U.PORTO
FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

Predicting and understanding the rankings of financial analysts

Artur Aiguzhinov

Tese de Doutoramento em Economia

Supervised by:
Carlos Soares
Ana Paula Serra

2015

Biographical Note

Artur Aiguzhinov was born on August 15th, 1976 in Shelekhov, Irkutsk region, USSR (now Russian Federation). In 1998 he graduated from the Russian State University for the Humanities, Moscow, with Bachelor degree in World Economy. In 2000, he pursued his Master degree in Business Administration from the Western Washington University (WWU), Bellingham, WA which he concluded in 2002.

While at WWU, he did his internship as a Financial Analyst with Port of Seattle's Seattle Tacoma International Airport and managed the budget analysis of the aeronautical line of business. A completion of this 3-month summer internship lead to a year-long post-graduate optional practical training with the same functions in 2002-2003.

By the recommendation of a professor at WWU, Artur pursued his doctorate studies in economics at the Department of Economics (FEP) from the University of Porto, Portugal. In 2008, he received a Research Grant from FEP under supervision of Prof. Carlos Soares and Prof. Ana Paula Serra which lead to the present thesis. In 2010, he received a PhD grand from FCT to conclude his doctorate studies.

Acknowledgment

I thank my advisors, Carlos Soares and Ana Paula Serra, for their support and exceptional guidance. Their insightful ideas and helpful discussions are invaluable and greatly acknowledged. I also thank Pavel Brazdil, Luis Torgo, Alipio Jorge, Germán Creamer for their comments and suggestions.

Any written acknowledgment is not enough for the love, patience and understanding of my wife, my parents, and my children. I hope that I made all of you proud.

Resumo

Os rankings dos analistas financeiros são amplamente utilizados por fornecedores de informação financeira para avaliar o desempenho dos analistas em termos das suas recomendações de investimento, preços-alvo, e estimativas de resultados. Por um lado, os analistas no topo do ranking recebem maior reconhecimento entre os seus pares e têm mais atenção de potenciais investidores. Por outro lado, os fornecedores da informação financeira utilizam os rankings para criar recomendações "inteligentes" e vendê-las aos investidores. Assim, os rankings dos analistas são importantes e relevantes para muitos participantes no mercado. Neste trabalho mostramos que é possível prever os rankings e usá-los para delinear uma estratégia de investimento que resulta em retornos ajustados pelo risco positivos acima dos que seriam gerados por uma estratégia de investimento no mercado ou baseada em estimativas de consenso. O trabalho realizado inclui ainda a caracterização do desempenho relativo dos analistas; e a identificação das variáveis que mais contribuíram para alterações nos rankings. Para responder a estas questões, adaptámos e aplicámos um algoritmo de Machine Learning que modeliza a relação entre variáveis económicas e os rankings. O algoritmo baseia-se no conceito de semelhança entre os rankings e utiliza uma probabilidade de "label ranking", obtido a partir da transformação Bayesiana, para prever o ranking mais provável dado um conjunto das variáveis independentes. Em resumo, a contribuição de nossa trabalho é a seguinte: primeiro, mostramos que as estratégias de investimento baseadas nos rankings dos analistas superam aquelas baseadas mas estimativas de consenso dos analistas; segundo, adaptamos um algoritmo de classificação para resolver um problema de previsão dos rankings; terceiro, analisamos a capacidade explicativa das variáveis económicas de estado e identificamos as que determinam alterações nos rankings; finalmente, conseguimos prever os rankings dos analistas e mostramos que as estratégias de investimento baseadas em previsões de rankings superam aqueles baseados em rankings previstas sem modelo.

Abstract

Rankings of financial analysts are widely used by financial research vendors to evaluate the analysts' performance in terms of their recommendations, price target accuracy, and earnings forecasts accuracy. On one hand, top analysts receive a greater acknowledgment among their peers and more attention from the potential investors. On the other hand, financial information vendors utilize the rankings to create smart recommendations and sell them to investors. Thus, rankings of financial analysts are important and relevant for many financial market players. In this work we show that it is possible to predict rankings of financial analysts and use these rankings in active trading strategies with risk-adjusted returns above market returns and those that would result from using consensus estimates. We also address the problem of characterizing the general behavior of analysts' relative performance and identifying the variables that contributed the most to changes in rankings. We solve these tasks by adapting and applying a Machine Learning label ranking algorithm that models a relation between the state variables and the rankings. The algorithm relies on the concept of ranking similarity and uses a label ranking probability, obtained from the Bayesian transformation, to predict the most probable ranking given a set of descriptive independent variables. In summary, the contribution of our research is four-fold. First, we show that trading strategies based on analysts' rankings outperform those based on analysts' consensus estimates; second, we adapt a *naive Bayes* classification algorithm to solve a label ranking problem; third, we analyze state variables and identify the most contributive to changes in rankings; finally, we predict the rankings of financial analysts and show that the trading strategies based on model-predicted rankings outperform those base on a non-model rankings.

Contents

List of Figures

List of Tables

Introduction

Rankings, by definition, show who is the best in the field. People subconsciously believe that those who are in the top rank are the best ones. This can be true if the top ranked achievement can be easily observed (like sport event).

The idea that financial analysts play an important role in financial markets is rather consensual (Cowles, 1933; O'Brien, 1990). Yet there is some debate on whether following the advice of analysts brings value to investors after transaction costs (Womack, 1996; Mikhail, Walther, and Willis, 2004; Li, 2005). Related to this is the difficulty in identifying the analysts with superior stock picking skills. In this paper, we show that the rankings of financial analysts are useful to investors because strategies based upon these rankings yield positive abnormal returns.

In recent years, some institutions have been very active in publishing and selling rankings of financial analysts. Some rankings are based on privately held surveys of buy-side analysts (e.g., the Institutional Investor's rankings of the All-America Research Teams[1] and Bloomberg's America's Best Stock Analysts[2]); others are based on the performance of sell-side analysts (ThomsonReuters' top StarMine analysts[3]). In any of these cases, the rankings aim at identifying the top analysts. However, aside from personal acknowledgment among peers, it is still arguable whether these are useful to investors (Desai, Liang, and Singh, 2000) or are merely "popularity contests" (Emery and Li, 2009).

We show that the top ranked analysts have stock picking skills. The contributions of our research is fourfold. First, we develop a trading strategy that transforms the rankings of financial analysts into inputs for the Black-Litterman model (Black and Litterman, 1992). Second, we show that annualized cumulative returns generated by some trading

[1] http://www.institutionalinvestor.com/Research/4560/First-Team.html

[2] http://www.bloomberg.com/news/2013-08-14/jpmorgan-top-stock-picker-with-equities-out-of-lockstep.html

[3] http://excellence.thomsonreuters.com/award/starmine

strategies based upon analysts' rankings outperform a passive strategy (e.g., buy-and-hold the general stock market index). Third, we show that the strategy based upon the perfect foresight of rankings yields the highest cumulative annualized return. Fourth, we find that investors are better off following analysts that issue the most accurate target prices, rather than those that issue the most accurate EPS forecasts.

Label ranking is an increasingly popular topic in the machine learning literature. It studies the problem of learning a mapping from instances to rankings over a finite number of predefined labels. In some sense, it is a variation of the conventional classification problem; however, in contrast to the classification settings, where the objective is to assign examples to a specific class, in label ranking we are interested in assigning a complete preference order of labels to every example (Cheng, Hühn, and Hüllermeier, 2009).

Many algorithms have been adapted to deal with label ranking such as: decision-trees for label ranking (Cheng et al., 2009), algorithm based on Plackett-Luce model (Cheng, Dembczynski, and Hüllermeier, 2010), pairwise comparison (Hüllermeier, Fürnkranz, Cheng, and Brinker, 2008), and k-NN for label ranking (Brazdil, Soares, and Costa, 2003). Table 2.5 outlines the recent developments in solving a label ranking problem.

The Efficient Market Hypothesis (EHM) (Fama, 1970) suggests that all public information available to investors is incorporated in prices and new information is immediately reflected in valuations. Yet there are information gathering costs and financial analysts are better than an average investor at processing this information which reflects in issued buy/sell recommendations. These recommendations, like other news about the general economy as well as about a particular company, influence investors' perception and beliefs.

Previous studies show that analysts stock recommendations have investment value (Womack, 1996; Barber, Lehavy, McNichols, and Trueman, 2001). The literature also suggests further that foreknowledge of analyst forecast accuracy is valuable (Brown and Mohammad, 2003; Aiguzhinov, Serra, and Soares, 2015a). In line with academic research findings, practitioners too pay attention to analyst forecast accuracy rankings. On an annual basis, firms such as The Institutional Investor and StarMine[4] publish analysts ratings according to how well they performed, based partly on past earnings forecast accuracy.

The importance of these ratings should not be ignored because the attention that the market gives to the recommendations of different analysts is expected to correlate with

[4]http://www.starmine.com

them. Typically, the performance of analysts is analyzed in terms of their individual characteristics (e.g., experience, background) (Clement, 1999). The disadvantage of this approach is that the collection of the necessary data is difficult and it is not always reliable. As for practitioners, they rely mostly on past accuracy to predict future accuracy.

Rankings of financial analysts is not new in finance. Many agencies develop their procedures to evaluate analysts based on their performance either in forecasting or stock recommendations. Some institutions even hold a "Red Carpet" event to recognize the top analysts. On one hand, for market participants, the rankings may signal who is the best analysts. On the other hand, studies have shown that following the best analysts' recommendations of buy-sell stocks have statistically insignificant benefits.

This work addresses issue of proving that there are experts in stock markets and following their advice would generate abnormal returns.

In this thesis we address the topics of understanding and predicting the rankings of financial analysts. The research is split into four self-contained papers each addressing a specific question.

In general, rankings of financial analysts are widely used by financial research vendors to evaluate the analysts' performance in terms of their recommendations, price target accuracy, and earnings forecasts accuracy. On one hand, top analysts receive a greater acknowledgment among their peers and more attention from the potential investors. On the other hand, financial information vendors utilize the rankings to create smart recommendations and sell them to investors. Thus, rankings of financial analysts are important and relevant for many financial market players.

The first paper investigates whether rankings of financial analysts could be beneficial to investors. Indeed, the financial literature still debates on whether following advice of financial analysts brings value to investors after transaction costs. If it does, then it is worth to try to identify the analysts with the superior stock picking skills. We show that by ranking the analysts on the basis of their accuracy in forecasting stock prices in one year horizon, we are able to develop trading strategies that generate positive abnormal returns.

In the first paper, with a simple setup, we develop a trading strategy that transforms the rankings of financial analysts into inputs for the Black-Litterman model Black and Litterman (1992). The model incorporates "views" in a CAPM framework, forming optimal portfolios in a mean-variance optimization setting. "Views" are expectations on individual stocks' future performance obtained from the weighting an individual analyst's according to her rank: the view of the analyst that has a rank of 1 is weighted by 100% and the further ranks down, the lesser is the weight. In this setup, we assume and further test if top ranked analysts have stock picking skills. To obtain the information for the ranking, we look at three possibilities. First, we look at last known rank of analysts; this way, we see if the *recent* development in the stock market contributes to the analysts' knowledge. Second assumption is that the *all-time* history of a stock is important for a stock valuation analysis. Third, is a hypothetical assumption where we use a perfect forecast and assume investors have perfect foresight of analysts' rankings. Each of these three settings resulted in three different sets of rankings that we use as inputs for the Black-Litterman model. The strategy based on the rankings with the perfect foresight yields of course the maximum annualized cumulative returns. Out of the feasible strategies, a strategy based on rankings of analysts who issue more accurate price targets and use the most recent information set outperforms all other feasible strategies. The evidence confirms our assumption that rankings are important and further proves that by obtaining the recent rankings and setting up a simple trading strategy, an investor gains a cumulative annualized return that is higher than that of a passive strategy of buy-and-holding the market. These results suggest further that investor should try to predict the rankings of financial analysts.

The second paper explores the possibility of predicting the rankings of items (labels) and introduces a Machine Learning algorithm for this purpose. The main research questions are: given observed rankings and independent variables that describe the differences in these rankings, what would be the next most likely (similar) ranking? To answer this question, we adapt the naive Bayes classification algorithm to deal with this problem. Despite its limitations, this algorithm demonstrates good results in many applications. Specifically, the Bayesian framework is well understood in many domains such as Financial Economics where Bayesian models are widely used (e.g., the Black-Litterman model for active portfolio management). An algorithm models, via calculating conditional probabilities, a mapping between independent variables and corresponding rankings and, once a new set of independent variables is observed, it predicts the ranking. The paper shows the results of the experiments in testing the performance of the algorithm compared to other label ranking algorithms. We show that it consistently outperforms a baseline method and is competitive with other algorithms.

The third paper analyzes the variables that drive the differences in analysts' rankings. In this paper, we investigate whether it is possible to identify factors that influence analysts' opinions by looking at states in which these analysts make decisions as opposed to analyze characteristics of individual analysts as done in previous literature. To achieve this goal, we first, build rankings of analyst based on their EPS forecasts accuracy. Then, we select the state variables that are responsible for the differences of analysts' ranks. Finally, we apply a Machine Learning label ranking algorithm to build a model that relates the rankings with the variables and calculates a discriminative power of a variable, i.e., the contribution of each variable to rankings. Our results suggest that the variation in rankings is due to the different ability of the analysts to interpret the information environment (e.g., whether the market is bull or bear). We, thus, select and analyze variables that characterize this environment. To capture the full spectrum of the analyst's decision making process, we select variables based on different levels of information availability: analyst-specific, firm-specific and about the general economy. For each level, we choose variables that information asymmetry and uncertainty. As for analyst-specific variables, we select *dispersion*, *asymmetry*, and *uncertainty* in forecast. On a stock-specific level, we use the following variables: *book-to-market*, *debt-to-equity*, *size*, *stock returns*, *accruals*, and *industry returns*. Finally, to characterize the general economy, we select *GNP*, *inflation*, *interest rate*, and general *market volatility*. We also control for different states of the variables across time and introduce three methods to capture it: the first-difference of variables, the random part of the time-series decomposition, and the rolling eight-quarter standard deviation. We report that, controlling for the dynamics in the variables, the variables that influence the rankings are particularly those that capture the conditions of the general economy. For the first difference method, the variable with the most discriminative power is the *interest rate*; for the random dynamic state, the most significant variables are *GNP* and *inflation*. Finally, for the rolling standard deviation, it is again *interest rate*. For the static state of the variables, the most significant variable is the *dispersion* in EPS forecasts among analysts. Given the identified state variables we can use them to predict the rankings of financial analyst applying label ranking algorithm.

In the fourth paper, we apply the selected discriminative variables and predict the analysts' rankings. We report that, for the case when analysts' rankings are based on the price target accuracy, our model can predicted rankings that are more accurate than those obtained from the simple model of using the *all-time* average analysts' rankings. For the case of rankings based on the EPS forecasts accuracy, our model is able to predict rankings that are better than those of the *all-time* and the *recent* baselines. Moreover, the supremacy of our model occurs when the variables that characterize analysts' information environment exhibit the first-difference dynamic state. We also perform a back-test of active trading using the predicted rankings as inputs for the Black-Litterman model. The results show that the strategies based on the analysts' rankings outperform, in terms of the annualized cumulative return, a strategy based on the analysts' consensus. Furthermore, out of the ranking based trading strategies, the maximum annualized cumulative return yields a strategy that is based on the predicted rankings with the first-difference of state variables. Thus, we show that based on the predictions of the rankings of financial analysts, it is possible to devise successful trading strategies.

The thesis is organized as follows Chapter 1 presents the first paper; Chapter 2 covers the second paper; paper three is discussed in Chapter 3; Chapter 4 is devoted to paper four; finally, conclusion summarizes the main results of the thesis, outlines its limitations, and proposes the ideas for the future work.

Chapter 1

Are rankings of financial analysts useful to investors?

Abstract

Several institutions issue rankings of financial analysts based on the accuracy of their price and EPS forecasts. Given that these rankings are *ex-post* they may not be useful to investors. In this paper we show that trading strategies based on perfect foresight and on past rankings outperform a passive strategy. In addition, we report that investors are better off following analysts that issue accurate price targets rather than following those with accurate EPS forecasts

keywords: financial analysts; rankings; target price forecasts; earnings forecasts; portfolio management
JEL: G11; G14; G24; G29

1.1 Introduction

The idea that financial analysts play an important role in financial markets is rather consensual (Cowles, 1933; O'Brien, 1990). Yet there is some debate on whether following the advice of analysts brings value to investors after transaction costs (Womack, 1996; Mikhail et al., 2004; Li, 2005). Related to this is the difficulty in identifying the analysts with superior stock picking skills. In this paper, we show that the rankings of financial analysts are useful to investors because strategies based upon these rankings yield positive abnormal returns.

In recent years, some institutions have been very active in publishing and selling rankings of financial analysts. Some rankings are based on privately held surveys of buy-side analysts (e.g., the Institutional Investor's rankings of the All-America Research Teams[1] and Bloomberg's America's Best Stock Analysts[2]); others are based on the performance of sell-side analysts (ThomsonReuters' top StarMine analysts[3]). In any of these cases, the rankings aim at identifying the top analysts. However, aside from personal acknowledgment among peers, it is still arguable whether these are useful to investors (Desai et al., 2000) or are merely "popularity contests" (Emery and Li, 2009).

We show that the top ranked analysts have stock picking skills. The contributions of our research is fourfold. First, we develop a trading strategy that transforms the rankings of financial analysts into inputs for the Black-Litterman model (Black and Litterman, 1992). Second, we show that annualized cumulative returns generated by some trading strategies based upon analysts' rankings outperform a passive strategy (e.g., buy-and-hold the general stock market index). Third, we show that the strategy based upon the perfect foresight of rankings yields the highest cumulative annualized return. Fourth, we find that investors are better off following analysts that issue the most accurate target prices, rather than those that issue the most accurate EPS forecasts.

The paper is organized as follows: Section 1.2 provides motivation to use rankings of financial analysts; Section 1.3 outlines our proposed trading strategies; Section 1.4 describes the sample and presents some preliminary results; Section 1.5 presents and discusses the results; and Section 1.6 concludes.

1.2 Industry Rankings of Financial Analysts

In the financial literature there has been a long debate on whether financial analysts produce valuable advice. Some argue that following the advice of financial analysts, translated as recommendations of buying, holding, or selling a particular stock, does not yield abnormal returns, i.e., returns that are above the required return to compensate for risk. The Efficient Market Hypothesis (Fama, 1970) states that financial markets are efficient and that any public available information regarding a stock would be immediately re-

[1] http://www.institutionalinvestor.com/Research/4560/First-Team.html

[2] http://www.bloomberg.com/news/2013-08-14/jpmorgan-top-stock-picker-with-equities-out-of-lockstep.html

[3] http://excellence.thomsonreuters.com/award/starmine

flected in prices; hence, it would be impossible to generate abnormal returns based upon past information.

Yet, several authors have since stressed that there are information-gathering costs and information is not immediately reflected on prices (Grossman and Stiglitz, 1980). As such, prices may not reflect all the available information at all time because if this were the case, those who spent resources to collect and analyze information would not have an incentive to do it, because there would not get any compensation for it.

Some authors show that financial analysts' recommendations create value to investors (Womack, 1996; Barber et al., 2001)[4]. Assuming that some analysts produce valuable advice it makes sense to rank analysts based on the accuracy of their recommendations.

StarMine rankings are based on financial analysts' accuracy either on TP or EPS forecasts. To rank analysts based on EPS forecasts, StarMine developed a proprietary metric called a Single-stock Estimating Score (SES). This score measures "... [a] relative accuracy; that is, analysts are compared against their peers. An analyst's SES can range from 0 to 100, with 50 representing the average analyst. To get a score higher than 50, an analyst must make estimates that are both significantly different from and more accurate than other analysts' estimates"[5].

As for target price ranking, StarMine's methodology compares the portfolios based on analysts recommendations. Portfolios are constructed as follows. For each "Buy" recommendation, the portfolio is one unit long the stock and simultaneously one unit short the benchmark. "Strong buy" gets a larger investment of two units long the stock and two units short the benchmark. "Hold" invests one unit in the benchmark (i.e., an excess return of zero). "Sell" recommendations work in the reverse way. StarMine rebalances its calculations at the end of each month to adjust for analysts revisions (adding, dropping or altering a rating), and when a stock enters or exits an industry grouping.

Recent evidence suggests that top ranked financial analyst affect market participants: prices seem to react more to the recommendations issued by the top-ranked analysts (Emery and Li, 2009). As such, StarMine ranking based models can be used to identify such an-

[4]Womack (1996) finds that post-recommendation excess returns are not mean-reverting, but are significant and in the direction forecast by the analysts. Barber et al. (2001) finds that over the period of 1986-1996 a portfolio of stocks with the most (least) favorable consensus analyst recommendations yields an average abnormal return of 4.13 (-4.91)%.

[5]http://excellence.thomsonreuters.com/award/starmine?award=Analyst+Awards&award_group=Overall+Analyst+Awards

alysts and generate superior estimates (e.g., SmartEstimates[6]).

The goal of our study is to evaluate if and how rankings add value to investors. With this purpose, we develop several sets of active trading strategies, selecting the stocks most favored by analysts. The first strategy is based on the consensus estimate (giving equal weights to analysts' recommendations). The second set of strategies takes in consideration the analysts' target price and EPS accuracy ranks to form "smart estimates". For the latter set of strategies, we analyse different time information sets to define the accuracy of the analysts.

We compare the performance of the strategies based upon two types of rankings (target price and EPS forecast accuracy). By doing this, we indirectly address the ongoing debate in the literature on whether analysts, when issuing the target price reports, rely on simple growth-based models or use more complex models (such as the residual income model of Ohlson (1995)). For example, Bradshaw (2004) suggests that analysts' EPS forecasts are consistent with their price targets and that analysts use growth models based on EPS forecasts to estimate stocks target prices. Differently, Simon and Curtis (2011) argue that the most accurate analysts rely on more complex models in setting their price targets[7].

1.3 Trading Strategies

Our trading strategy uses the framework for active portfolio management proposed by Black and Litterman (1992). The model incorporates "views" in a CAPM framework, forming optimal portfolios in a mean-variance optimization setting. "Views" are expectations on individual stocks' future performance.

While in the CAPM model expected returns are a function of systematic risk, in the BL model some stocks can be over- or under-priced and, therefore, their alphas are non-zero. The model blends the subjective views of investors about future performance of a stock with the market implicit returns given by CAPM.

Chen, Da, and Schaumburg (2015) apply the BL model and use the consensus expected returns as a proxy for views. They report that the resulting strategy outperforms a

[6]http://www.starmine.com/index.phtml?page_set=sm_products&sub_page_set=sm_professional&topic=analytical&section=accurate_estimates

[7]A further major development in the theoretical accounting literature on equity valuation models is the abnormal earnings growth (AEG) model of Ohlson and Juettner-Nauroth (2005), which relates share price to the level of expected earnings per share.

passive buy-and-hold strategy. Our approach is similar to theirs but we define views not only based on consensus estimates but also on smart estimates that account for previous analysts' TP and EPS accuracy.

Here below we kept the notation in Black and Litterman (1992). Q is the vector of expected returns for the eligible stocks; Ω matrix is the confidence of Q. Altogether these two reflect the views of a particular analyst or a set of analysts.

To proxy expected returns we compare the analyst' 12-month target price (TP) with today's stock price. Confidence Ω for stock is based on variation of forecasts across analysts which is similar to the measure of dispersion in analysts' opinions outlined in Diether, Malloy, and Scherbina (2002).

We define a trading strategy as follows (Figure 1.1):

1. At the beginning of quarter t for each stock i, we define Q and Ω (see Section 1.3.1 and Section 1.3.2);

2. Using the market price information available at the last day of quarter $t-1$, we obtain the market implicit returns for each stock i, and the variance/co-variance matrix;

3. We apply the BL model to get optimal portfolio weights on the basis of combining views and implicit returns. We buy or sell stocks accordingly. At the beginning of $t+1$, based on the new views, we set the new portfolio weights following steps 1–3.

1.3.1 Defining Q

For the consensus strategy, we use median of expected returns for a particular stock i:

$$Q_{cons,i} = \text{median}\left\{r_{j,i}\right\} \tag{1.1}$$

where $r_{j,i} = TP_{j,i}/P_i - 1$ is last known analyst's j expected return computed using the analyst price target $TP_{j,i}$ and stock price P_i[8].

[8]Consistent with the literature, we use stock price 3 days *ex-ante* the TP announcement. This is done to avoid any information leakage around new TP announcement day (Bonini, Zanetti, Bianchini, and Salvi, 2010)

For the strategies that weight the analysts' estimates of expected return the weight of each analyst j is based on his/her rank such that the top analyst has the weight of 1 and then the weights diminish as the rank increases.

$$w_{j,i} = 1 - \frac{rank_{j,i} - \min_i \{rank\}}{\max_i \{rank\}} \tag{1.2}$$

The expected rank-weighted return is thus:

$$Q_{rank,i} = \frac{\sum_{j=1}^{N} (w_{j,i} \times r_{j,i})}{\sum_{j=1}^{N} w_{j,i}} \tag{1.3}$$

N is the number of analysts.

As mentioned above, we use both target price and EPS accuracy rankings.

1.3.1.1 Target Price ranking

Analysts are ranked on the basis of Proportional Mean Absolute Forecast Error (PMAFE) that measures the accuracy of a forecast (Clement, 1999; Brown, 2001; Ertimur, Sunder, and Sunder, 2007). First, we define the forecast daily error $FE_{j,i}$ as the absolute value of the difference between analyst' target price TP_j and the daily stock price P for each stock i:

$$FE_{j,i}^{TP} = |P_i - TP_{j,i}| \tag{1.4}$$

The PMAFE is given as:

$$PMAFE_{j,i}^{TP} = \frac{FE_{j,i}^{TP}}{\overline{FE_i^{TP}}} \tag{1.5}$$

where $\overline{FE_i^{TP}}$ is the average forecasting error across analysts. The target price is fixed over the quarter unless it gets revised.

The rank that enters Equation (1.2) is average analyst's $PMAFE^{TP}$ over a particular quarter:

$$\overline{PMAFE_{j,i}^{TP}} = \frac{1}{D} \sum_{t=1}^{D} PMAFE_{j,t,i}^{TP} \tag{1.6}$$

$$rank_{j,i} = \text{rank}_{j=1}^{N} \left\{ \overline{PMAFE_{j,i}^{TP}} \right\} \tag{1.7}$$

where D are the number of trading days in a quarter and N is the number of equity research firms with a valid TP. Figure 1.2 shows an example.

1.3.1.2 EPS ranking

To compute the EPS rankings, we apply the same procedure as above:

$$FE_{j,i}^{EPS} = |ACT_i - PRED_{j,i}| \tag{1.8}$$

$$PMAFE_{j,i}^{EPS} = \frac{FE_{j,i}^{EPS}}{\overline{FE_i^{EPS}}} \tag{1.9}$$

$$rank_{j,i} = \text{rank}_{j=1}^{N} \left\{ PMAFE_{j,i}^{EPS} \right\} \tag{1.10}$$

where ACT_i and $PRED_{j,i}$ are the actual quarterly EPS and analyst j's EPS forecast for stock i.

1.3.2 Defining the confidence of expected returns Ω

The confidence of Q is given by the coefficient of variation (CV) of forecasting errors. For each stock i is given by:

$$CV_i = \frac{\sigma_i(FE_i)}{\overline{FE}_i} \tag{1.11}$$

where σ_i and $\overline{FE}_i$ are the standard deviation and the mean of the forecast errors across analysts for either TP or EPS. A low value of CV reflects consensual estimates of either future prices or EPS.

1.3.3 Information sets to define the views

To proceed with the trading strategy, we need to establish which information we will be using to build the rankings. These rankings will be the inputs to compute the weighted return estimates ("smart estimates"). Different analysts' ranks are obtained if we select different time horizons. If we use only the most recent information, we will capture the recent performance of the analysts. This, of course, is more sensitive to unique episodes (e.g., a quarter which has been surprisingly good or bad). If, alternatively, we opt to incorporate the entire analyst performance, the ranking is less affected by such events, yet it may not reflect the current analyst ability. We use two information sets: the first uses only the information about the analyst' performance in period $t-1$; the second, uses all the available information for that particular analyst. We call the former the *recent* set and the latter the *all-time* set.

In addition to these sets, we also create a hypothetical scenario that assumes we anticipate perfectly the future analyst accuracy performance that would only be available at the end of t. This represents the perfect foresight strategy. The perfect foresight refers to analyst rankings not stock prices. Therefore, it serves a performance reference point to evaluate the other trading strategies. We call this the *true* set.

Formalizing information sets considered are:

- the *true* set
$$\widehat{Q_t} = Q_t \tag{1.12}$$

- the *recent* set
$$\widehat{Q_t} = Q_{t-1} \tag{1.13}$$

- the *all-time* set
$$\widehat{Q_t} = \frac{1}{T}\sum_{t=1}^{T} Q_t \tag{1.14}$$

where Q_t is the analysts' expected rank-weighted stock return (Equation (1.3))

1.4 Data and preliminary results

1.4.1 Database and sample

We focus on the S&P500 stocks. We extract the target price information and the EPS forecasts from ThomsonReuters I/B/E/S detailed history database. The S&P500 constituents' list and the stock daily prices are from ThomsonReuters DataStream.

Over the sample period, the total number of equity research firms[9] in TP dataset is 351, covering 498 S&P500 stocks. Given the fact that financial analysts commonly issue

[9] We use words "analyst" and "equity research firm" interchangeably.

TP with the one year horizon[10], we assume that analysts keep their TP forecasts valid for one calendar year unless it is revised. After one year we assume that TP recommendation expires.

Consistent with other studies on analysts' expected returns that work with price targets (Bradshaw, 2002; Brav and Lehavy, 2003; Da and Schaumburg, 2011), we truncate the sample of $TP/P-1$ at the 5^{th} percentile (values below -0.114) and at the 95^{th} percentile (values above 0.805). This is done due to occurrence of the extreme values. Most of these extreme values are driven by misalignment errors found on I/B/E/S data[11]. To implement ranking, we require that a stock had at least three equity research firms per quarter and that a equity research firm has to be active in covering a particular stock for at least 3 years (12 quarters). After all the data requirements, our final sample number of equity research firms issued target prices is 158 covering 448 S&P500 stocks. Overall, the number of observations (Stock $\times$ ERF $\times$ Quarter) is reduced from 131 068 (initial) to 100 974 (filtered).

In the case of EPS forecasts, the initial file of quarterly EPS forecast consists of 560 ERFs covering 3517 stocks. Considering the ranking data requirement, our final sample of EPS forecasts consist of 157 ERFs covering 402 S&P500 stocks. The total number of observation is 80 185.

Given that we have two different ranking datasets (based on TP and EPS), we, further, consider S&P500 stocks that are part of both datasets. We call this sample of stocks as the *same* and the full sample of S&P500 stocks as the *all*.

Table 1.1 shows the distribution of the final sample of target price and EPS data. Panel A shows the number of quarterly target prices per stock. For the sample period (1999-2009), we report that each stock in the *same* and the *all* stock sets had on average of 6.915 and 6.71, respectively, quarterly target price reports. Panel B presents similar statistics of the number of EPS forecasts. The average number of quarterly forecasts for the *same* and the *all* is 6.563 and 6.571, respectively.

We apply the ranking procedure outline in Section 1.3.1 to the two datasets. For target price rankings, we use the average daily errors within one quarter as the measure of analysts' forecasting ability (Equation (1.5)).

[10] According to Wharton Research Data Services (WRDS), 92.33% of all price targets reported in I/B/E/S have a 12-month horizon (Glushkov, 2009).

[11] We found some differences between the DataStream and I/B/E/S the databases. In some cases the stock-splits and the dividends were not properly adjusted.

Table 1.2 and Figure 1.2 illustrate an example illustrate how we estimate the *PMAFE*. Four analysts had valid target prices for Amazon for second quarter of 1999. We plot the daily Amazon price against the ERFs' target prices. Table 1.2 shows the resulting TP and EPS rankings. On the bases of the average daily errors, LEGG is the most accurate in forecasting stock price and DLJ is the least accurate. For the EPS case (panel B), PACCREST is the most accurate in EPS forecasting and RBRTSON is the least.

1.4.2 Ranking contingency results

We consider three terciles (*top*, *medium*, *bottom*). In one particular quarter (t), we place analysts at one of these bins which corresponds to a tercile. We, then, check analysts position at the immediate next quarter ($t+1$) and after one year ($t+4$).

Beforehand, we convert the rankings into scores as follows:

$$score_{j,i} = \frac{rank_{j,i}}{\max rank_i} \tag{1.15}$$

To get the cross-sectional values of scores across different stocks, we take the average of $score_{j,i}$

$$\overline{score_j} = \frac{1}{k}\sum_{s=1}^{k} score_{j,i} \tag{1.16}$$

where k is number of stocks followed by a particular analyst j.

Table 1.3 shows a contingency analysis of the ranks. Panel A shows the dynamics of each tercile for rankings based on target price accuracy. We observe that analysts exhibit strong ranking consistency as, on average, they stay at the same tercile after one quarter. For the *all* stocks, of the top (bottom) most accurate (inaccurate) analysts in the previous quarter 67.659% (69.358%) remain in that same tercile after one quarter. After one year the corresponding figures are lower respectively 46.717% and 40.877% for the top and bottom terciles[12].

In the case of EPS ranking (panel B) 49.087% and 28.387% (47.286% and 31.964%) of the analysts remained in the top and bottom terciles, respectively, after one quarter (year) in the *all* stock samples.

These results are consistent with the recent findings of Hilary and Hsu (2013) on analyst forecast consistency.

[12] For the case of the *same* stock sample, the results of ranking consistency analysis are similar.

1.4.3 Views: descriptive statistics

Table 1.4 presents the descriptive statistics of the analysts' expected returns conditional on the different information sets.

The expected returns are computed comparing TP estimates with actual prices. To form the smart strategies we compute rank-weighted estimates where weights are given either by the TP or the EPS ranks.

Bradshaw (2002) reports analyst average expected returns for the period of 2000–2009 and 206 ERFs of 24%. Da and Schaumburg (2011) report an average expected return of 40% for the period of 1996–2004. Zhou (2013) finds an average expected return of 96% for the sample period of 2000–2009. These figures suggest that analysts are overly optimistic.

Panel A of Table 1.4 show the statistics for the consensus expectations as defined in Equation (1.1). As mentioned above in Section 1.3.3, the consensus views have equal weights among the analysts, regardless of their ranks; thus, for the case of *true*, *recent*, and *all-time*, the median is the same regardless of knowing or not the present or past rankings (Q_{cons} in Equation (1.1)). As such, the mean, median, and standard deviation in the *true*, *recent*, and *all-time* information sets are the same. For the sample of *all* stocks the mean expected return is 18.61%. However, since views also include the confidence (Equation (1.11)), which is based on analysts past performance, the results of the trading strategy based on consensus expectations will be different for the *recent* and *all-time* information sets.

Panel B of the table shows the TP accuracy weighted average expected returns. For the sample of *all* stocks the weighted average returns for *true*, *recent*, and *all-time* information sets are respectively 14.876%, 15.742%, and 12.459%.

Panel C shows the EPS based weighted expected return. The average return for the *true*, *recent*, and *all-time* information sets are respectively 14.689%, 14.884%, and 12.843%.

The statistics for the subsample of stocks that integrate both the TP of the EPS samples are similar.

Overall compared to the consensus the ranked weighted expected returns ("smart estimates") are less optimistic. The *all-time* information set shows the lowest values of expected returns among all information sets. For the different stock sets, the *same* sample of stocks has higher values of expected returns compared to the *all* stocks.

Table Table 1.5 shows the number of active stocks for each of the trading strategies (*CONS*, *TP*, and *EPS*) conditional on considered information sets.

1.5 Empirical Results

We report the results from different trading strategies in Table Table 1.6. We split the table into four panels. Panel A shows the performance for the passive (market) strategy *Market*. Panels B to D compare the consensus , the TP rank weighted and the EPS rank weighted trading strategies for each of the information availability sets.

1.5.1 Passive strategy

The passive strategy generates an annualized cumulative return of -3.032% with a Sharpe Ratio of -0.182 over the period 1999-2001. The average number of stocks held per quarter was 499.975 and the turnover ratio was 0.053, which reflects solely the inclusion and deletions of the S&P 500 constituent list.

1.5.2 Perfect foresight strategy

Panel B presents the results for the case of the *true* information set. The annualized cumulative returns for each of the active strategies (*TP* and *EPS*) are, respectively, for the *all* stock sample: 4.325% and 0.574%. For the *same* sample of stocks these are respectively 4.549% and 0.719%. The two smart strategies outperform the passive benchmark (-3.032%). The consensus strategy annualized returns for the *all* and the *same* sample of stocks are respectively 0.116% and 0.434%

The results show, as expected, that investors would better off if they knew in advance who the top analysts in terms of TP or EPS accuracy would be. In any case, the results suggest that the advice of analysts, as a group, are valuable.

The *TP* ranking strategy dominates also when we look at the risk-adjusted returns. The Sharpe Ratio for the *all* (*same*) stock sample is 0.294 against 0.037, 0.007, and -0.182, respectively for the *EPS*, *CONS* and the *Market* strategies. In addition, the *TP* strategy dominates the others if we consider the shorter trading periods (5 years). Table 1.7 shows the Sharpe Ratio for six 5-year holding periods. The *TP* strategy wins over the others in every period. The results for the subsample of stocks that integrate both the TP of the EPS datasets are similar.

While this is an hypothetical setting, given that it is not possible to know in advance which analyst will rank first, it suggests that if we can predict the rankings with some accuracy this will be a useful investment trading tool. One of the possibilities is using methods developed in the Machine Learning literature (e.g., Aiguzhinov, Soares, and Serra (2010); Brazdil et al. (2003)), where this type of problem (referred as a label ranking problem) has been broadly studied. For example, Aiguzhinov et al. (2010) propose a label ranking algorithm using Bayesian approach to predict the rankings.

1.5.3 Feasible strategies

Panel C of Table 1.6 shows the performance of the different smart strategies and the consensus strategy in the *recent* information set. We report the results of forming portfolios with the available stocks within each dataset (*all*) and the subsample of stocks that include both the TP or the EPS datasets (*same*).

The *TP* and *CONS* active strategies outperform the *Market* (-3.032%) and show positive cumulative annualized returns for the *all* (*same*) sample of stocks of 0.282% (0.621%) and 0.116% (0.434%) respectively. The active strategy based on EPS forecasts, has negative annualized cumulative returns of -0.303% and -0.349% for the *all* and the *same* samples respectively.

The risk-adjusted results for this information set shows that, as in the case with the *true*, the dominant trading strategy is the *TP* strategy regardless of forming portfolios with *all* stocks or with the *same* subsample. The results in panel B of Table 1.7 show as well that this strategy outperforms the others for all of the shorter trading periods.

Panel D of Table 1.6 shows interesting and slightly different findings. On one hand, when all the analyst forecast performance track record is included to set the rankings, we observe an increase in annualized cumulative returns and risk-adjusted returns for the smart strategies. Particularly in the case of the EPS strategy, the results suggest that strategies that weight the estimates with accuracy rankings obtained using more information show better performance: when we consider the *all* sample of stocks, the strategy based on the accuracy of EPS forecasts outperforms the other strategies (annualized cumulative returns of 0.746%, 0.689%, 0.314%, respectively for the *EPS*, *TP*, and *CONS*). The TP strategy dominates when we consider the subsample of stocks that are included in both the TP and EPS datasets (0.717%, 1.056%, 0.686% respectively for the *EPS*, *TP* and *CONS* strategies) but the returns and risk-adjusted improve as well when compared to the figures in Panel C.

The analysis of the sub-periods performance of the *TP* and *EPS* -based strategies depicted in Table 1.7 shows that the latter outperforms the former in terms of the annualized cumulative return only for the first two periods: in 2000Q1/2004Q4 3.501% vs. 3.133% in 2001Q1/2005Q4 3.954% vs. 3.727% respectively for the *EPS* and the *TP* strategies. For the all the other sup-periods, the annualized cumulative returns of the *EPS* strategy are lower than those of the *TP* strategy. Table 1.8 shows the sub-period results when we consider only the sample of stocks that integrates the TP and EPS dataset. In this setting (*same*), the *TP* strategy dominates the *EPS* and *CONS* strategies in every sub-period.

To further investigate the prevailance of the strategies based on smart estimates as opposed to consensus (and *Market*), we perform a pairwise hypothesis test with *null*-hypothesis stating that the difference between the annualized cumulative returns based on smart estimate strategies and those of the consensus strategy (and *Market*) is equal to zero. Table 1.9 presents t-statistic and the corresponding p-values of this test. Panel A shows the results for the case of "All vs. *Market*". We report that all active strategies resulted in the statisitcally significant (at 1% level) prevailance over the *Market* for different information sets (*true*, *recent*, and *all-time*) as well as different stock sets (*all* and *same*).

Panel B of the table presents the t-statistic for the case of "All vs. *CONS*". Given the results, we reject the *null*-hypothesis in all of the experiment instances (informations sets and stock sets). In terms of the positive gains, the *TP* strategy demonstrated a statistically significant positive performance over the consensus in all information sets and for both stock sets. On the other hand, the *EPS* strategy resulted in the statistically significant negative performance over the *CONS* strategy except for the case of *all-time* information and *all* stocks. The results of the test confirms that the strategy based on the rankings of the analysts who issue more accurate target prices outperforms, in terms of the annualized cumulative returns, the strategy based on the consensus among analysts regarding stock target prices.

Figure 1.3 shows the graphical representation of the cumulative portfolio wealth for the passive and smart strategies in all the information sets. The y-axis is the dollar value of wealth and the x-axis is the time starting at January 2000 and ending at December 2009. The active investment management strategies in the *true* panel outperform the *Market* and the final value of the portfolio of the *TP* strategy is well above those of the other alternative strategies.

In sum the results of the feasible information sets outlined above suggest that it is worthwhile to follow the analysts, particularly the top ranked analysts, and are supportive of Desai et al. (2000) in that smart strategies based upon analyst accuracy rankings are beneficial for investors.

Further the results show that the values of the annualized cumulative returns are higher in the *all-time* information set when compared with those yield by the strategies that use only the most *recent* ranking information set. This seems to suggest that investors should estimate analysts forecasting skills over a long period of time rather than focusing on the most recent analyst accuracy performance.

Finally, the results suggest that, from an investor's point of view, following analysts who are accurate in setting price targets is more valuable than following those that are good at forecasting EPS. This result contradicts the findings of Bradshaw (2004) and supports the argument of Simon and Curtis (2011) that stock recommendations of the most accurate analysts are not based upon simple valuation models.

1.6 Conclusions

Some institutions, such as StarMine (ThomsonReuters), rank financial analysts based on EPS and target price accuracy. These rankings are published and are relevant: stocks favored by top ranked analysts will probably receive more attention from investors. Therefore, there is a growing interest in understanding the relative performance of strategies based upon analysts with different forecast accuracy.

We use the Black-Litterman model. The views are TP or EPS rank-weighted means of analysts forecasted returns. We developed simulations of trading strategies using different information sets to compute the ranks. If we consider that only the information known prior to time t is used to obtain the ranks, investors would be better off following the strategy that weights more heavily the estimates issued by the most accurate TP forecast analysts and considering the whole performance tracking record of the of the analysts.

For future work we will developed new methods to forecast analysts rankings so as to get closer to the upper bound of perfect foresight of rankings.

Table 1.1: Sample Statistics

This table shows the average number of target prices (panel A) and EPS forecasts (panel B) per stock per quarter. Stocks in the *all* sample are subsamples of the S&P 500, stocks in the *same* sample integrate both the EPS and TP datasets.

	Min		Mean		Median		Max		Std.dev	
	all	*same*	*all*	*same*	*all*	*same*	*all*	*same*	*all*	*same*
Panel A: TP										
1999	3	3	4.286	4.326	4	4	11	11	1.657	1.668
2000	3	3	4.873	4.944	4	4	14	14	2.009	2.017
2001	3	3	5.537	5.691	5	5	16	16	2.453	2.481
2002	3	3	6.411	6.611	6	6	19	19	3.108	3.145
2003	3	3	7.021	7.252	6	6	21	21	3.524	3.573
2004	3	3	7.477	7.728	7	7	22	22	3.671	3.726
2005	3	3	7.667	7.946	7	7	24	24	3.736	3.744
2006	3	3	7.754	8.037	7	7	22	22	3.561	3.532
2007	3	3	7.394	7.691	7	7	22	22	3.494	3.429
2008	3	3	6.708	6.973	6	6	18	18	3.023	2.964
2009	3	3	5.510	5.627	5	5	19	19	2.412	2.433
Total	3	3	6.710	6.915	6	6	24	24	3.336	3.356
Panel B: EPS										
1999	3	3	5.673	5.643	5	5	17	17	2.991	2.936
2000	3	3	5.433	5.426	5	5	17	17	2.774	2.733
2001	3	3	6.020	6.048	5	5	18	18	2.972	2.987
2002	3	3	6.338	6.357	5	5	21	21	3.264	3.274
2003	3	3	6.523	6.532	6	6	24	24	3.430	3.443
2004	3	3	6.947	6.978	6	6	22	22	3.749	3.771
2005	3	3	7.205	7.216	6	6	24	24	3.774	3.777
2006	3	3	7.494	7.467	7	7	26	26	3.843	3.820
2007	3	3	7.134	7.091	6	6	21	21	3.532	3.489
2008	3	3	6.378	6.353	6	6	22	22	3.062	3.039
2009	3	3	5.811	5.773	5	5	20	20	2.783	2.753
Total	3	3	6.571	6.563	6	6	26	26	3.423	3.412

Table 1.2: Example of ranking

This table shows target prices (panel A) and EPS forecasts (panel B) rankings for Amazon (AMZN) for the second quarter of 1999. We apply (Equation (1.7)) to obtain the ranks of the ERFs. TP are target prices; $\overline{PMAFE}^{TP}$ is the daily average proportional mean adjusted TP error. For the EPS case, $PRED$ are the EPS forecasts issued by the analysts; $PMAFE^{EPS}$ is the proportional mean-adjusted forecast error of quarterly EPS forecasts.

Panel A: TP

ERF/Analyst	TP	$\overline{PMAFE}^{TP}$	$rank^{TP}$
LEGG	58.50	0.05	1.00
MONTSEC	87.50	0.46	2.00
KAUFBRO	125.00	1.65	3.00
DLJ	140.00	1.97	4.00

Panel B: EPS

ERF/Analyst	$PRED$	$PMAFE^{EPS}$	$rank^{EPS}$
MONTSEC	-0.120	0.023	1.000
BEAR	-0.130	0.068	2.500
PACCREST	-0.130	0.068	2.500
BACHE	-0.135	0.091	4.000
RBRTSON	-0.140	0.114	5.000
FBOSTON	-0.145	0.137	6.000

Table 1.3: Analysts' accuracy consistency

This contingency table shows changes in analysts' *top*, *middle*, *bottom* ranking bins. Panel A (Panel B) depicts the dynamics of the analysts' ranks based on the accuracy in target prices (EPS forecasts). Stocks in the *all* sample are subsamples of the S&P 500, stocks in the *same* sample integrate both the EPS and TP datasets.

		top		*middle*		*bottom*		*Sum*	
		all	*same*	*all*	*same*	*all*	*same*	*all*	*same*
	Panel A: TP				$t+1$				
	top	67.7	67.8	22.1	22.0	10.2	10.2	100.0	100.0
	middle	30.6	29.9	47.6	48.2	21.7	21.9	100.0	100.0
	bottom	13.7	13.5	16.9	17.0	69.4	69.5	100.0	100.0
t					$t+4$				
	top	46.7	46.5	27.7	28.0	25.5	25.5	100.0	100.0
	middle	39.1	39.2	29.0	29.0	31.9	31.7	100.0	100.0
	bottom	32.8	32.2	26.4	26.4	40.9	41.4	100.0	100.0
	Panel B: EPS				$t+1$				
	top	49.1	49.1	25.6	25.6	25.8	25.8	100.5	100.5
	middle	48.5	48.6	26.0	25.8	25.9	25.8	100.4	100.2
	bottom	47.1	46.9	25.4	25.5	28.4	28.3	100.8	100.7
t					$t+4$				
	top	47.3	47.5	28.0	27.9	25.9	25.8	101.2	101.2
	middle	46.4	46.5	26.6	26.5	27.7	27.7	100.6	100.6
	bottom	43.2	43.2	27.5	27.5	32.0	32.1	102.7	102.7

Table 1.4: Descriptive statistics of views

This table shows the descriptive statistics of views (expected returns) based on the consensus (median) among the analysts (panel A); target price rankings (panel B); and EPS forecasts rankings (panel C). Stocks in the *all* sample are subsamples of the S&P 500, stocks in the *same* sample integrate both the EPS and TP datasets.

	Mean (in %)		Median (in %)		Std.dev	
	all	*same*	*all*	*same*	*all*	*same*
Panel A: Consensus						
true	18.610	18.985	16.889	17.185	0.120	0.119
recent	18.610	18.985	16.889	17.185	0.120	0.119
all-time	18.610	18.985	16.889	17.185	0.120	0.119
Panel B: TP						
true	14.876	15.191	13.380	13.604	0.096	0.096
recent	15.742	16.102	14.314	14.541	0.098	0.098
all-time	12.459	12.714	10.591	10.798	0.089	0.089
Panel C: EPS						
true	14.689	14.769	13.214	13.252	0.101	0.102
recent	14.884	14.959	13.392	13.420	0.103	0.103
all-time	12.843	12.916	11.410	11.439	0.089	0.089

Table 1.5: Number of active stocks

This table shows number of active stocks in each of the trading strategies conditional on different information sets: *true* (panel A), *recent* (panel B) and *all-time* (panel C). Stocks in the *all* sample are subsamples of the S&P 500, stocks in the *same* sample integrate both the EPS and TP datasets.

Year	*CONS*		*TP*		*EPS*	
	all	*same*	*all*	*same*	*all*	*same*
Panel A: *true*						
1999	71	67	71	67	42	42
2000	257	235	253	232	191	191
2001	310	281	302	273	244	244
2002	360	321	353	316	289	289
2003	374	334	369	330	299	299
2004	385	341	375	331	318	318
2005	401	352	395	347	337	337
2006	415	358	403	351	337	337
2007	421	361	418	359	351	351
2008	422	360	413	353	339	339
2009	405	351	346	306	306	306
Total	442	378	442	378	375	375
Panel B: *recent*						
1999	71	67	71	67	41	41
2000	257	235	257	235	185	186
2001	310	281	310	281	247	247
2002	360	321	360	321	287	287
2003	374	334	374	334	307	307
2004	385	341	385	341	318	318
2005	401	352	401	352	337	337
2006	415	358	415	358	343	343
2007	421	361	421	361	352	352
2008	422	360	422	360	349	349
2009	405	351	405	351	321	321
Total	442	378	442	378	376	376
Panel C: *all-time*						
1999	71	67	71	67	51	51
2000	257	235	257	235	204	204
2001	310	281	310	281	261	261
2002	360	321	360	321	302	302
2003	374	334	374	334	317	317
2004	385	341	385	341	332	332
2005	401	352	401	352	344	344
2006	415	358	415	358	353	353
2007	421	361	421	361	358	358
2008	422	360	422	360	359	359
2009	405	351	405	351	350	350
Total	442	378	442	378	377	377

Table 1.6: Trading strategies performance: entire period

This table shows the performance statistics of the different trading strategies. Panel A presents the results for the passive strategy. Panels B, C, and D show the results for the perfect foresight scenario (*true*), and, respectively, the scenarios for which we use the most recent (*recent*) and all ranking history of analysts (*all-time*) to weight the TP/EPS estimates. Within each panel, we show the strategy results of three views regarding expected return: *CONS* uses the median of the analysts estimates; *TP* is based upon TP accuracy ranking; and *EPS* is based upon EPS accuracy ranking. Stocks in the *all* sample are subsamples of the S&P 500, stocks in the *same* sample integrate both the EPS and TP datasets. The trading period goes from 2000Q1 until 2009Q4.

Strategy	Annualized return (in %)		Annualized Std. dev (in %)		Sharpe ratio		Average num. stock		Average turnover rate	
Panel A										
Market	-3.032		16.654		-0.182		499		0.053	
Panel B: ***true***										
	all	*same*	*all*	*same*	*all*	*same*	*all*	*same*	*all*	*same*
CONS	0.116	0.434	15.948	15.995	0.007	0.027	283	251	0.256	0.251
TP	4.325	4.549	14.697	14.794	0.294	0.307	283	251	0.345	0.327
EPS	0.574	0.719	15.528	15.505	0.037	0.046	205	205	0.496	0.494
Panel C: ***recent***										
CONS	0.116	0.434	15.948	15.995	0.007	0.027	283	251	0.256	0.251
TP	0.282	0.621	15.662	15.682	0.018	0.040	284	251	0.264	0.256
EPS	-0.303	-0.349	16.088	16.096	-0.019	-0.022	206	206	0.410	0.408
Panel D: ***all-time***										
CONS	0.314	0.686	15.773	15.825	0.020	0.043	283	251	0.228	0.223
TP	0.689	1.056	15.565	15.485	0.044	0.068	284	251	0.256	0.248
EPS	0.746	0.717	15.444	15.481	0.048	0.046	245	245	0.256	0.256

Table 1.7: Trading strategies performance: sub-periods, *all* sample

This table presents the annualized return (in %) and the Sharpe ratio of each of the trading strategies: the passive (*Market*) and the active (consensus and smart estimates) calculated for different holding periods. Panel A represents the perfect foresight information set; panels B and C show, respectively, the results of the strategies using the most recent and all history analysts' performance.

Period	*Market*		*CONS*		*TP*		*EPS*	
	Ann.ret	SR	Ann.ret	SR	Ann.ret	SR	Ann.ret	SR
Panel A: *true*								
2000Q1/2004Q4	-3.401	-0.201	2.663	0.167	5.844	0.395	1.890	0.124
2001Q1/2005Q4	-1.539	-0.093	2.669	0.170	6.119	0.425	3.945	0.270
2002Q1/2006Q4	2.567	0.196	5.734	0.447	9.124	0.757	5.877	0.492
2003Q1/2007Q4	7.919	0.925	9.548	1.106	14.715	1.915	10.681	1.364
2004Q1/2008Q4	-5.667	-0.435	-4.708	-0.343	0.908	0.070	-3.133	-0.232
2005Q1/2009Q4	-2.662	-0.158	-2.367	-0.146	2.827	0.189	-0.725	-0.045
All period	-3.032	-0.182	0.116	0.007	4.325	0.294	0.574	0.037
Panel B: *recent*								
2000Q1/2004Q4	-3.401	-0.201	2.663	0.167	2.620	0.168	1.172	0.074
2001Q1/2005Q4	-1.539	-0.093	2.669	0.170	3.266	0.214	2.659	0.171
2002Q1/2006Q4	2.567	0.196	5.734	0.447	6.075	0.490	5.485	0.455
2003Q1/2007Q4	7.919	0.925	9.548	1.106	10.068	1.248	9.337	1.155
2004Q1/2008Q4	-5.667	-0.435	-4.708	-0.343	-4.148	-0.305	-4.541	-0.332
2005Q1/2009Q4	-2.662	-0.158	-2.367	-0.146	-2.003	-0.125	-1.757	-0.106
All period	-3.032	-0.182	0.116	0.007	0.282	0.018	-0.303	-0.019
Panel C: *all-time*								
2000Q1/2004Q4	-3.401	-0.201	3.102	0.197	3.133	0.205	3.501	0.232
2001Q1/2005Q4	-1.539	-0.093	3.016	0.194	3.727	0.249	3.954	0.264
2002Q1/2006Q4	2.567	0.196	5.410	0.417	5.793	0.464	5.712	0.468
2003Q1/2007Q4	7.919	0.925	9.419	1.071	10.335	1.245	10.002	1.202
2004Q1/2008Q4	-5.667	-0.435	-4.860	-0.358	-3.941	-0.289	-4.223	-0.309
2005Q1/2009Q4	-2.662	-0.158	-2.398	-0.149	-1.697	-0.105	-1.936	-0.121
All period	-3.032	-0.182	0.314	0.020	0.689	0.044	0.746	0.048

Table 1.8: Trading strategies performance: sub-periods, *same* subsample

This table presents the annualized return (in %) and the Sharpe ratio of each of the trading strategies for the *same* set of stocks: the passive (*Market*) and the active (consensus and smart estimates) calculated for different holding periods. Panel A represents the perfect foresight information set; panels B and C show, respectively, the results of the strategies using the most recent and all history analysts' performance.

Period	*Market*		*CONS*		*TP*		*EPS*	
	Ann.ret	SR	Ann.ret	SR	Ann.ret	SR	Ann.ret	SR
Panel A: ***true***								
2000Q1/2004Q4	-3.401	-0.201	3.378	0.213	6.609	0.451	2.116	0.139
2001Q1/2005Q4	-1.539	-0.093	2.825	0.180	6.226	0.432	4.161	0.286
2002Q1/2006Q4	2.567	0.196	5.836	0.468	8.985	0.768	5.827	0.486
2003Q1/2007Q4	7.919	0.925	9.417	1.100	14.249	1.891	10.672	1.361
2004Q1/2008Q4	-5.667	-0.435	-4.826	-0.349	0.398	0.030	-3.126	-0.232
2005Q1/2009Q4	-2.662	-0.158	-2.426	-0.148	2.528	0.166	-0.658	-0.041
All period	-3.032	-0.182	0.434	0.027	4.549	0.307	0.719	0.046
Panel B: ***recent***								
2000Q1/2004Q4	-3.401	-0.201	3.378	0.213	3.331	0.216	1.073	0.067
2001Q1/2005Q4	-1.539	-0.093	2.825	0.180	3.523	0.231	2.598	0.167
2002Q1/2006Q4	2.567	0.196	5.836	0.468	6.218	0.516	5.436	0.450
2003Q1/2007Q4	7.919	0.925	9.417	1.100	9.972	1.244	9.325	1.153
2004Q1/2008Q4	-5.667	-0.435	-4.826	-0.349	-4.252	-0.311	-4.552	-0.332
2005Q1/2009Q4	-2.662	-0.158	-2.426	-0.148	-2.019	-0.124	-1.752	-0.105
All period	-3.032	-0.182	0.434	0.027	0.621	0.040	-0.349	-0.022
Panel C: ***all-time***								
2000Q1/2004Q4	-3.401	-0.201	3.909	0.250	3.852	0.257	3.470	0.230
2001Q1/2005Q4	-1.539	-0.093	3.254	0.210	4.014	0.271	3.919	0.261
2002Q1/2006Q4	2.567	0.196	5.461	0.433	5.915	0.488	5.686	0.465
2003Q1/2007Q4	7.919	0.925	9.297	1.067	10.199	1.253	9.988	1.199
2004Q1/2008Q4	-5.667	-0.435	-4.979	-0.364	-4.091	-0.299	-4.241	-0.310
2005Q1/2009Q4	-2.662	-0.158	-2.437	-0.150	-1.665	-0.102	-1.963	-0.122
All period	-3.032	-0.182	0.686	0.043	1.056	0.068	0.717	0.046

Table 1.9: Significance of cumulative returns

The table demonstrates a pairwise statisitical test in difference of the cumulative returns of all strategies vs. *Market* (Panel A) and vs. *CONS* strategy (Panel B). Case of *true* shows the known future information; *recent* is the case of ranking information know at $t-1$, and the *all-time* is the case of using all ranking information for up to $t-1$. *TP* is the strategy with rankings based on the accuracy in target prices, *CONS* is the strategy based on the consensus among the analysts regarding a stock's expected return. *EPS* is the strategy with rankings based on the accuracy of EPS forecasts. Stocks in the *all* sample are subsamples of the S&P 500, stocks in the *same* sample integrate both the EPS and TP datasets.

	all		*same*	
	t value	$\Pr(>\lvert t\rvert)$	t value	$\Pr(>\lvert t\rvert)$
Panel A: *Market*				
true				
CONS	15.247	0.000	16.857	0.000
TP	12.661	0.000	13.799	0.000
EPS	11.664	0.000	12.001	0.000
recent				
CONS	15.247	0.000	16.857	0.000
TP	14.015	0.000	15.481	0.000
EPS	12.655	0.000	12.640	0.000
all-time				
CONS	15.858	0.000	17.456	0.000
TP	14.314	0.000	15.833	0.000
EPS	15.625	0.000	15.690	0.000
Panel B: *CONS*				
true				
TP	9.284	0.000	9.625	0.000
EPS	-2.352	0.024	-6.358	0.000
Market	-15.247	0.000	-16.857	0.000
recent				
TP	2.032	0.049	1.706	0.096
EPS	-20.130	0.000	-27.074	0.000
Market	-15.247	0.000	-16.857	0.000
all-time				
TP	3.182	0.003	2.892	0.006

... continued next page

	all		*same*	
	t value	$\Pr(>\lvert t\rvert)$	t value	$\Pr(>\lvert t\rvert)$
EPS	10.492	0.000	-3.008	0.005
Market	-15.858	0.000	-17.456	0.000
		Panel C: *TP*		
true				
CONS	-9.284	0.000	-9.625	0.000
EPS	-12.876	0.000	-15.090	0.000
Market	-12.661	0.000	-13.799	0.000
recent				
CONS	-2.032	0.049	-1.706	0.096
EPS	-17.847	0.000	-23.272	0.000
Market	-14.015	0.000	-15.481	0.000
all-time				
CONS	-3.182	0.003	-2.892	0.006
EPS	18.770	0.000	-14.975	0.000
Market	-14.314	0.000	-15.833	0.000

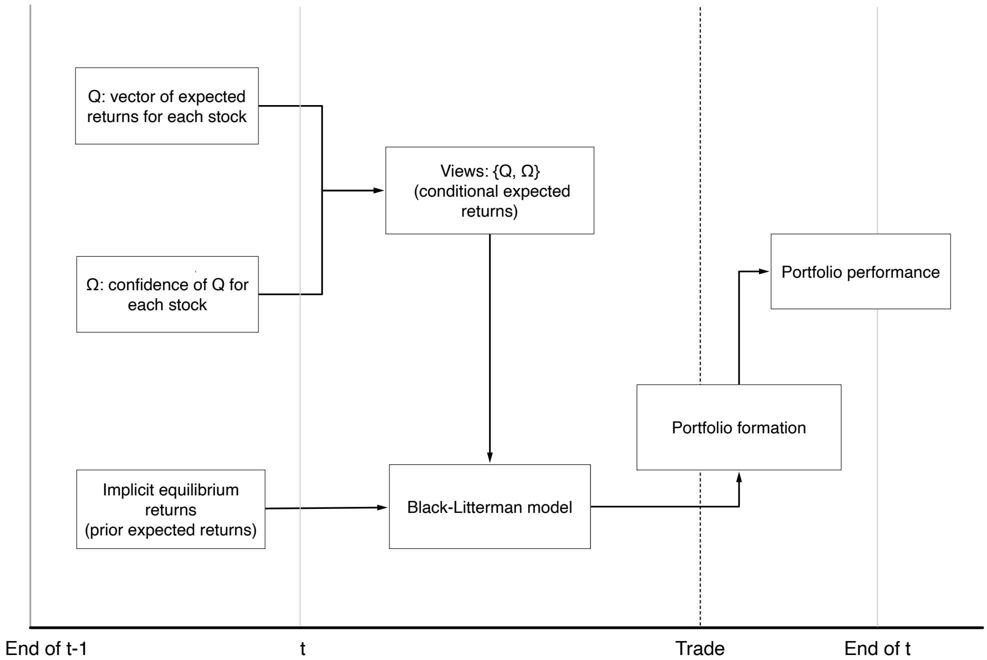

Figure 1.1: Trading strategy timeline
Black-Litterman model inputs are at the beginning of t we apply the BL model and form the active portfolio. At the end of t, we evaluate performance.

Figure 1.2: Amazon daily stock price and ERFs target prices
Target price and actual prices for Amazon the second quarter of 1999.

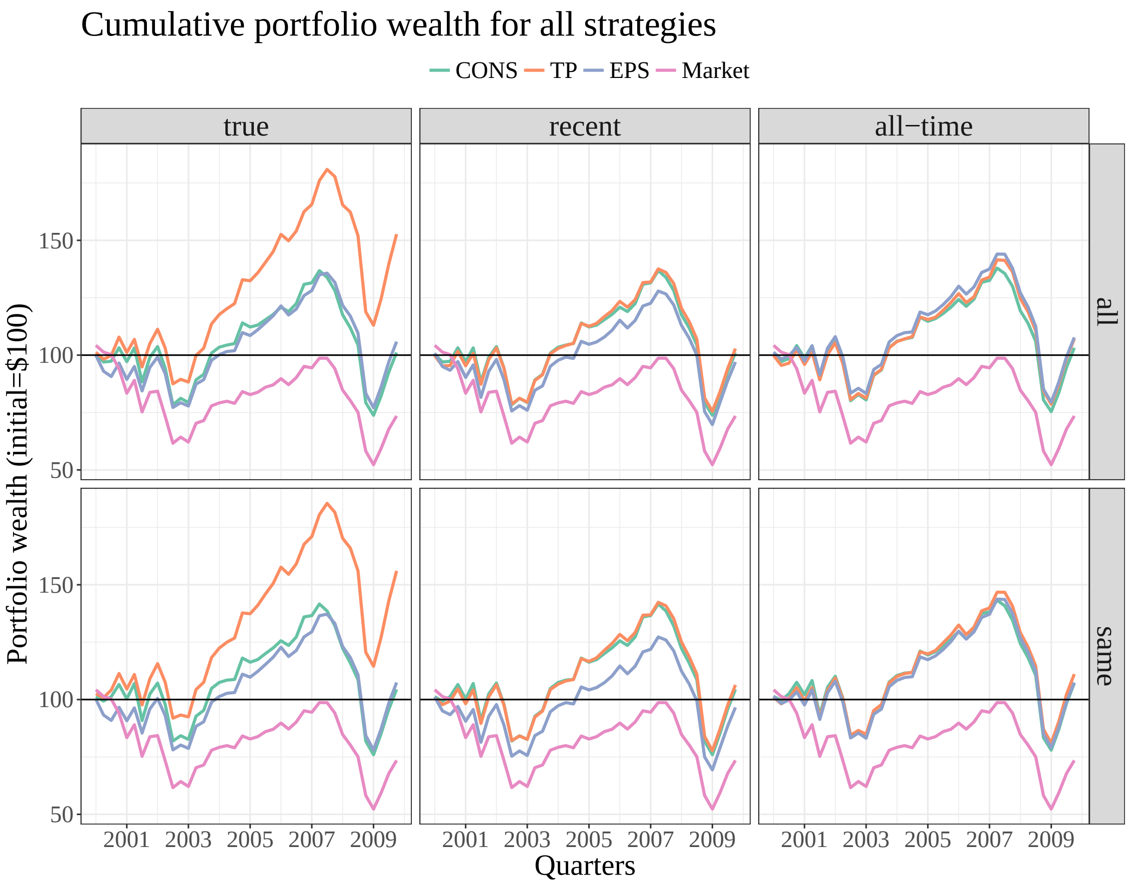

Figure 1.3: Performance of the BL model

Quarterly performance of the cumulative portfolio wealth for all strategies. Panel *true* shows the case of the known future information; *recent* is the case of ranking information know at $t-1$, and the *all-time* is the case of using all ranking information for up to $t-1$. *TP* is the strategy with rankings based on the accuracy in target prices, *CONS* is the strategy based on the consensus among the analysts regarding a stock's expected return. *EPS* is the strategy with rankings based on the accuracy of EPS forecasts. Stocks in the *all* sample are subsamples of the S&P 500, stocks in the *same* sample integrate both the EPS and TP datasets. The trading period ranges from 2000Q1 until 2009Q4.

Chapter 2

A similarity-based adaptation of naive Bayes for label ranking

Abstract

The problem of learning label rankings is receiving increasing attention from several research communities. A number of common learning algorithms have been adapted for this task, including k-Nearest Neighbours (k-NN) and decision trees. Following this line, we propose an adaptation of the naive Bayes classification algorithm for the label ranking problem. Our main idea lies in the use of similarity between the rankings to replace the concept of probability. We empirically test the proposed method on some metalearning problems that consist of relating characteristics of learning problems to the relative performance of learning algorithms. Our method generally performs better than the baseline indicating that it is able to identify some of the underlying patterns in the data.

keywords: label ranking; naive Bayes

2.1 Introduction

Label ranking is an increasingly popular topic in the machine learning literature. It studies the problem of learning a mapping from instances to rankings over a finite number of predefined labels. In some sense, it is a variation of the conventional classification problem; however, in contrast to the classification settings, where the objective is to assign examples to a specific class, in label ranking we are interested in assigning a complete preference order of labels to every example (Cheng et al., 2009).

Many different algorithms have been adapted to deal with label ranking such as: decision-trees for label ranking (Cheng et al., 2009), algorithm based on Plackett-Luce model (Cheng et al., 2010), pairwise comparison (Hüllermeier et al., 2008), and k-NN for label ranking (Brazdil et al., 2003). Table 2.5 outlines the recent developments in solving a label ranking problem.

In this paper, we introduce the ranking similarity approach. We propose an adaptation of the naive Bayes (NB) algorithm for label ranking. Despite its limitations, NB is an algorithm with successful results in many applications (Domingos and Pazzani, 1997). Additionally, the Bayesian framework is well understood in many domains. For instance, we apply this method on the problem of predicting the rankings of financial analysts since in the Financial Economics the Bayesian models are widely used (e.g., the Black-Litterman model for active portfolio management (Black and Litterman, 1992)).

The main idea lies in replacing the probabilities in the Bayes theorem with the distance between rankings. This can be done because it has been shown that there is a parallel between the concepts of distance and likelihood (Vogt, Godden, and Bajorath, 2007). We develop two versions of the algorithm: for discrete and continuous cases.

The paper is organized as follows: Section 2.2 provides the formalization of the label ranking problem; Section 2.3 briefly describes the naive Bayes algorithm for classification; Section 2.4 shows the adaptation of the NB algorithm for label ranking (NB4LR); Section 2.5 provides some extensions of NB4LR; namely, the scenario of features having a continuous values (Section 2.5.1) and a case when rankings are part of time series (Section 2.5.2); Section 2.6 outlines the datasets used for the experiments; Section 2.7 presents empirical results; finally, Section 2.8 concludes with the goals for future work.

2.2 Learning label rankings

The formalization of a label ranking problem is the following (Vembu and Gärtner, 2010). Let $\mathscr{X} \subseteq \{\mathscr{V}_1, \ldots, \mathscr{V}_m\}$ be an instance space of nominal variables, such that $\mathscr{V}_a = \{v_{a,1}, \ldots, v_{a,n_a}\}$ is the domain of nominal variable a. Also, let $\mathscr{L} = \{\lambda_1, \ldots, \lambda_k\}$ be a set of labels, and

$\mathscr{Y} = y_{\mathscr{L}}$ be the output space of all possible total orders[1] over $\mathscr{L}$ defined on the permutation space y. The goal of a label ranking algorithm is to learn a mapping $h : \mathscr{X} \to \mathscr{Y}$, where h is chosen from a given hypothesis space $\mathscr{H}$, such that a predefined loss function $\ell : \mathscr{H} \times \mathscr{Y} \times \mathscr{Y} \to \mathbb{R}$ is minimized. The algorithm learns h from a training set $\mathscr{T} = \{x_i, y_i\}_{i \in \{1,\ldots,n\}} \subseteq \mathscr{X} \times \mathscr{Y}$ of n examples, where $x_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,m}\} \in \mathscr{X}$ and $y_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,k}\} \in y_{\mathscr{L}}$. Furthermore, we define $y_i^{-1} = \{y_{i,1}^{-1}, y_{i,2}^{-1}, \ldots, y_{i,k}^{-1}\}$ as the order of the labels in example i. Given that we are focusing on total orders, y_i^{-1} is a permutation of the set $\{1, 2, \ldots, k\}$ where $y_{i,j}^{-1}$ is the rank of label λ_j in example i.

Unlike classification, where for each instance $x \in \mathscr{X}$ there is an associated class $y_i \in \mathscr{L}$[2], in label ranking problems there is a ranking of the labels associated with every instance x and the goal is to predict it. This is also different from other ranking problems, such as in information retrieval or recommender systems. In these problems the target variable is a set of ratings or binary relevance labels for each item, and not a ranking.

The algorithms for label ranking can be divided into two main approaches: methods that transform the ranking problem into multiple binary problems and methods that were developed or adapted to predict the rankings. An example of the former is the ranking by pairwise comparisons (Hüllermeier et al., 2008). Some examples of algorithms that are specific for rankings are: the predictive clustering trees method (Todorovski, Blockeel, and Dzeroski, 2002), the similarity-based k-Nearest Neighbor for label ranking (Brazdil et al., 2003), the probabilistic k-Nearest Neighbor for label ranking (Cheng et al., 2009) and the linear utility transformation method (Har-Peled, Roth, and Zimak, 2002; Dekel, Manning, and Singer, 2004).

To assess the accuracy of the predicted rankings relative to the corresponding target rankings, a suitable loss function is needed. In this paper we compare two rankings using the Spearman correlation coefficient (Brazdil et al., 2003; Vembu and Gärtner, 2010):

$$\rho(y, \hat{y}) = 1 - \frac{6 \sum_{j=1}^{k} (y_j - \hat{y}_j)^2}{k^3 - k} \tag{2.1}$$

[1]A total order is a complete, transitive, and asymmetric relation $\succ$ on $\mathscr{L}$, where $\lambda_i \succ \lambda_j$ indicates that λ_i precedes λ_j. In this paper, given $\mathscr{L} = \{A, B, C\}$, we will use the notation $\{A, C, B\}$ and $\{1, 3, 2\}$ interchangeably to represent the order $A \succ C \succ B$.

[2]Here, we use both y_i to represent the target class (label) in classification and the target ranking in label ranking to clarify that they are both the target of the learning problem. We will explicitly state the task we are dealing with when it is not clear from the context.

where y and $\hat{y}$[3] are, respectively, the target and predicted rankings for a given instance. Two orders with all the labels placed in the same position will have a Spearman correlation of $+1$. Labels placed in reverse order will produce correlation of -1. Thus, the higher the value of ρ the more accurate the prediction is compared to target. The loss function is given by the mean Spearman correlation values (Equation (2.1)) between the predicted and target rankings, across all examples in the dataset:

$$\ell = \frac{\sum_{i=1}^{n} \rho(y_i, \hat{y}_i)}{n} \tag{2.2}$$

An extensive survey of label ranking algorithms is given by Vembu and Gärtner (2010).

2.3 The Naive Bayes Classifier

We follow Mitchell (1997) to formalize the naive Bayes classifier. In classification, each instance $x_i \in \mathscr{X}$ is binded to class $y_i \in \mathscr{L}$. The task of a learner is to create a classifier from the training set $\mathscr{T}$.The classifier takes a new, unlabeled instance and assigns it to a class (label).

The naive Bayes method classifies a new instance x_i by determining the most probable target value, $c_{MAP}(x_i)$[4], given the attribute values that describe the instance:

$$c_{MAP(x_i)} = \underset{\lambda \in \mathscr{L}}{\arg\max}\, P(\lambda | x_{i,1}, x_{i,2}, \ldots, x_{i,m}) \tag{2.3}$$

where $x_{i,j}$ is the value of attribute j for instance i.

The algorithm is based on the Bayes theorem that establishes the probability of A given B as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.4}$$

Thus, the Bayes theorem provides a way to calculate the posterior probability of a hypothesis.

[3] In the following, we will use y_i and y_i interchangeably to represent the target ranking.
[4] MAP – Maximum A Posteriori

Using Equation (2.4), we can rewrite Equation (2.3) as

$$
\begin{aligned}
c_{MAP(x_i)} &= \underset{\lambda \in \mathscr{L}}{\arg\max} \frac{P(x_{i,1}, x_{i,2}, \ldots, x_{i,m}|\lambda)P(\lambda)}{P(x_{i,1}, x_{i,2}, \ldots, x_{i,m})} \\
&= \underset{\lambda \in \mathscr{L}}{\arg\max}\, P(x_{i,1}, x_{i,2} \ldots x_{i,m}|\lambda)P(\lambda)
\end{aligned} \tag{2.5}
$$

Computing the likelihood $P(x_{i,1}, x_{i,2}, \ldots, x_{i,m}|\lambda)$ is very complex and requires large amounts of data, in order to produce reliable estimates. Therefore, the naive Bayes classifier makes one simple, hence, naive, assumption that the attribute values are conditionally independent from each other. This implies that the probability of observing the conjunction $x_{i,1}, x_{i,2}, \ldots, x_{i,m}$ is the product of the probabilities for the individual attributes: $P(x_{i,1}, x_{i,2}, \ldots, x_{i,m}|\lambda) = \prod_{j=1}^{m} P(x_{i,j}|\lambda)$. Substituting this expression into Equation (2.5), we obtain the naive Bayes classifier:

$$
c_{nb}(x_i) = \underset{\lambda \in \mathscr{L}}{\arg\max}\, P(\lambda) \prod_{j=1}^{m} P\left(x_{i,j}|\lambda\right) \tag{2.6}
$$

2.4 Adapting NB to Ranking

Consider the classic problem of the "play/no play" tennis based on weather conditions. The naive Bayes classification algorithm can be successfully applied to this problem (Mitchell, 1997, chap. 6). For illustration purposes, we extend this example application to the label ranking setting by replacing the target with a ranking on the preferences of a golf player regarding three golf courts on different days (Table 2.1). The last three columns represent the ranks of the golf courts A, B and C.

As described earlier, the difference between classification and label ranking lies in the target variable, y. Therefore, to adapt NB for ranking we have to adapt the parts of the algorithm that depend on the target variable, namely:

- prior probability, $P(y)$
- conditional probability, $P(x|y)$

The adaptation should take into account the differences in nature between label rankings and classes. For example, if we consider label ranking as a classification problem, then the prior probability of ranking $\{A, B, C\}$ on the data given in Table 2.1 is

$P(\{A,B,C\}) = 3/6 = 0.5$, which is quite high. On the other hand, the probability of $\{A,C,B\}$ is quite low, $P(\{A,C,B\}) = 1/6 = 0.17$. However, taking into account the stochastic nature of these rankings (Cheng et al., 2009), it is intuitively clear that the observation of $\{A,B,C\}$ increases the probability of observing $\{A,C,B\}$ and vice-versa. This affects even rankings that are not observed in the available data. For example, the case of unobserved ranking $\{B,A,C\}$ in Table 2.1 would not be entirely unexpected in the future considering a similar observed ranking $\{B,C,A\}$.

One approach to deal with stochastic nature characteristic of label rankings is to use ranking distributions, such as the Mallows model (e.g., (Lebanon and Lafferty, 2002; Cheng et al., 2009)). Alternatively, we may consider that the intuition described above is represented by varying similarity between rankings.

Similarity-based label ranking algorithms have two important properties:

- they assign non-zero probabilities even for rankings which have not been observed. This property is common to distribution-based methods;

- they are based on the notion of similarity between rankings, which also underlies the evaluation measures that are commonly used. Better performance is naturally expected by aligning the algorithm with the evaluation measure.

Similarity and probability are different concepts and, in order to adapt NB for label ranking based on the concept of similarity, it is necessary to relate them. A parallel has been established between probabilities and the general Euclidean distance measure (Vogt et al., 2007). This work shows that maximizing the likelihood is equivalent to minimizing the distance (i.e., maximizing the similarity) in a Euclidean space. Although not all assumptions required for that parallel hold when considering distance (or similarity) between rankings, given that the naive Bayes algorithm is known to be robust to violations of its assumptions, we propose a similarity-based adaptation of NB for label ranking.

In the following description, we will retain the probabilistic terminology (e.g., prior probability) from the original algorithm, even though it does not apply for similarity functions. However, in the mathematical notation, we will use the subscript $_{LR}$ to distinguish the concepts. Despite the abuse, we believe this makes the algorithm easier to understand.

We start by defining $\mathscr{S}$ as a similarity matrix between the target rankings in a training set, i.e. $\mathscr{S}_{n\times n} = \rho(y_i, y_j)$. The prior probability of a label ranking is given by:

$$P_{LR}(y) = \frac{\sum_{i=1}^{n} \rho(y, y_i)}{n} \tag{2.7}$$

We say that the prior probability is the mean of similarity of a given rankings to all the others. We measure similarity using the Spearman correlation coefficient (Equation (2.1)). Equation (2.7) shows the average similarity of one ranking relative to others. The greater the similarity between two particular rankings, the higher is the probability that the next unobserved ranking will be similar to the known ranking. Take a look at panel A of Table 2.2 with the calculated prior probability for the unique rankings. We also added a column with prior probabilities considering the rankings as one class ($P(y)$). As stated above, the ranking $\{A,C,B\}$, due to its similarity to the other two rankings, achieves a higher probability (0.708)[5].

The similarity of rankings based on the value i of attribute a, ($v_{a,i}$), or conditional probability of label rankings, is:

$$P_{LR}(v_{a,i}|y) = \frac{\sum_{i:x_{i,a}=v_{a,i}} \rho(y, y_i)}{|\{i : x_{i,a} = v_{a,i}\}|} \tag{2.8}$$

Panel B of Table 2.2 demonstrates the logic behind the conditional probabilities based on similarity. Notice that there are no examples with $Outlook = Sunny$ and a target ranking of $\{A,C,B\}$; thus, $P(Outlook = Sunny|\{A,C,B\}) = 0.000$. However, in the similarity approach, the probability of $\{A,C,B\}$ depends on the probability of similar rankings, yielding $P_{LR}(Outlook = Sunny|\{A,C,B\}) = 0.412$.

Applying Equation (2.6), we get the estimated posterior probability of ranking y:

$$\begin{aligned} P_{LR}(y|x_i) &= P_{LR}(y) \prod_{a=1}^{m} P_{LR}(x_{i,a}|y) = \\ &= \frac{\sum_{j=1}^{n} \rho(y, y_j)}{n} \left[\prod_{a=1}^{m} \frac{\sum_{j:x_{j,a}=x_{i,a}} \rho(y, y_j)}{|\{j : x_{j,a} = x_{i,a}\}|} \right] \end{aligned} \tag{2.9}$$

The similarity-based adaptation of naive Bayes for label ranking will output the rank-

[5] Since we measure P_{LR} as a similarity between rankings, it would not sum to one as the in case of probability for classification.

ing with the higher $P_{LR}(y|x_i)$ value:

$$\hat{y} = \underset{y \in y_{\mathscr{L}}}{\arg\max} P_{LR}(y|x_i) = \\ = \underset{y \in y_{\mathscr{L}}}{\arg\max} P_{LR}(y) \prod_{a=1}^{m} P_{LR}(x_{i,a}|y) \tag{2.10}$$

2.5 Naive Bayes for label ranking: special cases

2.5.1 Continuous case

The naive Bayes algorithm for label ranking mentioned above requires nominal variables in order to calculate the probabilities. In this section we extend the adaptation for the continuous case.

We propose to modify conditional label ranking probability by utilizing Gaussian distribution of the independent variables; thus, applying traditional normal distribution approach. The naive Bayes for classification with continuous variables was implemented in Bouckaert (2005). We apply the same logic for conditional probability of label rankings and Equation (2.8) for the discrete case transforms to the continuous one as:

$$P_{LR}(x_i|y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{\frac{(x_i-\mu_y)^2}{2\sigma_y^2}} \tag{2.11}$$

where μ_y and σ_y^2 weighted mean and weighted variance for LR, defined as follows:

$$\mu_y = \frac{\sum_{i=1}^{n} \rho(y, y_i) x_i}{\sum_{i=1}^{n} \rho(y, y_i)}; \qquad \sigma_y^2 = \frac{\sum_{i=1}^{n} \rho(y, y_i)(x_i - \mu_y)^2}{\sum_{i=1}^{n} \rho(y, y_i)} \tag{2.12}$$

2.5.2 Time series of rankings

The time dependent label ranking (TDLR) problem takes the intertemporal dependence between the rankings into account. That is, rankings that are similar to the most recent ones are more likely to appear. To capture this, we propose the weighted TDLR prior probability:

$$P_{TDLR}(y_t) = \frac{\sum_{t=1}^{n} w_t \rho(y, y_t)}{\sum_{t=1}^{n} w_t} \tag{2.13}$$

where $w_t = \{w_1, \ldots, w_n\} \rightarrow \mathbf{w}$ is the vector of weights calculated from the exponential function $\mathbf{w} = b^{\frac{1-\{n\}_1^t}{n}}$. Parameter $b \in \{1 \ldots \infty\}$ sets the degree of the "memory" for the past rankings, i.e., the larger b, the more weight is given to the most recent rankings.

As for the conditional label ranking probability, the equation for the weighted mean (Equation (2.11)) becomes:

$$\mu(x_{t,m}|y_t) = \frac{\sum_{t=1}^{n} w_t \rho(y, y_t) x_{t,m}}{\sum_{t=1}^{n} \rho(y, y_t)} \tag{2.14}$$

and variance:

$$\sigma_w^2(x_{t,m}|y) = \frac{\sum_{i=1}^{n} w_t \rho(y, y_t)[x_{t,m} - \mu(x_{t,m}|y)]^2}{\sum_{i=1}^{n} \rho(y, y_t)} \tag{2.15}$$

2.6 Data

Given the novelty of the label ranking problem, it is hard to find the available data for LR experiments. We follow the convention to use the KEBI data set[6]. The description of individual set presented in Table 2.3. Notice that there are two types of data. The explanation of each of them is given in Cheng et al. (2009):

> [Type (A) data sets:] a naive Bayes classifier is first trained on the complete data set. Then, for each example, all the labels present in the data set are ordered with respect to the predicted class probabilities (in the case of ties, labels with lower index are ranked first) . . . [Type (B):] for regression data, a certain number of (numerical) attributes is removed from the set of predictors, and each one is considered as a label. To obtain a ranking, the attributes are standardized and then ordered by size.

[6]https://www.uni-marburg.de/fb12/kebi/research/repository/

2.7 Experiment Results

We empirically test the proposed adaptation of the naive Bayes algorithm for learning label rankings.

Given that the attributes in the datasets are numerical and the NB algorithm is for symbolic attributes, they must be discretized. We used a simple equal-width binning method using 10 bins. We also perform the experiments on continuous data applying the modified naive Bayes for continuous case outlined in Section 2.5.1. In addition, we compare the results with the state-of-the-art LR algorithm developed in Brazdil et al. (2003); namely, k-NN (k=3) for label ranking.

The baseline is a simple method based on the mean rank of each label over all training examples (Brazdil, Soares, Giraud-Carrier, and Vilalta, 2009).

$$\hat{y}_j^{-1} = \frac{\sum_{i=1}^{n} y_{i,j}^{-1}}{n} \tag{2.16}$$

where $y_{i,j}^{-1}$ is the rank of label λ_j on dataset i. The final ranking is obtained by ordering the mean ranks and assigning them to the labels accordingly. This ranking is usually called the *default ranking*, in parallel to the default class in classification.

The performance of the label ranking methods was estimated using a methodology that has been used previously for this purpose (Brazdil et al., 2003). It is based on 10-fold cross validation. The accuracy of the rankings predicted by methods was evaluated by comparing them to the target rankings (i.e., the rankings based on the observed performance of the algorithms) using the Spearman's correlation coefficient (Equation (2.1)). The code for all the examples in this paper has been written in R (R Development Core Team, 2008).

The results of the experiments are presented in Table 2.4. We report that the naive Bayes for label ranking for continuous case exhibits the maximum number of datasets for which it out-performed the baseline and is competitive with the state-of-the-art. The bold numbers in the table represent the value of the average Spearman correlation across 10-folds that are higher than that of the baseline.

For the continuous case, the naive Bayes for label ranking algorithm outperformed the baselines in 13 datasets out of 17. For the discretized case the number of outperformed datasets is 12. The state-of-the-art algorithm outperformed the baseline in continuous and discrete scenarios in 14 and 12 respectively. Observe, that given the different nature of datasets (types A or B) and different variable classes (continuous vs. nominal) both label ranking algorithms exhibit relatively great performance in predicting the rankings.

2.8 Conclusion

In this paper we presented an adaptation of the naive Bayes algorithm for label ranking that is based on similarities of the rankings taking advantage of a parallel that can be established between the concepts of likelihood and distance. We tested the new algorithm on label ranking datasets and conclude that it consistently outperforms a baseline method and is competitive with the state-of-the-art.

A number of issues remain open, which we plan to address in the future. Firstly, we are currently working on creating new datasets for ranking applications in different areas, including finance (e.g., predicting the rankings of the financial analysts based on their recommendations). These new datasets will enable us to better understand the behavior of the proposed algorithm. In addition, we assume that target rankings are total orders. In practice, this is often not true (Cheng et al., 2010; Brazdil et al., 2003). We plan to address the problem of partial orders in the future. Finally, we plan to compare the new method with existing ones.

Table 2.1: Example of preferences of a golf player

The table shows preferences of a golf player for the golf courts conditional on different weather conditions.

Day	Outlook	Temperature	Humidity	Wind	Ranks		
					A	B	C
1	Sunny	Hot	High	Weak	1	2	3
2	Sunny	Hot	High	Strong	2	3	1
3	Overcast	Hot	High	Weak	1	2	3
4	Rain	Mild	High	Weak	1	3	2
5	Rain	Mild	High	Strong	1	2	3
6	Sunny	Mild	High	Strong	3	2	1

Table 2.2: Comparison of probabilities

The table shows the comparison of values of prior (panel A) and conditional (panel B) probabilities of golf courts rankings from Table 2.1 as a classification (P) and as a label ranking (P_{LR}) problem.

Panel A: prior probability

y			$P(y)$	$P_{LR}(y)$
A	B	C	0.500	0.667
B	C	A	0.167	0.542
A	C	B	0.167	0.708

Panel B: conditional probability

y			$P(Outlook = Sunny \mid y)$	$P_{LR}(Outlook = Sunny \mid y)$
A	B	C	0.333	0.312
B	C	A	1.000	0.615
A	C	B	0.000	0.412

Table 2.3: Description of the label ranking dataset

The table depicts the data used in the label ranking experiments. Type A datasets is based on the naive Bayes classifier. Type B is from the regression data.

Datasets	type	Instances	Features	Labels
authorship	A	841	70	4
glass	A	214	9	6
iris	A	150	4	3
segment	A	2310	18	7
vehicle	A	846	18	4
vowel	A	528	10	11
wine	A	178	13	3
bodyfat	B	252	7	7
cpu-small	B	8192	6	5
housing	B	506	6	6
stock	B	950	5	5
wisconsin	B	194	16	16
cold	-	2465	24	4
diau	-	2465	24	7
dtt	-	2465	24	4
heat	-	2465	24	6
spo	-	2465	24	11

Table 2.4: Results of Label Ranking experiments on KEBI datasets

The table depicts the results of label ranking experiments applied on KEBI dataset sorted by type of the dataset. We use 10-fold cross validation. Bold fonts means the algorithm outperformed the baseline.

Datasets	type	baseline	nbr.cont	nbr.disc	knn.cont	knn.disc
authorship	A	0.643	0.365	**0.665**	**0.955**	**0.936**
glass	A	0.698	0.695	**0.764**	**0.901**	**0.846**
iris	A	0.150	**0.817**	**0.82**	**0.973**	**0.896**
segment	A	0.470	**0.756**	**0.742**	**0.978**	**0.952**
vehicle	A	0.216	**0.611**	**0.656**	**0.889**	**0.845**
vowel	A	0.250	**0.747**	**0.405**	**0.947**	**0.875**
wine	A	0.346	**0.781**	**0.483**	**0.942**	**0.892**
bodyfat	B	-0.074	**0.178**	**0.077**	**0.196**	**0.175**
cpu-small	B	0.259	**0.343**	**0.315**	**0.504**	0.126
housing	B	0.069	**0.604**	**0.557**	**0.819**	**0.628**
stock	B	0.075	**0.666**	**0.414**	**0.962**	**0.823**
wisconsin	B	-0.026	**0.555**	**0.184**	**0.601**	**0.454**
cold	-	0.050	**0.091**	0.02	**0.087**	**0.052**
diau	-	0.259	0.157	0.251	0.19	0.179
dtt	-	0.124	**0.143**	0.105	0.09	0.087
heat	-	0.035	**0.054**	0.029	**0.056**	0.033
spo	-	0.204	0.113	0.182	0.11	0.099

Table 2.5: Summary of models

Category	Label ranking methods	Description	References
Decomposition: The LR problem is decomposed into small, simpler sub-problems (binary classification problems) that, on average, achieve the great performance in experiments but requires an ensemble of binary models	Constraint classification (CC)	Turns the LR problem into single binary classification problem in an extended space and learns LR model from the classifier	Har-Peled et al. (2002)
	Log-linear model (LL)	Learns the utility function for each individual label	Dekel et al. (2004)
	Pairwise comparison (RPC)	Directly models individual preferences (without estimating utility function). An extension of pairwise classification	Hüllermeier et al. (2008)
Probabilistic: leverages statistical probability models to develop LR methods. Good: provides the measure of reliability of prediction. Bad: requires storing the all training data in memory	Instance-base (Mallows).	Distance-based probability model that defines the probability of ranking according to its distance to a center ranking.	Cheng and Hüllermeier (2009)

... continued next page

Table 2.5: (continued)

Category	Label ranking methods	Description	References
	Decision trees	Similar to conventional decision tree learning. The difference is that the split criterion is at inner nodes and different criterion for stopping the recursing partitioning.	Cheng et al. (2009)
	Instance-base (Plackett-Luce)	The probability is based on the scores of unassigned labels.	Cheng et al. (2010)
	Generalize linear models		Cheng et al. (2010)
	Gaussian mixture model	The model consist of mixtures defined by prototypes which are associated with preference judgment for each pair of labels.	Grbovic, Djuric, and Vucetic (2012)
Similarity: replaces probability with similarity between the rankings. Minimizing the distance is equivalent to maximizing the likelihood (maximizing the similarity). Good: assigns non-zero probabilities that are not observed in data. Bad: shows moderate predicting accuracy	Naive Bayes	Adaptation of naive Bayes for classification. Adapts the prior and conditional probabilities in the realm of LR	Aiguzhinov et al. (2010)
	Association rules	Adaptation of APRIORI. The goal is to discover frequent pairs of attributes associated with a ranking	de Sá, Soares, Jorge, Azevedo, and Costa (2011)

... continued next page

Table 2.5: (continued)

Category	Label ranking methods	Description	References
	Multilayer perception	Adaptation of MLP. Adapts the error functions that guide the back-propagation leaning process and the method to generate a ranking from the output layer.	Ribeiro, Duivesteijn, Soares, and Knobbe (2012)
	Rank distance	The LR model learns rankings from the nearest neighbor	Brazdil et al. (2003)
	Rule-based	The learning approach is based on reduction technique	Gurrieri, Siebert, Fortemps, Greco, and Słowiński (2012)

Chapter 3

Understanding rankings of financial analysts

Abstract

The prediction of the most accurate analysts is typically modeled in terms of individual analyst characteristics. This approach has its disadvantage that these data are hard to collect and often unreliable. We follow a different approach in which we characterize the general behavior of rankings of analysts based upon state variables rather than individual analyst characteristics or past accuracy. We use a Machine Learning label ranking algorithm that we adapt to model the relation between the selected variables and the rankings. By calculating a ranking-based discriminative power, we show that the uncertainty about future stock performance influences the rankings of the analysts while the macroeconomic variables have the most contribution to the changes in rankings.

keywords: financial analysts; rankings; state variables
JEL: G11

3.1 Introduction

The Efficient Market Hypothesis (EHM) (Fama, 1970) suggests that all public information available to investors is incorporated in prices and new information is immediately reflected in valuations. Yet there are information gathering costs and financial analysts are better than an average investor at processing this information which reflects in issued buy/ sell recommendations. These recommendations, like other news about the general economy as well as about a particular company, influence investors' perception and beliefs.

Previous studies show that analysts stock recommendations have investment value (Womack, 1996; Barber et al., 2001). The literature also suggests further that foreknowledge of analyst forecast accuracy is valuable (Brown and Mohammad, 2003; Aiguzhinov et al., 2015a). In line with academic research findings, practitioners too pay attention to analyst forecast accuracy rankings. On an annual basis, firms such as The Institutional Investor and StarMine[1] publish analysts ratings according to how well they performed, based partly on past earnings forecast accuracy.

The importance of these ratings should not be ignored because the attention that the market gives to the recommendations of different analysts is expected to correlate with them. Typically, the performance of analysts is analyzed in terms of their individual characteristics (e.g., experience, background) (Clement, 1999). The disadvantage of this approach is that the collection of the necessary data is difficult and it is not always reliable. As for practitioners, they rely mostly on past accuracy to predict future accuracy.

In this paper we follow an alternative approach. We model the general behavior of rankings of analysts by using variables that characterize the context (state variables) rather than individual analyst characteristics. The model we propose uses the state variables to distinguish which of them affects the rankings the most; hence, influence the analysts' forecast accuracy. In summary, our goal is not to understand relative performance of the analysts in terms of their characteristics but rather in terms of the characteristics of the context in which the analysts operate.

To achieve this goal, we, first, build rankings of analyst based on their EPS forecasts accuracy. Then, we select the state variables that are responsible in differences of analysts' rankings. Finally, we apply a Machine Learning label ranking algorithm to build a model that relates the rankings with the variables and calculates a discriminative power of a variable.

The paper is organized as follows: Section 3.2 provides the motivation for using rankings of the analysts; Section 3.3 outlines the state variables that characterize the context; Section 3.4 outlines the structure of the Machine Learning label ranking model and presents a methodology of building a "variable-ranking" relation; Section 3.5 describes

[1] http://www.starmine.com

the datasets used for the experiments; Section 3.6 summarizes the experiment setup; Section 3.7 presents and discusses the results; finally, Section 3.8 concludes this paper.

3.2 Rankings as a measure of accuracy

In spite of the Efficient Market Hypothesis, it is commonly accepted that the recommendations of financial analysts yield an economic value to investors (Womack, 1996); moreover, recommendations of superior analysts have impact on the market (Loh and Stulz, 2011). For this reason, researchers and practitioners have long been interested in understanding how financial analysts affect capital market efficiency (Ramnath, Rock, and Shane, 2008).

Most researchers conclude that financial analysts are better at making EPS forecasts than mathematical models. Specifically, Fried and Givoly (1982); Bouwman, Frishkoff, and Frishkoff (1987); Brown and Kim (1991) show that analysts are better at forecasting EPS values than any time series models (e.g., ARIMA). The analysts' superiority contributed to the fact that they utilize all available information at and after the date of time series model forecasts. Thus, the context in which analysts make decision matters for their accurate forecast.

In terms of the following advice of the most accurate analysts, it has been shown that the relative accuracy among financial analysts is more important than their absolute accuracy (Aiguzhinov et al., 2015a), e.g., in the context of analysts' turnover rate (Michaely and Womack, 1999), or in creating value to investors (Aiguzhinov et al., 2015a). In addition, financial analysts with superior past accuracy have a greater impact on the market (Park and Stice, 2000). It has also been shown that, under some assumptions, it is safe to assume that analysts with higher forecasting ability produce profitable stock recommendations (Loh and Mian, 2006). This fact is attributed to their deeper research and fundamental accounting knowledge. Furthermore, literature agrees that there is consistency in the superiority of these analysts over time (Li, 2005; Hilary and Hsu, 2013).

Many studies try to correlate the EPS forecasts accuracy of financial analysts with their intrinsic characteristics. However, existing academic research on the behavior of financial analysts have important limitations (Clement, 1999; Brown and Mohammad, 2003; Ramnath et al., 2008), namely an incomplete characterization of the analysts and their recommendations. For instance, Ramnath et al. (2008) address the question of what information affects the recommendations of analysts or how informative are their short-term earnings forecasts, using linear regression on a small sample of data. Despite the promising results, further work is necessary to improve both the methods and the characterization of the context of recommendations.

In this paper, we propose a novel approach in identifying variables that affect the rankings; hence, the relative accuracy of the analysts. The novelty lies in the new methodology of modeling the relationship between the analysts' rankings and the state variables. To build the model, we are required to select the state variables and build analysts rankings.

3.3 Ranking characterization variables

Several studies try to analyze factors that affect the performance of analysts (Clement, 1999; Brown and Mohammad, 2003; Jegadeesh, Kim, Krische, and Lee, 2004). However, most of these papers look at the individual characteristics of analysts such as their job experience, affiliation, education background, industry specializations. These variables are very important to characterize the relative performance of the analysts in general but they miss the "state-of-the-world" component, i.e., variables that affect all analysts at once. We believe that rankings of analysts capture this component in full.

A ranking is a result of differences in opinion among the analysts concerning the future performance of a company. This implies that there is a variability (dispersion) in analysts' forecasts for a given stock in a given quarter (Diether et al., 2002). Thus, we can analyze the analysts' forecasts dispersion in terms of its origin and factors that affect it; hence, assuming the same variables affect the rankings. It follows that the variation in rankings is due to the different ability of the analysts to interpret the informational environment (e.g., whether the market is bull or bear). We, thus, select and analyze variables that describe this environment.

To capture the full spectrum of the analyst's decision making process, we select variables based on different levels of information availability: analyst-specific, firm-specific and general economy. In each level, we want a variable to be responsible for information asymmetry and uncertainty. Thus, we believe that these two domains are responsible for the differences in analysts' opinions.

3.3.1 Analyst-specific variables

On an analyst level, we want to capture the asymmetry of information and uncertainty about the future among the analysts. Particularly, Barron, Stanford, and Yu (2009) point our that the reason for analysts' dispersion is either the uncertainty or the information asymmetry: prior to earnings announcement the uncertainty component prevail, whereas around the time of earnings announcement, information asymmetry is responsible for changes in analysts' opinions.

We use the same set of variables defined in (Barron et al., 2009, page 333):

$$
\begin{aligned}
SE_s &= (ACT_s - \overline{FE_s})^2 \\
disp_s &= \sum_{a=1}^{N} \frac{(FE_{a,s} - \overline{FE_s})^2}{(N-1)} && (3.1) \\
uncert_s &= \sum_{a=1}^{N} \left(1 - \frac{1}{N}\right) \times disp_s + SE_s && (3.2) \\
assym_s &= 1 - \frac{SE_s - \frac{disp_s}{N}}{uncert_s} && (3.3)
\end{aligned}
$$

where SE is the squared error in mean forecast; $\overline{FE}$ is the average per analyst a EPS forecast error (see Section 3.6.1); and N is the number of analysts in a given quarter for a given stock s.

Equation (3.1) calculates the dispersion among the analysts which is a variance of EPS forecasts of all analysts for a given stock. Equation (3.2) defines the uncertainty component of the dispersion per Barron et al. (2009). Equation (3.3) is the proxy for the information asymmetry.

3.3.2 Firm-based variables

To be consistent with the two paradigms that characterize the state of the analysts, we split the firm-based variables based on their influence on analysts' opinions. They are either the uncertainty or the information asymmetry.

3.3.2.1 Uncertainty

The following are the set of the variables and their definitions that we think are responsible for the information uncertainty component.

3.3.2.1.1 Business risk.

Business risk is associated with the uncertainty in operating results, especially, in operating earnings (Hill and Stone, 1980). An increase in business risk entails an increase in *ex-ante* volatility of the reported earnings (Parkash, Dhaliwal, and Salatka, 1995). We believe that book-to-market ratio can serve as a proxy for the business risk measurement.

$$btm_s = \frac{EQUITY}{MKT.CAP} = \frac{Tot.assts - Tot.liab}{Stocks \times Price} \tag{3.4}$$

where $Stocks$ is the number of stocks outstanding and $Price$ is the close stock price on last day of a quarter.

3.3.2.1.2 Financial risk.

Financial risk is responsible for the information uncertainty of the future earnings. More debt implies more variability in earnings as managers would try to maximize the value of a stock using the debt; thus, having high risk of default in the future or taking high risk investment projects. The debt-to-equity ratio is used to capture the financial risk (Parkash et al., 1995). We use short-term debt from the balance sheet (Notes payable) as a measure for debt.

$$dte_s = \frac{DEBT}{EQUITY} = \frac{ShortTermDebt}{Tot.assts - Tot.liab} \tag{3.5}$$

3.3.2.1.3 Size.

The firm size can be used as a proxy for amount of information available for a firm. Thus, a larger firm has more news coverage which reduces the uncertainty. An investor is likely to find private information about a larger firm more valuable than the same information about a smaller firm (Bhushan, 1989).

Size is measured as the market value (MV) of the firm as following:

$$size_s = \log(Price \times Stocks) \tag{3.6}$$

Consistent with the literature, we use log of market value.

3.3.2.1.4 Return variability.

Return variability influences the uncertainty regarding future earnings (Diether et al., 2002; Henley, Carruth, and Dickerson, 2003). An increase in variability of the abnormal returns is positively correlated with the uncertainty about the earnings; thus, affecting the dispersion among the analysts. To calculate the return variability, we use method outlined in Sousa and Serra (2008), where stock return volatility is decomposed into market and stock specific components as follow:

$$\begin{aligned} \sigma_{mkt}^2 &= \sum_{d \in q} (R_{mkt,d} - \mu_{mkt})^2 \\ \sigma_s^2 &= \sum_{d \in q} (R_{s,d} - R_{mkt,d})^2 \\ s.ret_s = Var(R_{s,q}) &= \sigma_{mkt}^2 + \sigma_s^2 \end{aligned} \tag{3.7}$$

where $R_{mkt,q}$ is the market return over sample period; μ_{mkt} is the mean of daily market returns; $R_{s,q}$ is an individual stock return; d is the number of trading days in period q.

3.3.2.2 Information asymmetry variables

3.3.2.2.1 Accruals.

Accruals, as a part of earnings, is one of the variables that cause the information asymmetry between managers of a firm and investors. Studies have shown that presence of asymmetry is a necessary condition for the earnings management (Trueman and Titman, 1988; Richardson, 2000). To be more specific, it is the discretionary part of the accruals that causes the information inefficiency in the earnings management (Richardson, 2000; Ahmed, Nainar, and Zhou, 2005). We calculated total accruals-to-total assets ratio defined in Creamer and Stolfo (2009):

$$accr_s = \frac{\Delta C.As - \Delta Cash - (\Delta C.Lb. - \Delta C.Lb.D) - \Delta T - D\&A_q}{(T.As. - T.As._{q-4})/2} \tag{3.8}$$

where $\Delta X = X_q - X_{q-1}$; $C.As$ – current assets; $C.Lb$ – current liabilities; $C.Lb.D$ – debt in current liabilities; T – deferred taxes; $D\&A$ – depreciation and amortization; and $T.A$ – total assets.

3.3.2.2.2 Sector-based variables.

The industry specific variables that cause the dispersion in the analysts' forecasts are also connected with the uncertainty concept. One of the variables that is suggested is the variability in the industry Producer Price Index (PPI) (Henley et al., 2003).

$$sec.ret = \sigma(\log PPI_{sec}) \tag{3.9}$$

where $\sigma(\log PPI_{sec})$ is the standard deviation of the log of SIC sectors' produce price index.

3.3.3 Macroeconomics variables

In the last set of the state variables, we capture the macroeconomic conditions which affect the analysts' dispersion. For example, different states of the economy are based on different levels of "GNP vs. inflation" combinations (Lev and Thiagarajan, 1993; Hope and Kang, 2005). When economy is booming, i.e. "high GNP-low inflation" state, Lev and Thiagarajan (1993) observe the significant increase in firms' Capital Expenditures coefficient. This implies that firms start enjoy capital investment due to the low cost of capital. This state of the economy produces less uncertainty. In the "medium GNP-high inflation" state of the economy, there is an increase in R&D expenditures, which, from the above mentioned analysis, may spur high level of information asymmetry based on the increase R&D activities. Finally, in the "low GNP-high inflation" state, Lev and Thiagarajan (1993) observe the Doubtful Receivables coefficient is the largest implying that at this recession state many firms go bankrupt or default on the loans – a signal of high uncertainty in the economy. All these states produce the dispersion of the analysts' forecasts.

We select the following set of the macroeconomic variables:

- gnp = Gross National Product;
- $infl$ = Inflation rate;
- $t.bill$ = Interest rate (90-days T-bill rate);
- $vix.ret$ = Market variability (CBOE VIX index)

3.4 Ranking-based discriminative power

We select the naive Bayes label ranking algorithm (Aiguzhinov et al., 2010) as a tool to calculate the discriminative power. The basic idea lies in the similarity among rankings conditional on a set of independent variables.

If we define $\mathscr{S}$ as a similarity matrix between the rankings ($\mathscr{S}_{n\times n} = \rho(y_i, y_j)$), then the prior probability of a label ranking is given by:

$$P(y) = \frac{\sum_{i=1}^{n} \rho(y, y_i)}{n} \tag{3.10}$$

where ρ is the Spearman ranking correlation.

The conditional probability of the value v of attribute x (x_v) given the ranking y is:

$$P(x_v|y) = \frac{\sum_{i:x_i=v} \rho(y, y_i)}{|\{i : x_i = v\}|} \tag{3.11}$$

We propose that, the discriminative power of x is based on the conditional ranking probabilities of values x and they should: 1) be different from each other; 2) be different from the prior probability. Thus, given the prior ranking $P(y)$ and conditional probability $P(x_v|y)$, the discriminative value of x can be found as follows:

$$DP_x = \frac{1}{n} \sum_{t=1}^{n} \min_{\forall p\neq q} \left\{ |P(x_{v_p}|y_t) - P(x_{v_q}|y_t))| \right\} \times \left\{ |P(x_{v_p}|y_t) - P(y_t)| \right\} \tag{3.12}$$

The multiplicand of Equation (3.12) finds the minimum absolute difference in conditional probabilities between different values of attribute x given ranking y. The multiplier checks that the conditional label ranking probability of x is different from the prior probability of ranking y_t. The case when $DP = 0$ means that x does not discriminate; thus, we consider the cases when $DP > 0$.

We also measure the discriminative variable contribution as a variable's discriminative power share in total discriminative power of all variables:

$$fracDP_x = \frac{DP_x}{\sum_{j=1}^{J} DP_{x=j}} \tag{3.13}$$

where J is the total number of independent variables.

Panel A of Table 3.1 shows an example where we have an artificial data for 5 quarters and rankings of 4 equity research firms (A,B,C,D). We assume that we identify some state variables $\{x_1, x_2\}$ that made the rankings as they are in a given quarter. Panel B (C) of the table shows the conditional ranking probabilities of x_1 (x_2).

We apply Equation (3.12) to calculate a discriminative power of the variable. For an example from Table 3.1, these values are 0 and 0.007 for variables x_1 and x_2 respectively. Based on this example, we conclude that the most discriminative variable is x_2. The result is intuitive as the variable x_2 takes only two values a and b with value a appearing 4 times; thus, it has more contributive effect on rankings.

3.5 Data and preliminary results

We selected companies that are publicly traded in either NYSE, NASDAQ, or AMEX. The stocks accounting data was obtained from the Thomson One/Reuters Fundamental database. The analysts[2] EPS forecasts data is from I/B/E/S for each company at study. The descriptive statistics of state variables is presented in Table 3.2.

We apply a number of requirements for our analysts' data. We select stocks with minimum 12 quarters of coverage by at least one analyst. In addition, for the computational purpose, we require at least 3 analysts per stock in each quarter. We call this a *filtered* set in contrast to *sample* set that includes all analysts and stocks in the sample.

Table 3.3 outlines the number of stocks, analysts and total forecasts in *sample* (Panel A) and *filtered* datasets (Panel B). For *sample* (*filtered*) data we report 560 (202) unique analysts covering 3517 (1059) stocks during 84 quarters from 1989Q1 until 2009Q4. For this period there were 698291 (164445) issued forecasts.

[2] We use words "analyst" even-though the database is for Equity Research Firms.

Table 3.4 presents the descriptive statistics of *sample* (Panel A) and *filtered* (Panel B) data from the "per analyst" perspective. Concretely, for the *sample* (*filtered*) data the total number of "$Analyst \times Forecasts$" observations is 11796 (7034). Each analyst, on average, issued 59.2 (41.15) forecasts per quarter, and, if we factor in stocks, the average forecasts per stock per quarter becomes 1.35 (1.65). We also report a share of analysts that revise their EPS forecasts within a quarter. For *sample* (*filtered*) data 76.16% (79.09)% of analysts revise their EPS forecasts. Finally, on average, analysts follow stocks for 4.49 (13.71) quarters.

The similar descriptive analysis but from the "per stock" perspective presented in Table 3.5. Namely, for the *sample* (*filtered*) data the total number of "$Stock \times Forecasts$" observations is 112992 (30073). Each stock, on average, receives 6.18 (9.62) forecasts per quarter. The average forecasts per analyst per quarters is 1.44 (1.67). On average, 61.6% (86.82%) of stocks receive a revision of EPS forecasts within a quarter for *sample* (*filtered*) dataset. Finally, on average, a stock is followed by analysts for 6.76 (14.8) quarters.

Figures on pages 77–79 depict some per quarter statistics. Figure 3.1 plots log of total number of EPS forecasts for both datasets. We observe that, while both datasets experience a constant growth in issuing forecasts, at the end of the sample period the *filtered* datasets shows a decline which can be contributed to the sub-prime crisis of 2007-2009. When looked at the per quarter forecast statistics in Figure 3.2, we observe that the analysts in *filtered* dataset issued fewer forecasts per quarter compared to those of the *sample* dataset. Figure 3.3 plots the average percent of analysts that revise their forecasts (revise from 1 time (top panel) to 5 times (bottom panel) per quarter). We observe that the analysts in *filtered* datasets, on average, are more active in revising their EPS forecasts. As we observe, despite the smaller number of stocks and total issued forecasts, the *filtered* dataset selects analysts that actively revise their forecasts and has a longer duration of a stock coverage when compared to those of the *sample* dataset. Because of this reason, we use *filtered* dataset to build the rankings.

3.6 Experiment setup

3.6.1 Rankings of financial analysts

Analysts are ranked on the basis of Proportional Mean Absolute Forecast Error (PMAFE) that measures the accuracy of a forecast (Clement, 1999; Brown, 2001; Ertimur et al., 2007). First, we define the forecast error $FE_{a,s}$ as an absolute value of the difference between an analyst's a forecasted EPS and actual EPS for each stock s:

$$FE_{a,s} = |ACT_s - PRED_{a,s}| \tag{3.14}$$

The PMAFE is given as:

$$PMAFE_{a,s} = \frac{FE_{a,s}}{\overline{FE_s}} \tag{3.15}$$

where ACT_s and $PRED_{a,s}$ are the actual quarterly EPS and analyst a's EPS forecast for stock s respectively.

Second, we rank analysts based on their PMAFE score:

$$rank_{a,s} = \mathrm{rank}_{a=1}^{N} \{PMAFE_{a,s}\} \tag{3.16}$$

3.6.1.1 Ranking contingency results

We analyze the analysts' ranking consistency based on the process outlined in Aiguzhinov et al. (2015a). Namely, we split the rankings into three terciles (*top*, *medium*, *bottom*). In one particular quarter (t), we place analysts at one of these bins which corresponds to a tercile. We, then, check analysts position at the immediate next quarter ($t+1$) and after one year ($t+4$).

Beforehand, we convert the rankings into scores as follows:

$$score_{a,s} = \frac{rank_{a,s}}{\max rank_s} \tag{3.17}$$

To get the cross-sectional values of scores across different stocks, we take the average of $score_{a,s}$

$$\overline{score_a} = \frac{1}{k}\sum_{s=1}^{k} score_{a,s} \tag{3.18}$$

where k is number of stocks followed by a particular analyst a.

Table 3.6 summarizes the resulted contingency table. We observe that analysts exhibit strong ranking consistency as, on average, they stay at the same tercile. The table demonstrates that 50.05% and 28.83% (46.75% and 30.83%) of the analysts remained in the top and bottom terciles, respectively, after one quarter (year).

3.6.2 Dynamic states

As we have mentioned above, we want to capture the state of the world in which the analysts operate. For this reason, it is necessary to take into account the dynamics of independent variables from one time period to another. We propose the following methods:

- `static`: no dynamics in the state of the variables, i.e., independent variables used as they are: $x_{\Delta t} = x_t$;

- `diff`: first-difference of the variables, i.e., $x_{\Delta t} = x_t - x_{t-1}$;

- `random`: in time series decomposition of the independent variables, it is an unobserved component: $x_{\Delta t} = T(t) + S(t) + \varepsilon(t)$, where $T(t)$- trend, $S(t)$ - seasonal part and $\varepsilon(t)$ - random part of time series decomposition.

- `roll.sd`: rolling 8 quarters standard deviation of the independent variables (Zivot and Wang, 2003):

$$\begin{aligned} \mu_t(8) &= \frac{1}{8}\sum_{j=0}^{7} x_{t-j} \\ \sigma_t^2(8) &= \frac{1}{7}\sum_{j=0}^{7} (x_{t-j} - \mu(8))^2 \end{aligned} \tag{3.19}$$

Each of these methods produces a different set of attributes. By building a discriminative model on each one of them separately, we get different sets of discriminative power of the variables.

3.7 Results

We report the result of the discriminative power of the variables in terms of their contribution in affecting the rankings of financial analysts (Equation (3.13)). Panel A of the Table 3.7 shows the case of analyst specific variables. We report that *uncert* is the most contributive variable for the all dynamic states with its maximum contribution occurring at the `static` state (14.25%). The least contributive variable of the analysts specific variables is the *assym*, the asymmetry of information, with the smallest share of contribution to the discriminative power (2.23%, 1.73%, 1.62%, 0.16% for the `static`, `diff`, `random`, and `roll.sd` states respectively). Thus, our model of the discriminative power suggests that in the `static` state, of all analysts' specific variables, the rankings are most affected by the earnings uncertainty.

Panel B of Table 3.7 presents the contribution of stock specific variables. We report that in each of the dynamic states the *s.ret* (*size*) showed the maximum (minimum) contribution to the difference in analysts' opinions regarding EPS forecasts (8.68% (0.23%), 9.39% (0.4%), 13.66% (0.76%), 12.76% (1.3%) for the `static`, `diff`, `random`, and `roll.sd` states respectively). As we defined above, the variability of stock returns is the measure of uncertainty about future earnings; thus, we report that, similar to the analyst-specific variables, the uncertainty is responsible for the rankings when consider stock-specific variables.

Finally, panel C shows the case of macroeconomic variables. Contrary to the previous variable types, the distribution of the most contributive individual variable differs across different states. For the `static` state, it is *infl* (27.59%) whereas for the all others states it is *t.bill* (25.69%, 21.98%, 26.02% for the `diff`, `random`, and `roll.sd` states respectively). The least contributive variable is the *vix.ret*: 0.38%, 0.45%, 1.64%, and 0.92% for the `diff`, `random`, and `roll.sd` states respectively.

The cross panel analysis of Table 3.7 show that the macroeconomic variables are the most contributive of all. Specifically, the `roll.sd` state accounts for 73.56% of total contribution in differences in rankings; for other states the share of these variables are (in a decreasing order) 61.32%, 59.98%, and 58.32% for the `diff`, `static`, and `random` states respectively. Figure 3.4 depicts the plot of total contribution of each of the variable conditional on variable types. Thus, we conclude that the condition of the economy represented by the GNP, inflation rate, stock market volatility, and interest rate is the one that influence the most analysts' opinions about future stocks' performance with respect to their earnings.

We also perform a hypothesis pairwise test of whether the discriminative power of variables in the dynamic states is significantly different from those in the `static` state. We report results in Table 3.8. Panel A shows significance of analysts specific variables in dynamic states. We reject the *null* hypothesis at 1% significance level in all variables of this category for the states `diff` and `random`, and variables *uncert* and *assym* for the `roll.sd` state. We fail to reject the *null* for the variable *disp* for the `roll.sd` state at 10% significance. Panel B presents the case of stock-specific variables and we report that we reject the *null* for all variables for all dynamic states except *accr* for the `roll.sd` state. Finally, panel C shows the case of macroeconomic variables and we also reject the *null* for all the variables and states except *infl* in the `diff` state.

Figure 3.5 plots the percentage split of the average *DP* between the analyst-, stock-specific, and macroeconomic variables per quarter for each of the states. We observe that, as mentioned above, macroeconomic variables are the most contributive components to the analysts' rankings. In the `diff` state, it is visible that at the end of the sample period (years 2006–2009) analyst-specific variables become more contributive.

3.8 Conclusion

Some institutions, such as StarMine, rank financial analysts based on their accuracy and investment value performance. These rankings are published and are relevant: stocks favored by top-ranked analysts will probably receive more attention from investors. Therefore, there is a growing interest in understanding the relative performance of analysts. Typical approaches are based on individual characteristics of those analysts or past analyst forecasting accuracy. Here, we follow an alternative approach that links the general behavior of rankings of analysts to variables that explain the uncertainty and information asymmetry on analyst-specific, stock-specific, and macroeconomic levels.

We introduce a new approach, based on the naive Bayes Label Ranking algorithm, in identifying the discriminative power of a variable; thus, its contribution to the rankings conditional on different states of the world: static state, first-difference, random part of time-series decomposition, and sliding standard deviation.

We report that for the analysts- and stock-specific variables, the uncertainty about future stock performance is the most contributive to the changes in rankings. The macroeconomic variables influence rankings the most considering all variables at once.

For the future work we would like to take the findings of this paper and apply them to the problem of predicting the actual rankings of the analysts.

Table 3.1: Example of Label Ranking problem

The table presents an example of label ranking problem (Panel A). Panel B (panel C) shows the conditional label ranking probabilities for variable x_1 (x_2) obtained from Equation (3.11)

Panel A: Example of LR ranking problem						
t	x_1	x_2	Ranks			
			A	B	C	D
1	a	b	1	2	3	4
2	b	a	2	1	3	4
3	c	a	3	2	1	4
4	d	a	4	3	2	1
5	e	a	4	1	2	3
Panel B: conditional LR probability o f x_1						
	a	b	c	d	e	priors
1	0.34	0.31	0.21	0.00	0.14	0.58
2	0.26	0.29	0.21	0.03	0.21	0.68
3	0.17	0.20	0.29	0.11	0.23	0.70
4	0.00	0.05	0.19	0.48	0.29	0.42
5	0.11	0.20	0.23	0.17	0.29	0.70
Panel C: conditional LR probability of x_2						
	b	a	a	a	a	priors
1	0.31	0.34	0.34	0.34	0.34	0.58
2	0.29	0.26	0.26	0.26	0.26	0.68
3	0.20	0.17	0.17	0.17	0.17	0.70
4	0.05	0.00	0.00	0.00	0.00	0.42
5	0.20	0.11	0.11	0.11	0.11	0.70

Table 3.2: Descriptive statistics of independent variable

The table presents the descriptive statistics of state variables that influence the ranking of the analysts.

Type	Variable	Stock	Median	Mean	std.dev	ACF (lag=1)
Analyst	uncert	988	0.001	0.629	10.881	0.983
	assym	988	0.155	0.403	0.484	0.978
	disp	988	0.000	0.006	0.125	0.971
Stock	btm	981	0.404	35.263	481.567	0.985
	size	981	20.856	20.795	1.577	0.978
	dte	608	0.004	0.131	0.651	0.984
	accr	960	-0.012	1.981	67.357	0.973
	s.ret	988	0.029	0.054	0.081	0.975
	sec.ret	988	0.021	0.008	0.119	0.977
Macro	gnp	988	0.014	0.013	0.005	0.975
	infl	988	0.007	0.008	0.005	0.975
	vix.ret	988	-0.049	0.013	0.309	0.980
	t.bill	988	0.049	0.048	0.023	0.979

Table 3.3: Summary of *sample* and *filtered* data

The table presents the total number of stocks, analysts and EPS forecasts for *sample* (Panel A) and *filtered* (Panel B) data.

Sector	# stocks	# analysts	# forecasts
Panel A: *sample* data			
Consumer Discretionary	567	336	144 754
Consumer Staples	155	215	28 593
Energy	277	205	88 289
Financials	650	228	93 656
Health Care	526	309	76 680
IT	697	413	159 012
Industrials	464	320	80 331
Materials	64	163	12 119
Telecom Services	14	93	1 882
Utilities	103	120	12 975
Total	3 517	560	698 291
Panel B: *filtered* data			
Consumer Discretionary	214	124	34 199
Consumer Staples	42	61	7 155
Energy	99	63	24 819
Financials	178	76	24 606
Health Care	118	91	16 791
IT	221	123	34 590
Industrials	146	101	17 900
Materials	25	46	3 183
Telecom Services	4	14	328
Utilities	12	18	874
Total	1 059	202	164 445

Table 3.4: Descriptive statistics of forecasts per analyst

The table presents the descriptive statistics for *sample* (Panel A) and *filtered* (Panel B) data. Namely, the table shows the total number of analyst-forecast observations, the average number of forecast per quarter, the average number of following stocks per analyst, the average number of forecasts per stock per analyst, share of analysts that make forecast revisions, and, finally, the average number of quarters a analyst follows a stock.

	Obsrv	Frcst/q	Stocks	Frcst/stock	Rev.	follow time, q
Panel A: *sample* data						
Consumer Discretionary	7 405	19.55	11.65	1.40	0.72	5.06
Consumer Staples	3 797	7.53	4.81	1.36	0.60	5.29
Energy	3 541	24.93	12.68	1.50	0.71	5.74
Financials	4 506	20.78	13.00	1.38	0.66	5.52
Health Care	5 812	13.19	8.87	1.29	0.63	4.14
IT	8 018	19.83	12.89	1.34	0.69	4.67
Industrials	6 290	12.77	8.47	1.32	0.65	5.04
Materials	2 531	4.79	3.00	1.44	0.56	6.26
Telecom Services	821	2.29	1.70	1.28	0.39	4.85
Utilities	1 662	7.81	5.53	1.27	0.52	4.34
Total	11 796	59.20	36.83	1.35	0.76	4.49
Panel B: *filtered* data						
Consumer Discretionary	3 966	15.81	8.62	1.70	0.76	13.71
Consumer Staples	1 967	6.07	3.64	1.58	0.65	13.90
Energy	2 212	23.78	11.22	1.96	0.80	17.84
Financials	2 719	15.66	9.05	1.66	0.70	15.19
Health Care	2 708	10.08	6.20	1.49	0.67	13.13
IT	4 007	14.13	8.63	1.56	0.74	13.62
Industrials	3 036	9.36	5.90	1.55	0.70	12.90
Materials	1 230	4.39	2.59	1.71	0.62	13.45
Telecom Services	236	1.89	1.39	1.40	0.40	11.93
Utilities	421	3.46	2.08	1.64	0.58	12.30
Total	7 034	41.15	23.38	1.65	0.79	13.71

Table 3.5: Descriptive statistics of forecasts per stock

The table presents the descriptive statistics per stock for *sample* (Panel A) and *filtered* (Panel B) data. Namely, the table shows the total number of stock-forecast observations, the average number of forecast per quarter per stock, the average number of following analysts per stock, the average number of forecasts per analyst per stock, share of stocks that got their forecast revised by analysts ,and, finally, the average number of quarters a stock being followed by a analyst.

	Obsrv	Frcst/q	analysts	Frcst/analyst	Rev.	follow time, q
Panel A: *sample* data						
Consumer Discretionary	20 772	6.97	4.15	1.47	0.66	7.19
Consumer Staples	5 321	5.37	3.43	1.42	0.63	6.56
Energy	8 325	10.61	5.39	1.67	0.76	7.62
Financials	17 222	5.44	3.40	1.49	0.58	7.20
Health Care	14 407	5.32	3.58	1.33	0.57	5.45
IT	22 469	7.08	4.60	1.40	0.64	6.63
Industrials	17 187	4.67	3.10	1.41	0.59	7.09
Materials	2 518	4.81	3.02	1.44	0.62	7.28
Telecom Services	399	4.72	3.51	1.33	0.56	5.41
Utilities	4 372	2.97	2.10	1.27	0.42	5.69
Total	112 992	6.18	3.84	1.44	0.62	6.76
Panel B: *filtered* data						
Consumer Discretionary	6 254	10.03	5.47	1.74	0.89	15.08
Consumer Staples	1 499	7.96	4.77	1.64	0.86	15.03
Energy	3 432	15.33	7.23	2.01	0.94	18.80
Financials	4 782	8.90	5.15	1.65	0.86	14.91
Health Care	3 235	8.44	5.19	1.58	0.84	13.95
IT	5 780	9.79	5.98	1.59	0.87	13.75
Industrials	4 017	7.08	4.46	1.57	0.83	14.23
Materials	760	7.10	4.19	1.65	0.85	14.47
Telecom Services	80	5.59	4.10	1.35	0.72	11.50
Utilities	234	6.23	3.74	1.65	0.82	11.08
Total	30 073	9.62	5.47	1.67	0.87	14.80

Table 3.6: Analysts' rankings contingency

The contingency table shows changes in analysts' *top*, *middle*, *bottom* ranking bins. Panel A (B) is the results of the rankings based on the *sample* (*filtered*) dataset.

		top	*middle*	*bottom*	*Sum*
			$t+1$		
	top	50.1	25.5	25.2	100.8
	middle	49.8	25.7	25.5	101.1
	bottom	46.0	26.3	28.8	101.1
			$t+4$		
t	*top*	46.8	28.4	27.0	100.8
	middle	46.0	27.7	28.3	101.1
	bottom	44.8	27.7	30.8	101.1

Table 3.7: Contribution of each of the variable to rankings, in %

The table shows the percent of contribution of the variable's *DP* value to total value of *DP*. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of variable sliding 8 quarters standard deviation.

Variable	`static`	`diff`	`random`	`roll.sd`
Panel A: Analyst				
uncert	14.25	12.49	9.71	1.25
assym	2.23	1.73	1.62	0.16
disp	9.00	8.81	4.79	0.67
Total	25.47	23.03	16.11	2.08
Panel B: Stock				
btm	0.78	1.91	3.04	3.37
size	0.23	0.40	0.76	1.30
dte	1.68	2.52	2.83	4.05
accr	1.74	0.77	1.96	1.31
s.ret	8.68	9.39	13.66	12.76
sec.ret	1.43	0.66	3.31	1.57
Total	14.54	15.65	25.57	24.37
Panel C: Macro				
gnp	22.03	17.53	18.12	24.26
infl	27.59	17.65	16.58	22.37
vix.ret	0.38	0.45	1.64	0.92
t.bill	9.99	25.69	21.98	26.02
Total	59.98	61.32	58.32	73.56

Table 3.8: Significance of dynamic states

This table shows the results of pairwise t-test of `static` state of the variables vs. the dynamic states. The null is that the difference in *DP* values is zero. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation.

Variable	`diff`		`random`		`roll.sd`	
	t value	$\Pr(>\lvert t\rvert)$	t value	$\Pr(>\lvert t\rvert)$	t value	$\Pr(>\lvert t\rvert)$
Panel A: Analyst						
uncert	9.31	0.00	15.88	0.00	7.61	0.00
assym	5.57	0.00	8.24	0.00	3.46	0.00
disp	2.53	0.01	9.74	0.00	1.61	0.11
Panel B: Stock						
btm	8.70	0.00	13.04	0.00	12.78	0.00
size	4.32	0.00	5.91	0.00	8.57	0.00
dte	8.61	0.00	6.80	0.00	4.61	0.00
accr	-3.16	0.00	3.63	0.00	-0.73	0.47
s.ret	11.46	0.00	22.24	0.00	17.68	0.00
sec.ret	-4.15	0.00	8.62	0.00	1.90	0.06
Panel C: Macro						
gnp	9.35	0.00	31.81	0.00	24.94	0.00
infl	-0.67	0.50	26.21	0.00	19.35	0.00
vix.ret	1.63	0.10	7.23	0.00	4.17	0.00
t.bill	40.58	0.00	29.88	0.00	43.02	0.00

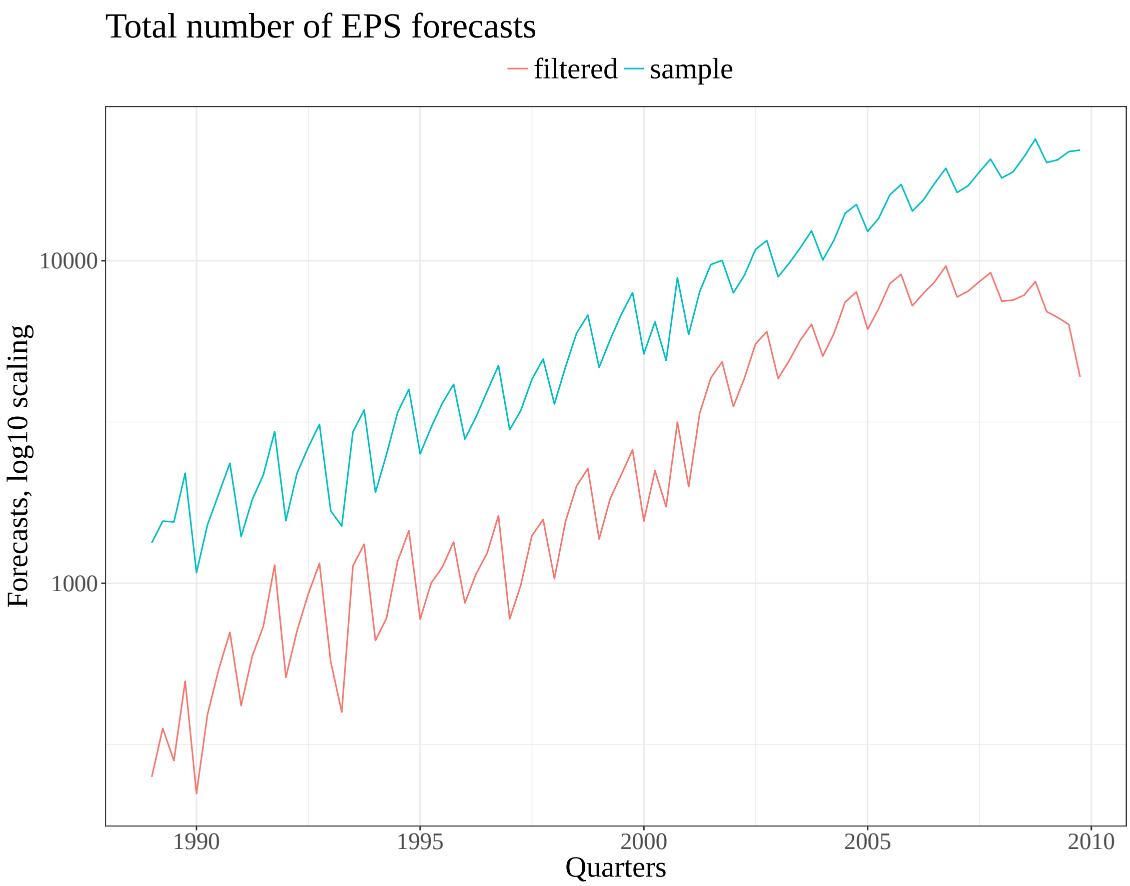

Figure 3.1: Total number of EPS forecasts.
The plot shows the log of total number of forecasts per quarter for *sample*d and *filtered* data sets.

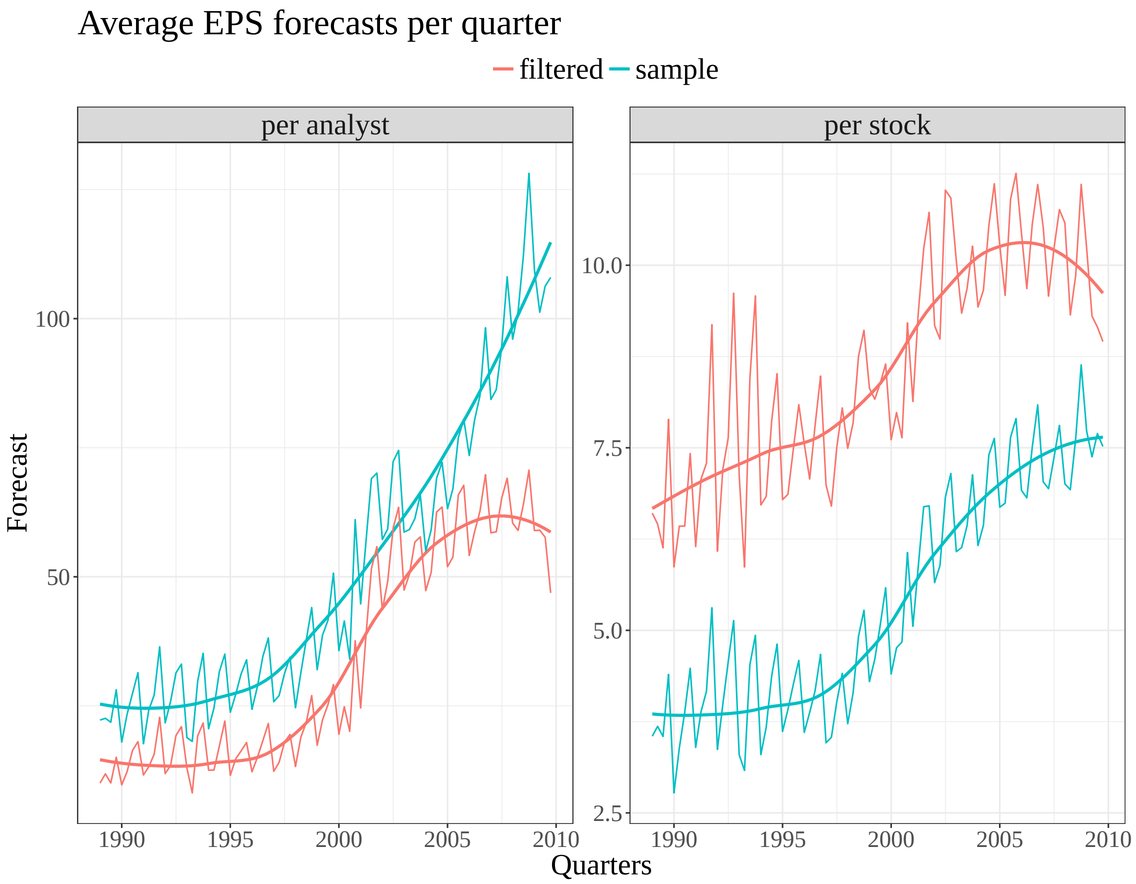

Figure 3.2: Average number of EPS forecasts

The plot depicts the average number of EPS forecasts per analyst (left panel) and per stock (right panel) for *sample* and *filtered* datasets.

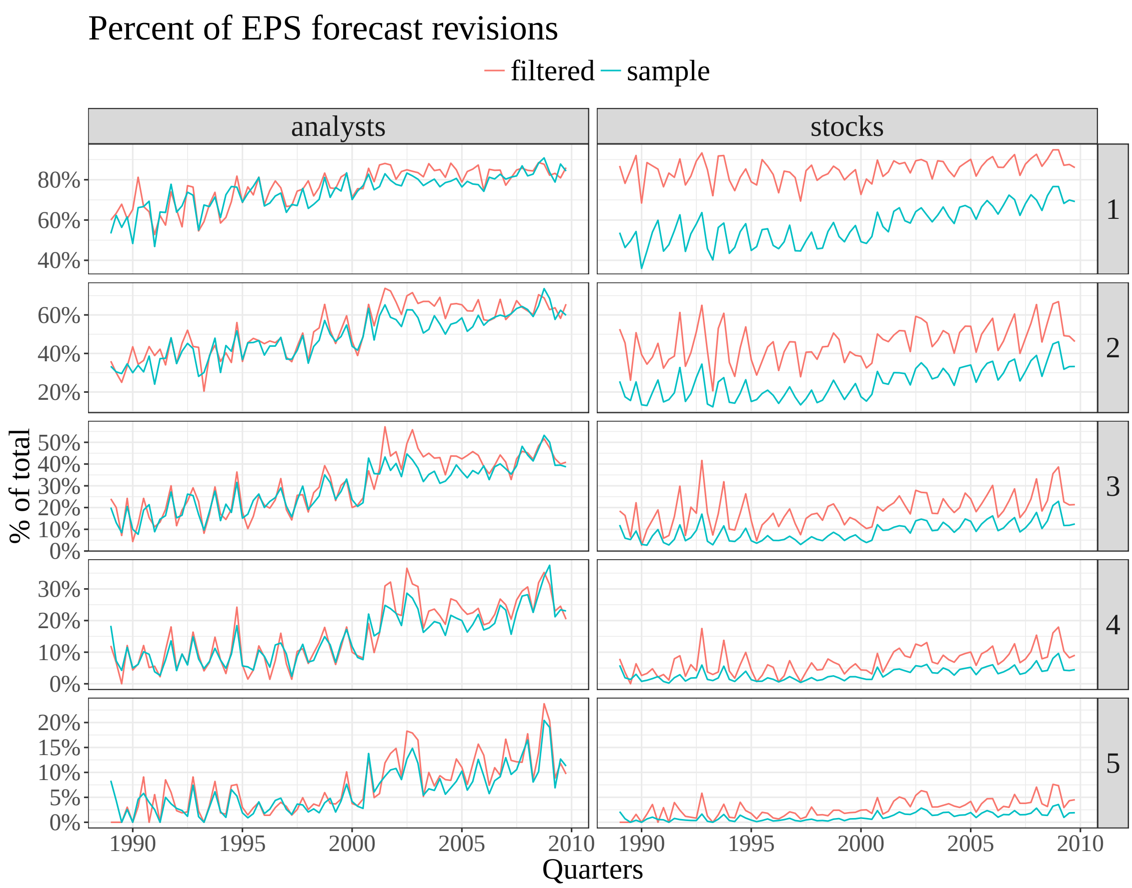

Figure 3.3: Revisions of forecasts

The plot shows the average percent of analysts (stocks) that revise (got revised) their forecasts for *sample* and *filtered* dataset. Horizontal panels shows the number of revisions per quarter from 1 revision per quarter (top panel) to 5 (bottom panel).

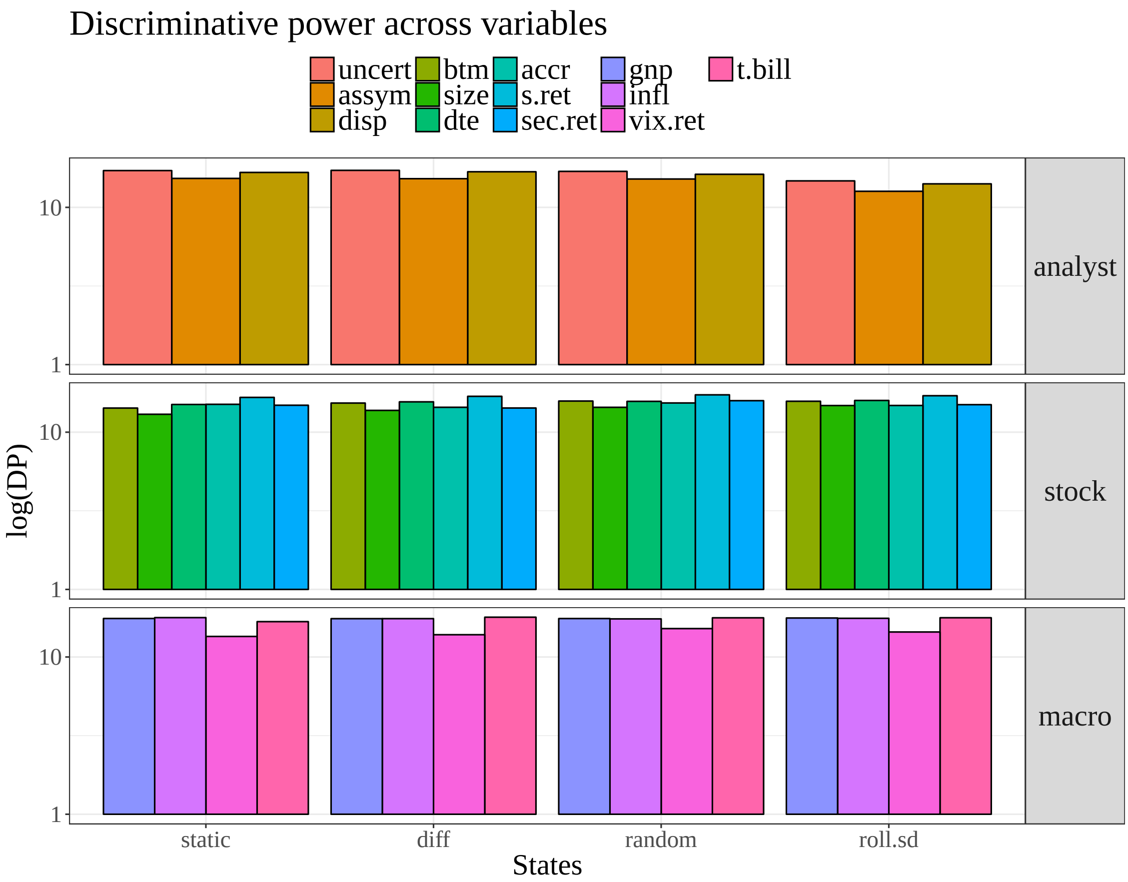

Figure 3.4: The discriminative power of the variables.
The plot depicts the discriminative power of variables. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of variable sliding 8 quarters standard deviation.

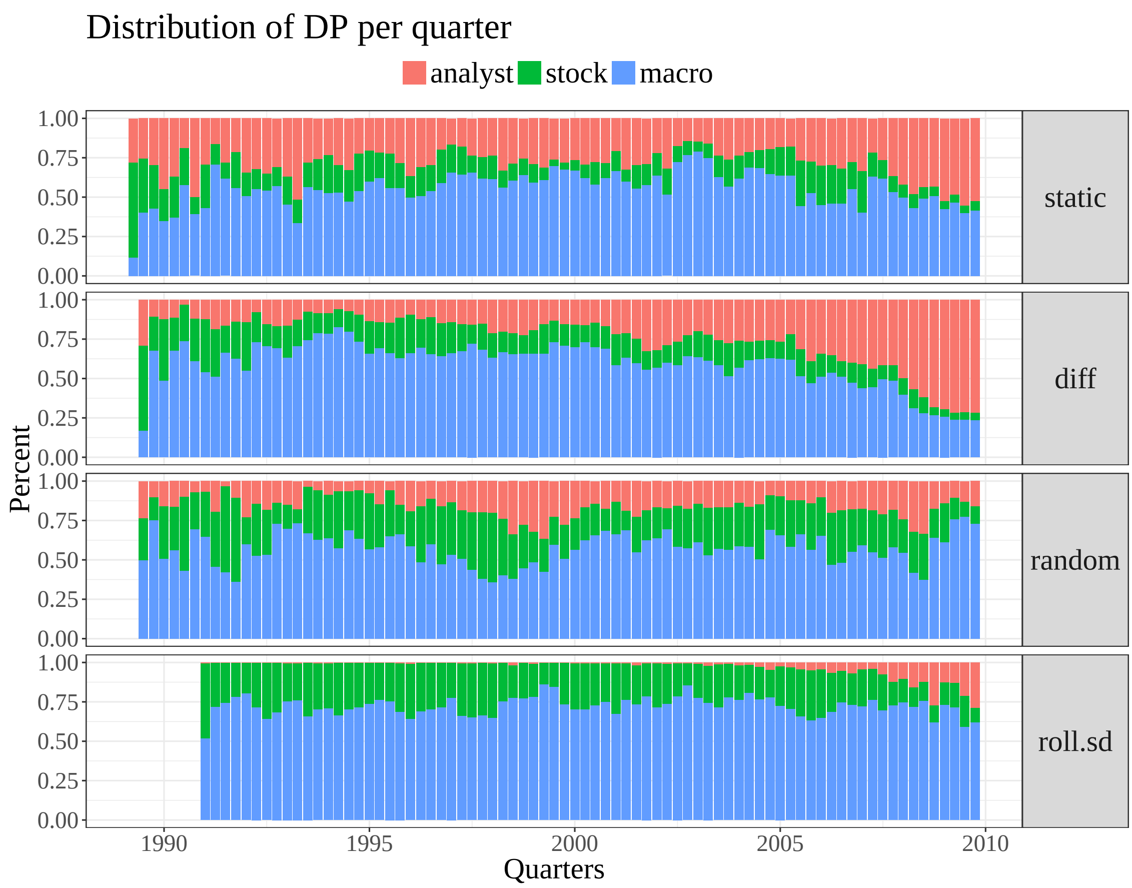

Figure 3.5: Distribution of discriminative power per quarter

The plot shows the composition of discriminative power across variable categories. The states are: `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of variable sliding 8 quarters standard deviation.

Chapter 4

Rankings of analysts as means to profits

Abstract

Financial analysts are evaluated based on the value they create for those who follow their recommendations. Some institutions use these evaluations to rank the analysts. Predictions of the most accurate analysts are typically modeled in terms of individual analyst's characteristics. The disadvantage of this approach is that these data are hard to collect. In this paper, we follow a different approach in which we characterize analysts' relative performance on the basis of state variables rather than individual characteristics. We apply a label ranking model to predict the rankings and use them to develop active trading strategies. The results show that it is possible to accurately model the relation between selected attributes and rankings of analysts. In addition, the strategy based on the predicted rankings applied to the S&P500 stocks generates higher returns than that based on the analysts' consensus.

4.1 Introduction

Rankings of financial analysts is not new in finance. Many agencies develop their procedures to evaluate analysts based on their performance either in forecasting or stock recommendations. Some institutions even hold a "Red Carpet" event to recognize the top analysts. On one hand, for market participants, the rankings may signal who is the best analysts. On the other hand, studies have shown that following the best analysts' recommendations of buy-sell stocks have statistically insignificant benefits.

In this paper we have an objective to show that rankings can serve as inputs for trading; that is, they can be a direct input for strategy. We base our research on the main assumption that analysts at the top ranks are the best analyst and they are worth to be followed. Naturally, instead of relying on personal expertise in selecting the expected returns, it is best to refer to the specialist in the field, namely, the financial analysts. Using the analysts' price target information we create a vector of expected returns and use it as a starting point for appropriate trading strategy.

Given the objective of the paper, we have to solve two problems. First, we need to predict the rankings of the analyst and, second, translate these rankings into an operational input for the trading strategy.

For the first problem, we take advantage of the algorithm used for prediction of rankings developed in Aiguzhinov et al. (2010) and adapt it for the case of analysts' rankings. In short, this algorithm is based on the Bayesian probability and the similarities between the rankings. The solution for the second problem relies on the Black-Litterman (BL) model (Black and Litterman, 1992). We are particularly confident in this choice of tools given that the BL model and ranking algorithm are both based on the Bayesian framework. Given the results form Aiguzhinov et al. (2015a), we base our rankings on analysts' target prices as it has been shown that strategies based on these rankings are the ones that yield the highest cumulative annualized return.

The paper is organized as follows: Section 4.2 provides motivation on use of rankings; Section 4.3 outlines label ranking algorithm; Section 4.4 summarizes the selection of the independent variables that affect the analysts' rankings; Section 4.5 describes a methodology of building the rankings; Section 4.6 discusses information environments that influence analysts' decisions; Section 4.7 outlines the steps of the trading strategy; Section 4.8 describes data used in the study; Section 4.9 discusses the results, and Section 4.10 concludes.

4.2 Rankings of financial analysts

In financial literature there has been a long debate on whether financial analysts produce valuable advice. Some argue that following the advice of financial analysts, translated as recommendations of buying, holding, or selling a particular stock, does not yield abnormal returns, i.e., returns that are above the required return to compensate for risk. The Efficient Market Hypothesis (Fama, 1970) states that financial markets are efficient and that any public available information regarding a stock would be immediately reflected in prices; hence, it would be impossible to generate abnormal returns based upon past information.

Yet, several authors have since stressed that there are information-gathering costs and information is not immediately reflected on prices (Grossman and Stiglitz, 1980). As such, prices may not reflect all the available information at all time because if this were the case, those who spent resources to collect and analyze information would not have an incentive to do it, because there would not get any compensation for it.

Many trading strategies try to forecast the price movements relying on the historical prices or estimate the intrinsic value of a company. Obviously, this type of research is associated with significant amount of up-front costs to acquire databases, software, etc. On the other hand, financial analysts have these tools and, presumably, skills to identify stocks that worth be invested. Thus, for an investor, it is cheaper to follow the recommendations of financial analysts rather than perform a proper stock market analysis.

Some authors show that financial analysts' recommendations create value to investors (Womack, 1996; Barber et al., 2001)[1]. Assuming that some analysts produce valuable advice it makes sense to rank analysts based on the accuracy of their recommendations.

StarMine rankings are based on financial analysts' accuracy either on TP or EPS forecasts. To rank analysts based on EPS forecasts, StarMine developed a proprietary metric called a Single-stock Estimating Score (SES). This score measures "... [a] relative accuracy; that is, analysts are compared against their peers. An analyst's SES can range from 0 to 100, with 50 representing the average analyst. To get a score higher than 50, an analyst must make estimates that are both significantly different from and more accurate than other analysts' estimates"[2].

As for target price ranking, StarMine's methodology compares the portfolios based on analysts recommendations. Portfolios are constructed as follows. For each "Buy" recommendation, the portfolio is one unit long the stock and simultaneously one unit short the benchmark. "Strong buy" gets a larger investment of two units long the stock and two units short the benchmark. "Hold" invests one unit in the benchmark (i.e., an excess return of zero). "Sell" recommendations work in the reverse way. StarMine re-

[1] Womack (1996) finds that post-recommendation excess returns are not mean-reverting, but are significant and in the direction forecast by the analysts. Barber et al. (2001) finds that over the period of 1986-1996 a portfolio of stocks with the most (least) favorable consensus analyst recommendations yields an average abnormal return of 4.13 (-4.91)%.

[2] http://excellence.thomsonreuters.com/award/starmine?award=Analyst+Awards&award_group=Overall+Analyst+Awards

balances its calculations at the end of each month to adjust for analysts revisions (adding, dropping or altering a rating), and when a stock enters or exits an industry grouping.

Recent evidence suggests that top ranked financial analyst affect market participants: prices seem to react more to the recommendations issued by the top-ranked analysts (Emery and Li, 2009). As such, StarMine ranking based models can be used to identify such analysts and generate superior estimates (e.g., SmartEstimates[3]).

The goal of our study is to predict StarMine rankings. With this purpose, we adapt a Machine Learning algorithm to predict rankings given a set of variables that characterize these rankings. We, further, apply the predicted rankings to build active trading strategies to evaluate quality of predictions against the consensus strategy (giving equal weights to analysts' recommendations).

4.3 Label ranking algorithm

The classical formalization of a label ranking problem is the following (Vembu and Gärtner, 2010). Let $\mathscr{X} = \{\mathscr{V}_1, \ldots, \mathscr{V}_m\}$ be an instance space of variables, such that $\mathscr{V}_a = \{v_{a,1}, \ldots, v_{a,n_a}\}$ is the domain of nominal variable a. Also, let $\mathscr{L} = \{\lambda_1, \ldots, \lambda_k\}$ be a set of labels, and $\mathscr{Y} = \Pi_{\mathscr{L}}$ be the output space of all possible total orders over $\mathscr{L}$ defined on the permutation space Π. The goal of a label ranking algorithm is to learn a mapping $h : \mathscr{X} \to \mathscr{Y}$, where h is chosen from a given hypothesis space $\mathscr{H}$, such that a predefined loss function $\ell : \mathscr{H} \times \mathscr{Y} \times \mathscr{Y} \to \mathbb{R}$ is minimized. The algorithm learns h from a training set $\mathscr{T} = \{x_i, y_i\}_{i \in \{1,\ldots,n\}} \subseteq \mathscr{X} \times \mathscr{Y}$ of n examples, where $x_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,m}\} \in \mathscr{X}$ and $y_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,k}\} \in \mathscr{Y}$. With time-dependent problem in rankings, we replace the i index with t; that is $y_t = \{y_{t,1}, y_{t,2}, \ldots, y_{t,k}\}$ is the ranking of k labels at time t described by $x_t = \{x_{t,1}, x_{t,2}, \ldots, x_{t,m}\}$ at time t.

Consider an example of a time-dependent ranking problem presented in Table 4.3. In this example, we have three brokers ($k = 3$), four independent variables ($m = 4$) and a period of 7 quarters. Our goal is to predict the rankings for period t, given the values of independent variables and rankings known up to period $t-1$; that is, to predict the ranking for time $t = 7$, we use $n = 6$ ($t \in \{1 \ldots 6\}$) examples to train the ranking model.

[3]http://www.starmine.com/index.phtml?page_set=sm_products&sub_page_set=sm_professional&topic=analytical&section=accurate_estimates

4.3.1 Naive Bayes algorithm for label ranking

The naive Bayes for label ranking (NBLR) will output the ranking with the higher $P_{LR}(y|x)$ value (Aiguzhinov et al., 2010):

$$\begin{aligned}\hat{y} &= \arg\max_{y \in \Pi_{\mathcal{L}}} P_{LR}(y|x) = \\ &= \arg\max_{y \in \Pi_{\mathcal{L}}} P_{LR}(y) \prod_{i=1}^{m} P_{LR}(x_i|y)\end{aligned} \tag{4.1}$$

where $P_{LR}(y)$ and $P_{LR}(x_i|y)$ are the prior and conditional label ranking probabilities af a nominal variable x of attribute a, (v_a), respectively, and they are given as follows:

$$P_{LR}(y) = \frac{\sum_{i=1}^{n} \rho(y, y_i)}{n} \tag{4.2}$$

$$P_{LR}(x_i|y) = \frac{\sum_{i:x_a=v_a} \rho(y, y_i)}{|\{i : x_a = v_a\}|} \tag{4.3}$$

where $\rho(y, y_i)$ is the similarity between rankings obtained from the Spearman ranking correlation:

$$\rho(y, y_i) = 1 - \frac{6\sum_{j=1}^{k}(y - y_{i,j})^2}{k^3 - k} \tag{4.4}$$

Similarity and probability are different concepts; however, a connection as been established between probabilities and the general Euclidean distance measure (Vogt et al., 2007). It states that maximizing the likelihood is equivalent to minimizing the distance (i.e., maximizing the similarity) in a Euclidean space. The predicted ranking for an example x_i is the one that will receive the maximum posterior label ranking probability $P_{LR}(y|x_i)$.

4.3.1.1 Continuous independent variables

In its most basic form, the naive Bayes algorithm cannot deal with continuous attributes. The same happens with its adaptation for label ranking (Aiguzhinov et al., 2010). However, there are versions of the naive Bayes algorithm for classification that support continuous variables (Bouckaert, 2005). The authors modify the conditional label ranking probability by utilizing the Gaussian distribution of the independent variables. We apply the same approach in defining the conditional probability of label rankings:

$$P_{LR}(x|y) = \frac{1}{\sqrt{2\pi}\sigma(x|y)} e^{\frac{(x-\mu(x|y))^2}{2\sigma^2(x|y)}} \tag{4.5}$$

where $\mu(x|y)$ and $\sigma^2(x|y)$ weighted mean and weighted variance, defined as follows:

$$\mu(x|y) = \frac{\sum_{i=1}^{n} \rho(y, y_i) x}{\sum_{i=1}^{n} \rho(y, y_i)} \tag{4.6}$$

$$\sigma^2(x|y) = \frac{\sum_{i=1}^{n} \rho(y, y_i)[x - \mu(x|y)]^2}{\sum_{i=1}^{n} \rho(y, y_i)} \tag{4.7}$$

4.3.2 Time series of rankings

The time dependent label ranking (TDLR) problem takes the inter-temporal dependence between the rankings into account. That is, rankings that are similar to the most recent ones are more likely to appear. To capture this, we propose the weighted TDLR prior probability:

$$P_{TDLR}(y_t) = \frac{\sum_{t=1}^{n} w_t \rho(y, y_t)}{\sum_{t=1}^{n} w_t} \tag{4.8}$$

where $w = \{w_1, \ldots, w_n\} \rightarrow \mathbf{w}$ is the vector of weights calculated from the exponential function $\mathbf{w} = b^{\frac{1-\{t\}_1^n}{t}}$. Parameter $b \in \{1 \ldots \infty\}$ sets the degree of the "memory" for the past rankings, i.e., the larger b, the more weight is given to the last known ranking (i.e, at $t-1$) and the weight diminishes to the rankings known at $t = 1$.

As for the conditional label ranking probability, the equation for the weighted mean (Equation (4.5)) becomes:

$$\mu(x_t|y_t) = \frac{\sum_{t=1}^{n} w_t \rho(y, y_t) x_t}{\sum_{t=1}^{n} \rho(y, y_t)} \tag{4.9}$$

$$\sigma^2(x_t|y_t) = \frac{\sum_{i=1}^{n} w_t \rho(y, y_t)[x_t - \mu(x_t|y)]^2}{\sum_{i=1}^{n} \rho(y, y_t)} \tag{4.10}$$

4.4 State variables

Several studies try to analyze factors that affect the performance of analysts (Clement, 1999; Brown and Mohammad, 2003; Jegadeesh et al., 2004). However, most of these papers look at the individual characteristics of analysts such as their job experience, their affiliation, education background, industry specializations, etc. These variables are very important to characterize the relative performance of the analysts in general but they miss the "state-of-the-world" component, i.e., variables that all analysts are affected. We believe that rankings of analysts capture this component in full.

Ranking means that there are differences in opinion among the analysts concerning the future performance of a company. This implies that there is a variability (dispersion) in analysts' forecasts for a given stock in a given quarter (Diether et al., 2002). Thus, we can analyze the analysts' forecasts dispersion in terms of its origin and factors that affect it; hence, assuming the same variables influence the rankings. It follows that the variation in rankings is due to the different ability of analysts to interpret informational from different states of the environment (e.g., whether the market is bull or bear). We, thus, select and analyze variables that describe these states of environment.

To capture the full spectrum of the analyst's decision making process, we select variables based on different levels of information availability: analyst-specific, firm-specific and general economy. In each level, we want a variable to be responsible for information asymmetry and uncertainty. Thus, we believe that these two domains are responsible for the differences in analysts' opinions.

4.4.1 Analysts-specific variables

On analysts' level, we want to capture the asymmetry and uncertainty among the analysts (Barron, Kim, Lim, and Stevens, 1998; Barron et al., 2009; Zhang, 2006; Sheng and Thevenot, 2012). Particularly, Barron et al. (2009) point our that the reason for dispersion is either uncertainty or information asymmetry. They find that prior to earnings announcement the uncertainty component prevail, whereas around the time of earnings announcement, information asymmetry is responsible for changes in analysts' opinions.

To capture the states of the dispersion, we use the same set of variables defined in (Barron et al., 2009, page 333):

$$
\begin{aligned}
SE &= (ACT - \overline{FE})^2 \\
disp &= \sum_{j=1}^{k} \frac{(FE_j - \overline{FE})^2}{(k-1)} && (4.11) \\
uncert &= \sum_{j=1}^{k} \left(1 - \frac{1}{k}\right) \times disp + SE && (4.12) \\
assym &= 1 - \frac{SE - \frac{disp}{k}}{uncert} && (4.13)
\end{aligned}
$$

where SE is the squared error in mean forecast; $\overline{FE}$ is the average per analysts EPS forecast error (see Section 4.5.1); and k is the number of analysts in a given quarter for a given stock.

Equation (4.11) calculates the dispersion among the analysts which is a variance of EPS forecasts of all analysts for a given stock. Equation (4.12) defines the Uncertainty component of the dispersion per Barron et al. (2009). Equation (4.13) is the proxy for information asymmetry which a function of dispersion, squared mean error, and a number of EPS forecasts.

4.4.2 Firm-based variables

To be consistent with the two paradigms that characterize the state of the analysts, we split the firm-based variables based on their influence on analysts' opinions.

4.4.2.1 Uncertainty

The following are the set of the variables and their definitions that we think are responsible for the uncertainty component.

4.4.2.1.1 Business risk.

Business risk is associated with the uncertainty in operating results, especially, in operating earnings (Hill and Stone, 1980). An increase in business risk entails an increase in *ex-ante* volatility of the reported earnings (Parkash et al., 1995). We select book-to-market ratio as a proxy for the business risk measurement.

$$btm = \frac{\text{EQUITY}}{\text{MKT.CAP}} = \frac{\text{Tot.assts} - \text{Tot.liab}}{\text{Stocks} \times \text{Price}} \tag{4.14}$$

where Stocks is the number of stocks outstanding and Price is the close stock price on last day of a quarter.

4.4.2.1.2 Financial risk.

Financial risk is responsible for the uncertainty of the future earnings. More debt implies more variability in earnings as managers would try to maximize the value of a stock using the debt; thus, having high risk of default in the future or taking high risk investment projects. The debt-to-equity ratio is used to capture the financial risk (Parkash et al., 1995). We use short-term debt from balance sheet (Notes payable) as a measure for debt.

$$dte = \frac{\text{DEBT}}{\text{EQUITY}} = \frac{\text{ShortTermDebt}}{\text{Tot.assts} - \text{Tot.liab}} \tag{4.15}$$

4.4.2.1.3 Size.

The firm size can be used as a proxy for amount of information available for a firm. Thus, larger firm has more news coverage which reduces uncertainty. An investor is likely to find private information about larger firm more valuable than the same information about smaller firm (Bhushan, 1989).

Size is measured as the market value (MV) of the firm as follows:

$$size = \log(\text{Price} \times \text{Stocks}) \tag{4.16}$$

Consistent with the literature, we use log of market value.

4.4.2.1.4 Return variability.

Return variability influence the uncertainty regarding future earnings (Diether et al., 2002; Henley et al., 2003). An increase in variability of the abnormal returns is positively correlated with the uncertainty about the earnings; thus, affecting the dispersion among the analysts. To calculate the return variability, we use method provided in Sousa and Serra

(2008), where stock return volatility is decomposed into market and stock specific components as follows:

$$
\begin{aligned}
\sigma^2_{mkt} &= \sum_{d\in t}(R_{mkt,d} - \mu_{mkt})^2 \\
\sigma^2 &= \sum_{d\in t}(R_d - R_{mkt,d})^2 \\
s.ret = Var(R_t) &= \sigma^2_{mkt} + \sigma^2
\end{aligned}
\tag{4.17}
$$

where $R_{mkt,t}$ is the market return over sample period; μ_{mkt} is the mean of daily market returns; R_t is an individual stock return; d is the number of trading days in period t.

4.4.2.2 Information Asymmetry variables

4.4.2.2.1 Accruals.

Accruals, as a part of earnings, is one of the variables that cause the information asymmetry between managers of a firm and investors. Studies have shown that presence of asymmetry is a necessary condition for the earnings management (Trueman and Titman, 1988; Richardson, 2000). To be more specific, it is the discretionary part of the accruals that causes the information inefficiency in the earnings management (Richardson, 2000; Ahmed et al., 2005). We calculated total accruals-to-total assets ratio defined in Creamer and Stolfo (2009):

$$
accr = \frac{\Delta\text{C.As} - \Delta\text{Cash} - (\Delta\text{C.Lb.} - \Delta\text{C.Lb.D}) - \Delta\text{T} - \text{D\&A}_t}{(\text{T.As.} - \text{T.As.}_{t-4})/2}
\tag{4.18}
$$

where $\Delta X = X_t - X_{t-1}$; C.As – current assets; C.Lb – current liabilities; C.Lb.D – debt in current liabilities; T – deferred taxes; D&A – depreciation and amortization; and T.A – total assets.

4.4.2.2.2 Sector-based variables.

The industry specific variables that cause the dispersion in the analysts' forecasts are connected with the uncertainty concept. One of the variables that is suggested to capture is the variability in the industry Producer Price Index (PPI) (Henley et al., 2003).

$$sec.ret = \sigma(\log \mathrm{PPI}_{sec}) \tag{4.19}$$

where $\sigma(\log \mathrm{PPI}_{sec})$ is the standard deviation of the log of SIC sectors' produce price index.

4.4.3 Macroeconomics variables

In the last set of the state variables, we want to capture the macroeconomic conditions which affect the analysts' dispersion. For example, different states of the economy are based on different levels of "GNP–inflation" combinations (Lev and Thiagarajan, 1993; Hope and Kang, 2005). When economy is booming, i.e. "high GNP-low inflation" state, Lev and Thiagarajan (1993) observe the significant increase in firms' Capital Expenditures coefficient. This implies that firms start enjoy capital investment due to the low cost of capital. This state of the economy produces less uncertainty. In the "medium GNP-high inflation" state of the economy, there is an increase in R&D expenditures, which, from the above mentioned analysis, may spur high level of information asymmetry based on the increase R&D activities. Finally, in the "low GNP-high inflation" state, Lev and Thiagarajan (1993) observe the Doubtful Receivables coefficient is the largest implying that at this recession state many firms go bankrupt or default on the loans – a signal of high uncertainty in the economy. All these states produce the dispersion of the analysts' forecasts.

We select the following set of the macroeconomic variables:

- gnp = Gross National Product;
- $infl$ = Inflation rate;
- $t.bill$ = Interest rate (90-days T-bill rate);
- $vix.ret$ = Market variability (CBOE VIX index)

4.5 Target rankings of financial analysts

Analysts are ranked on the basis of Proportional Mean Absolute Forecast Error (PMAFE) that measures the accuracy of a forecast (Clement, 1999; Brown, 2001; Ertimur et al., 2007). We use both target price and EPS accuracy to build the rankings.

4.5.1 Target Price ranking

We define the forecast daily error FE_j as the absolute value of the difference between analyst' target price TP_j and the daily stock price P for each stock:

$$FE_j^{TP} = |P - TP_j| \tag{4.20}$$

The PMAFE is given as:

$$PMAFE_j^{TP} = \frac{FE_j^{TP}}{\overline{FE^{TP}}} \tag{4.21}$$

where $\overline{FE^{TP}}$ is the average forecasting error across analysts. The target price is fixed over the quarter unless it gets revised.

The rank is average analyst's $PMAFE^{TP}$ over a particular quarter:

$$\overline{PMAFE_j^{TP}} = \frac{1}{d}\sum_{i=1}^{d} PMAFE_{j,i}^{TP} \tag{4.22}$$

$$rank_j = \mathrm{rank}_{j=1}^{k}\left\{\overline{PMAFE_j^{TP}}\right\} \tag{4.23}$$

4.5.2 EPS ranking

To compute the EPS rankings, we apply the same procedure as above:

$$FE_j^{EPS} = |ACT - PRED_j| \tag{4.24}$$

$$PMAFE_j^{EPS} = \frac{FE_j^{EPS}}{\overline{FE^{EPS}}} \tag{4.25}$$

$$rank_j = \text{rank}_{j=1}^{k} \left\{ PMAFE_j^{EPS} \right\} \tag{4.26}$$

where ACT and $PRED_j$ are the actual quarterly EPS and analyst j's EPS forecast for stock.

4.6 Analysts' information environment

To proceed with the ranking prediction, we need to establish which information we will be using to initially rank analysts.

4.6.1 Past information sets

Different analysts' ranks are obtained if we select different time horizons. If we use only the most recent information, we will capture the recent performance of the analysts. This, of course, is more sensitive to unique episodes (e.g., a quarter which has been surprisingly good or bad). If, alternatively, we opt to incorporate the entire analyst performance, the ranking is less affected by such events, yet it may not reflect the current analyst ability. We use two information sets: the first uses only the information about the analyst' performance in period $t-1$; the second, uses all the available information for that particular analyst. We call the former the *recent* rankings and the latter the *all-time* rankings.

In addition to these rankings, we also create a hypothetical scenario that assumes we anticipate perfectly the future analyst accuracy performance that would only be available at the end of t. We call this the *true* rankings.

Formalizing information sets considered are:

- the *true* rankings

$$\text{rank}_{j,t} = \text{rank}_{j,t} \tag{4.27}$$

- the *recent* rankings

$$\text{rank}_{j,t} = \text{rank}_{j,t-1} \tag{4.28}$$

- the *all-time* rankings

$$\text{rank}_{j,t} = \frac{1}{T} \sum_{t=1}^{T} \text{rank}_{j,t} \tag{4.29}$$

where $\text{rank}_{j,t}$ is analyst j rank at time t. These rankings will serve as baselines to assess the quality of the predicted rankings.

4.6.2 Dynamic states

For ranking predictions, the past information sets are no longer valid as we model the variables that affect analysts' performance with *true* rankings. For this reason, it is necessary to take into account the dynamics of independent variables from one time period to another. We propose the following methods:

- `static`: no dynamics in the state of the variables, i.e., independent variables used as they are: $x_{\Delta t} = x_t$;
- `diff`: first-difference of the variables, i.e., $x_{\Delta t} = x_t - x_{t-1}$;
- `random`: in time series decomposition of the independent variables, it is an unobserved component: $x_{\Delta t} = T(t) + S(t) + \varepsilon(t)$, where $T(t)$- trend, $S(t)$ - seasonal part and $\varepsilon(t)$ - random part of time series decomposition.
- `roll.sd`: rolling 8 quarters standard deviation of the independent variables (Zivot and Wang, 2003):

$$\begin{aligned} \mu_t(8) &= \frac{1}{8}\sum_{i=0}^{7} x_{t-i} \\ \sigma_t^2(8) &= \frac{1}{7}\sum_{i=0}^{7}(x_{t-i} - \mu(8))^2 \end{aligned} \tag{4.30}$$

Each of these methods produces a different set of attributes which corresponds to different predicted rankings. Overall, in each time period we would have seven different rankings: one is the perfect foresight ranking (*true*), two are based on the different sizes of analysts' past information (*recent* and *all-time*), and the rest are from the predicted model with different dynamic states (`static`, `diff`, `random`, and `roll.sd`).

We selected variables that describe the information environment consistent with Aiguzhinov, Serra, and Soares (2015b). We use variable that have more than 10% contribution to the rankings. Table 4.2 demonstrates the total discriminative power of state variables for different states for EPS ranking. The variables that contribute the most to the rankings are: {*uncert*; *disp*; *s.ret*; *gnp*; *infl*; *t.bill*} .

4.7 Trading Strategy

We follow the Black-Litterman procedure developed in Aiguzhinov et al. (2015a):

1. For each stock, at the beginning of quarter t, we use predicted rankings of all analysts that we expect to be at the end of the quarter t;

2. Based on these predicted rankings and analysts' price targets, we define Q_t and Ω_t (see (Section 4.7.1) and (Section 4.7.2));

3. Using market information available at the last day of quarter $t-1$, we obtain the market inputs;

4. Apply BL model to get optimized portfolio weights and buy/sell stocks accordingly;

The model requires form an investor two inputs: the vector of expected returns and the confidence of these returns. The vector of returns is where we rely on the knowledge of the analysts. We use two types of expected returns: 1) ones that are based on the consensus among analysts about future stock performance; 2) ones that are based on the rankings of analysts.

4.7.1 Defining Q

For the consensus strategy, we use median of expected returns for a particular stock:

$$Q_{cons} = \text{median}\left\{r_j\right\} \tag{4.31}$$

where $r_j = TP_j/P - 1$ is last known analyst's j expected return computed using the analyst price target TP_j and stock price P[4].

For the strategies that weight the analysts' estimates of expected return the weight of each analyst j is based on his/her rank such that the top analyst has the weight of 1 and then the weights diminish as the rank increases.

$$w_j = 1 - \frac{\text{rank}_j - \min\{\text{rank}\}}{\max\{\text{rank}\}} \tag{4.32}$$

[4]Consistent with the literature, we use stock price 3 days *ex-ante* the TP announcement. This is done to avoid any information leakage around new TP announcement day (Bonini et al., 2010)

where rank_j is the predicted analyst j rank (Section 4.6)

The expected rank-weighted return is thus:

$$Q_{rank} = \frac{\sum_{j=1}^{k}(w_j \times r_j)}{\sum_{j=1}^{k} w_j} \tag{4.33}$$

4.7.2 Defining the confidence of expected returns Ω

The confidence of Q is given by the coefficient of variation (CV) of forecasting errors:

$$\text{CV} = \frac{\sigma(FE)}{\overline{FE}} \tag{4.34}$$

where σ and $\overline{FE}$ are the standard deviation and the mean of the forecast errors across analysts for TP. A low value of CV reflects consensual estimates of future prices.

4.8 Data and experimental setup

4.8.1 Database and sample

We focus our sample on the S&P500 stocks. The period of the experiments runs from the first quarter of 2001 until the last quarter of 2009. We get the analysts price target and EPS forecast data from ThomsonReuters I/B/E/S dataset; the list of S&P constituents and stock daily prices data are from DataStream as well as the market capitalization data.

Over the sample period, the total number of Equity Research Firms (ERF)[5] in TP dataset is 477, covering 502 stocks. Given the fact that financial analysts commonly issue TP with the one year horizon[6], we assume that analysts keep their TP forecasts valid for one calendar year unless it is revised. After one year we assume that TP recommendation expires.

Consistent with other studies on analysts' expected returns that work with price targets (Bradshaw, 2002; Brav and Lehavy, 2003; Da and Schaumburg, 2011), we truncate the sample of $TP/P-1$ at the 5$^{\text{th}}$ percentile (values below -0.14) and at the 95$^{\text{th}}$ percentile (values above 0.99). This is done due to occurrence of the extreme values. Most of these

[5] We use words "analyst" and "Equity Research Firm" interchangeably.

[6] According to Wharton Research Data Services (WRDS), 92.33% of all price targets reported in I/B/E/S have a 12-month horizon (Glushkov, 2009).

extreme values are driven by misalignment errors found on I/B/E/S data[7]. To implement ranking, we require that a stock had at least three equity research firms per quarter and that a equity research firm has to be active in covering a particular stock for at least 3 years (12 quarters). After all the data requirements, our final sample number of equity research firms issued target prices is 152 covering 419 S&P500 stocks. Overall, the number of observations (Stock $\times$ ERF $\times$ Quarter) is reduced from 134336 (initial) to 90743 (filtered).

In the case of EPS forecasts, the initial file of quarterly EPS forecast consists of 437 ERFs covering 516 stocks. Considering the ranking data requirement, our final sample of EPS forecasts consist of 157 ERFs covering 402 S&P500 stocks. The total number of observation is 80185.

Table 4.4 presents descriptive statistics of the price targets (panel A) and EPS forecasts (panel B). We observe that, on average, the analysts issue 5.52 and 5.6 of price targets and EPS forecasts per quarter respectively.

4.8.2 Ranking contingency results

We check for analysts' ranking consistency as follows. In one particular quarter (t), we place analysts at one of the bins which corresponds to a tercile: *top*, *medium*, *bottom*. We, then, check analysts position at the immediate next quarter ($t+1$) and after one year ($t+4$).

Beforehand, we convert the rankings into scores as follows:

$$\text{score}_j = \frac{\text{rank}_j}{\max \text{rank}} \tag{4.35}$$

To get the cross-sectional values of scores across different stocks, we take the average of score_j

$$\overline{\text{score}_j} = \frac{1}{M}\sum_{i=1}^{M} \text{score}_{j,i} \tag{4.36}$$

where M is number of stocks followed by a particular analyst j.

Table 4.1 shows a contingency analysis of the ranks. Panel A shows the dynamics of each tercile for rankings based on target price accuracy for the *recent* and the *all-time* rankings. We observe that analysts exhibit strong ranking consistency as, on average,

[7] We found some differences between the DataStream and I/B/E/S the databases. In some cases the stock-splits and the dividends were not properly adjusted.

they stay at the same tercile after one quarter. For the *recent* case, of the *top* (*bottom*) most accurate (inaccurate) analysts in the previous quarter 67.79% (69.69%) remain in that same tercile after one quarter. After one year the corresponding figures are lower respectively 46.04% and 41.08% for the *top* and *bottom* terciles. In case of the *all-time*, the analyst consistency is even more profound with 92.35% (92.13%) of analysts that stayed on *top* (*bottom*) in previous quarter remained in the same tercile in the next quarter. Even after one year, the consistency does not change much with 81.76% of analysts stayed on *top* and 79.6% remained at the *bottom*.

In the case of EPS (panel B), for the *recent* ranking 67.87% and 54.86% (51.88% and 35.29%) of the analysts remained in the *top* and *bottom* terciles, respectively, after one quarter (year). For the case of *all-time* ranking, 67.19% and 53.03% (53.1% and 33.71%) of the analysts stayed on *top* and *bottom* respectively after one quarter (year).

These results are consistent with the recent findings of Hilary and Hsu (2013) on analyst forecast consistency.

4.8.3 Views: descriptive statistics

Table 4.5 presents the descriptive statistics of the analysts' expected returns. The expected returns are computed comparing TP estimates with actual prices. To form the smart strategies we compute rank-weighted estimates where weights are given by the TP rankings.

Bradshaw (2002) reports analyst average expected returns for the period of 2000–2009 and 206 ERFs of 24%. Da and Schaumburg (2011) report an average expected return of 40% for the period of 1996–2004. Zhou (2013) finds an average expected return of 96% for the sample period of 2000–2009. These figures suggest that analysts are overly optimistic.

Panel A of Table 4.5 show the statistics for the consensus expectations as defined in Equation (4.31). As mentioned above in Section 4.7.1, the consensus views have equal weights among the analysts, regardless of their ranks; thus, for the cases of *true*, *recent*, and *all-time*, the median is the same regardless of knowing or not the present or past rankings (Q_{cons} in Equation (4.31)). As such, the mean, median, and standard deviation are the same and independent of analysts' information environment. However, since views also include the confidence (Equation (4.34)), which is based on analysts past performance, the results of the trading strategy based on consensus expectations will be different for the *recent* and the *all-time* information sets.

Panel B of the table shows the TP accuracy weighted average expected returns. For each information environment (*true*, *recent*, *all-time*, `static`, `diff`, `random`, and `roll.sd`) the average expected return is respectively 0.15%, 0.16%, 0.12%, 0.14%, 0.14%, 0.14%, and 0.14%.

4.9 Empirical Results

4.9.1 Ranking predictions

We report the ranking prediction results in Table 4.6. Panel A (B) presents annual and total ranking accuracy measured as the average Spearman rank correlation between the *true* and all other label ranking methods obtained from the analysts' target price (EPS) forecasts. Specifically, the table shows if the label ranking model can better predict rankings than a "no-model" setup. In panel A, we observe that predicted rankings based on target prices are constantly outperformed the *all-time* baseline but fail the *recent*. Thus, our model can predict more accurately rankings of those analysts' whose relative performance in setting target prices is based on the whole history.

The average accuracy of the predicted rankings based on the EPS forecasts (panel B) demonstrates quite different results. We report that rankings predicted with the `diff` state of the variables are more accurate than those of the baselines. Thus, our label ranking model can predict rankings of analysts' who issue EPS forecast with higher accuracy than those obtained not just from the *all-time* analysts' past performance information (as in case of the target prices) but also from the *recent* set.

Figure 4.1 depicts the reported average Spearman correlations. We apply the local polynomial regression fitting (Cleveland, Grosse, and Shyu, 1992) to smooth the series for a better presentation. The top panel of Figure 4.1 plots the average accuracy of rankings based on the price targets. We observe the accuracy of rankings based on the *recent* information are constantly above all others; only by the end of the sample period it begins to drop. The *all-time* case, on the contrary, drops rapidly at the beginning of the sample period and continues to decrease in value until the end of the period. The accuracy of predicted rankings falls in between of the *recent* and the *all-time* cases. The cases of `diff` and `random` show a constant increase in accuracy starting from 2005-2006 and, by the end of the sample period, the `random` almost reaches the level of the accuracy of *recent* case while the `diff` is still in the upward trend.

The bottom panel of Figure 4.1 plots the case of the EPS based rankings. Looking at baselines rankings, i.e., rankings that obtained without label ranking model, we observe that for the first half of the sample period (from 2000 until 2004) the *recent* rankings demonstrate a constant increase in the average ranking accuracy. This means that during this period, rankings of the analysts, on average, did not change from one quarter to another. Starting from 2004, the average accuracy of the *recent* rankings begins to decrease until the end of the sample period. A somewhat interesting pattern shows the average accuracy of the *all-time* ranking: similar to the *recent*, it also has been increasing until 2005 but running below the *recent*. However, after 2005, the *all-time* ranking accuracy outran the *recent*. Moreover, in times of the financial crisis of 2007-2009, the *all-time* demonstrated an increasing trend in ranking accuracy. As far as the predicted rankings are concerned, we observe that the `diff` average accuracy has been constantly above the other methods and only during the period of crisis it dropped significantly. In fact, except for the `random`, all predicted methods reach the maximum average accuracy at around 2007, a pre-crisis time, and all, except for the *all-time*, show a downward trend afterwards.

Overall, the results of experiment of the predicting the rankings show that it is possible to model the independent variables under different dynamic states with the rankings. Moreover, the label ranking model can predict the rankings that outperform, in terms of average accuracy, the ones obtained from the past analysts' relative performance. Concretely, we report that the predicted rankings based on the EPS forecasts outperform the baselines under the `diff` dynamic state of independent variables.

4.9.2 Trading strategies

We perform a back-test of trading strategy consistent with Aiguzhinov et al. (2015a); namely, we build the "smart estimates" from rankings based on the price targets. The results are presented in Table 4.7. Panel A reports the performance of *Market* (passive strategy). This strategy showed annualized cumulative return of -3.03% and annualized Sharpe ratio of -0.18. The average number of stocks used per quarter is 499.98 and the turnover ratio of strategy is 0.05 which demonstrates the ins/outs of the S&P 500 constituents list.

Panel B of the table demonstrates the results of trading with consensus among analysts about price targets. The annualized cumulative return of this strategy under the *recent* (*all-time*) information set is 0.12% (0.31%) and the Sharpe ratio is 0.01 (0.02). This strategy outperforms the *Market* in both of the information sets with the *all-time* outrunning the *recent* in terms of the annualized cumulative returns.

Panel C of the table demonstrates the results of trading based on analysts' rankings. We observe that consistent with our assumption, the *true* resulted in the maximum possible annual cumulative return and the Sharpe ratio (4.32% and 0.29 respectively). Given the hypothetical assumption of the *true* set, it is not feasible to implement. Tthe next best feasible strategy is `diff` which is based on the predicted rankings from our label ranking model. This strategy yields an annualized cumulative return of 0.83% and the Sharpe ratio of 0.05 which are higher than those of the *CONS* and *Market* strategies. Moreover, we report that all strategies based on the predicted rankings with the dynamic states yield higher annualized cumulative returns than those based on the *recent* and the *all-time* rankings. The analysis of the sub-periods performance of the `diff` strategy is depicted in Table 4.8. The table shows the value of Sharp ratio for the 5-year periods. We observe that the `diff` strategy was dominant in most of the periods.

To test the significance of the annualized cumulative returns of strategies based on the predicted rankings we perform a null-hypothesis pairwise test when *null* is the difference in returns is zero. Table 4.9 presents the test results. We report that the returns of the strategies based on the predicted rankings with the `diff` and the `random` dynamic states are statistically significant at 1% when compared with the returns of all other strategies. The test accepts the *null* for the returns obtained for the strategies based on the predicted rankings of the `static` and the `roll.sd` states.

Figure 4.2 plots the graphical representation of the cumulative returns for all trading strategies. We observe that for the case of strategy based on the analysts' rankings, the *true* strategy is always on top of all others. This implies that in the settings where analysts' expected returns and rankings are based on price targets, an investor can gain a maximum results from trading strategy.

4.10 Conclusion

Some institutions, such as StarMine, rank financial analysts based on their accuracy and investment value performance. These rankings are published and are relevant: stocks favored by top-ranked analysts will probably receive more attention from investors. Therefore, there is a growing interest in understanding the relative performance of analysts.

In this paper we show that it is possible to model analysts' rankings and variables that affect them. With recent findings from Machine Learning body of research in label ranking, we adapted the algorithm to predict the rankings of financial analysts based on price targets and EPS forecasts. We report that, in case of price targets, our predicted rankings are more accurate than those obtained from using information of the whole history of analysts performance. For the rankings based on EPS forecasts, our model is able to predict rankings that are better than those of the *all-time* and the *recent* baselines. Moreover, the supremacy of our model above the baselines occurs when the variables that characterize analysts' information environment exhibit a stationary behavior expressed as the first-difference.

We also performed a back-test of active trading using the predicted rankings as inputs for the Black-Litterman model. The results showed that the strategies based on the analysts' rankings outperform, in terms of the annualized cumulative return, a strategy based on the analysts' consensus. Of the ranking based trading strategies, the maximum annualized cumulative return yields a strategy that is based on the predicted rankings with the first-difference of state variables.

The results of our work open many opportunities for future research. For example, in this paper we use the classical interpretation of the Black-Litterman model where risk is measured as a standard deviation. Recent work suggests utilizing more complex measures such as value-at-risk and high-moments approaches.

Table 4.1: Analysts' ranking consistency

This contingency table shows changes in analysts' *top*, *middle*, *bottom* ranking bins. Panel A (Panel B) depicts the dynamics of the analysts' ranks based on the accuracy in target prices (EPS forecasts). Rankings of the *recent* is the case of ranking information know at $t-1$ and the *all-time* is the case of using all ranking information for up to $t-1$.

	top_t			$middle_t$			$bottom_t$		
	top	mid	$bottom$	top	mid	$bottom$	top	mid	$bottom$
Panel A: TP									
					$t+1$				
recent	67.8	22.1	10.1	29.9	48.3	21.8	13.3	17.0	69.7
all-time	92.3	7.2	0.5	9.0	83.3	7.8	0.5	7.4	92.1
					$t+4$				
recent	46.0	28.1	25.9	39.2	29.4	31.4	32.3	26.7	41.1
all-time	81.8	15.2	3.0	19.2	64.3	16.5	2.9	17.5	79.6
Panel B: EPS									
					$t+1$				
recent	48.3	26.2	26.1	48.1	26.3	25.9	45.8	26.3	28.7
all-time	89.7	9.0	1.3	11.5	78.0	10.5	1.2	10.1	88.8
					$t+4$				
recent	47.0	28.2	26.1	45.0	27.2	28.5	43.5	27.9	31.4
all-time	79.3	16.6	4.1	20.9	60.2	18.9	4.2	18.9	76.9

Table 4.2: Discriminative power contribution

The table demonstrates the contributions (in %) of each of the variables to changes in analysts' rankings. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation.

Variable	`static`	`diff`	`random`	`roll.sd`
uncert	12.84	17.53	9.69	0.03
assym	3.88	4.79	3.17	0.01
disp	10.52	12.27	5.83	0.09
btm	2.34	2.47	5.72	4.29
size	0.39	0.32	1.30	1.54
dte	4.46	2.72	4.81	3.47
accr	0.90	0.77	1.21	1.42
s.ret	7.48	8.11	12.83	19.38
sec.ret	0.13	1.49	3.65	2.21
gnp	12.81	15.56	18.44	27.31
infl	18.16	24.90	13.43	20.83
vix.ret	0.08	0.22	1.43	0.19
t.bill	26.01	8.86	18.47	19.22
Total	100.00	100.00	100.00	100.00

Table 4.3: Example of label ranking problem

Period	$\mathscr{V}_1$	$\mathscr{V}_2$	$\mathscr{V}_3$	$\mathscr{V}_4$	Ranks		
					Alex	Brown	Credit
1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	1	2	3
2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	2	3	1
3	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	1	2	3
4	$x_{4,1}$	$x_{4,2}$	$x_{4,3}$	$x_{4,4}$	3	2	1
5	$x_{5,1}$	$x_{5,2}$	$x_{5,3}$	$x_{5,4}$	3	2	1
6	$x_{6,1}$	$x_{6,2}$	$x_{6,3}$	$x_{6,4}$	2	1	3
7	$x_{7,1}$	$x_{7,2}$	$x_{7,3}$	$x_{7,4}$	1	2	3

Table 4.4: Sample Statistics

This table shows the average number of target prices (panel A) and EPS forecasts (panel B) per stock per quarter.

	Min	Mean	Median	Max	Std.dev
Panel A: TP					
1999	3	4.144	4	10	1.549
2000	3	4.512	4	14	1.806
2001	3	4.873	4	16	2.188
2002	3	5.436	5	19	2.733
2003	3	5.763	5	21	3.073
2004	3	5.931	5	21	3.177
2005	3	6.042	5	21	3.207
2006	3	5.991	5	20	3.094
2007	3	5.755	5	20	2.953
2008	3	5.325	5	18	2.602
2009	3	4.535	4	18	1.958
Total	3	5.522	5	21	2.858
Panel B: EPS					
1999	3	5.165	4	17	2.758
2000	3	5.017	4	17	2.561
2001	3	5.418	5	18	2.758
2002	3	5.532	5	21	2.965
2003	3	5.643	5	24	3.034
2004	3	5.835	5	22	3.293
2005	3	5.933	5	24	3.328
2006	3	6.042	5	26	3.383
2007	3	5.841	5	22	3.103
2008	3	5.318	4	22	2.653
2009	3	4.998	4	20	2.425
Total	3	5.601	5	26	3.024

Table 4.5: Descriptive statistics of views

This table shows the descriptive statistics of views (expected returns) based on the consensus (median) among the analysts (panel A) and target price rankings (panel B). State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation.

	Mean (in %)	Median (in %)	Std.dev
	Panel A: Consensus		
true	18.610	16.889	0.120
recent	18.610	16.889	0.120
all-time	18.610	16.889	0.120
`static`	18.610	16.889	0.120
`diff`	18.610	16.889	0.120
`random`	18.610	16.889	0.120
`roll.sd`	18.610	16.889	0.120
	Panel B: TP		
true	14.876	13.380	0.096
recent	15.742	14.314	0.098
all-time	12.459	10.591	0.089
`static`	13.714	12.118	0.094
`diff`	13.701	12.042	0.095
`random`	13.910	12.267	0.094
`roll.sd`	13.690	12.119	0.095

Table 4.6: Average ranking accuracy

The table presents the average Spearman correlation between *true* and predicted rankings that are based on accuracy of price target (panel A) and on EPS forecasts (panel B) compared to baselines: *true* shows the case of the known future information; *recent* is the case of ranking information know at $t-1$, and the *all-time* is the case of using all ranking information for up to $t-1$. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation.

Year	*true*	*recent*	*all-time*	`static`	`diff`	`random`	`roll.sd`
Panel A: TP							
2000	1.0000	0.5988	0.4918	0.5695	0.5706	0.5312	0.5728
2001	1.0000	0.6600	0.4515	0.5355	0.6166	0.6058	0.6030
2002	1.0000	0.6092	0.3770	0.3960	0.4436	0.5310	0.4374
2003	1.0000	0.6168	0.3379	0.4812	0.4771	0.5372	0.5515
2004	1.0000	0.5500	0.2542	0.4031	0.2559	0.4610	0.2517
2005	1.0000	0.5862	0.2965	0.5114	0.4622	0.4100	0.5466
2006	1.0000	0.5908	0.3087	0.5323	0.4583	0.5804	0.4279
2007	1.0000	0.5787	0.2440	0.4823	0.2443	0.4703	0.3483
2008	1.0000	0.6573	0.3012	0.4720	0.5380	0.5247	0.6346
2009	1.0000	0.5720	0.2093	0.3902	0.4434	0.5507	0.3231
Total	1.0000	0.6007	0.3175	0.4746	0.4415	0.5167	0.4642
Panel B: EPS							
2000	1.0000	0.0344	0.0381	0.0420	0.0501	0.0461	0.0486
2001	1.0000	-0.0028	-0.0147	-0.0178	-0.0026	-0.0271	-0.0081
2002	1.0000	0.0386	0.0132	-0.0001	-0.0049	0.0089	0.0113
2003	1.0000	0.0489	0.0279	0.0380	0.0632	0.0263	0.0431
2004	1.0000	0.0255	0.0397	0.0021	0.0092	0.0184	-0.0414
2005	1.0000	0.0465	0.0369	0.0450	0.0297	0.0246	0.0380
2006	1.0000	0.0187	0.0416	0.0162	0.0367	0.0062	0.0332
2007	1.0000	0.0358	0.0246	0.0468	0.0535	0.0392	0.0469
2008	1.0000	0.0066	0.0222	0.0318	0.0110	0.0194	0.0110
2009	1.0000	0.0106	0.0240	0.0261	0.0198	0.0087	0.0228
Total	1.0000	0.0261	0.0264	0.0243	0.0268	0.0173	0.0205

Table 4.7: Trading strategy performance

The table presents the annualized cumulative statistics of the strategy performance based on PT rankings. *true* is actual ranking of the analysts. *recent* is the rankings from the last period. *all-time* is the average rank of an analyst for up to the last period. Trading period is from 2000Q1 until 2009Q4. Panel A presents the results from the passive strategy. Panel B summarizes the results of the strategy with rankings based on consensus in price targets. Panel C shows the case of the strategy with rankings based on price targets. State `static` is the state with no dynamics in values of the variables; `diff` is the state with first-difference in values; `random` is the state that captures the random part of values time-series decomposition; and `roll.sd` is the state of values sliding 8 quarters standard deviation.

Strategy	Annualized cum. return (in %)	Annualized Std. dev (in %)	Sharpe ratio	Average num. stock	Average turnover rate
Panel A					
Market	-3.032	16.654	-0.182	499	0.053
Panel B: CONS					
recent	0.116	15.948	0.007	283	0.256
all-time	0.314	15.773	0.020	283	0.228
Panel C: TP					
true	4.325	14.697	0.294	283	0.345
recent	0.282	15.662	0.018	284	0.264
all-time	0.689	15.565	0.044	284	0.256
`static`	0.547	15.759	0.035	251	0.266
`diff`	0.830	15.742	0.053	251	0.276
`random`	0.690	15.715	0.044	251	0.262
`roll.sd`	0.738	15.726	0.047	251	0.270

Table 4.8: Trading strategy performance: Sharpe ratio

This table presents the Sharpe ratio of each of the trading strategies: the passive (*Market*) and the active (consensus and smart estimates) calculated for different holding periods. Panel A represents the perfect foresight information set; panels B and C show, respectively, the recent and the all history analysts' performance. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation.

Periods	*Market*	*true*	*recent*	*all-time*	`static`	`diff`	`random`	`roll.sd`
2000Q1/2004Q4	-0.201	0.395	0.168	0.205	0.218	0.224	0.223	0.208
2001Q1/2005Q4	-0.093	0.425	0.214	0.249	0.232	0.242	0.239	0.222
2002Q1/2006Q4	0.196	0.757	0.490	0.464	0.525	0.538	0.529	0.511
2003Q1/2007Q4	0.925	1.915	1.248	1.245	1.248	1.305	1.250	1.303
2004Q1/2008Q4	-0.435	0.070	-0.305	-0.289	-0.308	-0.292	-0.308	-0.283
2005Q1/2009Q4	-0.158	0.189	-0.125	-0.105	-0.136	-0.106	-0.124	-0.106
All period	-0.182	0.294	0.018	0.044	0.035	0.053	0.044	0.047

Table 4.9: Significance of cumulative returns

The table demonstrates a pairwise test in difference of the cumulative returns of all strategies vs. those based on the predicted rankings. Case of *true* shows the known future information; *recent* is the case of ranking information know at $t-1$, and the *all-time* is the case of using all ranking information for up to $t-1$. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation.

	`static`		`diff`		`random`		`roll.sd`	
	t value	$\Pr(>\lvert t\rvert)$	t value	$\Pr(>\lvert t\rvert)$	t value	$\Pr(>\lvert t\rvert)$	t value	$\Pr(>\lvert t\rvert)$
true	8.329	0.000	8.383	0.000	8.237	0.000	8.698	0.000
recent	-28.526	0.000	-23.989	0.000	-29.860	0.000	-24.316	0.000
all-time	-5.521	0.000	-13.647	0.000	-8.831	0.000	-8.929	0.000
`static`	-	-	-6.934	0.000	-8.509	0.000	-0.910	0.368
`diff`	6.934	0.000	-	-	4.521	0.000	10.374	0.000
`random`	8.509	0.000	-4.521	0.000	-	-	2.731	0.009
`roll.sd`	0.910	0.368	-10.374	0.000	-2.731	0.009	-	-
Market	-15.357	0.000	-15.168	0.000	-15.361	0.000	-15.166	0.000

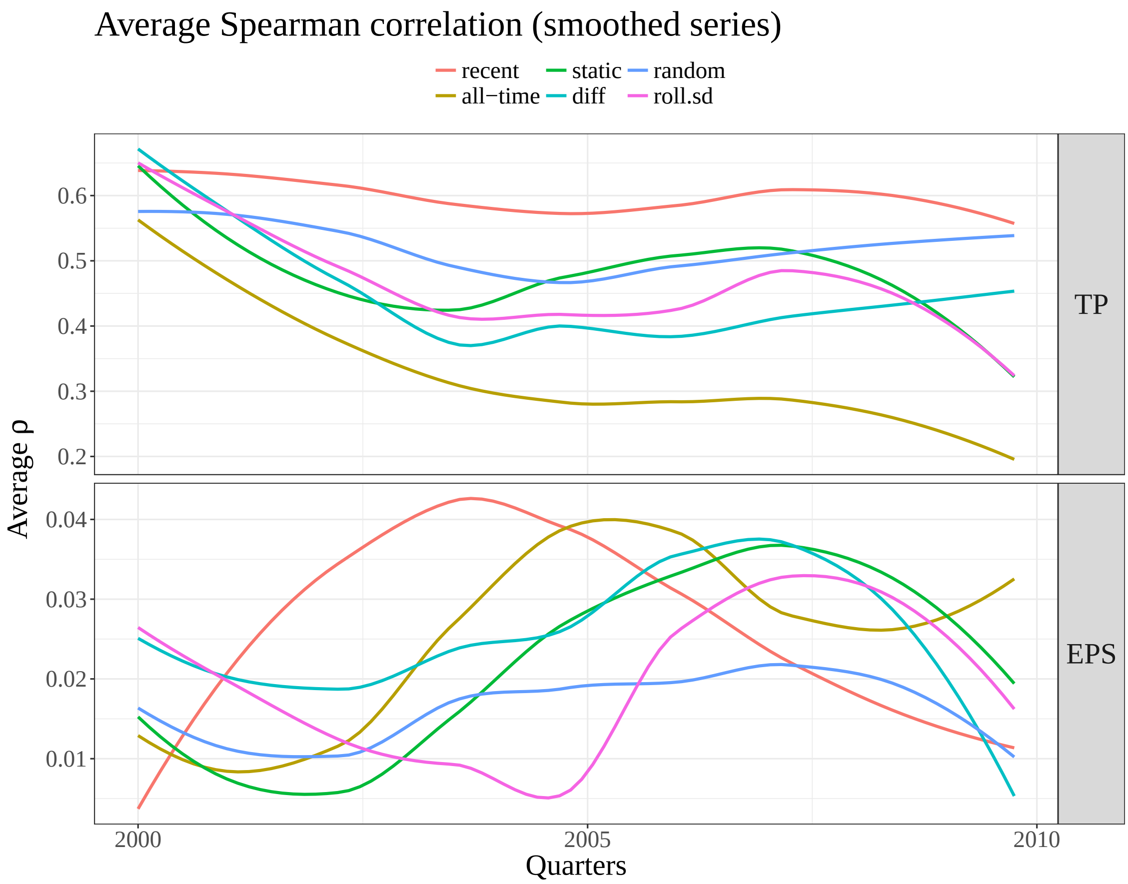

Figure 4.1: Average Spearman correlation (ρ)

The figure plots the smoothed series of the average Spearman correlation between *true* and predicted rankings resulted from static and dynamic states of variables: `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation. Smoothing is done by applying a local polynomial regression fitting (loess smoothing) (Cleveland et al., 1992). The top panel shows results of rankings based on the EPS forecast accuracy, the bottom are based on the price target accuracy.

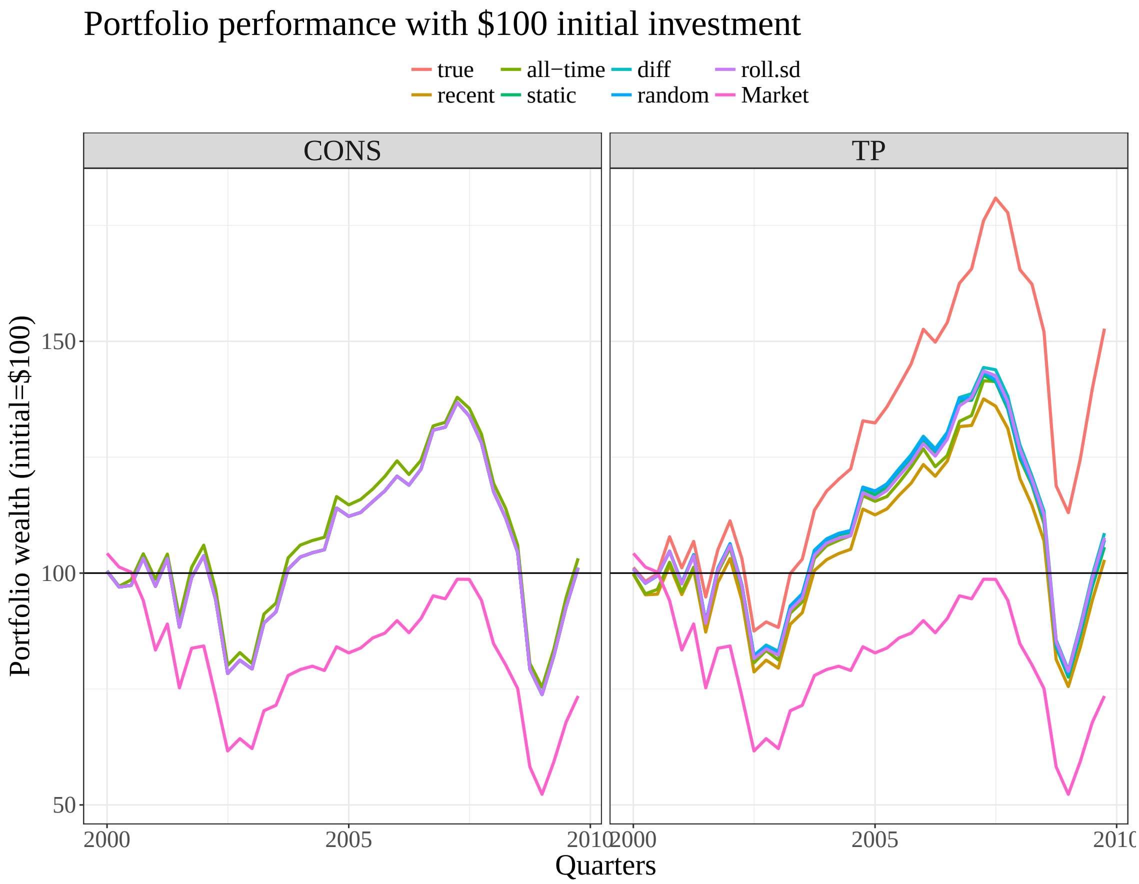

Figure 4.2: Performance of BL model

In this figure we show the quarterly performance of the cumulative portfolio wealth for all strategies. State `static` is the state with no dynamics in values of the variables, `diff` is the state with first-difference in values, `random` is the state that captures the random part of values time-series decomposition, and `roll.sd` is the state of values sliding 8 quarters standard deviation.

Conclusion

In this thesis we achieve multiple goals. First we show that rankings of financial analysts are important to investors as a trading strategy based on the rankings outperforms the ones that are based on the consensus or a buy and hold strategy. Second, we adapt the label ranking algorithm to map the rankings with the variables that affect these rankings (state variables) and, further, we identified a set of variables that affect the rankings the most. Finally, we successfully predict the rankings and we show that a strategy based on the predicted rankings outperforms the baselines.

Our work has the following limitations. First, our ranking model of financial analysts does not take into consideration analysts' verbal recommendations about a stock. In our work we assume that by setting an upward (downward) price target an analyst signals a "buy" ("sell") recommendation for investors. It would be interesting to incorporate all three components of analysts' report (explicit stock recommendations, EPS forecasts, and price targets) into a single ranking and apply models developed in this thesis. Second, we select a naive Bayes label ranking algorithm to predict the rankings mainly due to the Bayesian framework used in the Black-Litterman model. For the future work, we may extend the analysis to other label raking models, such as the nearest neighbor or association rules. Finally, our work could be applied in different ranking domains; for example, predicting the rankings of mutual funds based on their annual performance.

References

Ahmed, A., S. Nainar, and J. Zhou (2005). Do analysts' earnings forecasts fully reflect the information in accruals? *Canadian Journal of Administrative Sciences 22*(4), 329–342.

Aiguzhinov, A., A. Serra, and C. Soares (2015a). Are rankings of financial analysts useful for investors? http://ssrn.com/abstract=2680094. Last checked on 26/10/2015.

Aiguzhinov, A., A. Serra, and C. Soares (2015b). Understanding rankings of financial analysts. http://ssrn.com/abstract=2680301. Last checked on 26/10/2015.

Aiguzhinov, A., C. Soares, and A. Serra (2010). A similarity-based adaptation of naive Bayes for label ranking: Application to the metalearning problem of algorithm recommendation. In B. Pfahringer, G. Holmes, and A. Hoffmann (Eds.), *Discovery Science*, Volume 6332 of *Lecture Notes in Computer Science*, pp. 16–26. Springer Berlin Heidelberg.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance 56*(2), 531–563.

Barron, O., O. Kim, S. Lim, and D. Stevens (1998). Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review 73*(4), 421–433.

Barron, O., M. Stanford, and Y. Yu (2009). Further evidence on the relation between analysts' forecast dispersion and stock returns. *Contemporary Accounting Research 26*(2), 329–357.

Bhushan, R. (1989). Firm characteristics and analyst following. *Journal of Accounting and Economics 11*(2-3), 255–274.

Black, F. and R. Litterman (1992). Global portfolio optimization. *Financial Analysts Journal 48*(5), 28–43.

Bonini, S., L. Zanetti, R. Bianchini, and A. Salvi (2010). Target price accuracy in equity research. *Journal of Business Finance & Accounting 37*(9-10), 1177–1217.

Bouckaert, R. R. (2005). Naive Bayes classifiers that perform well with continuous variables. In *AI 2004: Advances in Artificial Intelligence*, Volume 3339 of *Lecture Notes in Computer Science*, pp. 85–116.

Bouwman, M., P. Frishkoff, and P. Frishkoff (1987). How do financial analysts make decisions? A process model of the investment screening decision. *Accounting, Organizations and Society 12*(1), 1–29.

Bradshaw, M. T. (2002). The use of target prices to justify sell-side analysts' stock recommendations. *Accounting Horizons 16*(1), 27–41.

Bradshaw, M. T. (2004). How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review 79*(1), 25–50.

Brav, A. and R. Lehavy (2003). An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *The Journal of Finance 58*(5), 1933–1968.

Brazdil, P., C. Soares, and J. Costa (2003). Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results. *Machine Learning 50*(3), 251–277.

Brazdil, P., C. Soares, C. Giraud-Carrier, and R. Vilalta (2009). *Metalearning Applications to Data Mining*. Springer.

Brown, L. (2001). How important is past analyst earnings forecast accuracy? *Financial Analysts Journal 57*(6), 44–49.

Brown, L. and K. Kim (1991). Timely aggregate analyst forecasts as better proxies for market earnings expectations. *Journal of Accounting Research 29*(2), 382–385.

Brown, L. and E. Mohammad (2003). The predictive value of analyst characteristics. *Journal of Accounting, Auditing and Finance 18*(4), 625–647.

Chen, L., Z. Da, and E. Schaumburg (2015). Implementing Black-Litterman using an equivalent formula and equity analyst target prices. *The Journal of Investing 24*(1), 34–47.

Cheng, W., K. Dembczynski, and E. Hüllermeier (2010). Label ranking methods based on the Plackett-Luce model. In J. Fürnkranz and T. Joachims (Eds.), *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. Omnipress.

Cheng, W., J. Hühn, and E. Hüllermeier (2009). Decision tree and instance-based learning for label ranking. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, New York, NY, USA, pp. 161–168. ACM.

Cheng, W. and E. Hüllermeier (2009). A new instance-based label ranking approach using the Mallows model. In W. Yu, H. He, and N. Zhang (Eds.), *Advances in Neural Networks – ISNN 2009*, Volume 5551 of *Lecture Notes in Computer Science*, pp. 707–716. Springer Berlin Heidelberg.

Clement, M. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics 27*(3), 285–303.

Cleveland, W. S., E. Grosse, and W. M. Shyu (1992). Local regression models. *Statistical models in S*, 309–376.

Cowles, A. (1933). Can stock market forecasters forecast. *Econometrica 1*(3), 309–324.

Creamer, G. and S. Stolfo (2009). A link mining algorithm for earnings forecast and trading. *Data Mining and Knowledge Discovery 18*(3), 419–445.

Da, Z. and E. Schaumburg (2011). Relative valuation and analyst target price forecasts. *Journal of Financial Markets 14*(1), 161–192.

de Sá, C., C. Soares, A. Jorge, P. Azevedo, and J. Costa (2011). Mining association rules for label ranking. In J. Huang, L. Cao, and J. Srivastava (Eds.), *Advances in Knowledge Discovery and Data Mining*, Volume 6635 of *Lecture Notes in Computer Science*, pp. 432–443. Springer Berlin Heidelberg.

Dekel, O., C. Manning, and Y. Singer (2004). Log-linear models for label ranking. In *Advances in Neural Information Processing Systems 16*.

Desai, H., B. Liang, and A. Singh (2000). Do All-Stars shine? Evaluation of analyst recommendations. *Financial Analysts Journal 56*(3), 20–29.

Diether, K., C. Malloy, and A. Scherbina (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance 57*(5), 2113–2141.

Domingos, P. and M. Pazzani (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning 29*(2), 103–130.

Emery, D. and X. Li (2009). Are the Wall Street analyst rankings popularity contests? *Journal of Financial and Quantitative Analysis 44*(2), 411.

Ertimur, Y., J. Sunder, and S. Sunder (2007). Measure for measure: The relation between forecast accuracy and recommendation profitability of analysts. *Journal of Accounting Research 45*(3), 567–606.

Fama, E. (1970). Efficient capital markets: A review of empirical work. *The Journal of Finance 25*, 383–417.

Fried, D. and D. Givoly (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics 4*(2), 85–107.

Glushkov, D. (2009). Overview of IBES on WRDS: Research and data issues. http://goo.gl/M0T1tw. Last checked on: 27/01/2014.

Grbovic, M., N. Djuric, and S. Vucetic (2012). Learning from pairwise preference data using gaussian mixture model. *Preference Learning: Problems and Applications in AI*, 33.

Grossman, S. and J. Stiglitz (1980). On the impossibility of informationally efficient prices. *The American Economic Review 70*(3), 393–408.

Gurrieri, M., X. Siebert, P. Fortemps, S. Greco, and R. Słowiński (2012). Label ranking: A new rule-based label ranking method. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. Yager (Eds.), *Advances on Computational Intelligence*, Volume 297 of *Communications in Computer and Information Science*, pp. 613–623. Springer Berlin Heidelberg.

Har-Peled, S., D. Roth, and D. Zimak (2002). Constraint classification: A new approach to multiclass classification. In N. Cesa-Bianchi, M. Numao, and R. Reischuk (Eds.), *Algorithmic Learning Theory*, Volume 2533 of *Lecture Notes in Computer Science*, pp. 365–379. Springer Berlin Heidelberg.

Henley, A., A. Carruth, and A. Dickerson (2003). Industry-wide versus firm-specific uncertainty and investment: British company panel data evidence. *Economics Letters 78*(1), 87–92.

Hilary, G. and C. Hsu (2013). Analyst forecast consistency. *The Journal of Finance 68*(1), 271–297.

Hill, N. and B. Stone (1980). Accounting betas, systematic operating risk, and financial leverage: A risk-composition approach to the determinants of systematic risk. *Journal of Financial and Quantitative Analysis 15*(03), 595–637.

Hope, O. and T. Kang (2005). The association between macroeconomic uncertainty and analysts' forecast accuracy. *Journal of International Accounting Research 4*(1), 23–38.

Hüllermeier, E., J. Fürnkranz, W. Cheng, and K. Brinker (2008). Label ranking by learning pairwise preferences. *Artificial Intelligence 172*(2008), 1897–1916.

Jegadeesh, N., J. Kim, S. Krische, and C. Lee (2004). Analyzing the analysts: When do recommendations add value? *The Journal of Finance 59*(3), 1083–1124.

Lebanon, G. and J. Lafferty (2002). Cranking: Combining rankings using conditional probability models on permutations. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pp. 370. Morgan Kaufmann Publishers Inc.

Lev, B. and S. Thiagarajan (1993). Fundamental information analysis. *Journal of Accounting research 31*(2), 190–215.

Li, X. (2005). The persistence of relative performance in stock recommendations of sell-side financial analysts. *Journal of Accounting and Economics 40*(1), 129–152.

Loh, R. and G. Mian (2006). Do accurate earnings forecasts facilitate superior investment recommendations? *Journal of Financial Economics 80*(2), 455–483.

Loh, R. and R. Stulz (2011). When are analyst recommendation changes influential? *Review of Financial Studies 24*(2), 593–627.

Michaely, R. and K. Womack (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *Review of Financial Studies 12*(4), 653–686.

Mikhail, M., B. Walther, and R. Willis (2004). Do security analysts exhibit persistent differences in stock picking ability? *Journal of Financial Economics 74*(1), 67–91.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.

O'Brien, P. C. (1990). Forecast accuracy of individual analysts in nine industries. *Journal of Accounting Research 28*(2), 286–304.

Ohlson, J. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research 11*(2), 661–687.

Ohlson, J. A. and B. E. Juettner-Nauroth (2005). Expected EPS and EPS growth as determinantsof value. *Review of Accounting Studies 10*(2-3), 349–365.

Park, C. and E. Stice (2000). Analyst forecasting ability and the stock price reaction to forecast revisions. *Review of Accounting Studies 5*(3), 259–272.

Parkash, M., D. Dhaliwal, and W. Salatka (1995). How certain firm-specific characteristics affect the accuracy and dispersion of analysts' forecasts:: A latent variables approach. *Journal of Business Research 34*(3), 161–169.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ramnath, S., S. Rock, and P. Shane (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting 24*(1), 34–75.

Ribeiro, G., W. Duivesteijn, C. Soares, and A. Knobbe (2012). Multilayer perceptron for label ranking. In A. Villa, W. Duch, P. Érdi, F. Masulli, and G. Palm (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2012*, Volume 7553 of *Lecture Notes in Computer Science*, pp. 25–32. Springer Berlin Heidelberg.

Richardson, V. J. (2000). Information asymmetry and earnings management: Some evidence. *Review of Quantitative Finance and Accounting 15*(4), 325–347.

Sheng, X. and M. Thevenot (2012). A new measure of earnings forecast uncertainty. *Journal of Accounting and Economics 53*(1), 21–33.

Simon, A. and A. Curtis (2011). The use of earnings forecasts in stock recommendations: Are accurate analysts more consistent? *Journal of Business Finance & Accounting 38*(1-2), 119–144.

Sousa, S. and A. Serra (2008). What drives idiosyncratic volatility over time? *Portuguese Economic Journal 7*(3), 155–181.

Todorovski, L., H. Blockeel, and S. Dzeroski (2002). Ranking with predictive clustering trees. In T. Elomaa, H. Mannila, and H. Toivonen (Eds.), *Machine Learning: ECML 2002*, Lecture Notes in Computer Science, pp. 444–455. Springer Berlin Heidelberg.

Trueman, B. and S. Titman (1988). An explanation for accounting income smoothing. *Journal of Accounting Research 26*, 127–139.

Vembu, S. and T. Gärtner (2010). Label ranking algorithms: A survey. In J. Fürnkranz and E. Hüllermeier (Eds.), *Preference Learning*, pp. 45–64. Springer.

Vogt, M., J. Godden, and J. Bajorath (2007). Bayesian interpretation of a distance function for navigating high-dimensional descriptor spaces. *Journal of Chemical Information and Modeling 47*(1), 39–46.

Womack, K. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance 51*(1), 137–168.

Zhang, X. (2006). Information uncertainty and stock returns. *The Journal of Finance 61*(1), 105–137.

Zhou, J. (2013). Analysts' target prices and stock recommendations. *Accounting and Finance Research 2*(1), p1.

Zivot, E. and J. Wang (2003). *Modeling financial time series with S-PLUS*, Volume 191. Springer Berlin Heidelberg.