (2001) presents five key motivations for research on earnings forecasts. First of all, valuation models often employ either directly or indirectly earnings forecasts. Discounted cash flow model uses forecasted earnings as proxy for future cash flows. In addition, the residual-income valuation model requires knowledge of net “normal” earnings. Secondly, in the calculations of security returns, researchers frequently use earnings forecasts as a benchmark for unexpected earnings. Thirdly, earnings forecasts are often analyzed in the evaluation of market efficiency. Fourthly, the research requires knowledge of the “normal” earnings for a comparison to manipulated earnings. Earnings management is based on the choices of the management in accounting procedures. The last motivation Kothari (2001) lists concerns the informative input of analysts’ and management’s forecasts for the capital market.

2.1.2 Development of earnings forecasting research

Forecasting earnings has been a research topic for a rather long time and there can be some main trends recognized. The research started from the time-series analysis of earnings by exploiting the random-walk and random-walk with a drift methods. Then the research headed towards more complex models taking into account different accounting variables. Univariate and time-series models turned into multivariate and cross-sectional models. Also new aspects in earnings’ behavior including the mean reversion of profitability and earnings were discovered. Next, these development phases are discussed more in detail.

Earnings forecasting literature is usually linked to capital market research and, particularly, to its usefulness in stock valuation. Brown (1993) surveys the studies on earnings forecasting mainly from the 1970s and 1980s. It has been argued that, prior late 1970s, most earnings forecasting studies lacked implications for capital market research. Previously, the focus was solely in the accuracy of the earnings expectation models. Researchers concentrated to provide evidence whether the annual earnings followed a random walk process. Several times the researchers failed to disprove the power of the random walk process. For instance, autoregressive integrated moving average (ARIMA) models were extensively tested but the findings indicated that the earnings forecasts of ARIMA models were not more accurate than the random walk models. (Brown 1993)

Brown (1993) uses a term “association approach” for describing the research trend which was developed especially in the 1980s. The association approach linked the capital market performance and earnings forecasts. In capital market research, the earnings forecast models have been used to measure earnings expectations (Brown 1993). Foster (1977) was amongst the first ones to focus on the accuracy of earnings expectation models in relation to abnormal stock returns. Brown (1993) questions whether the best earnings forecast model in terms of accuracy actually has the strongest association with abnormal returns. After finding the association between earnings forecasts and capital market reactions, the main trend in the research was to combine two worlds, accounting and finance, in different ways. For example, in the 1980s, the researchers evaluated analysts' earnings forecasts and, in fact, found that analysts' forecasts were actually more accurate than the time-series models at that time (see e.g. Fried & Givoly 1982, Brown et al. 1987). Earnings forecasting research maintained the link to capital market research but examined also the time-series properties of earnings and pure earnings forecasting. (Brown 1993)

Kothari (2001) reviews the capital market research in accounting, in other words, the relation between capital markets and financial statement information, focusing on the studies of the late 1980s and 1990s. The researcher suggests that a major academic challenge lies in the absence of rigorous theory for the time-series properties of earnings (Kothari 2001, 124). Kothari (2001) categorizes earnings forecasting to the methodological research of capital markets. Earnings forecasting research includes the research on the properties of time series, the management's and analysts' earnings forecasts and the earnings growth rates. Forecasting earnings is useful also in defining firms' cost of capital. Analysts' earnings estimates are often employed as an earnings expectation in the cost of capital models. Hou et al. (2010) presents a new method to estimate earnings expectations in their cost of capital model. Hou et al. (2010) use a cross-sectional analysis to estimate an earnings forecast model based on accounting variables. Also Lee et al. (2011) have employed the model of Hou et al. (2010). This earnings forecast model is tested in this study. The model is discussed in detail in Chapter 2.3.1 and in the empirical part in Chapter 4.2.

Richardson et al. (2010) survey the recent research, papers published after 2000, on forecasting future earnings and returns. Richardson et al. (2010) emphasize the significance of financial statement information for investors. "An investor can use information in these statements to forecast earnings for the reporting entity, estimate the risk of these earnings and ultimately make an assessment of the intrinsic value of the firm that can be compared to observed market prices" (Richardson et al. 2010, p. 410). Richardson et al. (2010) recognize that there are still interesting topics for future research in this field of studies. For instance, the researchers state that to date there are still very little study on forecasting future earnings or returns which combines multiple accounting factors or uses macroeconomic data. In fact, Richardson et al. (2010) document that there is lack of generally agreed variables which should be included to earnings forecast models. However, Manegold (1981) reminds that the accuracy of multivariate earnings forecasting models appears not to be substantially better than the accuracy of less sophisticated univariate models. Also the additional costs of developing a complex multivariate model have to be taken into consideration.

2.2 Main trends in earnings forecasting research

2.2.1 Mean reversion of earnings and profitability

Predicting annual earnings is often linked to the mean reversion of earnings and profitability. Stigler (1963, 54) states that “ -- under competition, the rate of return on investment tends toward equality in all industries. Entrepreneurs will seek to leave relatively unprofitable industries and enter relatively profitable industries”. Furthermore, above-normal profitability should not be sustainable due to competitive markets. In a competitive environment, profitability should be mean reverting within as well as across industries (Fama & French 2000). The mean reversion of profitability and earnings would lead to a situation where the changes in profitability and earnings are to some extent predictable (Fama & French 2000). However, accounting conservatism and litigation risk result in negative correlation in earnings since managers tend to recognize losses more quickly than positive outcomes (Kothari 2001). Furthermore, managers are motivated to take “big baths” in earnings and firms often recognize anticipated losses which make losses less permanent (Kothari 2001). Large changes in earnings are documented to reverse more rapidly than small changes and negative changes are documented to reverse faster than positive changes (Fama & French 2000).

Fama and French (2000) study the mean reversion of earnings and profitability. The researchers note that existing forecasting literature is focused primarily on the mean reversion of earnings instead of the mean reversion of profitability. Fama and French (2000) find in their cross-sectional analysis that profitability is mean reverting. There appears to be a predictable variation in earnings due to the observed mean reversion of profitability (Fama & French 2000). Also Allen and Salim (2005) examine the mean reversion of profitability in the UK companies during 1982-2000. Allen and Salim (2005) apply the same method as Fama and French (2000). The findings in the British sample support the results of Fama and French (2000). Fama and French (2000) state that the earnings forecasts should exploit the observed mean reversion in profitability. However, Kothari (2001) notes that, according to prior literature, the predictive ability of mean reversion is not better than a random walk model in out-of-sample testing.

2.2.2 Time-series estimation

Forecasting earnings has often been studied as a time-series analysis. Brown (1993) surveys that the random walk model of earnings is extensively discussed in the research of earnings properties. According to prior research, “a random walk or random walk with drift is a reasonable description of the time-series properties of annual earnings” (Kothari 2001, 145). However, Kothari (2001) points out that the random walk in earnings is not supported by any economic theory. In addition, the time-series approach in earnings forecasting is argued to be lacking important information which could be achieved via cross-sectional estimation. Fama and French (2000) assert that one of the key weaknesses in the time-series forecasting of earnings is the shortage of time-series observations available for several companies. Often there is no required data for the whole time-series for each sample company. If these companies with incomplete data are excluded from the research, the results may exhibit survivor bias which implies that there may be more observations of positive earnings changes (Fama & French 2000). The survivor bias issue is especially discussed in the high-tech research since the survival of these smaller, young and unprofitable companies is questioned (see e.g. Demers & Joos 2007). In addition, Kothari (2001) argues that the time-series estimation exhibit low explanatory power due to both large standard errors and survivor bias. Kothari (2001) even proposes that these weaknesses of time-series estimations lead to preferring the random walk in annual earnings forecasts.

2.2.3 Cross-sectional estimation

In a cross-sectional analysis, the estimation is performed annually across all firms. Conclusions are based on the annual parameter estimates from the cross-sectional regressions. Cross-sectional method originates from Fama and MacBeth's (1973) research and Fama and French (2000) introduced the method in earnings forecasting. There are several benefits discovered in using the cross-sectional method in forecasting. Cross-sectional analysis allows exploiting a large sample data throughout a long time period. Kothari (2001) states that generally cross-sectional analysis enhances the power in estimations compared to time-series estimation and reduces the issue of survivor bias. Furthermore, economic determinants, for example inflation, may be incorporated in the cross-sectional analysis.

There are also observed weaknesses in a cross-sectional estimation. One of the most significant drawbacks is the loss of the firm-specific time-series properties (Kothari 2001). Allen and Salim (2005) as well as Fama and French (2000) argue that the correlation of regression residuals across firms is not often taken into account in the standard errors of the regression. This correlation has been a problem in the prior research employing cross-sectional estimation (Allen & Salim 2005). Furthermore, the researchers question the underlying assumption that "there is no correlation across firms in current changes in the dependent variables earnings or profitability driven by common macroeconomic or industry shocks beyond those picked up by lagged predictor variables" (Allen & Salim 2005, 2010). The assumption has to be carefully analyzed especially in the case of high-tech firms and the high-sentiment periods of Internet companies, for example, the Internet boom in the beginning of the 2000s. Moreover, financial press is now discussing the rise and possible second bubble created by the social media companies (e.g. LinkedIn, Facebook, and Twitter, see e.g. Helsingin Sanomat 13 July, 2011).

2.3 Multivariate earnings forecasting models

2.3.1 Recent models using accounting variables

This study is focused on the forecast models using accounting information. The studies of Fama and French (2000, 2001, 2006) Hou et al. (2010), Hou and van Dijk (2010), and Hou and Robinson (2006) have developed multivariate earnings and profitability forecasting models which are mostly based on accounting variables. These studies employ either profitability or earnings forecasts as tools for further analysis of stock returns. For instance, Hou and van Dijk (2010) and Hou and Robinson (2006) both examine the explanatory power of profitability shocks in relation to stock returns. In Hou et al. (2010), the researchers exploit their earnings forecasting model in calculating the implied cost of capital.

The recent studies often utilize the findings of Fama and French (2000). Fama and French (2000) investigate the mean reversion of earnings and profitability. Forecasting earnings and profitability are commonly linked to the studies of mean reversion of earnings and profitability. Fama and French (2000) construct their model of very basic accounting variables. The explanatory variables include market value to total assets, dividend dummy variable for firms not paying dividends and dividends to book value of equity. Dividends are expected to have information about future earnings and firms not paying dividends are found to be less profitable than the dividend paying companies (Fama & French 2001). In fact, Fama and French (2001) document that the larger and more profitable firms are more likely to pay dividends whereas firms with heavy investments, efforts in R&D and higher market value of assets to book value are less likely to pay dividends. However, the number of firms paying dividends has declined dramatically from the late 1970s to late 1990s (Fama & French 2001). Market value is chosen for the model since market value should reflect the current value of future net cash flows. Market value also completes the model by picking up the variation in expected profitability missed by the dividend variables (Fama & French 2000). The researchers chose this parsimonious model after testing different additional variables, for example, the logarithm of total assets as a measure of firm size. Fama and French (2000) find that all the chosen explanatory variables including the dividend dummy, dividends paid and market value have explanatory power to future profitability. Dividends to book value of equity are positively related to expected profitability. Also market value is found to have a strong positive association with profitability.

Fama and French (2006) analyze expected profitability in the study of expected stock returns. The researchers suggest that expected profitability is one component explaining the future stock returns. The constructed proxy for expected future profitability employs accounting fundamentals: a dummy variable for negative earnings, lagged profitability, accruals to book equity separately for positive and negative accruals, an investment variable including the change in assets to lagged assets, a dummy variable for firms not paying dividends and dividends to book value. The model is rather similar to Fama and French (2000) but it is completed with additional accounting variables. Lagged profitability is added since it tends to be persistent over time. Also the logarithm of market capitalization is included in the model because smaller firms tend to be less profitable.

Fama and French (2006) find that lagged profitability has the strongest forecasting power and it is considerably persistent. Profitability is found to be mean reverting similarly to Fama and French (2000). Also the results on dividends are in line with Fama and French (2000) meaning that firms not paying dividends are less profitable in the future. Firms with higher book-to-market ratio are found to be less profitable. For the additional variables, Fama and French (2006) document that lagged asset growth is negatively related to future profitability and growth in earnings. Negative coefficients for accrual variables support the prior literature findings that accruals result in transitory variation in earnings. In detail, the reversal of positive accrual occurs faster but both negative and positive accruals have long-run transitory effects. Also Sloan (1996) documents that higher current accruals indicate lower future profitability. Richardson et al. (2010) argue that investors do not understand the time-series impact of accruals when constructing earnings forecasts and, therefore, accruals are reflected in market prices. The phenomenon where the investors incorrectly weight the accrual information is called accrual anomaly and it is widely studied in prior research (see e.g. Sloan 1996). However, Richardson et al. (2010) argue that the anomaly has gradually decreased in recent years.

Motivated by the argument that analysts do not often follow and, thus, forecast earnings of small and distressed companies, Hou and van Dijk (2010) construct a cross-sectional model based on accounting variables to forecast profitability. The model is employed to examine whether the disappearance of size effect in realized returns can be explained by unexpected shocks in profitability. Profitability shock is considered to be the difference between the realized and expected profitability. Also Hou and Robinson (2006) examine cash flow shocks with almost similar method as Hou and van Dijk (2010) but Hou and Robinson (2006) consider the profitability shock as a regression error of the constructed model for expected profitability.

Hou and van Dijk (2010) replicate the models used in Fama and French (2000, 2006). The dependent variable earnings at year t+1 is divided by lagged total asset. The explanatory variables include market value to total assets, a dummy variable for dividend payers (value of 0), dividends to book equity and lagged earnings scaled by total assets at the time of t-1. The variables are winsorized at the 0.5 and 99.5 percentiles in order to avoid the interference of extreme values especially in the scaled variables. However, the robust tests indicate that the results are in line with and without winsorizing the variables. Following Fama and French (2006), Hou and van Dijk (2010) test also the effect of additional variables, such as asset growth, negative earnings dummy, and positive and negative accruals. The test results indicate statistically significant coefficients for most of these additional variables but those are ignored in the further examination as they do not enhance the overall explanatory power.

Hou and van Dijk (2010) argue that, according to prior studies, small firms have experienced large negative profitability shocks after the early 1980s whereas bigger firms have encountered large positive shocks compared with expected profitability. The results confirm these expectations. All explanatory variables including market value to total assets, a dummy variable for dividend payers, dividend payments to book equity and lagged earnings scaled by total assets are statistically significant at the level of 0.01 and adjusted R^2 for the model is approximately 60 percent. Hou and Robinson (2006) note that the inclusion of lagged profitability increases the regression R^2 values compared with Fama and French (2000). Otherwise the results follow the studies of Fama and French (2000, 2006) and Hou and Robinson (2006) since market-to-book and dividend-to-book ratios, and lagged profitability are all positively related to profitability whereas the dividend dummy is negatively associated with profitability. However, the results differentiate between the two periodical subsamples because the coefficient for market-to-book is statistically significant for the period before the 1980s but not from the 1980s onwards. In addition, the coefficient for the dividend dummy is statistically significant from the 1980s onwards and the reverse holds true to the earlier period from 1963 to the 1980s. Fama and French (2004) argue that the variation in market-to-book ratios is less significant in the latter period because of the flow of young, low or zero profitability firms with substantial growth opportunities to the markets. Hou and van Dijk (2010) document a decline in profitability and survival rate which is also in line with the results of Fama and French (2004). Moreover, since the number of companies paying dividends has declined, the dividend dummy has become a more powerful tool in explaining the future profitability from the 1980s onwards (Fama & French 2001). Hou and van Dijk (2010, 25) note that "significant proportion of profitability shocks can be attributed to new lists". In addition, several industries including high-technology sectors have experienced significant development steps and those industries are leading in the number of IPOs (Hou & van Dijk 2010).

Hou et al. (2010) introduce a different approach to estimate the implied cost of capital (ICC). ICC estimates are generally based on the analysts' earnings forecasts. Instead of analysts' earnings forecasts, Hou et al. (2010) construct a cross-sectional model of accounting variables completed with market value to forecast future earnings one, two and three years ahead. The model developed by Hou et al. (2010) is replicated in this study and the formula is presented in Chapter 4.2. As discussed in Chapter 2.2.3, there are several advantages using the cross-sectional approach whereas there are several disadvantages, including the recognized biases and narrow coverage of firms and time-series, in using analysts' forecasts. Moreover, there is documented lack of reliability in the estimated ICCs based on analysts' forecasts. Therefore, accounting based proxies for future earnings are demanded (Hou et al. 2010). Hou et al. (2010) insert the calculated expected earnings into the discounted residual income model to estimate the ICC for each company. The dependent variable used in Hou et al.'s (2010) earnings forecast model differs from the other models discussed in this chapter. Hou et al. (2010) has chosen absolute dollar earnings as the dependent variable due to the comparability with analysts' forecasts. In addition, the forecasting horizon in Hou et al. (2010) is longer whereas the other previously mentioned studies calculate the expected profitability only for the next year. Variables employed are in line with the previous studies but those are not scaled by any size variable, for example market capitalization or total assets. The dependent variable is net income before extraordinary items and the explanatory variables are market value, total assets, dividends paid, a dividend dummy for non-paying firms (value 0), lagged earnings, a negative earnings dummy (value 1) and operating accruals. The variables are winsorized at the 0.5 and 99.5 percentiles and the results are robust to scaling all variables with total assets, market equity, sales and net operating assets.

The results of Hou et al. (2010) are consistent with Fama and French (2000, 2006), Hou and Robinson (2006), and Hou and van Dijk (2010). In addition, all the explanatory variables except the negative earnings dummy have the same sign consistently throughout the forecasting horizons of one to three years. Especially lagged earnings are highly persistent and statistically significant. Lagged market value is positively and lagged total asset are negatively associated with future earnings. Furthermore, firms that pay out more dividends and firms with lower operating accruals tend to have higher future earnings. The negative earnings dummy is negative for the year t+1 and positive for the years t+2 and t+3 but the dummy is statistically significant only for the third year. The overall explanatory power of the model is high with adjusted R^2 of 87%, 81% and 77% for the years one to three ahead, respectively. The model seems to capture a substantial part of the variation in future earnings by using ex ante accounting variables. Moreover, the researchers considered and tested several additional earnings predictors, for example capital expenditure, R&D, and firm age, but those were not included in the model because they did not improve the quality of the earnings forecasts or the reliability of the calculated ICC estimates (Hou et al. 2010). When examining the forecast bias and accuracy, the researchers find that both the cross-sectional model and analysts' forecasts tend to be overoptimistic but analysts' forecasts are documented to exhibit even more severe negative bias. However, when considering the forecasting accuracy defined as the absolute value of the forecast bias, analysts' forecasts were more accurate than the model-based forecasts. For example, Lee et al. (2011) employ the cross-sectional model of Hou et al. (2010) to avoid problems associated with analysts' earnings forecasts in calculating future earnings in their study of ICC estimates.

2.3.2 The DuPont model in earnings forecasting

The DuPont model is one method of modeling present and future earnings. Figure 1 illustrates the composition of ROE. The DuPont model is often used as a basic tool for ratio analyses and analyses of firms' performance. The DuPont model splits the company's return on equity (ROE) into its key value drivers: net profit margin, asset turnover, and financial leverage (Palepu et al. 2007). Return on equity answers to a question of "how well managers are employing the funds invested by the firm's shareholders to generate returns" (Palepu et al. 2007, 199). Trotta (2003) argues that the main advantage of the DuPont model is its simplicity as it is solely based on basic accounting variables. Palepu et al. (2007) explain that the three components of DuPont capture different aspects of firms' management. Net profit margin reflects the effectiveness of companies' operating management, asset turnover the effectiveness of investment management and financial leverage the effectiveness of liability management. Soliman (2008) suggests a slightly different approach for the terms of ROE. The researcher argues that profit margin provides information about firms' pricing power, for example, about product innovation and positioning, brand name recognition as well as first mover advantage. On the other hand, asset turnover results from the asset utilization and the efficiency derived from efficient processes in using property, plant and equipment and inventory. Hence, the management of working capital is vital for firms' profitability. (Soliman 2008)

Profit Margin	×	Asset Turnover	=	Return on Assets (ROA)
$\frac{\text{Net income}}{\text{Sales}}$	×	$\frac{\text{Sales}}{\text{Total assets}}$	=	$\frac{\text{Net income}}{\text{Total assets}}$
Return on Assets (ROA)	×	**Financial Leverage**	=	**Return on Equity (ROE)**
$\frac{\text{Net income}}{\text{Total assets}}$	×	$\frac{\text{Total assets}}{\text{Common equity}}$	=	$\frac{\text{Net income}}{\text{Common equity}}$

Figure 1. Composition of ROA in connection to ROE (Trotta 2003, 153).

Palepu et al. (2007) demonstrate the link between valuation and return on equity. The authors explain that “in the long-run value of the firm’s equity is determined by the relationship between its ROE and its cost of equity capital” (Palepu et al. 2007, 199). In other words, firms for which the forecasted ROEs exceed the cost of equity capital in the long-term should have valuations in excess of book value.

ROE is also a tool for forecasting firms’ future performance. Beaver (1998) argue that current earnings provide information to forecast earnings. Furthermore, prior literature documents a positive relation between the components of DuPont model and equity returns (see e.g. Soliman 2008). For instance, asset turnover is documented to be positively associated with future changes in earnings and both, asset turnover and profit margin are documented to have explanatory power for changes in future profitability (Soliman 2008). Moreover, asset turnover is found to be more persistent in nature than profit margin (Soliman 2008). However, Fairfield and Yohn (2001) document that profit margin and asset turnover do not have predictive power but change in asset turnover is positively associated with future changes in earnings. Palepu et al. (2007) add that comparing the cost of capital with ROE is useful also in analyzing future profitability.

2.3.3 Industry-level forecasting models

The need for the industry-specific forecasts of financial performance has been questioned. When analyzing the usefulness of industry-level forecasts, the aforementioned Stigler’s (1963, 54) statement have to be kept in mind: “ -- under competition, the rate of return on investment tends toward equality in all industries. Entrepreneurs will seek to leave relatively unprofitable industries and enter relatively profitable industries”. In other words, as prior studies (e.g. Fama & French 2000, Allen & Salim 2005) document, profitability and earnings across industries tend to be mean reverting. However, there are also studies (see e.g. Fairfield et al. 2009) which question the economywide thinking of mean reversion and suggest that mean reversion would be more of an industry-specific phenomenon. Fairfield et al. (2009) argue that there are systematic differences between industries which hinder the mean-reversion across the industries. For example, barriers to entry, product demand and business risk may have influence on the persistence and level of firm performance. Fairfield et al. (2009) explain that industry membership is one of the key factors affecting the cross-sectional differences and firms' performance. For instance, Fama and French (2000) suggest that average industry profitability may be higher in industries with higher risk.

Fairfield et al. (2009) test whether the long-term firm profitability or sales growth converge to industry or economywide benchmarks. Fairfield et al. (2009, 148) state that "if mean reversion contributes to the predictability of firm performance and the mean reversion parameters differ systematically across industries, then it follows that industry-level prediction models will provide more accurate forecasts of firm performance than economy-wide models". The researchers criticize that the prior research on mean reversion is mainly focused on year-ahead firm performance. Thus, Fairfield et al. (2009) examine the long-term forecasts since the effect of industry membership on firms' performance may not be immediate.

Fairfield et al. (2009) find that industry-specific models are rather accurate in predicting firm growth but not profitability. Industry-specific forecast models are both in short (one year ahead) and long-term (five years ahead) more accurate than the economywide models when estimating sales growth. Fairfield et al. (2009) argue that growth is more dependent of product demand which is connected to industry level factors. Rather surprisingly, Fairfield et al. (2009) find that the industry-level forecasts of profitability (ROE and RNOA) are not more effective than the economywide forecasts either in short-term or in long-term analysis. In fact, the long-run ROEs are found to be closer to economywide than to industry-specific benchmarks. Moreover, Porter (1979) argues that firms' profitability depends on the structure within industries and differs between strategic groups within an industry. Fairfield et al. (2009) also suggest that profitability measures may converge to industry benchmarks gradually over time as firms leave less profitable industries and new firms enter more profitable industries. This kind of movement in the economy takes time to adjust. In a conclusion, as Fairfield et al. (2009, p. 148) state, "heterogeneity within industries and/or homogeneity across industries will reduce or eliminate the incremental benefits of industry-level forecasting models". Cost structures exhibit even greater variance than revenues within industries (Fairfield et al. 2009).

2.4 Forecasting errors and uncertainty factors

Uncertainty about the future scenario affects the predictions of future performance and makes predicting future performance more difficult (Pastor & Veronesi 2003). Especially in the case of young and small firms uncertainty contributes to the high valuations (Pastor & Veronesi 2003, Baker & Wurgler 2006). The increase of new lists has impacted on the average volatility of expected profitability and decreased the number of dividend paying firms (Pastor & Veronesi 2003). Gu and Wang (2005) find that more uncertain prospects are related to innovative technologies. Earnings forecasting is documented to be especially difficult for high-intangible firms (Barron et al. 2002). Uncertainty not only describes the negative outcomes and riskiness of the company but also positive unexpected performance. There are several studies focusing on the analysts' forecasting errors and the factors affecting the precision and bias of the forecasts.

Several researchers argue that the increase in the number of intangible intensive firms and the complexity of intangible information have made predicting future performance even more difficult (Higgings 2011, Gu & Wang 2005, Barron et al. 2002). Forecasting errors produced by analysts and empirical models, for example random-walk and extrapolative models, are documented to be positively related to the magnitude of intangible assets and R&D costs (Barron et al. 2002, Gu & Wang 2005, Higgings 2011). However, analysts' forecasting accuracy is found to be better than the accuracy of empirical models especially for high-R&D firms (Barron et al. 2002). Still, the analysts' consensus is lower and associated with more uncertainty when analyzing the high-technology manufacturing companies, for instance electronics, drug and software, because of relatively high R&D expenditures (Barron et al. 2002). Due to the information complexity of intangibles, it has been argued that analysts place greater relative emphasis on their private information when deriving estimations for intangible intensive firms (Higgings 2011, Gu & Wang 2005, Barth et al. 2001, Barron et al. 2002). Nevertheless, analysts have an incentive to cover the firms with higher intangibles and R&D costs because the market prices are less informative (Barth et al. 2001, Gu & Wang 2005). Pastor and Veronesi (2003) suggest that analysts' coverage could be one proxy for uncertainty. However, IBES mainly covers big and established firms and there is lack of analysts following small firms (Pastor & Veronesi 2003).

Uncertainty about the future performance is often linked to the age of the firm. Learning phenomena cause a decline in uncertainty when firms mature. Investors are, on average, too optimistic about the future profitability of young companies (Pastor & Veronesi 2003). Earnings volatility tends to be higher for younger firms. However, profitability is assumed to be mean reverting over time as discussed in Chapter 2.2.1. Pastor and Veronesi (2003) argue that the high valuations of young firms are an indication of overoptimism. Baker and Wurgler (2006) analyze investor sentiment and firms' characteristics compared with firms' valuations. The researchers emphasize that the firms linked to subjective valuations are typically small, young, unprofitable, highly volatile, not paying dividends, distressed and extreme growth stocks with lack of earnings history. Investor sentiment has larger effects on these kinds of securities and increases uncertainty and speculative trade (Baker & Wurgler 2006). Several uncertainty characteristics are recognized in the firms operating in high-tech industry. Typical high-tech firm-specific factors are discussed in Chapter 3.2.

3 Prior research on high-tech's future performance

The goal of this chapter is to review the prior research on high-tech companies' future performance. Prior literature is mainly focused on the stock performance rather than operating performance. In this chapter, the industry specific factors and prior empirical research on high-tech firms' performance are covered. Firstly, the development of high-tech IPO markets, the most common high-tech specific factors and the financial reporting of high-tech firms are described. Then, the prior literature on financial statement information linked to the operating performance and the valuation of high-tech firms in IPO and post-IPO situations is reviewed.

3.1 High-tech IPO markets

> *"The high-tech IPO is a mysterious beast. It's attractive, seductive and irresistible. But it's also fickle, temperamental and not always well-behaved. Still, investors have a difficult time resisting the high-tech IPO even when the fundamentals aren't solid or even exist."*
>
> *(Evans 2010)*

> *"US IPO markets were rejuvenated by small-cap high-tech and energy companies"*
>
> *(Ernst & Young, Global IPO trends 2011)*

Social media rushing to markets, speculations about the listings of Facebook, LinkedIn, Twitter, Hulu and Skype, second IT bubble, another sign of the technology apocalypse, Silicon gold rush 2.0: The next wave of high-tech IPOs. All the previous wordings describe the recent opinions of financial press. After couple years of downshifting the high-tech industry is rising again. Discussions about the performance of high-technology industries are often connected with the economic growth of economies, the source of employment growth, profits, innovation and technological opportunities in products and production processes (see e.g. OECD Science, Technology and Industry Scoreboard 2005, Tong 2005, Kask & Sieber 2002). Technological markets are characterized by high growth potential but also by a high degree of uncertainty. Uncertainty makes forecasting the performance of high-tech companies exceptionally difficult. The challenge to predict future performance was remarked especially during the dot.com bubble from the late 1990s to 2000 when the industry exhibited investors' extreme high-sentiment period and the subsequent crash. High-technology and particularly Internet stocks were popping to public trade. The dot.com bubble was a rather unique event in the industry and also extensively studied (see e.g. Trueman et al. 2001, Penman 2001, Demers & Lev 2001).

The increase in the listings of high-tech companies has driven also the whole IPO markets. In fact, Demers and Joos (2007) document that the proportion of high-technology and especially Internet IPOs has been increasing rapidly relative to nontechnology firms in the U.S., particularly since 1995. Technology stocks began their stroke in IPO markets and increased the amount of new lists already since the 1980s but high-tech industries have jumped into picture after the born of New Economy (Ritter & Welch 2002). Schultz and Zaman (2001) argue that one possible motive behind the numerous listings of Internet companies is that the firms attempt to grab market share in industries where there are large economies of scale. The significant number of high-tech IPOs has led to a situation where high-technology IPO offerings have become more important key drivers for economy-wide growth (Ritter & Welch 2002, Fama & French 2004).

The listing of Netscape (an Internet browser company) in 1995 started a new era in business called the New Economy. Especially in North America, the financial market has facilitated the growth of new lists. Venture capitalist have helped the startup companies by financing companies' innovative ideas and IPO market participants have been supporting by buying the future shares. In fact, it has been argued that the possibility of big gains from innovations has changed the investment incentives towards investing in high-risk new businesses. The financial market has made these kinds of investments possible by offering a liquid stock market also for the small and risky stocks. Financing has pushed the innovation and productivity forward. (Mandel 2001) However, high-technology industry is widely recognized as hard-to-value and forecast due to several special challenges in accounting data, performance and company characteristics. These characteristics are described below in Chapter 3.2.

The market has exhibited several waves of IPO "hot issue markets". Hot issue market is defined as a situation where there is a significantly greater number of new lists and higher average initial returns. The situation leads to a chain effect where there is an excessive investor demand for the IPOs due to the higher initial returns and, on the other hand, the high demand attracts new firms to enter the market. Regardless, Schultz and Zaman (2001) find only weak support for the assumption that Internet firms are going public to sell overpriced stock. It has been argued that also new lists with poorer economic condition ran into markets. This kind of rush of several new lists and high valuations were experienced at the turn of the 21st century in the technology and especially Internet industry. However, the bubble burst after the boom and several companies were in big trouble. (See e.g. Loughran & Ritter 2004)

3.2 High-tech firm specific characteristics and financial reporting

High-tech business has its own special characteristics which separate it from the more established industries. Even though there are several different definitions for high-technology industry or company, common features can be recognized. Nevertheless, some of the industries included in the high-technology sector vary widely in their characteristics and performance (Kask & Sieber 2002). The shared characteristics and their occurrence in the financial statements are described in this chapter.

3.2.1 Firm specific characteristics

<u>Research and development expenditures and investments in intangible assets</u>

First of all, the research and development based business model is one of the most commonly linked features to high-tech. Not only the high level of expenditures on research and development but also the substantial investments in intangible assets are engaged in the development of high-tech companies (Demers & Joos 2007, Hand 2005, Jain et al. 2008). Keating (2000, 163) lists that Internet firms are characterized by "high degree of innovation, and reliance on intangible assets, such as patents, software, marketing and promotions, brand name, reputation, customer satisfaction, and knowledge management". Investments in intellectual property are difficult to interpret in the current financial reporting environment. R&D costs are generally expensed and not capitalized. This results in heavy losses especially in the early stages of high-tech companies' life span (see e.g. Demers & Joos 2007). Demers and Joos (2007) document that accumulated deficits and net losses represent relatively greater proportion in high-tech than non-technology companies a year before the IPO. Spending on research and development can also be recognized in the cash flow statement as negative cash flow from operations even if there were successive years behind (Demers & Joos 2007). However, investors expect the high intensity of intangible assets and research and development costs to create returns in the future (Trueman et al. 2001, Kask & Sieber 2002). Accounting and reporting matters related to the industry specific characteristics are discussed separately in the next Chapter 3.2.2.

<u>Age and size</u>

Typically high-technology companies, especially IT companies, are young firms with lack of historical financial information available. Lack of financial statement information poses difficulties for time-series analysis (Trueman et al. 2000, 2001, Demers & Joos 2007). Schultz and Zaman (2001) document that Internet companies are going public earlier in their life-cycle than other firms. Demers and Joos (2007) find that the average age at the time of IPO for high-tech companies was 10 years and for non-technology companies 17 years.

Bartov et al. (2002) conclude that Internet IPO firms are also smaller than their counterparties measured by sales and total assets. In addition, Internet companies engage fewer employees due to their virtual operation environment (Bartov et al. 2002).

Growth and profitability

Even though the high-tech companies are typically rather small when listing to public markets, they are reporting increase in operations and are characterized by substantial growth opportunities. For instance, Internet industry reports significantly greater annual sales growth than other industries (Bartov et al. 2002). However, the rapid and unpredictable growth pace complicate the forecasting of future performance (see e.g. Trueman et al. 2000, 2001, Bartov et al. 2002, Kask and Sieber 2002). In addition to substantial growth rates, particularly Internet companies exhibit high cash burn rates (Jain et al. 2008). Cash burn was particularly high during the Internet boom (Mudambi & Zimmerman Treichel 2005). Eisenmann (2006) examine the Internet firms' growth strategies and find evidence that heavy early investments were economically rational and provided long-term returns. Also Demers and Lev (2001) document that cash burn is significantly and positively related to price-to-sales ratios of Internet companies. Still, the outflowing cash may question the long-term economic viability of these firms (Jain et al. 2008). Bartov et al. (2002) argue that the lack of profitability creates uncertainty about the future. Internet companies are generating significant and often growing losses (Bartov et al. 2002) and they are predominantly unprofitable at the time of going public (Jain et al. 2008). In fact, Hand (2000) finds that the past, present and expected future profitability of Internet firms is dramatically less than the profitability of both seasoned non-Internet firms and IPO-matched non-Internet firms.

Risk and capital structure

As high-tech firms are rather young but developing in a fast pace, the companies require significant support for financing their operations (Jain et al. 2008). Demers and Joos (2007) document that high-tech companies are predominantly equity-financed. The high-tech start-up companies demand significant up-front financing to establish the technological architecture and achieve the critical mass of clientele in order to attain profitability (Demers & Lev 2001, Bartov et al. 2002). Firms often report negative cash flows and they are not self-sufficient, hence, firms are very dependent on external financing (Jain et al. 2008). The usage of external financing results also in substantial capital expenditures (Jain et al. 2008). Regardless, Demers and Joos (2007) document that the leverage ratio was higher for the nontechnology than high-technology firms a year before the IPO.

Involvement of venture capitalists is significant especially in the earlier stages of high-tech firms' life cycle. Demers and Joos (2007) document that in their sample 62 % of high-tech and 24 % of nontechnology companies were backed by venture capitalists. Thus, "venture capitalists play a significant role in shifting the risk of financing developmental firms from the private equity to public equity markets" (Jain et al. 2008, 190). Mudambi and Zimmermen Treichel (2005) argue that IPO might be an exit mechanism to move to the next project or "cashing in" for the company entrepreneurs. However, Hand (2005) states that venture-backed technology firms involve more risk. Especially Internet companies struggle for survival during the post-IPO phase as their profitability is questioned and growth rates are especially high (Jain et al. 2008). In fact, Demers and Joos (2007) made a comparison of forecasting failure risk between high-tech and nontech companies. The researchers report that in their sample approximately 17 % of nontech IPOs and 9 % of high-tech IPOs failed within five years of going public. This finding is slightly surprising as often, especially Internet companies, are claimed to be risky in their nature (see e.g. Hand 2005).

3.2.2 Financial reporting

Keating (2000, 163) captures the essential idea of analyzing the financial performance of high-tech companies: "New economy firms are of particular interest because of concerns about whether the traditional accounting model can be used to value firms in instances where most of that value consists of growth opportunities". Of course, the primary assumption is that the reporting standards (e.g. US GAAP) are designed to apply equally to all types of businesses (Hand 2005). However, financial statements are typically more informative for seasoned firms (Penman 2001). Informativeness indicates that the current figures in financial statements are describing also the future steady state (Penman 2001).

Bowen et al. (2002) examine the financial reporting of Internet companies. The researchers focus on the motives and factors affecting accounting choices and revenue-recognition policies of Internet companies. Financial press and reporting regulators (e.g. the Securities and Exchange Commission and Financial Accounting Standards Board) have expressed concern about the pressure on Internet firms to report high levels of sales (Bowen et al. 2002). In addition, Jain and Kini (1994) note that one of the potential explanations for the decline in post-IPO operating performance is related to managers' possible window-dressing of their accounting figures prior going public. However, Bowen et al. (2002) state that unlike most profitable companies, Internet firms have been alleged to manage the components of earnings rather than the net income itself. Owner-controlled firms are more carefully monitored and, hence, appear to engage in less earnings management (Bowen et al. 2002, Warfield et al. 1995). Bowen et al. (2002) recognize several alarming revenue-accounting practices used by Internet firms, including reporting barter revenue, reporting grossed-up (as opposed to net) revenue, and excluding the effect of coupons, discounts, and loss leaders from revenue. For instance, barter revenue typically originates from the exchange of advertising space. In other words, the firm reports equal amounts of barter revenue and advertising expense, hence, the practice does not affect the net income of the company.

Bowen et al. (2002, 526) lists the incentives why managers tend to report relatively higher revenues:

1) to raise additional equity,
2) to use their stock to acquire other companies and
3) to hire and retain key employees with stock-based compensation.

However, Bowen et al. (2002) note that the incentive to manipulate revenues and not the net income likely stems from the assumption that revenues play an important role in the valuation of Internet firms. Recent academic research has provided evidence of linking the valuation of Internet firms to revenues (see e.g. Hand 2000, Bowen et al. 2002). Fast growth in revenues is perceived as proxy for future financial success while most Internet companies are reporting negative earnings. As a result, analysts report and follow price-to-sales ratios (Demers and Lev 2001). Regardless, Bowen et al. (2002) find that coefficient for earnings is statistically significant and positive in relation to the firm value indicating that also earnings are associated with the valuation of Internet firms.

Hand (2005) examines the value relevance of financial statements in the venture capital market of the U.S. based biotechnology firms. The study clarifies the reporting condition of firms before entering the public market. The researcher compares nonfinancial information with financial information as well as financial reporting in pre-IPO venture capital market with reporting in post-IPO equity market. Hand (2005) finds that financial information is highly value-relevant in the venture capital market. In addition, equity values and financial statement information are similarly associated in the venture capital market and public equity market. The value relevance of financial statements increases whereas the value relevance of nonfinancial information decreases as firms mature (Hand 2005). Hand (2005) explains the development by the transform of intangible assets to assets-in-place. In the earlier stages of firms' life cycle when operating in the venture capital market, the nonfinancial information is more value relevant than the financial statement reporting (Hand 2005). However, the value relevance of financial statement and nonfinancial information reverse in importance at the time of IPO. Hence, nonfinancial information is momentarily more value-relevant than financial information. Hand (2005) argues that the IPO investors are less sophisticated than venture capitalists in pre-IPO situation and investors investing in seasoned post-IPO firms. In addition, Amir and Lev (1996) document that nonfinancial indicators are value-relevant particularly in industries characterized by high growth opportunities and substantial intangibles. Growth opportunities and intangible assets are asserted to be poorly captured by the accounting system (Amir & Lev 1996).

3.3 Operating performance and valuation of high-tech companies

3.3.1 Operating performance

IPO literature is typically more focused on the valuation and post-issue stock price performance than on the operating performance (Jain & Kini 1994). IPO is a moment when the investors have the first opportunity to value the company’s set of assets and future prospects (Aggarwal et al. 2009). The survival rate of newly listed companies has declined dramatically over the past several decades, both in absolute terms and relative to seasoned firms (Fama & French 2004) and failure risk is one of the key topics in the literature. Eight most popular topics or measures of performance in the new ventures literature include efficiency, growth, profit, size, liquidity, success/failure, market share, and leverage (Murphy et al. 1996).

Several studies find that initial public offerings underperform after the issue (see e.g. Fama & French 2000). Jain and Kini (1994) investigate the change in operating performance when firms make the transition from private to public ownership. The data is collected from 1976 to 1988 and the sample consists of firms across industries, thus, the research is not high-technology specific. Jain and Kini (1994) document a significant decline in post-issue operating performance relative to pre-IPO levels as measured by the operating return on assets and operating cash flows on assets. The results are robust to industry adjustments. IPO firms exhibit high growth in sales and capital expenditures relative to seasoned firms in the same industry (Jain & Kini 1994).

Jain and Kini (1994) find no relation between post-issue changes in operating performance and initial returns at the IPO. In fact, the decline in post-issue operating performance is inconsistent with the fact that IPO firms are initially priced at high price-earnings (P/E) multiples. High initial P/Es indicate that investors have expectations of high earnings growth in the future. IPO firms start with high market-to-book (M/B) and P/E ratios relative to their seasoned industry counterparties but experience a decline in the ratios after the IPO. In addition, earnings per share (EPS) decline over time and even the pre-IPO profit margins are not sustained. (Jain & Kini 1994)

Jain and Kini (1994) argue with possible explanations for the decline in operating performance. Change in the ownership structure when going public may increase the agency costs of the company. In addition, prior going public, the managers may have an incentive to window-dress their accounting numbers. As a result of an IPO, the company receives additional money for their operations. There is a risk of using the proceeds from the IPO in non-value maximizing projects. Entrepreneurs might time the listing to a period of unusually good performance and it is questionable if this level of performance can be sustained in the future. In fact, Ritter (1991), and Loughran and Ritter (1995) document IPOs' long-run underperformance and suggest that the decline in operating performance is not anticipated. Investors are constantly surprised by the poor performance of IPO firms.

High-tech companies go public earlier in their life cycle when there is only a promise of profitability (see e.g. Ritter & Welch 2002, Jain et al. 2008, Janey & Folta 2006). The lack of profitability creates uncertainty about the future prospects amongst the investors (Janey & Folta 2006). However, prior studies have widely discussed if losses in the earlier stages of firms' life cycle can actually be an indication of future success (see e.g. Jain et al. 2008). Profitability is strongly linked to earnings forecasting and long-term economic viability. Jain et al. (2008) study the factors influencing the probability and timing of post-IPO profitability of Internet firms. Mudambi and Zimmermen Treichel (2005) argue that pre-IPO profit cannot be employed because pre-IPO profit is found to be an unreliable measure of a new venture's financial position. The risk of post-IPO failure is especially high for unprofitable firms. Jain et al. (2008) find that the same factors that impact on the probability of post-IPO profitability also tend to impact on the survival probability of Internet IPO firms. Failure risk of high-tech IPOs will be discussed later in this chapter.

Jain et al. (2008) find several factors affecting the probability and timing of profitability. The probability of post-IPO profitability for Internet firms decreases with an increase in the size of the offering, valuation uncertainty (measured by divergence in the opinions of the institutional investors on the future prospects of the IPO firm), venture capitalist participation, and proportion of outside board members. Interestingly, the proceeds raised in the IPO are negatively related to the probability of profitability but positively related to the time-to-profitability. However, extra cash is found to increase incentives to invest in negative net present value projects resulting in a lower probability of attaining profitability and a longer time-to-profitability. In contrast, an increase in firm age and number of employees, pre-IPO investor demand and presence of insiders in the board are all associated with a higher probability of post-IPO profitability. (Jain et al. 2008)

Motivated by the facts that a dramatic decline has been reported in the survival rates of newly listed firms (Fama & French 2004) and that little has been documented concerning firm-specific accounting variables that are associated with IPO firm failure, Demers and Joos

(2007) examine factors related to IPO failures and develop an IPO failure prediction model. The purpose of the study is to test whether financial statement information captures the risk factors of different sectors. There are significant structural differences in failure models for nontech and high-tech IPO firms. Demers and Joos (2007) identify that heavy investments in intangible assets, equity-financed business, and losses from the high R&D expenditures differentiate high-tech from nontech sector. Interestingly, the time period for the realization of ultimate failing is longer for high-tech firms because the outcome of R&D activities reveals only after a longer period of time. (Demers & Joos 2007)

Accounting variables increase significantly the explanatory power of the developed failure forecasting model for both nontech and high-tech samples, especially for the latter (Demers & Joos 2007). For example, leverage has been recognized to be an important predictor of firm failure. Adequate financing is vital for high-tech firms because they are predominantly suffering from negative cash flows (Jain et al. 2008). Inability to raise additional capital may quickly lead to delisting or even bankruptcy. However, as the high-tech companies are more equity-financed, the leverage ratio may differ substantially from the one of non-high-tech companies (Demers & Joos 2007). Demers and Joos (2007) find that post-IPO abnormal returns, logAge, gross margin and logSales are significantly negatively associated with failure probability. For example, younger firms are more likely to fail. Demers and Joos (2007) argue that sales might indicate of the firm's stage of development and/or sales are proxy for size. Larger firms tend to have a lower risk to fail. Debt, logSGA and accumulated deficits are significantly and positively associated with the likelihood of failure. Demers and Joos (2007) assert that accumulated deficits capture more of the increased uncertainty in the firms' business than the reduced risk deriving from past successful investments in intangible assets. High spending on R&D in pre-IPO situation which is very typical for high-tech industry is associated with a significantly lower failure probability in the high-tech sample but not within nontech firms. Firms listed during hot-issue market tend to have higher probability to fail within five years of the listing.

Few researchers have studied earnings forecasting in the high-tech context. Trueman et al. (2001) examine whether past revenues, web usage data, and analysts' estimates are useful in forecasting the future revenues of Internet firms during the years 1998-2000. The sample consists of 95 companies. The research is executed as a comparison of several time-series forecasting models, each of which is based on historical revenues. Trueman et al. (2001) divide the Internet companies into two groups: e-tailers who sell goods or commodities to consumers electronically in Internet, and portal and content/community providers (the p/c firms). In general, the forecast model assuming constant change in quarterly revenue outperforms the other time-series models tested. Current and past revenue growth has significant incremental predictive power for the p/c firms but not for the e-tailers. Furthermore, significant relation is also found between the current revenue growth and the growth in web traffic measured by unique visitors, pageviews and minutes spent at the internet firm's website. The relation is stronger for the e-tailers than for the p/c firms. The most interesting finding is the correlation between the analysts' forecasts and realized revenues. Trueman et al. (2001) find that analysts' forecasts are highly correlated with realized revenues but with positive mean bias. This means that analysts generally underestimate the Internet firms' revenues. The researchers assert that the behavior can be explained by two possible ways; either the fast growth of Internet firms impedes forecasting the future prospects or then the analysts have motivation to underestimate their forecasted results in order to allow firms to report positive revenue surprises.

3.3.2 Financial statement information in relation to valuation

High-tech companies are generally recognized as hard-to-value companies due to, for example, the rapid changes in the industry and significant growth rates. These companies are also young with few years of historical data available (Trueman et al. 2000, 2001, Demers & Joos 2007). High growth rates and young age cause difficulties in the valuation and forecasting future profitability of high-tech firms compared with more established, seasoned firms (Trueman et al. 2000). The market values of Internet stocks before the year 2000 were on average several times greater than the residual income intrinsic valuations suggested (Hand 2000).

Penman (2001) calls for new approaches in stock valuation due to the changes brought by the New Economy and the following dot.com bubble. The high valuations attached to Internet firms with little or no sales or earnings have motivated several studies to explain the pricing of Internet stocks (Schultz & Zaman 2001, see e.g. Bartov et al. 2002, Trueman et al. 2000, Demers & Lev 2001, Trueman et al. 2001). Demers and Lev (2001) argue that traditional financial statement information might not be relevant in the valuation of Internet stocks. Also Trueman et al. (2001) assert that past revenues may have only limited usefulness for forecasting purposes. The prior research findings on the usefulness of accounting information in valuation are mixed (see e.g. Demers & Lev 2001, Hand 2000, Bowen et al. 2002, Trueman et al. 2000). Hand (2000), Bowen et al. (2002) and Aggarwal et al. (2009) document that financial statement variables, not only sales, are significantly associated with the market values of Internet companies. Non-financial information, for example pageviews and stickiness of customers (see e.g. Demers&Lev 2001), has been raised as an alternative approach to assess the future performance especially in Internet industry. However, this study focuses on the usefulness of financial statement information and non-financial indicators are only mentioned.

The traditional evaluation of stocks includes the analysis of the firm's price-to-earnings ratio (P/E-ratio) but, in the case of high-tech stocks, price-to-sales is commonly referred metric by the analysts and investment professionals (Demers & Lev 2001, Trueman et al. 2001). The researchers argue that P/E-ratio is not applicable for the high-tech industry, especially IPOs, since most companies are not yet profitable (Demers & Lev 2001, Trueman et al. 2001). Similarly, the usage of book-to-equity ratio has been questioned because high-tech companies have few tangible assets. Market-to-book ratio tends to "blow up" because the high-tech companies have weaker balance sheet resulting in a small denominator (Demers & Lev 2001). Contrary to Demers and Lev (2001) and Trueman et al. (2001), Bowen et al. (2002) find that the coefficient for earnings is positive and statistically significant indicating that earnings are also associated with the valuation of Internet firms. Also Penman (2001) reminds that investors are interested in the earnings and/or cash flows that the sales will generate in the future. Therefore, profitability and earnings are still valuable in forecasting the future performance.

Bartov et al. (2002) examine the association between the valuation of Internet IPOs and a set of financial and non-financial variables. The researchers aim to recognize the differences in IPO valuations between Internet and non-Internet firms. The sample consists of Internet IPOs from 1997 to 1999. Earnings and cash flows are most commonly used financial variables in valuation but due to the lack of historical data of profits and weak book values, investors are suggested to rely more on revenues, for example annual sales growth or sales per share (Bartov et al. 2002). However, there are problems found in the revenue reporting of Internet firms, for instance grossed up and barter revenue (see e.g. Bowen et al. 2002). Bartov et al. (2002) examine also the effect of book value of equity. Prior research on Internet valuation has documented that negative book values may have different implications for Internet firms than for the firms of other industries. Negative book values may indicate of successful investments in R&D and other intangibles and, thus, valued by the stock market (see e.g. Aggarwal et al. 2009). In fact, Bartov et al. (2002) find R&D expenditures and negative cash-flow value-relevant whereas earnings are not priced in the case of Internet firms.

Aggarwal et al. (2009) examine how IPO valuations have changed over time by comparing three periods: 1986 – 1990, January 1997 to March 2000 (designated as the boom period) and April 2000 to December 2001 (designated as the crash period). The researchers divide their sample of US IPOs to technology and nontechnology as well as to Internet and non-

Internet firms. Aggarwal et al. (2009) include the variables of income, book value of equity, sales, R&D, industry price-to-sales ratio, insider retention, and investment banker prestige ranking in their valuation model. The valuation model applied consists of three components: the replacement cost of the firm's physical assets in place, the net present value of the firm's expected future cash flows from assets in place, and the value of growth options associated with a future technological upgrade (Aggarwal et al. 2009, 254). The researchers expect that the Internet boom period was generally a period of rapid technological upgrades and negative earnings are likely stemming from investments in these technological upgrades. R&D spending is predicted to be associated with technological upgrade.

Aggarwal et al. (2009) document that the replacement cost of physical capital (book value) is associated with IPO valuation only during the crash period. Tangible assets became more important after the Internet bubble. Income is found to be weighted more whereas sales are weighted less in the valuation during the boom period compared with the late 1980s. These results contradict with the argument that sales are more significant than earnings in the valuation of the high-tech companies (see e.g. Demers & Lev 2001, Trueman et al. 2001). Findings of growth options are mixed. In general, the proxies of growth options, such as R&D expenditures or industry price-to-sales comparables, are associated with greater valuation. However, these measures of growth options are not consistently correlated with any specific industry or with valuation over time. More significant correlation with the boom period or with the Internet or technology firms would have been expected. Negative earnings are considered to be one proxy for growth options especially with Internet companies and, interestingly, higher stock values have been reported for firms with larger negative earnings. In fact, Aggarwal et al. (2009) document a V-shaped relation between firm value and earnings. In other words, firms with more negative earnings have higher valuations than firms with less negative earnings and firms with more positive earnings have higher valuations than firms with less positive earnings. These results are consistent with the findings of Bartov et al. (2002).

Also Hand (2000) examines the relevancy of financial statement variables in stock valuation. The researcher investigated the quarterly data of 167 publicly traded Internet firms from Q1 1997 to Q2 1999. Hand (2000) focuses on the traditional valuation principles where earnings are priced in the stock value. In fact, Hand (2000) finds that basic accounting data is highly value-relevant in a simple non-linear manner. Hand (2000) examines particularly the relation of earnings, book equity, revenues, profitability and R&D as well as marketing expenses to stock value.

One of the most essential findings of Hand (2000) is that Internet firms' market values are concave and increasing with positive net income while concave and decreasing with negative core net income. This is an interesting finding compared with Aggarwal et al.'s (2009) observation of V-shaped relation between the firm value and earnings. Contrary to the suggested usefulness of sales in valuation, Hand (2000) finds only weak association between revenues and firm value. However, revenues are positively priced in a concave manner. When it comes to selling and marketing expenses, those are positively and concavely related to market value, particularly during the first two quarters after the IPO but only when negative earnings are reported. Similarly, R&D expenditures are priced in a positive and concave manner but the relation is more sustainable over time than with the selling and marketing costs. Hand (2000, 5) states that "Net firms' lack of profitability has its roots in, but is not entirely explained by, their huge investments in intangible marketing brand assets aimed at rapidly seizing a dominant market-share position". Profitability only weakly improves when the Internet companies are maturing. (Hand 2000)

Trueman et al. (2000) document controversial results compared with Hand's (2000) findings but consistent results with the argument that financial statement information can only be limitedly used in the valuation of Internet companies. Trueman et al. (2000) find no significant association between core net income and market prices. However, when net income is decomposed, Trueman et al. (2000) document a significant and positive association of gross profits with prices. The researchers explain the results with the effect of large transitory items, for example merger-related costs, and R&D costs included to bottom line. Also Keating (2000) notes that most Internet firms are reporting substantial current losses and conservative accounting choices, for instance merger and intangible related, may shift the earnings between current and future periods. Investors tend to consider these R&D costs as investments rather than expenses. On the other hand, gross profits are often considered reflecting the current operating performance and perceived as a more permanent component (Keating 2000). Trueman et al. (2000) analyze also the role of marketing costs and they conclude that marketing costs are viewed more as an expense than as an investment. Marketing costs received a negative coefficient related to stock price. Keating (2000) adds that the proceeds from the recent IPO are often spent on marketing and R&D which creates a correlation between the book value and net income components.

Demers and Lev (2001) note that significant up-front capital expenditures are required in high-tech and especially in Internet business. Companies have to build the technological architecture and achieve the sufficient mass of customers in order to attain profitability. Firms need funds to meet these goals. Demers and Lev (2001) examine the role of R&D and cash burn in companies' operations and compare these actions with the behavior of the investors.

The researchers find that investors turned to more skeptical attitude towards R&D expenses as the industry matured. Before, when the prospects were brighter, the investors tend to more capitalize the R&D and customer acquisition costs, for instance advertising costs. After IT bubble burst, approximately in the spring of 2000, only the R&D costs were capitalized into value. Demers and Lev (2001) note similar kind of behavior also in the attitudes towards cash burn. Market was favorable to aggressive cash expenditures in 1999 but turned to more critical in 2000. However, Demers and Lev (2001) document that cash burn is significantly and positively associated with price-to-sales ratio.

In a conclusion, the papers reviewed mainly focus on Internet industry, the U.S. based firms and a certain short period of time. Furthermore, the results on the usefulness of financial statement information in valuation were mixed. There are also some statistical problems identified which have to be taken into account in the interpretation of the prior studies. Keating (2000) argue that Trueman et al. (2000) do not raise the issue of sample selection bias since there are only 63 publicly traded Internet companies in the sample. Also Demers and Lev (2001) examine whether the financial and nonfinancial variables explain the market values with a sample of only 84 Internet companies during 1999 and 2000. Aggarwal et al. (2009) note that in the prior IPO studies either the sample sizes are small or very industry-specific or only IPOs with positive earnings are taken into consideration. Penman (2001, 363) notes that "in a small, time-dependent samples, there is a probability of getting contrary results for one partition of the sample that may disappear under further replication". Keating (2000) continues that the stock price data may be cross-sectionally dependent in a small sample of Internet firms meaning that the stock prices can be expected to move simultaneously in response to specific earnings announcements or other news events creating cross-correlations in the prices. In addition, prior studies are generally focused on certain very unusual period of time such as the Internet bubble. Hence, the generalizability of these studies has to be carefully rethought.

4 Research design

4.1 Predictions

In this chapter, the key predictions of the study are constructed. Firstly, financial statement variables are examined whether those explain the future earnings of high-technology IPO companies. The forecasting horizons are set to one, three and five years ahead. Accounting is argued to poorly capture the economics of high-technology firms (see e.g. Lev and Zarowin 1999, Francis and Schipper 1999). There are also mixed results whether the financial statement information has explanatory power for the valuation of high-technology companies (see e.g. Trueman et al. 2000, 2001, Hand & Lev 2000, Bartov et al. 2002, Demers & Lev 2001, Hand 2000, Bowen et al. 2002). In this study, the model of Hou et al. (2010) is tested. The earnings forecasting model is mainly based on accounting variables and employs cross-sectional data. The model and samples applied in this study are presented in the following Chapters 4.2 and 4.4, respectively. In addition, the expected signs of the variables are discussed in Chapter 4.3.

Fama and French (2000, 2006) Hou et al. (2010), Hou and van Dijk (2010), and Hou and Robinson (2006) document that future earnings are attributable to basic accounting variables. Based on the results of Fama and French (2000, 2006) Hou et al. (2010), Hou and van Dijk (2010), and Hou and Robinson (2006), the explanatory power of accounting variables in association with future earnings is studied with a following prediction.

P_1: Financial statement information has explanatory power in relation to the future earnings of newly listed high-tech companies.

The second approach of the thesis is to study the predictive power of accounting variables and analyze factors impacting on the observed forecasting errors. At this point, there are two subsamples constructed: high-tech IPO companies and all listed high-tech companies. Forecasting accuracy of the model is then compared between the two samples. The analysis includes examination of the bias and precision of the forecasts. The performance of IPO companies is recognized to be more difficult to predict than the performance of seasoned firms since the IPO firms are predominantly young with lack of historical data and report more volatile earnings (see e.g. Pastor & Veronesi 2003).The predictive power of accounting variables is examined with the following prediction.

P_2: Forecasts of earnings are less accurate for newly listed high-tech companies.

After analyzing the predictive power of financial statement information, the firm-specific uncertainty factors about future performance are tested whether the factors explain the forecasting errors detected. Uncertainty about the future affects the predictions of future performance (Pastor & Veronesi 2003). Uncertain prospects may also result in unexpected positive performance. Future performance of small, young, unprofitable and highly volatile growth stocks with lack of earnings history is typically difficult to forecast (see e.g. Pastor & Veronesi 2003, Baker & Wurgler 2006). The relation of uncertainty factors to forecasting errors is examined with the following prediction.

P_3: Forecasts of future earnings are less accurate for high-tech companies with more uncertainty about the future performance.

4.2 Empirical model

This study replicates the earnings forecast model in Hou et al. (2010). The model of Hou et al. (2010) is based on the prior research of several earnings forecast models. The model is strongly linked to studies of Fama and French (2000, 2006), Hou and van Dijk (2010), and Hou and Robinson (2006). The model is selected since it covers a large number of accounting variables which presumably have explanatory and predictive power in relation to future earnings. In addition, the model produces absolute dollar earnings forecast which are comparable with analysts' estimates. However, compared with the analysts' estimates, the applied technique does not exhibit the biases documented in analysts' forecasts and the model is found to have higher levels of earnings response coefficients (Hou et al. 2010).

The model of Hou et al. (2010) is not yet tested in the context of high-technology industry. Regardless, it has received good results in terms of forecasting accuracy and explanatory power (adjusted R^2) in a large pooled cross-sectional analysis of U.S. companies during 1967-2005. Also Lee et al. (2011) have employed the model of Hou et al. (2010) to estimate the future earnings in their study. Hou et al. (2010) document that in their sample the average adjusted R^2 of the regressions for earnings of one, two, and three years ahead are 87%, 81%, and 77%, respectively. Based on these findings, the model seems to capture most of the variation in future earnings. Furthermore, one benefit of the model is that the variables are known at the time of the forecast. The equation used in the thesis and modified from Hou et al. (2010, 7) is presented below.

$$IncExt_{i,t+\tau} = \alpha_0 + \alpha_1 MCap_{i,t} + \alpha_2 TAssets_{i,t} + \alpha_3 DIV_{i,t} + \alpha_4 DD_{i,t} + \alpha_5 IncExt_{i,t} + \alpha_6 NegIncExt + \alpha_7 TAccr + \varepsilon_{i,t+\tau} \quad (1)$$

where IncExt is the net income before extraordinary items, α_0 is the intercept of the model, MCap is the market capitalization, TAssets is the total assets, DIV is the dividends paid, DD is the dividend dummy for companies not paying dividends, IncExt is the lagged earnings, NegIncExt is the dummy for negative earnings, TAccr is the total accruals and ε is the error term. All the variables are measured at the end of the fiscal year. Variable details are presented next in Chapter 4.3 and in Appendix 1. The studies related to the model used by Hou et al. (2010) were presented in Chapter 2.3.1.

4.3 Variable descriptions

Richardson et al. (2010) emphasize that the selection of the explanatory variables should be based on a theory. However, as recognized in Chapters 2 and 3 on relevant literature, no single theory exists to explain the company performance, especially in the field of high-technology. This study is carried out as a fundamental analysis of accounting variables and their ability to explain and forecast future earnings. Richardson et al. (2010, 419) state: "Consistent with this investor perspective, an important objective of empirical archival research is to understand the properties of financial accounting information and how this information might help generate better forecasts of those investment inputs." Especially in the context of high-technology industry there is lack of studies concentrating on forecasting or explaining future operating performance. Therefore, finding reference variables and models for forecasting future earnings is more challenging.

The explanatory power of accounting variables is examined by analyzing total assets, dividends, lagged earnings and total accruals in respect to future earnings. Following Hou et al. (2010) market capitalization is added to complete the model. Also two dummy variables are constructed. Firms not paying dividends receive the value of one and firms reporting negative earnings receive the value of one. As discussed in the previous Chapter 4.2, the model replicates the model in Hou et al. (2010). Hou et al. (2010) collected the variables for their earnings forecasting model by analyzing the findings in Fama and French (2000, 2001, 2006), Hou and van Dijk (2010), and Hou & Robinson (2006). In this study, only few simplifications are made to Hou et al.'s (2010) variable definitions due to data collection reasons. Hou et al. (2010) have used Compustat which has small differences in variable definitions compared with Thomson Reuters which is used in this study. Variable details with variable codes in databases are reported in Appendix 1.

Firstly, the variable descriptions of the chosen dependent and explanatory variables were analyzed content-wise in both Reuters and Worldscope databases in Thomson Reuters. Then, the different options were compared with Nokia's (ticker: nok1v-he) actual financial statement figures. This kind of analysis was made to identify the variables which are closest to real reported figures and with fewer adjustments made. In a conclusion of the comparison, Reuters variables matched better with Nokia's reported figures. Therefore, Reuters was chosen as a primary source for the variable retrieval.

Next, the expected signs of the variables are discussed.

Future earnings ($IncExt_{i,t+T}$)

The dependent variable of the future earnings is derived from the net income before extraordinary items ($IncExt_{i,t+T}$). Hou et al. (2010) predicted future earnings for one, two and three years ahead. In this thesis, the forecasting horizon is split to three phases: to short-term period considered as one-year ahead forecast, to mid-term period considered as three-year ahead forecast and to long-term period considered as five-year ahead forecast. Motivation for this approach is that presumably high-technology companies are evolving fast and, for instance, five-year period is rather long compared with firms' average age. Therefore, it is interesting to see if the development can be predicted by the accounting figures.

Market value (MCap)

Market value (MCap) of the company is derived from market capitalization on equity. Market value should reflect the current value of future net cash flows and expected earnings. However, the possible mispricing of high-tech, especially Internet, stocks during investors' high-sentiment periods and hot-issue market has been widely studied (see e.g. Ritter 2001). Negative earnings are considered to be one proxy for growth options especially with Internet companies. For instance, Aggarwal et al. (2009) document a V-shaped relation between the firm value and earnings. In other words, firms with more negative earnings have higher valuations than firms with less negative earnings and firms with more positive earnings have higher valuations than firms with less positive earnings. These results are consistent with the findings of Bartov et al. (2002). Also based on the results in Fama and French (2000, 2006) Hou et al. (2010), Hou and van Dijk (2010), and Hou and Robinson (2006), market value is expected to be positively associated with future earnings. In a contradiction, Lee et al. (2011) document a negative relation of market value to the future earnings of four and five years ahead.

Total assets (TAssets)

Total assets (TAssets) are often considered as an indicator of firm size. In the context of high-tech industry, investments in intangible assets and R&D are in a significant role (Demers & Joos 2007, Hand 2005, Jain et al. 2008). However, R&D costs are not recognized on balance sheet. Investors may experience difficulties to value knowledge based assets. The proportion of normal assets-in-place is generally less than in more established companies and industries. Also typically at the beginning of high-technology companies' life span the accumulated deficits have impact on the balance sheet (Demers & Joos 2007).

Based on the results in Hou et al. (2010), total assets are expected to be negatively associated with the future earnings of one to three years but positively associated with the earnings of five years ahead according to Lee et al. (2011).

Dividends paid (DIV) and dividend dummy (DD)

There are two variables in the model for dividends: dividends paid (DIV) and a dummy (DD) for the dividend paying companies receiving the value of zero and firms not paying dividends receiving the value of one. Dividends paid are retrieved from the cash flow statement and not reported dividends are handled as not paid dividends. Dividends are considered having information about future earnings. Fama and French (2001) document that the larger and more profitable firms are more likely to pay dividends whereas firms with heavy investments, efforts in R&D and higher market value of assets to book value are less likely to pay dividends. Furthermore, the number of firms paying dividends has declined dramatically from the late 1970s to the late 1990s (Fama & French 2001). Based on the results in Fama and French (2000, 2001, 2006), Hou et al. (2010), Hou and van Dijk (2010), and Hou and Robinson (2006) dividends paid are expected to be positively associated with future earnings. The prior research document a negative slope for the dividend dummy compared with the future earnings or profitability (Hou & Robinson 2006, Fama & French 2000, 2006, Lee et al. 2011). Consistently with these findings, non-dividend paying is expected to be negatively associated with future earnings.

Lagged earnings (IncExt) and negative earnings dummy (NegIncExt)

Lagged earnings (IncExt) and profitability are documented to be strongly and positively linked to future earnings (Fama & French 2006, Hou et al. 2010, Hou & van Dijk 2010, Hou & Robinson 2006). Profitability and earnings are recognized as very persistent attributes explaining future earnings (Fama & French 2006). Based on the findings in Fama and French (2006), Hou et al. (2010), Hou and van Dijk (2010), Hou and Robinson (2006) and Lee et al. (2011), lagged earnings are expected to be positively associated with future earnings.

Negative earnings of high-technology companies are often explained to be caused by substantial investments in research and development and, thus, to be investments in future earnings (Trueman et al. 2001, Kask & Sieber 2002). Therefore, current negative earnings are expected to imply of future welfare. Negative earnings (NegIncExt) is a dummy variable. When negative earnings are reported the dummy receives the value of one and when positive or zero earnings are reported the dummy receives value of zero. The prior findings on this variable are mixed. Fama and French (2006) document a negative relation of negative earnings to future profitability for the first year whereas Hou et al. (2010) document a positive relation for two and three years ahead but the result is statistically significant only for the regression of three years ahead. Lee et al. (2011) find negative earnings with positive relation to future earnings in two first years but negative relation in the years from three to five. Due to the mixed findings, no expected sign is set for the negative earnings.

<u>Total accruals (TAccr)</u>

Depart from Hou et al. (2010), accruals (TAccr) are calculated as a difference between cash flow from operations and net income before extraordinary items (total accruals). Hou et al. (2010, 6) employ the concept of operating accruals using “indirect balance sheet method as the change in non-cash current assets less the change in current liabilities excluding the change in short-term debt and the change in taxes payable minus depreciation and amortization expense”. This description is not applied due to the low availability of aforementioned financial items in Thomson Reuters.

Accruals are often considered to be a tool of earnings management. Sloan (1996) finds that firms with higher operating accruals tend to have lower future earnings and returns. Fama and French (2006) document that the negative coefficient of the accrual variable compared with future profitability confirms the prior literature findings that accruals result in transitory variation in earnings. Also Hou et al. (2010) and Lee et al. (2011) find that firms with lower operating accruals tend to have higher future earnings. Based on these findings, total accruals are expected to be negatively associated with future earnings. The expected signs of the variables are summarized below.

Variable		Exp. sign
Market value	MCap	+
Total assets	TAssets	-/+
Dividends paid	DIV	+
Non-paying dividends	DD	-
Lagged earnings	IncExt	+
Negative earnings	NegIncExt	+/-
Total accruals	TAccr	-

Table 1. Expected signs of the variables.

4.4 Definition of high-tech industries and sample selection

4.4.1 Definition of high-tech industries

The sample selection and information collection in this study can be divided into five main steps. These steps are illustrated in the figure below (Figure 2) and the text follows these steps.

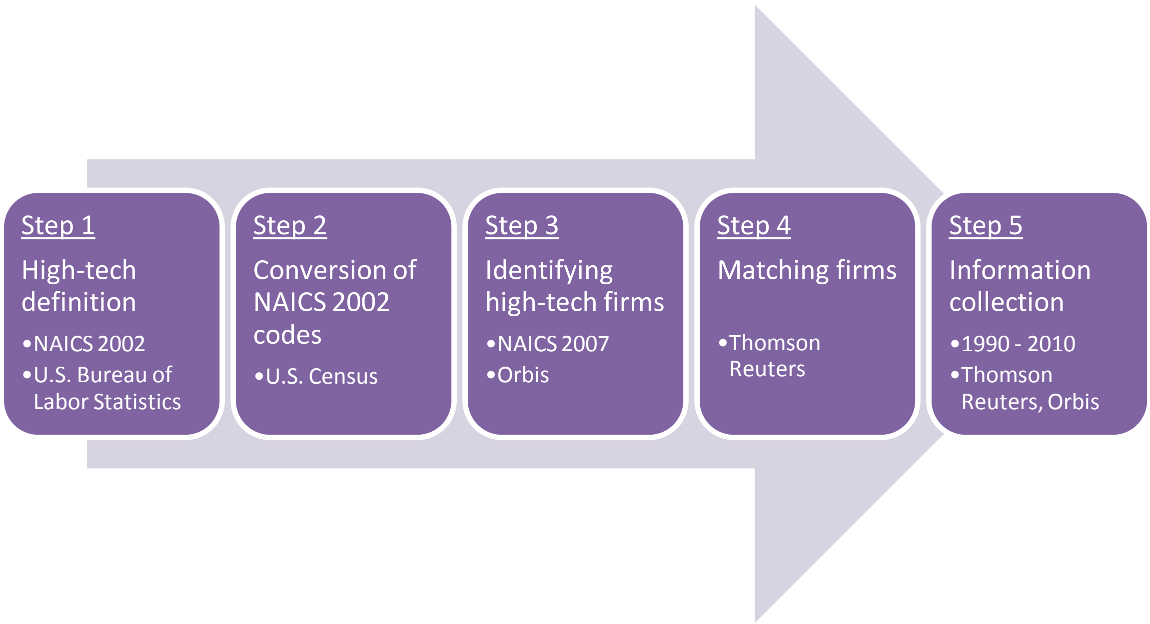


Figure 2. Sample selection and information collection process.

In the first step of the process, high-tech industries are defined. There is no universal definition for high-tech industries. Prior literature on high-tech has employed several different sample choice criteria. High-tech can be defined, for example, by analyzing the input of the services or products that a company produces. Input is generally considered as employees' efforts to perform the services or products, for instance research and development contribution in the company. Output based approach focuses on the products and services which firms produce to their customers. Output point of view analyzes companies from the end-user perspective. Results of the production efforts are examined if those can be considered as high-technology products or services.

In this thesis, the high-tech definition is derived from the research of technology-oriented worker intensity in the industry. The study is conducted by the U.S. Bureau of Labor Statistics (Hecker 2005). The input approach is the key identifier whether the industry is included in high-tech industries. An industry is considered to be a member of high-technology industries if the employment in technology-oriented occupations accounted for at least twice the 4.9-percent average of all industries. The proportion of technology-oriented occupations is based on the total employment of the industry. In this thesis, only Level 1 industries are taken into the sample in order to capture the essentials of the core high-technology businesses. For Level 1 industries, the proportion of the technology-oriented occupations accounted for at least 5 times the average and constituted 24.7 percent or more of the industry employment. Below, Table 2 lists the Level 1 high-tech industries. (Hecker 2005, 58)

NAICS 2002	Level 1 High-Technology Industries
5415	Computer systems design and related services
5112	Software publishers
5413	Architectural, engineering, and related services
5417	Scientific research and development services
5181	Internet service providers and web search portals
3341	Computer and peripheral equipment manufacturing
5161	Internet publishing and broadcasting
3345	Navigational, measuring, electromedical, and control instruments manufacturing
5182	Data processing, hosting, and related services
3364	Aerospace product and parts manufacturing
3342	Communications equipment manufacturing
3344	Semiconductor and other electronic component manufacturing
3254	Pharmaceutical and medicine manufacturing
5179	Other telecommunications

Table 2. Level 1 high-technology industries (Hecker 2005, 64).

High-tech industries are identified by the 4-digit industry classification of the North American Industry Classification System 2002 (NAICS). NAICS is a standard used by American Federal statistical agencies and it is replacing the Standard Industrial Classification (SIC) system (North American Industry Classification System (NAICS), Introduction). In the step two, the NAICS 2002 codes are converted to 4-digit NAICS 2007 codes. In this study, NAICS 2007 codes are used for data retrieval in Orbis database. The conversion is based on the conversion table provided by U.S. Census (2002 NAICS to 2007 NAICS, see Appendix 2).

4.4.2 Sample selection

The sample selection was conducted in two steps including steps 3 and 4 in the process flow. In the step 3, high-tech companies were identified in OECD Orbis database by employing the converted NAICS 2007 high-tech industry codes. Public, active and inactive companies were included in the sample whereas American Depositary Receipts (ADRs) were excluded. Double observations were eliminated. This resulted in 12 989 active and inactive high-technology companies during 1990-2010.

After the companies were identified in Orbis database (step 3) then, in the step 4, the companies were matched in Thomson Reuters database. The goal of the matching was to find ws.entity key in Thomson Reuters for each firm identified in Orbis as a member of high-technology industry. Lack of universal identification code across different databases made the company matching between Orbis and Thomson Reuters rather challenging. BvD number in Orbis database and ws.entity key in Thomson Reuters are the primary identification codes in the databases. Nowadays, after the changes in Thomson Reuters database during the summer 2011, the firm identification code is called Company Permanent ID.

Several different matching strategies were applied in order to match the companies between these two databases. All companies (not only high-tech firms) were retrieved from Thomson Reuters database and then the sample companies identified in Orbis were matched against the retrieved Thomson Reuters companies. First, the companies were matched by using specific identification codes. ISIN number (ws.ISIN in Thomson Reuters) was identified to have most matches and considered to be rather reliable identifier for matching purposes. Therefore, ISIN number was used as a primary matching criterion. 7 148 companies were matched by using ISIN number, thus, 5 841 companies had to be matched by other means. Secondly, the firms were matched by using SEDOL number (ws.SEDOL in Thomson Reuters) and then by using Ticker (Ticker symbol in Orbis and ws.Ticker in Thomson Reuters). VALOR number could not be used as the values were basically missing from Thomson Reuters. Firms found via matching were controlled to have consistent information with every identifier (ISIN, SEDOL and Ticker) meaning that each ISIN, SEDOL and Ticker in Orbis database resulted in the same ws.entity key in Thomson Reuters. If the result was not consistent with each identifier (ISIN, SEDOL and Ticker), the proposed ws.entity key was hand checked by the name and main exchange. The rest companies left out from the matching with identification codes were hand checked by comparing the company name and primary exchange. Only companies with exactly the same name and primary exchange were taken into the sample. Hand check resulted in total of 3 494 company matches. Thomson Reuters match was not found for 2 347 Orbis high-technology companies. In total, 10 642 inactive and active high-technology companies were identified in Orbis and for which ws.entitykeys were found in Thomson Reuters.

4.4.3 Information collection

In the step 5, the data of model variables was collected in Thomson Reuters database. OECD Orbis database was used only as a supplemental information source, for example for listing and firm foundation information. However, the sample selection was made first in Orbis as described in the previous Chapter 4.4.2. Thomson Reuters is preferred to ensure the availability of financial statement information and Thomson Reuters is the main source database in this study. Orbis database offers less historical financial information. All financial statement information is retrieved from Thomson Reuters. The data is gathered from the years 1990–2010 and the analysis is done by using cross-sectional approach. The longer time period (20 years) is chosen because the industry has evolved and experienced economic expansions and downturns as well as investors' high- and low-sentiment times during the time period. All companies identified either as active or inactive in the period are included in the sample in order to avoid survivor bias. One strength of this study is the number of observations with a sample of 10 642 firms and a data collection of 20 years. The quality of information retrieved from the Thomson Reuters is ensured by performing the data runs twice for each variable. Then the results were compared and corrected if needed. The statistical analysis is made by using software IBM SPSS PASW version 18.

4.4.4 Subsamples of the explanatory and estimated forecast models

<u>Subsample of the explanatory model</u>

Two different samples were gathered for the explanatory and estimated earnings forecast models. The focus of the explanatory model is to study whether the accounting variables at the time of listing explain the future earnings. Analysis is performed by examining the newly listed high-tech IPO companies between the years 1991-2010. The newly listed firms were identified by using several approaches in order to ensure that these companies are genuine first IPOs. Companies with several listings were excluded from the analysis. The primary requirement in identifying the correct listing year was the first appearance of market capitalization in Thomson Reuters database. In addition, the year of the first appearance of market capitalization had to match with either of the two Thomson Reuters instrument active dates at year level (TF.PR.PriceDateFirst or TF.InstrumentActiveDate). These two Thomson Reuters variables were analyzed to bring the most results of the listing indicators also compared with Orbis information. This matching of different listing indicators was performed because there was inconsistent information found in the database about listing information. By using the first appearance of market capitalization can be ensured that the firm is certainly listed and for sensitivity analysis the result is then confirmed by the instrument active dates. Since year 1990 is the first year in the time-series, year 1991 is the first possible listing year identified with this technique of the first appearance of market capitalization. Descriptive statistics of the subsample are presented in Chapter 5.1.

<u>Subsample of the estimated forecast model</u>

The subsample of the estimated earnings forecast model differs from the subsample of the explanatory model. The model is first estimated in one sample and then tested in a hold-out sample. The time period of the data collection was divided into two parts: 1990-2000 was chosen for the estimation period and the time period of 2001-2010 for testing the model in a hold-out sample. The cut between the years 2000 and 2001 is interesting because of the experienced IT bubble in the high-tech markets.

The estimated forecast model is also divided into the general model and IPO model. In the general model, all listed high-tech companies, including IPOs, are taken into the sample. Listing is ensured by the requirement of reported market capitalization in the year in question. For age analysis purposes, there is an additional requirement set that the age of the company has to be zero or more. The IPO model examines the predictive power of the accounting variables in forecasting future earnings of newly listed high-tech IPOs. The time of listing is defined similarly as described in the sample of the explanatory model. The IPO and general models are estimated and tested separately.

5 Empirical results

The results of the empirical analysis are reported in this chapter. The analysis can be divided into four main steps. These steps are illustrated in the figure below (Figure 3). In the first step, the descriptive statistics of the sample companies are reported. The results of the explanatory model in the IPO sample are discussed in the second section. Then the model is estimated for both IPO and all listed high-tech firms based on the information gathered from the years 1990-2000 and tested in the independent hold-out sample between the years 2001 to 2010. Forecasted earnings are then compared with the actual earnings. In the last step, the forecasting errors are analyzed and tested if the errors can be explained by the firm specific uncertainty factors.

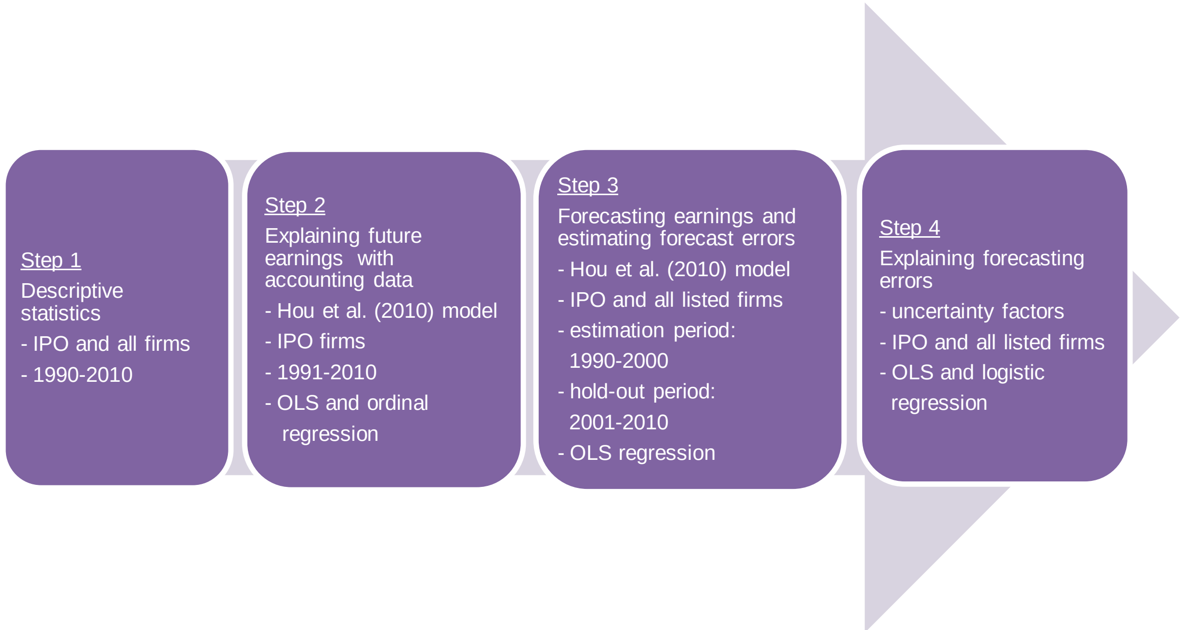


Figure 3. Analysis of the empirical results.

5.1 Descriptive statistics

5.1.1 Sample companies

In total 5 341 high-technology companies were identified as newly listed IPOs during 1991-2010 (see Figure 4). The number of new lists has steadily increased from 1990 to 2000. The peak of new lists, the IT bubble, is observed during 1999-2000. The number of new listings was cut to half in 2001 compared with the year 2000. The financial crisis reflected negatively in the number of new lists during 2008-2009 but in 2010 the level of IPOs has almost reached the number of listings before the crisis.

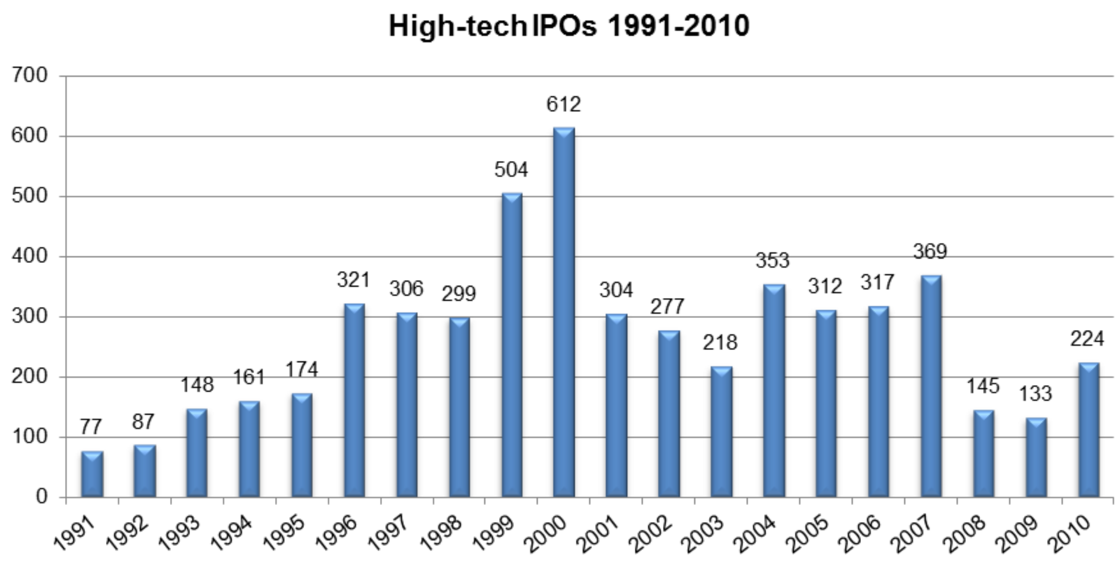


Figure 5. High-tech IPOs during 1991-2010.

The three most common industries amongst all listed high-tech companies are related to electronic component manufacturing, computer systems services and pharmaceutical manufacturing (see Figure 5 and Table 3). These top three industries with “Other Telecommunications” added constitute 59% of the high-tech companies. Internet firms can be categorized into four main industries (NAICS codes): 5191, 5179, 5171, 5182. In total, 814 Internet firms were listed during the period and constituting 15.2 % of the newly listed companies.

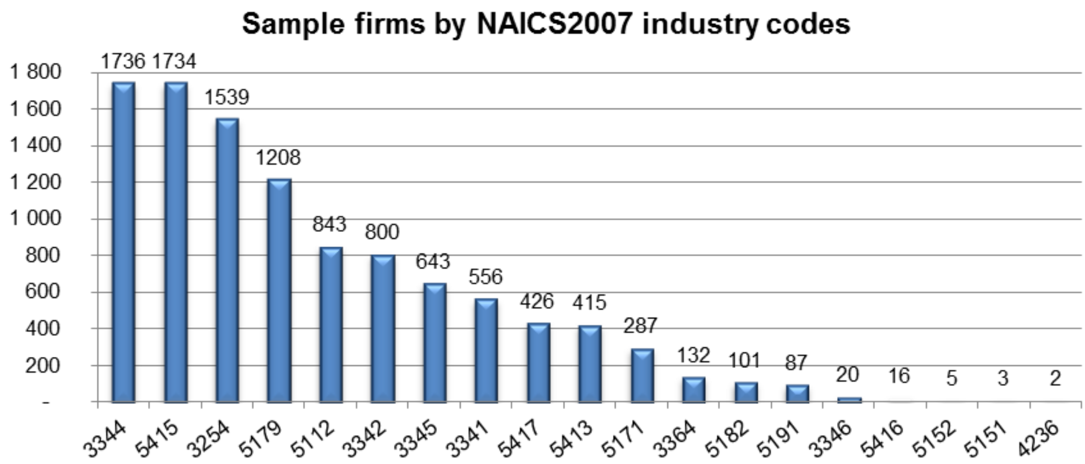


Figure 4. Industries of all listed high-tech companies 1990-2010.

NAICS2007	Title	Frequency	Percentage
3344	Semiconductor and Other Electronic Component Manufacturing	1 736	16,45 %
5415	Computer Systems Design and Related Services	1 734	16,43 %
3254	Pharmaceutical and Medicine Manufacturing	1 539	14,58 %
5179	Other Telecommunications	1 208	11,45 %
5112	Softw are Publishers	843	7,99 %
3342	Communications Equipment Manufacturing	800	7,58 %
3345	Navigational, Measuring, Electromedical, and Control Instruments Manufacturing	643	6,09 %
3341	Computer and Peripheral Equipment Manufacturing	556	5,27 %
5417	Scientific Research and Development Services	426	4,04 %
5413	Architectural, Engineering, and Related Services	415	3,93 %
5171	Wired Telecommunications Carriers	287	2,72 %
3364	Aerospace Product and Parts Manufacturing	132	1,25 %
5182	Data Processing, Hosting, and Related Services	101	0,96 %
5191	Other Information Services	87	0,82 %
3346	Manufacturing and Reproducing Magnetic and Optical Media	20	0,19 %
5416	Management, Scientific, and Technical Consulting Services	16	0,15 %
5152	Cable and Other Subsription Programming	5	0,05 %
5151	Radio and Television Broadcasting	3	0,03 %
4236	Electrical and Electronic Goods Merchant Wholesalers	2	0,02 %
Total		10 553	100,00 %

Table 3. Industries of all listed high-tech companies 1990-2010.

5.1.2 Variables

Variable descriptive statistics are reported in Table 5. The table presents the descriptive statistics of newly listed high-tech companies during 1991-2010 and, also for a comparison, all listed high-tech companies with market value above zero during 1990-2010. The amounts are reported in million US dollars. The variables are winsorized at level of 1 and 99 percentiles in order to control the influence of outliers. Thus, the mean value of Cook's distance declined to below 1 in all regressions of the explanatory model. For sensitivity analysis, variables were both truncated at the level of 1 and 99 and deflated by total assets and market capitalization separately. Neither truncation nor deflation reduced the degree of outliers. Total of 5 341 valid IPO observations were identified. However, the number of the valid observations in future earnings declined throughout the forecasting horizons.

Observations of the dependent variable concentrate close to median values. The variable distribution is very peaky with fat tails. The fat tails indicate of extreme values in both ends of the distribution. Also the distributions of explanatory variables differ from normal distribution with similar peaky, close to median observations and fat tails. The distribution with peaky, close to median observations and fat tails is called a leptokurtic distribution. Also highly positive Kurtosis values compared with the standard error of the Kurtosis indicate that the distributions are statistically significantly kurtic for almost all variables (see Table 4).

Skewness statistics compared with the standard error of the Skewness are also indicating of skew distributions for all variables (see Table 4). Different data transformations were considered in order to achieve more normally distributed data. Logarithms were taken of market capitalization and total assets and, in fact, logarithms leveled the distributions. However, logarithms could not be taken of all variables due to the requirement of positive core value. When the logarithmic variables were analyzed together with the other variables, the results did not improve. Therefore, the idea of employing logarithmic variables was rejected.

Skewness and Kurtosis

	N		Skewness				Kurtosis			
			Statistic		Std. Error		Statistic		Std. Error	
	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All
IncExt_Y1	3 696	56 344	4.699	6.300	0.040	0.010	31.094	42.761	0.081	0.021
IncExt_Y3	3 482	46 505	4.379	6.349	0.041	0.011	28.365	43.209	0.083	0.023
IncExt_Y5	3 082	37 231	5.031	6.259	0.044	0.013	31.472	41.640	0.088	0.025
MCap	5 341	88 849	5.437	6.271	0.034	0.008	32.296	41.883	0.067	0.016
TAssets	3 627	61 083	6.132	6.226	0.041	0.010	40.287	40.992	0.081	0.020
DIV	5 341	88 849	7.917	7.280	0.034	0.008	64.549	54.580	0.067	0.016
DD	5 341	88 849	-2.168	-1.545	0.034	0.008	2.700	0.387	0.067	0.016
IncExt	3 640	61 210	4.389	6.289	0.041	0.010	28.091	42.735	0.081	0.020
NegIncExt	3 640	61 210	0.216	0.334	0.041	0.010	-1.955	-1.889	0.081	0.020
TAccr	3 504	59 660	-6.199	-6.260	0.041	0.010	41.540	41.408	0.083	0.020
Valid N (listwise)	2 180	28 828								

Table 4. Kurtosis and Skewness statistics after winsorizing.

When analyzing the descriptive statistics in Table 5, 44.6 percent of the IPO firms at the listing year and 41.8 percent of all listed high-tech companies reported negative earnings. The median earnings of the IPO firms stay rather constant over the five years from the listing but the mean values increase gradually with the forecasting horizon. The future earnings of all listed high-tech companies are larger in terms of mean and median values than the respective of IPOs. There is an increasing trend observed in earnings from the first year to fifth year. However, the standard deviation of the future earnings of all listed high-tech companies is rather high for all forecasting horizons.

When it comes to the size variables, total assets and market capitalization, there are substantial differences between the sample companies. It is interesting to note that the median MCap for all listed high-tech firms is smaller than MCap of IPOs whereas the mean MCap of all listed high-tech companies is bigger than the one of IPOs. The standard deviations of MCap for both groups are rather high when compared with the mean values.

However, the mean and median of TAssets of all listed high-tech firms are substantially bigger than TAssets of IPOs.

Both sample groups are rather reluctant to pay dividends as 87 percent of IPOs and 81 percent of all listed high-tech firms are not paying dividends on average. Only 13.3 percent of the newly listed firms paid dividends to their shareholders at the listing year. Companies are predominantly young at the time of listing. The median listing age is five years. The median listing year of IPOs, year 2000, reflects the Internet boom in the late 1990s. In general, both groups are rather young with the average age of close to eight and median age of five.

5.1.3 Correlations

The correlation analysis is performed only for the IPO sample which is the sample group of the explanatory model. Variable correlations were tested with both Pearson and Spearman correlation metrics. Based on the analysis of the descriptive statistics, the distributions of variables are very leptokurtic and do not obey the normal distribution. Especially, Pearson's correlation results have to be cautiously interpreted due to the possible non-linear relations of variables. Therefore, Spearman correlation analysis may be better since it based on the ordinal scale. The future earnings of one, three and five years ahead are also split into 20 equal sized groups based on the size of the future earnings. The smallest future earnings are assigned to the group 1 and the largest to the group 20. The split is done for each forecasting horizon separately. The grouping of the dependent variables is made because of the violation of the residuals' normality assumption when the continuous variable is employed in OLS regression. Correlations of the explanatory model with the continuous (Appendix 3) and grouped (Appendix 4) dependent variables are reported in Appendices. Spearman correlations are shown above the diagonal and Pearson correlations are shown below the diagonal.

When Pearson correlations are analyzed, all the explanatory variables are statistically significantly correlated at the level of 0.01 in relation to the continuous dependent variables. Also Spearman correlation coefficients are in line with the Pearson correlations except TAccr is not statistically significantly correlated with the IncExt_Y3 and IncExt_Y5. Metsämuuronen (2006) argues that often when the sample size is large as in this study the correlations of variables are statistically significant even though the correlations might not be that high. When analyzing the correlations of the grouped dependent variables, the results are very similar to the continuous dependent variables. Only the strength of correlation varies but the signs are in line with the continuous dependent variables.

Descriptive statistics: dependent and explanatory variables

		IncExt_Y1		IncExt_Y3		IncExt_Y5		MCap		TAssets		DIV		DD		IncExt		NegIncExt		TAccr	
		IPO	All	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All
N	Valid	3 696	56 344	3 482	46 505	3 082	37 231	5 341	88 849	3 627	61 083	5 341	88 849	5 341	88 849	3 640	61 210	3 640	61 210	3 504	59 660
	Missing	1 645	32 505	1 859	42 344	2 259	51 618	0	0	1 714	27 766	0	0	0	0	1 701	27 639	1 701	27 639	1 837	29 189
Mean		6,79	46,95	8,89	57,07	12,29	68,97	340,83	872,07	242,16	996,87	1,01	7,59	0,87	0,81	5,63	43,26	0,45	0,42	-16,55	-63,35
Median		0,32	0,88	0,20	1,06	0,39	1,35	68,52	52,92	42,16	69,81	0,00	0,00	1,00	1,00	0,48	0,84	0,00	0,00	-0,66	-2,17
Std. dev.		61,93	249,74	75,88	297,62	80,42	348,99	954,55	3 487,47	837,78	3 975,97	6,04	43,39	0,34	0,40	52,69	232,29	0,50	0,49	78,86	282,66
Min		-164,56	-240,80	-219,41	-262,37	-168,60	-276,18	0,18	0,00	0,03	0,15	0,00	0,00	0,00	0,00	-137,49	-231,68	0,00	0,00	-631,07	-2 232,34
Max		452,33	1 988,00	535,85	2 377,70	587,25	2 739,62	7 117,85	27 922,09	6 665,53	31 356,91	54,25	370,47	1,00	1,00	372,86	1 850,92	1,00	1,00	40,86	78,54
Sum		-	-	-	-	-	-	-	-	-	-	-	-	4 633	71 583	-	-	1 625	25 569	-	-
Percentiles																					
	5	-36,43	-37,80	-38,76	-39,09	-39,03	-40,86	2,63	0,42	1,18	1,67	0,00	0,00	0,00	0,00	-37,03	-37,28	0,00	0,00	-66,51	-258,60
	25	-4,69	-3,46	-5,45	-3,33	-4,51	-3,09	21,33	11,78	13,92	18,48	0,00	0,00	1,00	1,00	-3,38	-3,42	0,00	0,00	-4,79	-13,14
	50	0,32	0,88	0,20	1,06	0,39	1,35	68,52	52,92	42,16	69,81	0,00	0,00	1,00	1,00	0,48	0,84	0,00	0,00	-0,66	-2,17
	75	5,65	10,50	6,02	12,24	6,99	14,15	231,04	248,37	116,37	275,09	0,00	0,00	1,00	1,00	5,21	9,84	1,00	1,00	1,06	0,16
	95	52,35	214,91	67,48	249,13	74,89	296,29	1 347,77	3 481,62	938,47	4 025,01	2,54	15,62	1,00	1,00	48,56	198,36	1,00	1,00	13,04	13,47

Table 5. Descriptive statistics of the explanatory model.

The correlation coefficients of all the explanatory variables in relation to dependent variables remain consistently positive (or negative) through the period of five years in both Pearson and Spearman correlations and with both continuous and grouped dependent variables. Only the sign of TAccr in relation to dependent variables is different when comparing the results from Pearson and Spearman correlations. Pearson coefficients indicate that accruals are negatively associated with future earnings whereas Spearman rho's results in positive association with both continuous and grouped dependent variables. However, only the Spearman correlation of TAccr to IncExt_Y1 is statistically significantly associated with the future earnings of one-year ahead with the continuous and grouped dependent variables. Market capitalization, total asset, dividends paid, lagged earnings are positively and non-dividend paying and negative earnings reporting are negatively related to the future earnings of one, three and five years ahead. Variables of non-dividend paying and negative earnings reporting are dummy variables.

Lagged earnings have highly positive correlation in association with all the future earnings throughout the forecasting horizons. On the other hand, NegIncExt rises with the high negative correlation. All other variables remain close to the same correlation level throughout forecasting horizons in relation to the dependent variables. Only TAccr has rather low correlation over the forecasting horizons. Correlations also between the explanatory variables are statistically significant. Regression analysis requires that the multicollinearity should not be high within the explanatory variables (Metsämuuronen 2006). However, when moving to the regression analysis and analyzing the variance inflation factors (VIFs) of the variables, the VIF values are not alarming and the model should not be weakened by multicollinearity.

5.2 Explaining future earnings

In this chapter, the goal is to examine whether the chosen accounting variables have explanatory power for future earnings. The sample consists of IPO companies listed during 1991-2010. There are three forecasting horizons applied: one, three and five years ahead. The variables are investigated with two different statistical methods: OLS linear regression and ordinal regression. The continuous dependent variable, future earnings, is first analyzed with the OLS linear regression. Then the dependent variable is split into 20 equal sized groups based on the size of the future earnings. The split was made since there were statistical issues encountered in OLS regression residuals. The categorical dependent variable is analyzed first with OLS regression and then with ordinal regression. The ordinal regression method is discussed in Chapter 5.2.2 and the results are reported in Chapter 5.2.3.

5.2.1 Linear regression results

First, the explanatory model with the continuous dependent variables is analyzed with OLS linear regression (Model 1). The regression results of the model explaining future earnings one, three, and five years ahead are reported in Table 6. The residual distributions of Model 1 are alarming. The S-shaped line in the normal probability plot of regression standardized residuals indicates of non-normal distribution of residuals. The normal distribution of residuals is one of the basic assumptions of OLS regression (Metsämuuronen 2006). Thus, the terms of using OLS regression are severely violated. There were several tests and analyses made in order to reduce the violation of residuals' normality assumption. The deflation, truncation and taking logarithms of the variables were tested and even the sample selection was rethought to focus on the small or nano-cap firms. Variables were also squared as Metsämuuronen (2006) suggests for mitigating the normality issues. These tests did not remove the violation of residuals' normality assumption. Hence, the results of the variable coefficients and statistical significance as well as the explanatory power of the Model 1 cannot be reliably interpreted and, therefore, the results are not discussed in this chapter.

Model 2 was constructed because of the violation in residuals' normality assumption in Model 1. In Model 2, the dependent variables are divided into 20 groups as described in Chapter 5.1.3. However, the applicability of OLS regression for Model 2 may be questioned as the dependent variables are categorical (from 1 to 20). Nevertheless, also ordinal data is often first analyzed by using OLS regression. The analysis is problematic because the assumptions of OLS are violated when a non-interval dependent variable is used (Metsämuuronen 2006, SPSS Data Analysis Examples: Ordinal Logistic Regression). Hence, the results of OLS regression for Model 2 have to be cautiously interpreted. Model 2 is therefore analyzed also with an ordinal regression in the next Chapter 5.2.3 (Model 3).

Adjusted R squares of Model 2 decrease throughout the forecasting horizons (Y1: 44 %, Y3: 23 %, Y5: 18 %) but remain rather high. The significance of F values changes throughout the forecasting horizons indicates that the model is statistically significant at level of less than 1 %. The explanatory variables are not deflated by any size measures and, hence, there is a risk that the model explains also the size variance of the companies. Durbin-Watson statistics remain close to 2 which indicate that there is no autocorrelation detected.

When analyzing the statistical significance of the variable coefficients in Model 2, non−dividend paying (negatively), lagged earnings (positively) and negative earnings (negatively) are statistically significant at the level of 0.01 over the forecasting period, except IncExt is insignificant for the fifth year. The signs are consistent with the predictions. TAssets and TAccr remain insignificant. MCap is statistically significant for the years of one and five but the coefficient is 0.000 implying that MCap is insignificant in economic terms also compared with the mean MCap. In a conclusion, the results of Model 2 are used as a reference for the further analysis with ordinal regression. Ordinal regression analysis is needed because OLS regression should be used for continuous dependent variables and not for categorical dependent variables.

5.2.2 Ordinal regression method

Ordinal regression is a modification of the binary logistic regression model. In ordinal regression, there is more than one individual event and the events can be ordered. Multinomial regression captures the differences between all possible pairs of the dependent variable groups but ignores the ordering of the dependent variable. Ordinal logistic regression in SPSS is called PLUM (Polytomous Universal Model). (Norusis 2010) Similarly to other logistic regressions, ordinal regression calculates the changes in the log of odds of the dependent variable and not the changes in the dependent variable as in the OLS regression. The coefficients of a logit model indicate of the logit changes. There are advantages in using a logistic model to analyze the explanatory power of accounting variables. Logistic regression does not require normally distributed residuals which was the identified issue in analyzing the results of the linear regression. In addition, logistic regression does not assume a linear relationship between the explanatory and dependent variables nor homoscedasticity of observations. (Garson 2011) Ordinal regression assumes that the relationship between the explanatory variables and the logits are the same for all the logits. This null hypothesis is called parallel lines assumption. (Norusis 2010)

Even though the interpretation of ordinal regression is considered as tricky (Kennedy 2008), there are quite many similarities compared with OLS regression. For instance, the coefficient of continuous explanatory variables is interpreted rather similarly as in a linear regression. A positive coefficient indicates that when the value of the continuous explanatory variable increases, the probability of higher score in the dependent variable category increases. The ordinal regression model is also called the proportional odds model. The name is derived from the existence of separate threshold values (also called cut values, α) for each logit. The thresholds are similar to the intercept value in linear regression and are generally used only in calculating predicted values. However, according to the assumption of parallel lines which is also called proportional odds assumption, each logit should have the same coefficient (β) for explanatory variables. (Norusis 2010, Kennedy 2008, SPSS Data Analysis Examples: Ordinal Logistic Regression) The test of parallelism analyzes the assumption of parallel lines. It tests that the relationships between independent variables and logits actually are the same for all logits. The significance level for the Chi-Square in the test of parallelism should be large (insignificant) in order not to reject the null hypothesis of parallel lines. When the null hypothesis is not rejected, there is only one set of variable coefficients and multinomial regression analysis is not required. (Norusis 2010, SPSS Data Analysis Examples: Ordinal Logistic Regression)

5.2.3 Ordinal regression results

The sample, the dependent variable and the forecast horizons are equal to the ones of the Model 2. Ordinal regression is preferred over OLS regression when the dependent variable is categorical and the categories can be organized into ordinal order. In addition, the sample was ranked by the size of the dependent variable, future earnings, but with the sample size over 3 000 companies, running the ordinal regression with the ranked data is not technically possible.

First, the basic assumption of parallel lines is tested by analyzing the significance level of Chi-Square (see Table 7). Because the difference of - 2 log-likelihood in null hypothesis and in the general model is statistically significant over the forecasting period, the null hypothesis of parallel lines has to be rejected. The rejection means that it cannot be ensured that the relationships between the explanatory variables and logits are same for all logits. Therefore, the multinomial regression should also be examined. The results of the multinomial regression are discussed later in this chapter.

Explanatory model IPO OLS regression

Variable	Intercept			Mcap (+)			TAssets (-/+)			DIV (+)			DD (-)			IncExt (+)			NegIncExt (-/+)			TAccr (-)		
	Coef.	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*
Model 1: Dependent variable: continuous IncExt (t+1)																								
Y1	0.383	0.238	0.812	-0.001	-1.097	0.273	0.001	0.398	0.691	0.579	4.635	0.000	-0.739	-0.394	0.694	0.807	41.720	0.000	-2.161	-1.457	0.145	-0.161	-9.969	0.000
Y3	-0.611	-0.225	0.822	0.001	0.476	0.634	0.006	1.840	0.066	1.632	7.513	0.000	-1.109	-0.347	0.729	0.629	18.359	0.000	0.081	0.032	0.975	-0.144	-5.326	0.000
Y5	3.803	1.113	0.266	0.005	2.245	0.025	0.025	6.122	0.000	1.525	5.344	0.000	-2.645	-0.658	0.511	0.468	10.676	0.000	-2.071	-0.643	0.520	-0.017	-0.495	0.621
Model 2: Dependent variable: grouped IncExt (t+1) to 20 groups																								
Y1	14.252	74.092	0.000	0.000	-3.528	0.000	0.000	1.727	0.084	0.014	0.917	0.359	-1.515	-6.750	0.000	0.019	8.302	0.000	-6.234	-35.118	0.000	-0.002	-0.940	0.348
Y3	13.275	54.167	0.000	0.000	0.324	0.746	0.000	-0.298	0.766	0.040	2.022	0.043	-1.397	-4.843	0.000	0.011	3.679	0.000	-4.356	-18.884	0.000	-0.004	-1.463	0.144
Y5	13.359	47.281	0.000	0.000	2.491	0.013	0.000	1.417	0.157	0.028	1.202	0.230	-1.895	-5.701	0.000	0.002	0.445	0.656	-3.640	-13.663	0.000	0.002	0.609	0.543

	Adj R^2	*Durbin-Watson*	*Model F-value*	*Sig. F-value*	*N*
Model 1: Dependent variable: continuous IncExt (t+1)					
Y1	0.666	1.963	927.841	0.000	3 254
Y3	0.462	2.023	337.294	0.000	2 742
Y5	0.410	2.033	219.315	0.000	2 197
Model 2: Dependent variable: grouped IncExt (t+1) to 20 groups					
Y1	0.441	1.902	367.724	0.000	3 254
Y3	0.231	1.925	118.884	0.000	2 742
Y5	0.178	1.977	69.099	0.000	2 197

Explanatory model IPO ordinal regression

Variable	Mcap (+)			TAssets (-/+)			DIV (+)			DD = 0 (+)			IncExt (+)			NegIncExt = 0 (-/+)			TAccr (-)		
	Coef.	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*
Model 3: Dependent variable: grouped IncExt (t+1) to 20 groups																					
Y1	0.000	13.984	0.000	0.001	20.952	0.000	0.033	7.162	0.007	0.379	16.493	0.000	0.058	554.531	0.000	1.941	484.624	0.000	-0.008	36.895	0.000
Y3	0.000	1.225	0.268	0.000	1.034	0.309	0.059	22.893	0.000	0.298	8.942	0.003	0.024	166.987	0.000	1.198	199.648	0.000	-0.005	22.832	0.000
Y5	0.000	14.644	0.000	0.000	3.449	0.063	0.035	9.516	0.002	0.503	21.336	0.000	0.011	40.834	0.000	1.006	122.629	0.000	-0.002	2.048	0.152

	Cox and Snell	*Nagelkerke*	*N*
Model 3: Dependent variable: grouped IncExt (t+1) to 20 groups			
Y1	0.567	0.568	3 255
Y3	0.318	0.319	2 742
Y5	0.244	0.245	2 197

Table 6. OLS and ordinal regression results of the IPO explanatory model.

When analyzing the model-fitting information, a statistically significant result is observed over all forecasting horizons. This means that the tested multivariate model is better than a simple model with only an intercept. However, the results of the two goodness-of-fit measures are mixed. The goodness-of-fit statistics compare the observed and expected frequencies. The insignificant results of Pearson measures over the forecasting horizons indicates that the model is fitting well whereas the statistically significant result of the Deviance measure imply that the model is fitting poorly. There are a high number of empty cells observed in the crosstab between the explanatory and dependent variables. Norusis (2010) argues that the high number of empty cells normally affects the interpretation of goodness-of-fit measures and may imply of an impaired and unstable model. Therefore, the goodness-of-fit results cannot be reliably analyzed.

Model Fitting

	Intercept only				Model 3			
	-2 Log Likelihood	Chi-Square	df	Sig.	-2 Log Likelihood	Chi-Square	df	Sig.
Y1	19 489.05	-	-	-	16 764.54	2 724.51	7	0.000
Y3	16 405.86	-	-	-	15 355.89	1 049.98	7	0.000
Y5	13 141.35	-	-	-	12 526.50	614.85	7	0.000

Goodness-of-Fit

	Pearson			Deviance		
	Chi-Square	df	Sig.	Chi-Square	df	Sig.
Y1	939 861 074 687 861	61 401	0.000	16 762.91	61 401	1.000
Y3	228 106 148	51 673	0.000	15 354.26	51 673	1.000
Y5	3 189 545	41 451	0.000	12 524.38	41 451	1.000

Table 7. Model fitting and Goodness-of-Fit results of the ordinal regression.

Ordinal regression results are reported in Table 6 together with the regression results of Model 1 and 2. When it comes to the overall explanatory of the ordinal model, the pseudo-R^2s follow the same pattern as in OLS regression. In fact, the pseudo R^2 values are higher than Model 2 R^2 values but decreasing with the forecast horizon. The values of the more commonly used Nagelkerke R^2 are 57%, 32% and 25% for one, three and five years ahead, respectively. There is no substantial difference to Cox and Snell R^2 values. However, it has to be reminded that the pseudo-R^2 values do not reflect the percent of variance explained as the R^2 in OLS regression.

The threshold values for each group and forecasting horizon are reported in Appendix 5. The threshold values are commonly used only for calculations of predicted values and the thresholds are not applied in this study. When analyzing the variable results, all the explanatory variables are statistically significant at the level of 0.01 in explaining the first year earnings. However, it has to be noted that the coefficients of MCap and TAssets remain zero or very close to zero throughout the analyzed period of five years. TAssets is statistically significant only for the first year whereas MCap is statistically significant for the first and fifth year.

Both dividend variables, dividends paid and the dummy for non-dividend paying, are statistically significant at the level of 0.01 over the forecasting period. Both are positively associated with future earnings. This means that when a one unit increase in dividends paid, the expected ordered log odds increases 0.033 when moved to the higher future earnings category. The results are in line with Fama and French (2000, 2001, 2006), Hou et al. (2010), Hou and van Dijk (2010), and Hou and Robinson (2006). Also dividends paying companies (DD=0) are more likely to assign higher future earnings than firms not paying dividends (DD=1). The observed coefficients of DIV are in line with the predictions. Also the observed coefficients of DD are in line with the expectations and Model 2 OLS regression results.

Both income variables (IncExt and NegIncExt) are positively and statistically significantly at the level of 0.01 associated with future earnings. The results confirm that the current earnings explain the future earnings. Both variables remain persistent over the forecasting period. These results are in line with several studies (Fama & French 2006, Hou et al. 2010, Hou & van Dijk 2010, Hou & Robinson 2006). Firms reporting positive or zero earnings at the time of IPO (NegIncExt = 0) are more likely to assign higher future earnings. The finding contradicts with the argument that the current negative earnings of high-tech companies imply of future positive earnings and investment in the future welfare (Trueman et al. 2001, Kask & Sieber 2002). These investments do not seem to pay off in the next five years from the listing but perhaps with a longer forecasting horizon.

Total accruals are negatively and statistically significantly (0.01) related to future earnings for the first and third year. The results are following the expected sign. The results also confirm Sloan's (1996) findings that firms with higher operating accruals tend to have lower future earnings. However, it seems that the total accruals lose their effect on transitory variation in earnings in the fifth year when the total accruals are not statistically significant.

For sensitivity analysis, Model 3 was run also with multinomial regression. The run was made due to the violation of parallel lines assumption. The results are not attached but are, however, approximately in line with the results of ordinal regression. When it comes to the significance and goodness of the whole model, the results are very similar to the results of ordinal regression. The model has high pseudo-R^2s (Cox and Snell, Nagelkerke) and the likelihood ratio tests indicate that all variables except TAccr should be included to the model. In addition, MCap and TAssets are also in the multinomial regression statistically significant at the level 0.01 but only with very low or zero coefficients. There are differences found in the explanatory variables between the logits. Therefore, the test results of parallel lines appear to be correct. For example, dividends paid and dividend dummy are statistically significant for the groups of higher future earnings and insignificant for the groups of lower future earnings. In addition, the most significant difference lies in IncExt variable which seems not to be consistently statistically significant. However, the model fitting information indicates that the tested model is better than a model with only an intercept.

In a conclusion of the ordinal regression results, there is some evidence found that financial statement variables provide information about the future earnings. The data with the continuous dependent variable could not be analyzed with the traditional OLS linear regression due to the violation in the assumptions of OLS. However, there was proxy of future earnings constructed based on the size of the earnings. The proxy was constructed to mitigate the statistical issues encountered in OLS regression analysis. Future earnings were divided into 20 groups for each forecasted year separately. Analysis was performed with the ordinal logistic regression. Ordinal regression results suggest that lagged earnings (positively), dividends (positively) and accruals (negatively) are explaining the future earnings of high-tech IPOs whereas the size variables of market capitalization and total assets are not significant in economic terms. These ordinal regression results are consistent with the expected variable signs.

5.3 Forecasting future earnings

5.3.1 Estimation of the model

The next step is to analyze the predictive power of the model in comparison with actual earnings. The estimation model is similar to the explanatory Model 1 with the continuous dependent variable. The model can still be used for estimating future earnings despite the violation of the residuals' normality assumption. The estimation is executed with OLS regression. Testing the performance of forecasting earnings proceeds in two steps: estimation of the model and testing the model in a hold-out sample. As described in the sample selection Chapter 4.4.4, the time period 1990-2010 is divided into two periods. The model is estimated based on the information collected between the years 1990 and 2000 and tested in the independent hold-out sample with the information from 2001-2010. In addition, the analysis is separated to concern

IPO high-tech companies (Model 4) and all listed high-tech companies (Model 5). After the model estimation, the expected earnings and forecasting errors are calculated and analyzed in the hold-out sample.

The regression results from the model estimation for both IPO and all listed high-tech companies are reported in Table 8. All the dependent and explanatory variables are winsorized at level of 1 and 99 percentiles in order to decrease the impact of outliers. The adjusted R^2s in both models remain rather high over the forecasting horizons. For the Model 4, adjusted R^2s are Y1: 57 %, Y3: 41 %, Y5: 24 % and for the Model 5 74 %, 64 % and 70 %, respectively. Especially the adjusted R^2s of the Model 5 (all listed high-tech companies) are persistent throughout the forecasting horizons. It has to be noted that the Durbin-Watson statistics in Model 5 are less than two for the third and fifth year (Y3: 1.16, Y5: 1.01) and are indicating positive autocorrelation. However, the Durbin-Watson statistics are not below one which is considered as a limit for alarming autocorrelation.

5.3.2 Forecasting errors

The expected values of future earnings in the hold-out sample are calculated based on the estimated models reported in Table 8. Then the expected values are compared with the actual earnings. The descriptive statistics of the forecasting errors are documented in Table 9. The forecasts are analyzed with two approaches: precision and bias. Precision describes the absolute deviation of the expected value from the actual earnings (calculated as abs(actual – expected)). Bias takes into account also the direction of the error. Bias is considered as the raw difference of the expected value from the actual earnings (calculated as actual – expected). Both statistics are scaled by market capitalization at the end of the forecasting period. Market capitalization is chosen as the denominator instead of the actual earnings because there are plenty of close to zero earnings which tend to ‘blow up’ the errors (small denominator issue).

Estimated earnings forecast model by OLS regression with the estimation period 1990-2000.

Variable	Intercept			Mcap (+)			TAssets (-/+)			DIV (+)			DD (-)			IncExt (+)			NegIncExt (-/+)			TAccr (-)		
	Coef.	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*	*Coef.*	*t-value*	*Prob.*
Model 4: IPO high-tech firms																								
Y1	2,218	0,595	0,552	-0,003	-1,846	0,065	-0,012	-4,052	0,000	1,641	6,149	0,000	-1,481	-0,347	0,728	0,694	21,117	0,000	-5,178	-1,753	0,080	-0,217	-9,453	0,000
Y3	0,767	0,147	0,883	0,003	1,638	0,102	0,009	2,319	0,021	2,989	8,012	0,000	-4,623	-0,775	0,439	0,349	7,466	0,000	0,164	0,039	0,969	-0,010	-0,297	0,767
Y5	5,601	0,692	0,489	0,009	2,835	0,005	0,024	3,910	0,000	3,501	6,192	0,000	-5,871	-0,638	0,523	0,067	0,936	0,350	-5,621	-0,877	0,381	0,135	2,769	0,006
Model 5: All high-tech firms																								
Y1	-4,259	-1,516	0,130	0,007	14,987	0,000	-0,008	-12,916	0,000	0,603	14,566	0,000	2,610	0,787	0,431	0,690	72,896	0,000	-1,919	-0,702	0,483	-0,164	-22,709	0,000
Y3	1,204	0,298	0,766	0,013	18,838	0,000	-0,009	-9,754	0,000	0,395	6,636	0,000	-6,866	-1,438	0,150	0,651	47,934	0,000	6,750	1,696	0,090	-0,261	-25,126	0,000
Y5	4,220	0,859	0,390	0,024	29,366	0,000	-0,002	-1,718	0,086	0,142	1,974	0,048	-7,862	-1,351	0,177	0,722	43,437	0,000	6,952	1,421	0,155	-0,198	-15,682	0,000

	Adj R^2	*Durbin-Watson*	*Model F-value*	*Sig. F-value*	*N*
Model 4: IPO high-tech firms					
Y1	0,566	1,997	255,299	0,000	1 366
Y3	0,409	2,011	131,047	0,000	1 319
Y5	0,240	2,021	57,912	0,000	1 264
Model 5: All high-tech firms					
Y1	0,739	2,099	5187,834	0,000	12 838
Y3	0,635	1,163	3103,491	0,000	12 488
Y5	0,669	1,013	3508,351	0,000	12 123

Table 8. OLS regression results of the estimated earnings forecast model.

Forecasting errors: Bias and Precision

	Bias						Precision					
	Y1		Y3		Y5		Y1		Y3		Y5	
	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All
N Valid	1 644	22 085	1 225	14 917	790	8 860	1 644	22 085	1 225	14 917	790	8 860
N Missing	811	14 833	1 230	22 001	1 665	28 058	811	14 833	1 230	22 001	1 665	28 058
Mean	0.128	0.137	-0.004	0.067	-0.083	0.072	0.297	0.407	0.355	0.414	0.400	0.466
Median	0.011	0.014	0.024	0.024	-0.008	0.008	0.067	0.075	0.112	0.098	0.109	0.097
Std. dev.	0.96	6.35	1.03	6.68	1.26	7.19	0.92	6.34	0.97	6.67	1.20	7.17
Min	-13.41	-636.97	-14.88	-448.56	-22.20	-33.31	0.000	0.000	0.000	0.000	0.000	0.000
Max	14.86	668.94	13.18	668.56	6.06	666.72	14.86	668.94	14.88	668.56	22.20	666.72
Percentiles												
5	-0.357	-0.393	-0.712	-0.533	-0.814	-0.756	0.005	0.005	0.007	0.007	0.008	0.006
25	-0.051	-0.051	-0.089	-0.068	-0.154	-0.111	0.024	0.025	0.044	0.036	0.043	0.034
50	0.011	0.014	0.024	0.024	-0.008	0.008	0.067	0.075	0.112	0.098	0.109	0.097
75	0.090	0.110	0.126	0.127	0.076	0.086	0.189	0.243	0.277	0.265	0.312	0.292
95	0.977	1.137	0.757	0.886	0.552	0.856	1.183	1.451	1.343	1.340	1.705	1.592

All forecasting errors scaled by MCapYt.

One-Sample T-test

	Bias						Precision					
	Y1		Y3		Y5		Y1		Y3		Y5	
	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All	IPO	All
t	5.425	3.199	-0.127	1.221	-1.856	0.947	13.088	9.539	12.860	7.592	9.369	6.113
df	1 643	22 084	1 224	14 916	789	8 859	1 643	22 084	1 224	14 916	789	8 859
Sig.	0.000	0.001	0.899	0.222	0.064	0.344	0.000	0.000	0.000	0.000	0.000	0.000

Test value = 0

Table 9. Bias, precision and one-sample t-test results of the forecasts.

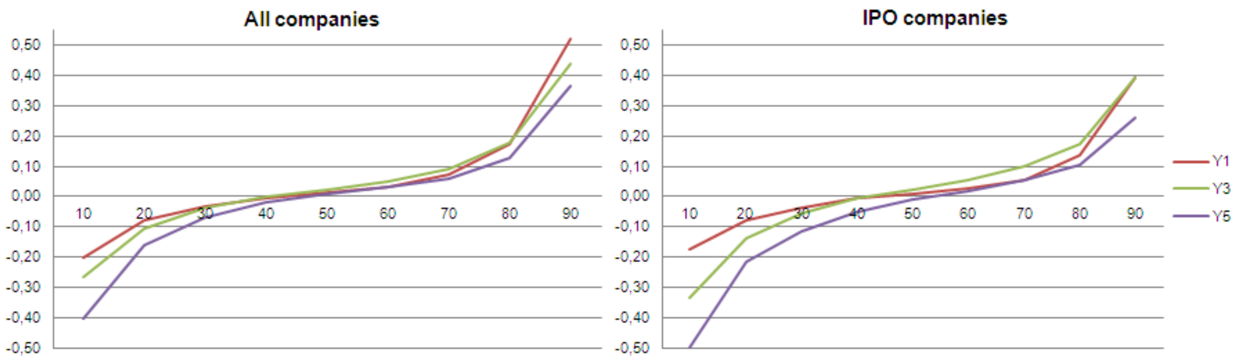


Figure 6. Bias of the forecasts for all and IPO companies.

First, the direction of the forecasting errors is analyzed. Generally a good forecasting model is considered to have neutral bias but especially the overly optimistic forecasting models should be cautiously employed. For example, Hou et al. (2010) document that both the cross-sectional model and analysts' forecasts tend to be overoptimistic but analysts' forecasts exhibit even more severe negative bias. In Figure 6, the bias of forecasts with the samples of all listed and IPO companies is reported in deciles. The extreme negative forecasting errors tend to be more negative in the IPO sample than in the sample of all listed firms. On the other hand, forecasting errors in IPO sample do not reach the levels of all companies' forecasting errors in the positive extreme values. Otherwise, the patterns over the deciles and forecasting horizons seem to be rather similar in both samples. Mostly the raw forecasting errors are concentrated between 0.10 and -0.10. Figure 7 reports the percentage of forecasts with positive bias (pessimistic forecast). In the IPO sample, the tested model underestimates the earnings of one and three years ahead but overestimates the earnings of the fifth year. With all listed high-tech companies, the model underestimates the earnings throughout the forecasting horizons. In addition, also the mean and median forecast errors in bias reported in Table 9 are positive throughout the forecasting horizons. However, in the IPO sample, the forecast errors in bias are positive only for the first year but negative for the fifth year in terms of both median and mean errors. For the third year, the mean error is negative whereas the median forecast error is positive. In a conclusion, the observed bias of the models' forecasts remains rather neutral and the model is not overly optimistic (or pessimistic).

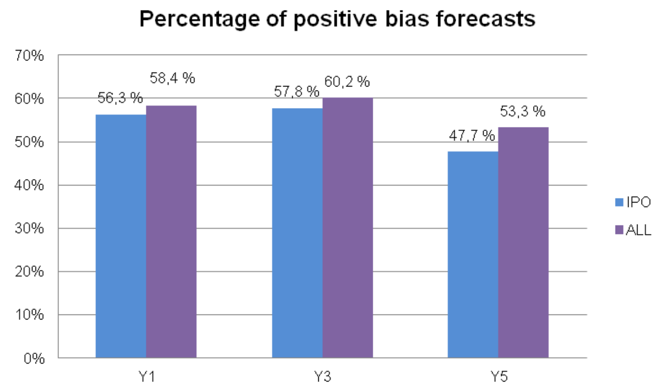
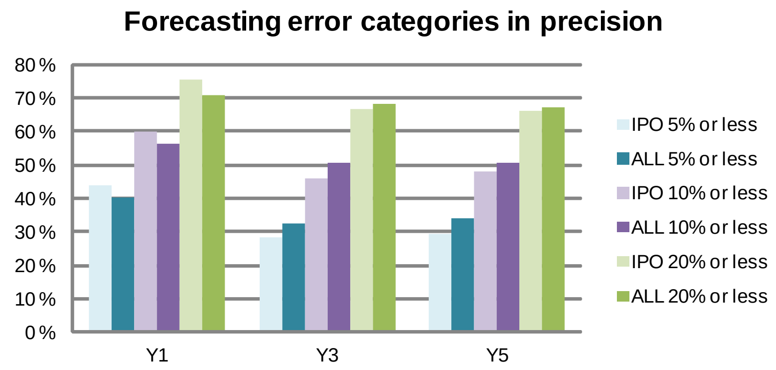


Figure 7. Percentage of positive bias forecasts.

When analyzing the precision of the forecasts in both samples, there is an interesting pattern observed. Contrary to the expectations, the model performs better in terms of the mean error in precision with the IPO sample than with the sample consisting of all listed high-tech companies. However, in terms of median errors in precision, the model performs better with the sample of all companies for the third and fifth year. All in all, the differences in the forecasting performance between the samples remain rather low. With IPO companies, the forecasting errors increase in precision on average throughout the forecasting horizons as expected whereas the forecasting errors of all companies remain rather stable but increasing towards the fifth year. In Figure 8, the forecasting errors are divided into three groups based on the precision of the forecasts: errors which are 5% or less, 10% or less or 20% or less. In other words, the errors belonging to the first category constitute 5% or less of the market value of the company at the end of the forecasting period. Interestingly, the model in the IPO sample produces more errors within 5% of the market capitalization when analyzing the first year forecasts. However, the high valuations related to IPOs might affect the percentage errors. In the third and fifth year forecasts, the model performs better for all companies than IPOs when analyzing the 5% error category. Interestingly, the mid-term forecasts of three years are not any better than the fifth year forecasts. Overall, the model performs rather well in terms of precision. The percentage of errors within the best 5% category decrease when the forecasting horizon increases but still approximately half of the forecasts in both samples reach the 10% category throughout the forecasting horizons.



	Precision	Y1	Y3	Y5
IPO	5% or less	43,8 %	28,2 %	29,2 %
	10% or less	60,0 %	46,2 %	48,1 %
	20% or less	75,7 %	66,8 %	66,2 %
ALL	5% or less	40,3 %	32,3 %	33,9 %
	10% or less	56,5 %	50,6 %	50,6 %
	20% or less	70,9 %	68,5 %	67,3 %

Figure 8. Forecasting error categories in precision.

The forecasting errors were analyzed also with one-sample t-test comparing whether the observed forecasting errors are statistically significantly different from zero. T-test results are reported in Table 9 with the descriptive statistics of the precision and bias of the forecasts errors. T-test results indicate that the errors differ statistically significantly from zero in precision throughout the forecasting horizons and both samples. In bias, the findings are statistically significantly different at the level of 0.01 in the first year for both samples. On the other hand, the errors are not statistically significantly different from zero for the third and fifth year.

The differences between the two samples can also be analyzed with independent samples t-test. The test evaluates if there are statistical differences in the observed mean values between the samples of IPO and all listed high-tech companies. The results of the t-test for the equality of means (not attached in the study) indicate that there are no statistically significant differences in the mean forecasting errors in bias and precision between IPO and all listed high-tech companies.

5.4 Explaining forecasting errors

5.4.1 Descriptive statistics

The future earnings of companies with uncertain future prospects are expected to be more difficult to forecast. In this chapter, uncertainty factors are tested whether those explain the forecasting errors calculated in the previous Chapter 5.3. The future performance of small, young, unprofitable and highly volatile growth stocks with lack of earnings history is typically difficult to forecast (see e.g. Pastor & Veronesi 2003, Baker & Wurgler 2006, Demers & Joos 2007). The chosen firm-specific uncertainty factors are age (AGE), sales growth (SalesG), analyst coverage (COV), research and development costs (RD) and intangible assets (Intang). Detailed variable descriptions are reported in Appendix 1. With this analysis of firm-specific uncertainty factors can also be tested whether the uncertainty variables improve the model of Hou et al. (2010). The model of Hou et al. (2010) already includes several variables which are recognized as uncertainty factors and are taken into account already in the estimated forecast model. Baker and Wurgler (2006) document that the performance of small (MCap, TAssets), unprofitable (NegIncExt, IncExt) and non-dividend paying (DD, DIV) companies is more difficult to predict.

The descriptive statistics of the chosen uncertainty factors are reported in Table 10. When it comes to the age of companies in the hold-out sample, there are no substantial differences between all and IPO high-tech companies. Sales growth in terms of mean values is rather high in both samples. However, in terms of median values, the sales growth remains rather subtle but higher for IPO companies. Standard deviation in sales growth is high. Only few analysts are following the sample companies. The median values of analysts coverage is zero for both samples. There are quite large differences found in R&D expenditures in terms of standard deviation. Few firms are reporting their R&D costs. Also intangibles are less reported. The both mean and median values of intangibles are almost double for all companies compared with IPO companies.

Descriptive Statistics

		AGE		SalesG		COV		RD		Intang	
		IPO	All	IPO	All	IPO	All	IPO	All	IPO	All
N	Valid	2 455	36 918	1 672	28 114	2 455	36 918	1 426	21 041	1 383	20 265
	Missing	0	0	783	8 804	0	0	1 029	15 877	1 072	16 653
Mean		8.51	8.41	2.77	7.05	1.26	2.08	13.90	37.32	49.00	97.01
Median		6	5	0.24	0.12	0	0	1.75	2.82	1.13	2.03
Std. dev.		9.66	10.43	42.22	872.07	2.65	4.48	128.12	438.39	701.02	1 317.42
Min		0	0	-1.00	-21.82	0	0	0.00	-0.93	-0.08	-465.90
Max		161	161	1 260.69	145 569.81	38	46	4 571.73	51 773.50	24 223.00	79 541.38
Percentiles											
	5	0	0	-0.40	-0.51	0	0	0.00	0.00	0.00	0.00
	25	3	2	0.04	-0.07	0	0	0.44	0.60	0.24	0.39
	50	6	5	0.24	0.12	0	0	1.75	2.82	1.13	2.03
	75	11	11	0.58	0.34	1	2	6.98	13.48	4.80	9.34
	95	24	27	2.32	1.46	5	11	48.34	96.17	49.15	131.40

Table 10. Descriptive statistics of the uncertainty factors.

Correlations between the explanatory and dependent variables are reported in Appendix 6 (IPO sample) and Appendix 7 (all companies). The precision of the forecasts is chosen as the dependent variable. Pearson correlations indicate that AGE (negatively) and SalesG (positively) and COV (negatively) are statistically significantly at the level of 0.01 related to the precision of forecasts in the first year forecasts. For the first year forecasts, all the explanatory variables, AGE (-), SalesG (-), COV (-), RD (-) and Intang (-), are statistically significant at the level of 0.01 in terms of Spearman rhos. Only RD and Intang in Spearman correlations remain statistically significant throughout the forecasting horizons. The signs of coefficients remain consistent over the forecasting horizons in both, Pearson and Spearman correlations. Only the signs of SalesG are different when the Pearson and Spearman correlations are compared. Positive coefficient would have been expected for SalesG since volatility in operations normally indicates of the uncertainty in future outcome. The negative coefficient for AGE seems to be rational because when the firm matures the uncertainty generally decreases and, therefore, forecasting errors decrease. Analysts have an incentive to follow firms with high intangibles and R&D expenditures because the market prices are less informative (Barth et al. 2001, Gu & Wang 2005). On the other hand, analysts do not normally follow smaller firms. Correlation for analyst coverage is negative but statistically significant (0.01) only for first and third year in Pearson and Spearman correlations. RD and Intang are negatively and statistically significantly related to the precision of the forecasts only in Spearman correlations. The finding contradicts with the expectations that the future performance is more difficult to predict due to the complexity of intangible information.

5.4.2 Regression results

Explanatory power of the uncertainty factors in relation to forecasting errors is first analyzed with OLS regression. Differences between countries, industries and observation years were controlled with dummy variables. The dummy variables with most frequent observations were left out from the regressions. For sensitivity analysis, the regressions were performed also without RD and Intang which have a substantial number of missing observations. The exclusion of RD and Intang did not improve the results. OLS regressions explaining the precision and bias of the forecasts violate the residuals' normality assumption. Similar S-shaped line in the normal probability plot of regression standardized residuals as in the explanatory Model 1 is recognized also in the regression results of uncertainty factors with the sample of IPOs and all listed firms. Normal distribution of residuals is one of the basic assumptions of OLS regression (Metsämuuronen 2006) and, therefore, the statistical significance of the model and variables cannot be reliably measured.

OLS regressions were tested with different sample selections in order to mitigate the violation of residuals' normality assumption. The samples were based on the percentage of the precision in forecasting errors. Similar categorization was made in Chapter 5.3.2 when analyzing the precision of the forecasts. For instance, the companies were split into different samples if the forecasting error was less or more than 20 % of the market value in the end of the forecasting period. Residuals' normality issues remained in the regressions for the samples of forecasting errors less than and more than 20% of the market value in the end of the forecasting period. For the samples of forecasting errors less or equal to 5% and less and equal to 10%, the residuals' normality issues were mitigated. However, the models remained statistically insignificant with the explanatory power close to zero. These analyses made with OLS regression are not reported in this study. It was concluded that OLS regression cannot reliably analyze the explanatory power of the uncertainty factors in relation to the forecasting errors.

Logistic regressions do not pose the similar requirements of residuals' normal distribution as the OLS regression. Therefore, similar analyses as made with OLS regression were run with binominal logistic regression. Logistic regression results are reported in Appendix 8. Several logistic regressions were executed in order to identify whether the uncertainty factors explain the differences in the precision and bias of the forecasts. When analyzing the precision, forecasting errors were split similarly to the OLS regression analyses into three categories: errors which are 5% or less, 10% or less and 20% or less. Also the errors over 20% were analyzed whether the uncertainty factors explain the biggest forecasting errors. There were no consistent factors identified explaining the errors in certain error categories. The overall explanatory power of the model in both samples remained low even though there were several statistically significant variables found. In addition, the coefficients of the statistically significant variables were close to zero (0.000) and remained insignificant in economic terms.

Only the results of the forecasting errors falling to the 10% category are reported in Appendix 8. In this study, the earnings forecast model is considered performing well if the forecasting errors are less or equal to 10% of market value of the forecasting period. Analyst coverage is the only statistically significant factor also in economic terms explaining the precision of the forecasts in both samples (IPO and all firms). The coefficient for analyst coverage is consistently positive and statistically significant at level of 0.01 throughout the forecasting horizons. Also intangible assets are recognized to be statistically significant over the forecasting horizons in all companies sample but the coefficients are insignificant (0.000) in economic terms. When it comes to Nagelkerke and Hosmer-Lemeshow statistics, the uncertainty factor model seems to have overall explanatory power in IPO sample. However, in the sample of all companies, Hosmer-Lemeshow statistics for the first and third year are statistically significant indicating that the model is not fitting well. The dummy control variables are not reported in this analysis but there is no systematic pattern identified in the statistical significance of country or industry dummies.

When analyzing the bias, forecasting errors were split into positive or zero errors receiving the value of 1 and negative errors receiving the value of 0. The goal of this split was to analyze if there are uncertainty factors found which would explain the positive bias found in the errors. The results are reported in Appendix 8. Hosmer-Lemeshow statistic is insignificant over the forecasting horizon in the IPO sample meaning that the model is fitting well. In the sample of all listed firms, Hosmer-Lemeshow is not consistently insignificant; hence, the model is not fitting well. Even though the logistic model appears to be explaining the positive bias in terms of Nagelkerke statistics, there are no consistent statistically significant explanatory variables identified. Also in this analysis, dummy control variables are not reported but there is no systematic pattern found in the statistical significance of country or industry dummies.

In a conclusion, there was no conclusive evidence found that the chosen uncertainty factors explain either the precision or bias of the forecasts in the samples of IPO and all listed high-tech firms. The variables were not consistently statistically significant and in many analyses the model itself was not statistically significant. Therefore, the model could not be reliably estimated and statistically significant factors could not be identified. The uncertainty factors do not improve the model of Hou et al. (2010). Further analysis is needed to achieve conclusive evidence which factors explain the forecasting errors produced by the earnings forecast model in the high-tech sample. These findings impact also on the applicability of Hou et al.'s (2010) model in the high-tech sample. Even though the model seems to perform rather well in terms of the precision and bias of the forecasts, the observed forecasting errors could not be explained by the generally recognized uncertainty factors. Usefulness of financial statement information requires that the riskiness of forecasted earnings should be reliably evaluated (Richardson et al. 2010). Therefore, it would have been important to identify the factors affecting the forecasting errors. The root causes of the forecasting errors should be identified in order to improve the forecasting model of Hou et al. (2010). Meanwhile, the model should be cautiously applied in the sample of high-tech companies.

6 Conclusion

In this chapter, the study is concluded with the overview of the results, limitations of the study and suggestions for future research. The objective of this study was to analyze whether financial statement information has explanatory and predictive power in relation to future earnings. Especially the usefulness of financial statement information in high-tech industries has been questioned in the prior literature. In addition, the study further analyzed whether the uncertainty factors recognized by the prior literature were associated with the forecasting errors of the estimated earnings forecast model.

The study was carried out as a cross-sectional analysis of global listed active and inactive high-tech companies from 1990-2010. The empirical part of the study was split into three phases and there were three key predictions constructed. The first part analyzed whether financial statement information explains the future earnings of newly listed high-tech companies during 1991-2010. The earnings forecasting model was replicated from Hou et al. (2010). The dependent variable consisted of absolute dollar earnings of one, three and five years ahead. The independent variables included basic accounting variables, for instance total assets, dividends paid, lagged earnings and accruals, completed with market capitalization. The first prediction suggested that financial statement information has explanatory power in relation to the future earnings of newly listed high-tech companies. The results indicate that financial statement information provides valuable information regarding the future earnings of newly listed high-tech companies. These findings are based on the ordinal regression analysis where the future earnings were divided into 20 equal sized groups based on the size of the future earnings. Ordinal regression was employed because of the statistical problems found in OLS regression. Especially lagged earnings, dividends and accruals were identified explaining the future earnings whereas the association of the size variables market capitalization and total assets with future earnings was weak. Lagged earnings and dividends paid were positively and total accruals negatively related to future earnings. In a conclusion, financial statement information is useful in explaining the future earnings of newly listed high-tech firms.

The second part tested the predictive power of the earnings forecast model. The Hou et al.'s (2010) model was estimated with OLS regression based on the data from 1990-2000 and tested in the hold-out sample from 2001-2010. The model was estimated and tested separately with the samples of all listed high-tech and IPO high-tech firms. The split was made to analyze the second key prediction that the forecasts for IPOs are less accurate than the forecasts for seasoned firms. Forecasting horizons were set to one, three and five years. The performance of the model was analyzed in terms of the precision and bias of the estimated forecasts. Overall, the model performed rather well in terms of the precision and bias of the estimated forecasts in the hold-out sample. Contrary to the second prediction, the earnings forecast model did not perform better with the sample consisting of all listed high-tech companies than with the IPO sample. In fact, the performance of the model was rather similar with both samples in terms of the mean and median errors in precision. The bias of the forecasts in both samples was close to neutral or slightly positive with an average bias of 0.045 for IPOs and 0.09 for all listed firms over all forecasting horizons. The forecasting errors tend to increase with the forecast horizon. However, the increase was not substantial and, actually, the forecasts of the third and fifth year were almost equal in terms of the mean and median errors in precision with the sample of all listed high-tech companies. Approximately half of the forecasts in both samples and throughout the forecasting horizons produced forecasting errors within 10% of the market value. In a conclusion, financial statement information is useful in forecasting the earnings of high-tech firms.

Lastly, in the third part, the study examined the association of uncertainty factors with the estimated forecasting errors of the hold-out period to improve the model of Hou et al.'s (2010). Uncertainty factors included firm age, sales growth, analyst coverage, research and development expenditures and intangible assets. The association of the uncertainty factors with the forecasting errors was analyzed with OLS and binominal logistic regressions with the samples of IPOs and all listed companies. Forecasts of future earnings were expected to be less accurate for high-tech companies with more uncertainty about the future performance. There was no conclusive evidence found that the chosen uncertainty factors explain either the precision or bias of the forecasts in the hold-out sample. These findings impact on the applicability of Hou et al.'s (2010) model in high-tech context. Even though the model seems to perform rather well in terms of the precision and bias of the forecasts, the observed forecasting errors could not be explained by the generally recognized uncertainty factors. In order to improve the forecasting model of Hou et al. (2010), the root causes of the forecasting errors should be identified. Meanwhile, the model should be cautiously applied in high-tech samples. In a conclusion, the uncertainty factors do not improve the Hou et al.'s (2010) earnings forecast model.

The findings of this study have several practical implications. The prior literature has lacked studies on earnings forecasting in high-tech industries. The studies have mostly been focused on U.S. based companies, Internet firms and on exceptional periods (IT bubble). Firstly, the findings of the study provide confirmatory evidence that financial statement information is useful in analyzing high-tech firms' future performance. Basic accounting variables are documented to have explanatory power for the future earnings of IPOs in high-tech context which is generally recognized as hard to predict. Secondly, the study provides some answers to Richardson et al.'s (2010) question which variables should be included in earnings forecasting models. The findings of the study indicate that lagged earnings, dividends and accruals have explanatory power in relation with future earnings. Thirdly, the results benefit especially the investors and analysts evaluating the performance of high-tech firms. Based on the findings of the study, Hou et al.'s (2010) cross-sectional multivariate model does fairly good job in forecasting of the earnings of high-tech firms. The forecasts of absolute dollar earnings can also be compared with the analysts' estimates. The model performed rather well in terms of the bias and precision of the forecasts in an out-of-sample test. Also the variance in the earnings with longer forecast horizon was captured by the model. The financial statement information seems to take into account the uncertainty factors recognized by the prior literature because the findings indicate that none of the uncertainty factors explained the forecasting errors.

There are few limitations recognized in this study. First of all, the group of high-tech companies is identified to be rather heterogenic. Heterogeneity within industries may reduce the incremental benefits of industry-level forecasting models (Fairfield et al. 2009). Observations of several variables were spread to leptokurtic distributions. This caused issues in the OLS regression analyses of the explanatory model. Multiple transformations of variables were tested but still the normality problems were detected in the distributions of model residuals. Therefore, the OLS regression results cannot be reliably interpreted. For mitigating the normality issues, the sample companies were split to 20 equal sized groups based on the reported earnings. Ordinal logistic regression was used to analyze the ordinal, categorical data and the results of the explanatory model are based on the ordinal regression results. In this study, the categorical dependent variable (20 groups) is considered to be a sufficient proxy for the future earnings.

The second limitation concerns the link between the variable selection and theoretical studies. There is documented lack of generally agreed variables which should be included in earnings forecast models (Richardson et al. 2010). The variables for this study were selected based on the prior empirical studies and, in fact, the tested model replicates the model introduced in Hou et al. (2010). Hence, the variable selection is not based on any theory. Also the third limitation is related to the variables in the explanatory and estimated forecast models. Neither the dependent nor the explanatory variables were scaled by any size measures. Thus, there is a possibility that the results indicate the variation in the firm sizes rather than the variation of the earnings. However, the variables were not scaled since there are also benefits in the analysis of absolute figures, for example the comparability of the forecasts with analysts' estimates.

There are interesting topics left for the future research in this area of study. Alternative explanatory variables should be tested to explain the fluctuations in the forecasts which were not be identified by the uncertainty factors. It should be important to identify which factors cause especially the biggest errors in earnings forecasts because these forecasting errors may be costly. Also other complementary variables could be tested to enhance the performance of the Hou et al.'s (2010) model with the high-tech sample. The most interesting future research topic would be to test the performance of the forecasting model in comparison with the analysts' estimates.

REFERENCES

2002 NAICS to 2007 NAICS. North American Industry Classification System (NAICS). http://www.census.gov/eos/www/naics/concordances/concordances.html, 10 March 2011.

Aggarwal, R,. Bhagat, S., Rangan, S. 2009. The Impact of Fundamentals on IPO Valuation. Financial Management, summer 2009, 253-284.

Allen, D. E., Salim H. M. 2005. Forecasting profitability and earnings: a study of the UK market (1982 - 2000). Applied Economics, 37, 2009-2018.

Amir, E., Lev, B. 1996. Value-relevance of nonfinancial information: The wireless communications industry. Journal of Accounting and Economics, 22, 3-30.

Baker, M., Wurgler, J. 2006. Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, 61, 4, 1645-1679.

Barron, O. E., Byard D., Kile, C., Riedl, E. J. 2002. High-Technology Intangibles and Analysts' Forecasts. Journal of Accounting Research, 40, 2.

Barth, M. E., Kasznik, R., McNichols, M. F. 2001. Analyst Coverage and Intangible Assets. Journal of Accounting Research, 39, 1, 1-34.

Bartov, E., Mohamram, P., Seethamraju, C. 2002. Valuation of Internet Stock – An IPO Perspective. Journal of Accounting Research, 40, 2, 321 – 346.

Beaver, W. H. 1998. Financial reporting: an accounting revolution. Third edition. Prentice Hall. Upper Saddle River, New Jersey.

Bowen, R. M., Burgstahler, D., Daley, L. A. 1986. Evidence on the Relationships Between Earnings and Various Measures of Cash Flow. The Accounting Review, 61, 4, 713-725.

Bowen, R. M., Davis, A. K., Rajgopal, S. 2002. Determinants of Revenue-Reporting Practices for Internet Firms. Contemporary Accounting Research, 19, 4, 523-562.

Brown, L. D. 1993. Earnings forecasting research: its implications for capital markets research. International Journal of Forecasting, 9, 295-320.

Brown, L. D., Griffin, P. A., Hagerman, R. L., Zmijewski, M. E. 1987. Security analysts superiority relative to univariate time-series models in forecasting quarterly earnings. Journal of Accounting and Economics, 9, 61-87.

Demers, E., Joos, P. 2007. IPO Failure Risk. Journal of Accounting Research 45, 2, 333 – 371.

Demers, E., Lev, B. 2001. A Rude Awakening: Internet Shakeout in 2000. Review of Accounting Studies, 6, 331-359.

Eisenmann, T. R. 2006. Internet companies' growth strategies: determinants of investment intensity and long-term performance. Strategic Management Journal, 27, 1183-1204.

Ernst & Young, Global IPO trends 2011.
http://www.ey.com/Publication/vwLUAssets/Global-IPO-trends_2011/$FILE/Global%20IPO%20trends%202011.pdf, 22 January, 2012.

Evans, M. 2010. Is the High-Tech IPO Really Back?
http://www.markevanstech.com/2010/08/29/is-the-high-tech-ipo-really-back/, 22 January, 2012.

Fairfield, P. M., Ramnath, S., Yohn, T. L. 2009. Do Industry-Level Analyses Improve Forecasts of Financial Performance?. Journal of Accounting Research, 47, 1, 147 – 178.

Fairfield, P. M., Yohn, T. L. 2001. Using Asset Turnover and Profit Margin to Forecast Changes in Profitability. Review of Accounting Studies, 6, 371–385.

Fama, E. F., French, K. R. 2000. Forecasting Profitability and Earnings. Journal of Business, 73, 2, 161 – 175.

Fama, E. F., French, K. R. 2001. Disappearing dividends: changing firm characteristics or lower propensity to pay? Journal of Financial Economics, 60, 3-43.

Fama, E. F., French, K. R. 2004. New lists: Fundamentals and survival rates. Journal of Financial Economics, 73, 229-269.

Fama, E. F., French, K. R. 2006. Profitability, investment and average returns. Journal of Financial Economics, 82, 491-518.

Fama, E. F., MacBeth, J. 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy, 81, 607-636.

Foster, G. 1977. Quarterly accounting data: time-series properties and predictive-ability results. Accounting Review, 52, 1-21.

Francis, J., Schipper, K. 1999. Have financial statement lost their relevance? Journal of Accounting Research, 37, 2, 319-352.

Fried, D., Givoly, D. 1982. Financial analysts' forecasts of earnings: a better surrogate for market expectations. Journal of Accounting Research, 30, 185- 209.

Garson, G. D. 2011. Logistic Regression. North Carolina State University. www.faculty.chass.ncsu.edu/garson/PA765/logistic.htm, 18 January 2012.

Gu, F., Wang W. 2005. Intangible Assets, Information Complexity, and Analysts' Earnings Forecasts. Journal of Business Finance & Accounting, 32, 9 & 10, 1673-1702.

Hand, J. R. M. 2000. Profits, losses and the non-linear pricing of Internet stocks. Working paper, January, 10, Kenan-Flagler Business School.

Hand, J.R.M. 2005. The Value Relevance of Financial Statements in the Venture Capital Market. The Accounting Review, 80, 2, 613-648.

Hecker, D. E. 2005. High-technology employment: a NAICS-based update. Monthly Labor Review, July, 57–72.

Helsingin Sanomat 13 July, 2011. Aleksi Moisio. Niin katoaa digitaalinen kunnia.

Higgings, H. N. 2011. Can Securities Analysts Forecast Intangible Firms' Earnings? Working Paper, August, 11, Worcester Polytechnic Institute School of Business.

Hou, K., Robinson, D. T. 2006. Industry Concentration and Average Stock Returns. Journal of Finance, 61, 4, 1927-1956.

Hou, K., van Dijk, M. 2010. Profitability shocks and the Size Effect in the Cross-Section of Expected Stock Returns. Working paper, January, 14, Fisher College of Business.

Hou, K., van Dijk, M. A., Zhang, Y. 2010. The Implied Cost of Capital: A New Approach. Working Paper, February, 10. Fisher College of Business.

IFRS Framework, Conceptual Framework for Financial Reporting. September 2010. www.iasplus.com/standard/framewk.htm, 30 March, 2011.

Jain, B. A., Jayaraman, N., Kini, O. 2008. The path-to-profitability of Internet IPO firms. Journal of Business Venturing, 23, 165–194.

Jain, B. A., Kini, O. 1994. The Post-Issue Operating Performance of IPO firms. The Journal of Finance, 49, 5, 1699 – 1726.

Janey, J.J., Folta, T.B. 2006. Moderating effects of investor experience on the signaling value of private placements. Journal of Business Venturing, 21, 27–44.

Kask, C., Sieber, E. 2002. Productivity growth in 'high-tech' manufacturing industries. Monthly Labor Review, March, 16–31.

Keating, E. K. 2000. Discussion of The Eyeballs Have It Searching for the Value in Internet Stocks. Journal of Accounting Research, 38, Supplement, 163-169.

Kennedy, P. 2008. A guide to econometrics. 6th edition. Blackwell Publishing. Malden, USA.

Kinnunen, J. 1988. The time series properties of accrual versus cash-based income variables: Empirical evidence from listed Finnish firms. Series A-58. The Helsinki School of Economics and Business Administration.

Kothari, S. P. 2001. Capital markets research in accounting. Journal of Accounting and Economics, 31, 105–231.

Lee, C. M. C., So E. C., Wang, C. C. Y. 2011. Evaluating Implied Cost of Capital Estimates. Working paper, September, 7.

Lev, B., Zarowin, P. 1999. The boundaries of financial reporting and how to extend them. Journal of Accounting Research, 37, 2, 353-385.

Loughran, T., Ritter, J. R. 2004. Why Has IPO Underpricing Changed Over Time? Financial Management, 33, 5-37.

Mandel, M. 2001. The coming internet depression. Prentice Hall, Pearson Education Limited. London, Great Britain. First Edition.

Manegold, J. G. 1981. Time-Series Properties of Earnings: A Comparison of Extrapolative and Component Models. Journal of Accounting Research, 19, 2, 360-373.

Metsämuuronen, J. 2006. Tutkimuksen tekemisen perusteet ihmistieteissä. 2. laitos, 3. Painos. International Methelp Ky, Helsinki.

Mudambi, R., Zimmermen Treichel, M. 2005. Cash crisis in newly public Internet-based firms: An empirical analysis. Journal of Business Venturing, 20, 543 -571.

Murphy, G., Trailer, J., Hill, R. 1996. Measuring performance in entrepreneurship research. Journal of Business Research, 36, 15 – 23.

North American Industry Classification System (NAICS), Introduction. http://www.census.gov/eos/www/naics/, 10 March, 2011.

Norusis, Marija 2010. PASW statistics 18.0 Advanced Statistical Procedures Companion: 'CHAPTER 4: Ordinal Regression' pp. 69 – 89. www.norusis.com/pdf/ASPC_v13.pdf. 18 January 2012.

OECD Science, Technology and Industry Scoreboard 2005. Executive summary. OECD Publishing, 2005. http://www.oecd.org/dataoecd/59/52/35465901.pdf, 30 March, 2011.

Palepu, K. G., Healy, P. M., Bernard, V.L., Peek, E. 2007. Business Analysis and Valuation. IFRS Edition. South-Western Cengage Learning EMEA. London, UK.

Pastor, L., Veronesi, P. 2003. Stock Valuation and Learning about Profitability. The Journal of Finance, 58, 5, 1749-1789.

Penman, S. H. 2001. Discussion of "Back to Basics: Forecasting the Revenues of Internet Firms" and "A Rude Awakening: Internet Shakeout in 2000". Review of Accounting Studies, 6, 361-364.

Porter, M. E. 1979. The Structure Within Industries and Companies' Performance. The Review of Economics and Statistics, 61, 2, 214–227.

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.

Ritter, J. R. 1991. The long-run performance of initial public offerings. Journal of Finance 46, 3-27.

Ritter, J. R. 2001. Some Factoids About the 2000 IPO Market (A brief history of the internet bubble). www.bear.warrington.ufl.edu/Ritter/ipos2000.pdf, 7 April 2012.

Ritter, J. R., Welch, I. 2002. A review of IPO activity, pricing, and allocations. Journal of Finance, 58, 1795–1828.

Schultz, P., Zaman, M. 2001. Do the individuals closest to internet firms believe they are overvalued? Journal of Financial Economics, 59, 347-381.

Sloan, R. G. 1996. Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings? The Accounting Review, 71, 3, 289-315.

Soliman, M. T. 2008. The Use of DuPont Analysis by Market Participants. The Accounting Review, 83, 3, 823-853.

SPSS Data Analysis Examples: Ordinal Logistic Regression. UCLA Academic Technology Services, Statistical Consulting Group, University of California. www.ats.ucla.edu/stat/spss/dae/ologit.htm, 24 January 2012.

Stigler, G. J. 1963. Capital and Rates of Return in Manufacturing Industries. Princeton, New Jersey. Princeton University Press.

Tong, Jian 2005. High-tech and high capability in a growth model. International Economic Review, 46, 1, 215–243.

Trotta, R.J. 2003. Translating Strategy into Shareholder Value, 'CHAPTER 7: The DuPont Model' pp. 149-162. American Management Association International Business Source Complete, EBSCOhost, 30 October 2011.

Trueman, B, Franco Wong, M. H., Zhang, X-J. 2000. The Eyeballs Have It: Searching for the Value in Internet Stocks. Journal of Accounting Research, 38, Supplement, 137 - 162.

Trueman, B, Franco Wong, M. H., Zhang, X-J. 2001. Back to Basics: Forecasting the Revenues of Internet Firms. Review of Accounting Studies, 6, 305-329.

Warfield, T.D., Wild, J.J., Wild, K.L. 1995. Managerial Ownership, Accounting Choices, and Informativeness of Earnings. Journal of Accounting and Economics, 20, 61-91.

APPENDICES

Appendix 1. Variable details.

Name	Abbreviation	Variable code in database	Calculations	Database
Net income before extraordinary items	IncExt	TF.RF.NetIncomeBeforeExtraItems		Reuters
Market capitalization	MCap	TF.PR.MarketCap		Pricing
Total assets	TAssets	TF.RF.TotalAssets		Reuters
Dividends paid	DIV	TF.RF.CashDividendsPaidCommon		Reuters
Dividend dummy	DD	TF.RF.CashDividendsPaidCommon	1 = firm not paid dividends, 0 = firm paid dividends	Reuters
Negative earnings dummy	NegIncExt	TF.RF.NetIncomeBeforeExtraItems	1 = firm reports negative earnings, 0 = otherwise	Reuters
Total accruals	TAccr	TF.RF.NetIncomeBeforeExtraItems; TF.RF.CashFromOperatingActivities	Net income before extraordinary items minus Cash from operating activies	Reuters
Cash from operating activities	CFO	TF.RF.CashFromOperatingActivities		Reuters
Founded year combined	FoundY	TF.OrganizationFoundedYear; TF.FN.CompanyFoundedDate; TF.FN.CompanyIncorporatedDate	Min(TF.OrganizationFoundedYear; TF.FN.CompanyFoundedDate; TF.FN.CompanyIncorporatedDate)	Worldscope
Listing year combined	List1	TF.PR.PriceDateFirst; TF.InstrumentActiveDate	First appearance of market capitalization matched with TF.PR.PriceDateFirst or TF.InstrumentActiveDate	Pricing & GEM
NAICS2007 code		4-digit primary NAICS		Orbis
Age	Age	TF.OrganizationFoundedYear; TF.FN.CompanyFoundedDate; TF.FN.CompanyIncorporatedDate	Year-FoundY	Worldscope
Sales growth	SalesG	TF.FN.Sales	$(Sales_{t+1}-Sales_t)/Sales_t$	Worldscope
Analyst coverage	COV	TF.ES.EPS.NumEsts		Unknown
Research and development costs	RD	TF.RF.ResearchAndDevelopment; TF.FN.ResearchAndDevelopmentExpense; TF.RF.RDExpSupplemental	Max(TF.RF.ResearchAndDevelopment; TF.FN.ResearchAndDevelopmentExpense; TF.RF.RDExpSupplemental)	Reuters, Worldscope
Intangible assets	Intang	TF.RF.IntangiblesNet		Reuters

Appendix 2. High-technology Level 1 NAICS 2002 conversion table to NAICS 2007.

Source: http://www.census.gov/eos/www/naics/concordances/concordances.html

HIGH-TECHNOLOGY INDUSTRIES

Bureau of Labor Statistics

Employment

#	Level I NAICS2002	NAICS2007
1	3254	3254
2	3341	3341
3	3342	3342
		3345
4	3344	3344
5	3345	3345
6	3364	3364
7	5112	5112
8	5161	5191
9	5179	5179
10	5181	5171
		5179
		5191
11	5182	5182
12	5413	5413
13	5415	5415
14	5417	5417

Appendix 3. Correlations of the continuous dependent variables in the explanatory model.

Pearson and Spearman Correlations

* significant at level 0.05
** significant at level 0.01

		IncExt_Y1	IncExt_Y3	IncExt_Y5	MCap	TAssets	DIV	DD	IncExt	NegIncExt	TAccr		
IncExt_Y1	Pearson		,575**	,452**	,199**	,326**	,329**	-,313**	,735**	-,626**	,080**	Spearman	IncExt_Y1
	Sig		.000	.000	.000	.000	.000	.000	.000	.000	.000	Sig	
	N		3146	2563	3696	3393	3696	3695	3416	3416	3285	N	
IncExt_Y3	Pearson	,709**		,581**	,197**	,286**	,257**	-,243**	,511**	-,434**	.016	Spearman	IncExt_Y3
	Sig	.000		.000	.000	.000	.000	.000	.000	.000	.405	Sig	
	N	3146		2872	3482	2864	3482	3482	2882	2882	2769	N	
IncExt_Y5	Pearson	,596**	,743**		,239**	,315**	,246**	-,236**	,430**	-,371**	.000	Spearman	IncExt_Y5
	Sig	.000	.000		.000	.000	.000	.000	.000	.000	.999	Sig	
	N	2563	2872		3082	2303	3082	3082	2316	2316	2216	N	
MCap	Pearson	,535**	,493**	,494**		,772**	,123**	-,106**	,250**	-,202**	-,206**	Spearman	MCap
	Sig	.000	.000	.000		.000	.000	.000	.000	.000	.000	Sig	
	N	3696	3482	3082		3627	5341	5340	3640	3640	3504	N	
TAssets	Pearson	,626**	,573**	,588**	,728**		,224**	-,193**	,395**	-,352**	-,202**	Spearman	TAssets
	Sig	.000	.000	.000	.000		.000	.000	.000	.000	.000	Sig	
	N	3393	2864	2303	3627		3627	3626	3604	3604	3472	N	
DIV	Pearson	,542**	,490**	,460**	,445**	,565**		-,997**	,381**	-,347**	.021	Spearman	DIV
	Sig	.000	.000	.000	.000	.000		.000	.000	.000	.206	Sig	
	N	3696	3482	3082	5341	3627		5340	3640	3640	3504	N	
DD	Pearson	-,197**	-,169**	-,162**	-,103**	-,137**	-,429**		-,361**	,345**	-,040*	Spearman	DD
	Sig	.000	.000	.000	.000	.000	.000		.000	.000	.018	Sig	
	N	3695	3482	3082	5340	3626	5340		3639	3639	3503	N	
IncExt	Pearson	,796**	,627**	,580**	,546**	,645**	,592**	-,224**		-,861**	,265**	Spearman	IncExt
	Sig	.000	.000	.000	.000	.000	.000	.000		.000	.000	Sig	
	N	3416	2882	2316	3640	3604	3640	3639		3640	3504	N	
NegIncExt	Pearson	-,265**	-,179**	-,162**	-,112**	-,124**	-,167**	,345**	-,351**		-,255**	Spearman	NegIncExt
	Sig	.000	.000	.000	.000	.000	.000	.000	.000		.000	Sig	
	N	3416	2882	2316	3640	3604	3640	3639	3640		3504	N	
TAccr	Pearson	-,496**	-,465**	-,447**	-,678**	-,834**	-,494**	,095**	-,404**	.005		Spearman	TAccr
	Sig	.000	.000	.000	.000	.000	.000	.000	.000	.761		Sig	
	N	3285	2769	2216	3504	3472	3504	3503	3504	3504		N	

Pearson correlation coefficients are presented in the left hand side and Spearman rho's are presented on the right hand side.

Dependent variable: continuous IncExt (t+1)

Appendix 4. Correlations of the grouped dependent variables in the explanatory model.

Pearson and Spearman Correlations

* significant at level 0.05
** significant at level 0.01

		IncExt_Y1	IncExt_Y3	IncExt_Y5	MCap	TAssets	DIV	DD	IncExt	NegIncExt	TAccr		
IncExt_Y1	Pearson		,573**	,450**	,199**	,326**	,329**	-,313**	,733**	-,625**	,081**	Spearman	IncExt_Y1
	Sig		.000	.000	.000	.000	.000	.000	.000	.000	.000	Sig	
	N		3146	2563	3696	3393	3696	3695	3416	3416	3285	N	
IncExt_Y3	Pearson	,573**		,581**	,196**	,287**	,257**	-,243**	,511**	-,434**	.016	Spearman	IncExt_Y3
	Sig	.000		.000	.000	.000	.000	.000	.000	.000	.407	Sig	
	N	3146		2872	3482	2864	3482	3482	2882	2882	2769	N	
IncExt_Y5	Pearson	,448**	,580**		,238**	,315**	,243**	-,234**	,427**	-,369**	-.002	Spearman	IncExt_Y5
	Sig	.000	.000		.000	.000	.000	.000	.000	.000	.907	Sig	
	N	2563	2872		3082	2303	3082	3082	2316	2316	2216	N	
MCap	Pearson	,161**	,166**	,171**		,772**	,123**	-,106**	,250**	-,202**	-,206**	Spearman	MCap
	Sig	.000	.000	.000		.000	.000	.000	.000	.000	.000	Sig	
	N	3696	3482	3082		3627	5341	5340	3640	3640	3504	N	
TAssets	Pearson	,218**	,191**	,193**	,728**		,224**	-,193**	,395**	-,352**	-,202**	Spearman	TAssets
	Sig	.000	.000	.000	.000		.000	.000	.000	.000	.000	Sig	
	N	3393	2864	2303	3627		3627	3626	3604	3604	3472	N	
DIV	Pearson	,243**	,210**	,183**	,445**	,565**		-,997**	,381**	-,347**	.021	Spearman	DIV
	Sig	.000	.000	.000	.000	.000		.000	.000	.000	.206	Sig	
	N	3696	3482	3082	5341	3627		5340	3640	3640	3504	N	
DD	Pearson	-,313**	,210**	-,234**	-,103**	-,137**	-,429**		-,361**	,345**	-,040*	Spearman	DD
	Sig	.000	.000	.000	.000	.000	.000		.000	.000	.018	Sig	
	N	3695	3482	3082	5340	3626	5340		3639	3639	3503	N	
IncExt	Pearson	,398**	,293**	,244**	,546**	,645**	,592**	-,224**		-,861**	,265**	Spearman	IncExt
	Sig	.000	.000	.000	.000	.000	.000	.000		.000	.000	Sig	
	N	3416	2882	2316	3640	3604	3640	3639		3640	3504	N	
NegIncExt	Pearson	-,628**	-,437**	-,372**	-,112**	-,124**	-,167**	,345**	-,351**		-,255**	Spearman	NegIncExt
	Sig	.000	.000	.000	.000	.000	.000	.000	.000		.000	Sig	
	N	3416	2882	2316	3640	3604	3640	3639	3640		3504	N	
TAccr	Pearson	-,112**	-,123**	-,114**	-,678**	-,834**	-,494**	,095**	-,404**	.005		Spearman	TAccr
	Sig	.000	.000	.000	.000	.000	.000	.000	.000	.761		Sig	
	N	3285	2769	2216	3504	3472	3504	3503	3504	3504		N	

Pearson correlation coefficients are presented in the left hand side and Spearman rho's are presented on the right hand side.

Dependent variable: grouped IncExt (t+1) to 20 groups

Appendix 5. Ordinal regression thresholds.

Thresholds

	Y1			Y3			Y5		
	Coef.	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*
Group 1	-3.157	860.69	.000	-2.483	681.76	.000	-2.291	507.82	.000
Group 2	-2.043	705.65	.000	-1.687	509.73	.000	-1.534	363.78	.000
Group 3	-1.369	434.94	.000	-1.164	305.74	.000	-1.056	212.29	.000
Group 4	-0.876	208.48	.000	-0.770	150.46	.000	-0.706	104.87	.000
Group 5	-0.443	57.10	.000	-0.457	56.08	.000	-0.427	40.42	.000
Group 6	-0.069	1.41	.235	-0.149	6.11	.013	-0.178	7.16	.007
Group 7	0.287	23.39	.000	0.080	1.76	.185	0.086	1.66	.197
Group 8	0.609	98.78	.000	0.317	27.07	.000	0.311	21.73	.000
Group 9	0.951	220.79	.000	0.532	73.93	.000	0.533	62.14	.000
Group 10	1.264	358.26	.000	0.772	149.85	.000	0.742	116.80	.000
Group 11	1.596	521.28	.000	1.004	242.04	.000	0.960	188.01	.000
Group 12	1.928	698.34	.000	1.243	352.05	.000	1.183	272.31	.000
Group 13	2.257	885.73	.000	1.510	487.79	.000	1.420	371.93	.000
Group 14	2.611	1 095.48	.000	1.776	631.45	.000	1.680	487.79	.000
Group 15	2.987	1 322.11	.000	2.097	808.76	.000	1.969	619.38	.000
Group 16	3.417	1 569.51	.000	2.451	998.58	.000	2.287	759.95	.000
Group 17	3.965	1 831.89	.000	2.888	1 202.30	.000	2.682	914.68	.000
Group 18	4.775	2 015.95	.000	3.523	1 388.44	.000	3.265	1 070.19	.000
Group 19	6.453	1 595.93	.000	4.723	1 291.54	.000	4.430	1 035.83	.000

Appendix 6. Pearson and Spearman correlations of the uncertainty factors in relation to the precision of the forecasts in the sample of IPO companies.

Pearson and Spearman Correlations: IPO companies

* significant at level 0.05
** significant at level 0.01

		PrecisionY1	PrecisionY3	PrecisionY5	AGE	SalesG	COV	RD	Intang		
PrecisionY1	Pearson		,353**	,326**	-,168**	-,082**	-,172**	-,194**	-,134**	Spearman	PrecisionY1
	Sig		.000	.000	.000	.004	.000	.000	.000	Sig	
	N		1218	786	1644	1228	1644	1059	1205	N	
PrecisionY3	Pearson	,266**		,367**	-,082**	-,094**	-,113**	-,207**	-,143**	Spearman	PrecisionY3
	Sig	.000		.000	.004	.005	.000	.000	.000	Sig	
	N	1218		783	1225	894	1225	791	883	N	
PrecisionY5	Pearson	,201**	,506**		-.054	-.017	-.028	-,102*	-,145**	Spearman	PrecisionY5
	Sig	.000	.000		.132	.687	.438	.025	.001	Sig	
	N	786	783		790	563	790	487	543	N	
AGE	Pearson	-,075**	-.012	-.035		-,086**	.005	.011	-.033	Spearman	AGE
	Sig	.002	.669	.324		.000	.804	.665	.219	Sig	
	N	1644	1225	790		1672	2455	1426	1383	N	
SalesG	Pearson	,144**	.003	,185**	-.023		,075**	,118**	.020	Spearman	SalesG
	Sig	.000	.939	.000	.353		.002	.000	.520	Sig	
	N	1228	894	563	1672		1672	1063	1049	N	
COV	Pearson	-,099**	-,069*	-.055	-.001	-.009		,484**	,342**	Spearman	COV
	Sig	.000	.016	.120	.969	.704		.000	.000	Sig	
	N	1644	1225	790	2455	1672		1426	1383	N	
RD	Pearson	-.046	-.032	-.028	-.006	.001	,493**		,345**	Spearman	RD
	Sig	.134	.376	.535	.812	.967	.000		.000	Sig	
	N	1059	791	487	1426	1063	1426		867	N	
Intang	Pearson	-.004	-.007	-.033	.016	-.005	,248**	,141**		Spearman	Intang
	Sig	.895	.841	.441	.554	.859	.000	.000		Sig	
	N	1205	883	543	1383	1049	1383	867		N	

Pearson correlation coefficients are presented in the left hand side and Spearman rho's are presented on the right hand side.
Dependent variable: Precision (t+1) of forecasts.

Appendix 7. Pearson and Spearman correlations of the uncertainty factors in relation to the precision of the forecasts in the sample of all companies.

Pearson and Spearman Correlations: all high-tech companies

* significant at level 0.05
** significant at level 0.01

		PrecisionY1	PrecisionY3	PrecisionY5	AGE	SalesG	COV	RD	Intang		
PrecisionY1	Pearson		,463**	,409**	-,024**	-,147**	-,250**	-,312**	-,249**	Spearman	PrecisionY1
	Sig		.000	.000	.000	.000	.000	.000	.000	Sig	
	N		14823	8799	22085	19540	22085	14639	16733	N	
PrecisionY3	Pearson	,383**		,459**	-.005	-,112**	-,213**	-,295**	-,258**	Spearman	PrecisionY3
	Sig	.000		.000	.564	.000	.000	.000	.000	Sig	
	N	14823		8824	14917	12920	14917	9684	10986	N	
PrecisionY5	Pearson	,025*	.009		,022*	-,069**	-,153**	-,272**	-,241**	Spearman	PrecisionY5
	Sig	.019	.389		.035	.000	.000	.000	.000	Sig	
	N	8799	8824		8860	7631	8860	5610	6244	N	
AGE	Pearson	-.012	-.013	-.013		-,022**	-,049**	-,065**	-,112**	Spearman	AGE
	Sig	.067	.105	.224		.000	.000	.000	.000	Sig	
	N	22085	14917	8860		28114	36918	21041	20265	N	
SalesG	Pearson	,058**	,088**	.021	-.002		,096**	,045**	,096**	Spearman	SalesG
	Sig	.000	.000	.065	.755		.000	.000	.000	Sig	
	N	19540	12920	7631	28114		28114	18672	18145	N	
COV	Pearson	-,021**	-,018*	-.018	-.010	-.004		,592**	,452**	Spearman	COV
	Sig	.002	.028	.099	.063	.526		.000	.000	Sig	
	N	22085	14917	8860	36918	28114		21041	20265	N	
RD	Pearson	-.006	-.006	-.006	.008	-.001	,244**		,479**	Spearman	RD
	Sig	.464	.583	.665	.248	.916	.000		.000	Sig	
	N	14639	9684	5610	21041	18672	21041		13369	N	
Intang	Pearson	-.002	-.002	-.003	.003	-.001	,303**	,344**		Spearman	Intang
	Sig	.768	.835	.812	.664	.934	.000	.000		Sig	
	N	16733	10986	6244	20265	18145	20265	13369		N	

Pearson correlation coefficients are presented in the left hand side and Spearman rho's are presented on the right hand side.

Dependent variable: Precision (t+1) of forecasts.

Appendix 8. Logistic regression results explaining forecasting errors.

Logistic explanatory model: uncertainty factors vs precision. Error category: within 10%.

Variable	Intercept			AGE			SalesG			COV			RD			Intang		
	Coef.	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*
IPO firms																		
Y1	114,338	0,000	0,999	0,019	1,857	0,173	-0,007	0,665	0,415	0,151	6,188	0,013	0,003	0,245	0,620	0,002	0,340	0,560
Y3	-71,280	0,000	1,000	-0,027	3,291	0,070	0,014	0,292	0,589	0,126	4,368	0,037	0,008	1,278	0,258	-0,002	0,589	0,443
Y5	-22,059	0,000	1,000	-0,032	1,965	0,161	-0,013	0,244	0,622	0,324	7,773	0,005	0,004	0,157	0,692	0,000	0,002	0,961
All firms																		
Y1	-10,635	0,000	1,000	0,008	10,237	0,001	-0,005	3,939	0,047	0,125	308,487	0,000	0,001	2,846	0,092	0,000	34,621	0,000
Y3	33,798	0,000	1,000	0,005	2,155	0,142	-0,004	1,157	0,282	0,097	141,711	0,000	0,002	6,825	0,009	0,000	13,379	0,000
Y5	203,847	0,000	0,999	0,005	1,562	0,211	-0,014	1,439	0,230	0,096	72,709	0,000	0,004	9,012	0,003	0,000	8,939	0,003

	Cox and Snell	*Nagelkerke*	*Hosmer-Lemeshow*	*N*
IPO firms				
Y1	0,150	0,212	0,105	556
Y3	0,217	0,289	0,356	400
Y5	0,315	0,420	0,914	225
All firms				
Y1	0,098	0,133	0,000	9 583
Y3	0,101	0,135	0,045	5 983
Y5	0,105	0,140	0,116	3 281

Dependent variable receiving value of 1 if the observed error is within 10% of the firm's market value.

Logistic explanatory model: uncertainty factors vs bias. Explaining positive forecasting errors.

Variable	Intercept			AGE			SalesG			COV			RD			Intang		
	Coef.	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*	*Coef.*	*Wald*	*Prob.*
IPO firms																		
Y1	69,531	0,000	1,000	-0,001	0,005	0,942	-0,009	0,669	0,413	-0,016	0,118	0,731	0,007	2,246	0,134	-0,004	2,430	0,119
Y3	62,437	0,000	1,000	-0,025	2,380	0,123	-0,041	1,695	0,193	0,023	0,151	0,698	0,012	2,915	0,088	-0,017	5,671	0,017
Y5	51,730	0,000	1,000	0,004	0,022	0,882	-0,465	3,608	0,058	-0,191	2,475	0,116	0,057	4,858	0,028	-0,003	0,222	0,637
All firms																		
Y1	-97,691	0,000	0,999	-0,003	1,281	0,258	-0,002	2,222	0,136	-0,004	0,578	0,447	0,000	0,364	0,547	0,000	5,602	0,018
Y3	98,822	0,000	0,999	0,001	0,212	0,645	-0,002	0,978	0,323	0,000	0,003	0,958	0,000	7,169	0,007	0,000	4,245	0,039
Y5	1,879	0,000	1,000	0,008	3,953	0,047	-0,016	1,998	0,158	0,024	9,085	0,003	0,000	1,700	0,192	0,000	6,946	0,008

	Cox and Snell	*Nagelkerke*	*Hosmer-Lemeshow*	*N*
IPO firms				
Y1	0,157	0,210	0,601	556
Y3	0,242	0,323	0,964	400
Y5	0,363	0,488	0,721	225
All firms				
Y1	0,053	0,072	0,115	9 583
Y3	0,085	0,114	0,030	5 983
Y5	0,124	0,166	0,378	3 281

Dependent variable receiving value of 1 if the observed forecasting error is positive or zero.