DEGREE PROJECT, IN COMPUTER SCIENCE , SECOND LEVEL
STOCKHOLM, SWEDEN 2015

New Algorithms for Evaluating Equity Analysts' Estimates and Recommendations

FREDRIK BÖRJESSON

KTH ROYAL INSTITUTE OF TECHNOLOGY

SCHOOL OF COMPUTER SCIENCE AND COMMUNICATION (CSC)

New Algorithms for Evaluating Equity Analysts' Estimates and Recommendations

Nya algoritmer för att utvärdera aktieanalytikers estimat och rekommendationer

Fredrik Börjesson

June 2015

Master's thesis in Computer Science
Examensarbete 30hp
Datalogi 2D1021

Supervisor: Karl Meinke
Examinator: Stefan Arnborg

Abstract

The purpose of this study is to find improved algorithms to evaluate the work of equity analysts. Initially the study describes how equity analysts work with forecasting earnings per share, and issuing recommendations on whether to invest in stocks. It then goes on to discuss techniques and evaluation algorithms used for evaluating estimates and recommendations found in financial literature. These algorithms are then compared to existing methods in use in the equity research industry. Weaknesses in the existing methods are discussed and new algorithms are proposed. For the evaluation of estimates the main difficulties are concerned with adjusting for the reducing uncertainty over time as new information becomes available, and the problem of identifying which analysts are leading as opposed to herding. For the evaluation of recommendations, the difficulties lie mainly in how to risk-adjust portfolio returns, and how to differentiate between stock-picking ability and portfolio effects. The proposed algorithms and the existing algorithms are applied to a database with over 3500 estimates and 7500 recommendations and an example analyst ranking is constructed. The results indicate that the new algorithms are viable improvements on the existing evaluation algorithms and incorporate new information into the evaluation of equity analysts.

Sammanfattning

Syftet med denna studie är att hitta förbättrade algoritmer för att utvärdera aktieanalytikers arbete. I studien beskrivs inledningsvis hur aktieanalytiker arbetar med att ta fram prognoser för vinst per aktie och rekommendationer för att köpa eller sälja aktier. Därefter diskuteras tekniker och algoritmer för att utvärdera analytikers vinstprognoser och rekommendationer som hämtats från finansiell litteratur. Dessa algoritmer jämförs därefter med befintliga utvärderingsmetoder som används inom aktieanalys-branschen. Svagheter i de befintliga utvärderingsmetoderna diskuteras och nya algoritmer föreslås. För utvärderingen av vinstprognoser diskuteras svårigheterna i att justera för minskande osäkerhet allteftersom ny information blir tillgänglig, samt svårigheter att identifiera vilka analytiker som är ledande och vilka som är efterföljande. För utvärderingen av rekommendationer ligger svårigheterna främst i risk-justering av avkastningar, samt i att skilja mellan förmåga att bedöma enskilda aktiers utveckling och portföljeffekter. De föreslagna algoritmerna och de befintliga algoritmerna tillämpas på en databas med över 3500 vinstestimat och 7500 rekommendationer och ett exempel på ranking av analytiker tas fram. Resultaten indikerar att de nya algoritmerna utgör förbättringar av de befintliga utvärderingsalgoritmerna och integrerar ny information i utvärderingen av aktieanalytiker.

Contents

1 Introduction

1.1 Background

The subject of this thesis in Computer Science was conceived in collaboration with the equity research department of a bank. The bank had identified that their methods and techniques for evaluating their analysts could be improved upon and wanted to build a new evaluation tool to this end. Although to build a practical implementation for a specific company was always an aim of this thesis, the techniques described herein have general qualities and can be applied to evaluate estimations and recommendations elsewhere in the financial market.

In a general sense, evaluating the work of equity analysts presents us with the same problems as any evaluation: we want to make sure that the evaluation is done in a manner which is as objective and fair as possible. By fair we here mean to be able to discern between skill and luck, and reward the former. To be able to identify how to do this in this specific evaluation problem, it is necessary to first understand the context in which the problem exists. Therefore, we will begin by taking a look at how the world of equity research works. Hopefully this approach will offer the reader a more complete understanding of the problem, give a motivation as to why it deserves our attention and at the same time provide for a more interesting read.

Let us begin with a basic concept in finance – equity. Equity is the capital due to the shareholders of a company. Together with debt capital - the other principal form of capital – the equity forms the total capital available to a company. The term equity research thus refers to analysts' work on determining the value of the part of a company's capital which is due to its shareholders. In other words – the value of the stocks of a company. From the whole universe of companies, equity analysts occupy themselves with analyzing a limited subset of companies; namely those companies which are public and listed on a stock exchange. Consequently, the stocks that are analyzed by equity analysts are all freely available to buy or sell for anyone at the market price (given that there can be found another party prepared to sell or buy, respectively, the same number of stocks for that price).

Clients of an investment bank may use research provided by equity analysts – together with other sources of information – to decide whether or not they wish to own a particular stock. Based on such investment decisions, these investors will perform trades, i.e. buy or sell stocks with the intent of maximizing their returns. Clients usually do not pay directly for access to equity analysts' research reports. Instead, an investment bank will normally distribute its research reports freely to clients, but will in return expect clients to do a number of trades, from which the bank's stockbrokers will earn commission. Equity analysts working for banks like this are said to be "sell-side" analysts. There are also people analyzing stocks working on the so-called "buy-side", which means that they work with money management in one form or other - for example portfolio managers working for mutual funds or insurance companies. One simplified way to look at it is that “sell-side” equity analysts publish research reports that “buy-side” fund managers will read to support their decisions whether to hold a certain stock in their portfolios or not. There are certainly also equity analysts employed on the “buy-side” but they do not publish their research and for the purpose of this thesis, we restrict ourselves to discussing evaluation of “sell-side” equity analysts.

Let us now try to describe in more detail what it is that equity analysts do. The work of an equity analyst involves above all two main activities, which are separate yet intrinsically linked together. One of these activities is to produce estimates for certain economical key parameters in the accounting figures which each company must publish regularly, usually once every quarter at so-called earnings announcements. The most important estimate is without question earnings per share (EPS). There are plenty of other figures and ratios, which are also commonly found in analysts’ forecasts, but none of these are typically considered as important as EPS. Estimates are usually done on a yearly basis, i.e. analysts generally do not produce separate estimates for every quarter, only one figure for the whole year. As companies release their earnings reports, the uncertainty about the final figure for the full year is reduced and upon publication of the annual report the estimates are compared to actual outcomes. Analysts continually incorporate new information by revising their estimates as the year progresses.

In principle, the estimates are used by the analysts themselves as input parameters in equity valuation models. Such valuations can be expressed in terms of a price per stock – a target price. A difference between the target price and the current market price, with proper adjustments made for dividends (profits paid out to shareholders), is perceived by analysts as an upside or downside potential in the current stock price – i.e. a mispricing by the market discovered with the help of superior analytical abilities. This mispricing is assumed to be corrected by the market at some point, which would lead to an opportunity to earn an expected return. Based on this expected return – together with any relevant additional information which may be hard or impossible to quantify – analysts will then issue a recommendation for the stock. There is usually a pre-defined scale for recommendations, such as for example “Buy”, “Outperform”, “Hold”, “Underperform” and “Sell”.

All the estimates and recommendations for a stock are collected by providers of financial data and presented as an average called consensus. There are basically two types of recommendations: absolute recommendations, where the expected return is the only considered parameter, and relative recommendations, which are based on the expected return compared to other comparable stocks or the stock market in general. In other words, absolute recommendations implicitly consider each stock in isolation, whereas relative recommendations look at a particular stock as one of several alternative investment opportunities. It has become a de-facto industry standard that equity analysts’ recommendations on a stock should be considered relative to its peers within the same industry sector. We will expand on this in the next chapter.

Let us have a brief look at the equity analyst role as such. In most respects, the equity analyst is an individual specialist. There is always a lead analyst who has the ultimate responsibility for the coverage of a given company. Equity analysis is a competitive business and analysts are periodically evaluated and ranked, both internally and by external firms. Although these rankings surely are a source of rivalry, co-operation among colleagues is necessary. Equity analysts usually work in industry-specific teams. A high degree of specialization is necessary for analysts to develop a sufficiently deep understanding of the business in general and the minutiae of particular companies. Moreover, companies often report their results during a short space of time and work division is necessary to cope with the heavy workload during these reporting periods. Companies are usually categorized first by industry sector (and sometimes subsectors), then by countries or regions. Dividing up the work by industry and then region – rather than the other way around – is natural since companies of the same industry share more similarities than companies belonging to the same geographical market but different industries. For example, stocks can be first categorized into industries such as financials, consumer goods, health care etc., and then once more by geographical markets such as Germany, the UK, the Nordic region and so on.

At this point it might be useful to introduce a perspective which puts the significance of equity analysts' work into the wider context of the workings of the capital markets in general. This perspective builds on a theory called the efficient market hypothesis (EMH), a theory which is usually deemed important enough to warrant a chapter of its own in introductory finance textbooks (see e.g. chapter 13 in Brealey & Myers, 2000). Market efficiency is a concept that deals with the mechanisms allowing new information to disseminate into the market and affect prices. An efficient market is, in principle, a market where any informational advantages are instantly neutralized by the market as it incorporates the information into prices, and thus investors cannot exploit any such advantages to consistently make abnormal returns. Consistently in this case means that there needs to be an element of predictability over time in the ability of investors to earn these abnormal returns, and by abnormal we mean that the returns obtained must be superior to those from alternative investments which carry the same risk.

For the purpose of this study, risk can generally be thought of as a statistical measure of how much the price of a financial asset, such as a stock, has moved over time historically: the greater the variance (or standard deviation) of the price of a stock, the greater its risk. Moreover, in financial literature it is generally postulated that investors are risk-averse – i.e. in choosing between two assets with identical expected returns, a risk-averse investor will always prefer the asset with lower risk. Thus, under these assumptions, lowering the risk is desirable and investors might be willing to give up some expected return to accomplish that, or equivalently, such investors require a so-called risk-premium (higher expected return) to accept an uncertain outcome over a certain one. Introducing risk, then, makes the concept of abnormal profits a bit more difficult to grapple. In fact, how to correctly relate return to risk is one of the longest-standing debates among academics in the field of finance (see e.g. chapter 8 in Brealey & Myers, 2000). The important thing to keep in mind is that to compare the returns of two financial assets, we should also take into account the risks associated with respective asset.

The EMH comes in three different flavors: the weak, the semi-strong and the strong form. The weak form of the hypothesis entails that prices accurately reflect all the information in historical series of stock prices. In other words, investors cannot exploit patterns in prices such as e.g. predictable seasonal variations in the stock market to make abnormal returns. The semi-strong form states that prices reflect all publicly available information. That means that it is impossible for investors to earn abnormal profits simply by reading news articles, scrutinizing the company's annual accounts etc. The strong form, finally, – and this is where equity analysts are most concerned – states that stock prices effectively contain all available information. That includes even that information which is laboriously produced by equity analysts in an effort to help their clients outsmart the market. "It [the strong form of the hypothesis] tells us that superior information is hard to find because in pursuing it you are in competition with thousands, perhaps millions, of active, intelligent, and greedy investors." (Brealey & Myers 2000, p. 377). Thus, under the strong form of market efficiency, equity analysts have essentially no hope of consistently contributing any valuable advice to their clients, and our attempts of developing a methodology for evaluating the work of equity analysts would then be a meaningless effort right from the outset. After numerous efforts to test the EMH, results are quite mixed. There seems to be widespread agreement among researchers that consistently earning abnormal returns is indeed difficult, but few researchers would be prepared to go so far as to argue that markets are strong-form efficient. Several researchers have also found so-called anomalies (e.g. the January-effect where a general increase in stock prices during the month of January has been observed, or the so-called post-earnings-announcement drift where markets are seemingly slow to discount new information after earnings surprises), which would suggest that markets are indeed not efficient at all (for a review of some of the evidence see e.g. Hawawini & Keim, 1995).

1.2 Purpose

The purpose of this thesis is to improve on existing techniques and algorithms used for evaluating the estimations and recommendations of equity analysts. There are techniques and algorithms to evaluate equity analysts in place already, as we will describe in later chapters, but they do not always fully take into account certain problems, which can give a bias and distort the true picture of who is the better analyst.

The problem was approached by researching techniques for evaluating estimations and recommendations described in finance literature and looking into what is considered 'best practice' in the industry of equity research evaluation. Based in this, new techniques and algorithms are proposed, which address some weaknesses in the existing techniques and algorithms, with the aim to improve the ability of these tools to help reliably distinguish between good analysts and not-so-good analysts.

1.3 Contribution

We may ask ourselves why this is a worthwhile topic for a thesis? There are several reasons. Firstly because equity analysts perform an important task in a market economy and therefore it is in everyone's interest that they are evaluated in an unbiased way. Secondly, equity research can generate important business for a bank, and so it is of great commercial importance to a bank to measure analyst performance as correctly as possible to ensure a high quality service to clients. Finally, a sound and unbiased evaluation procedure might prove a valuable tool for analysts themselves if they can take advantage of it to improve their work.

2 Evaluating equity analysts in theory

This section aims to survey previous research and introduce some of the basic metrics and terminology. After carefully reviewing the previous research available, equity analyst evaluation methodology appears to be a relatively scarcely researched subject in academic literature. Nevertheless, researchers have indeed indirectly developed methods for evaluating analysts' estimates and recommendations, although they have defined the problem in a slightly different way. For example, a number of researchers have investigated the information content of equity research reports. In other words, they have tried to determine whether investors can profit from following the recommendations of equity analysts – in which case they draw the conclusion that the average recommendation does indeed hold new information. Our approach is somewhat similar in that we wish to measure recommendation profitability (as one of the relevant dimensions of evaluation), yet quite different in that we do not look at an "average" recommendation but instead wish to differentiate between analysts. Thus, even though the following section is to a large extent based on research, which may be only indirectly related to our problem, it still gives us some firm ground to build our own analysis on. Algorithms will be presented throughout in pseudo-code.

2.1 Methods for evaluating estimates

2.1.1 Measuring estimation accuracy

O'Brien (1990) investigates whether observed distribution of analyst forecast accuracies differs from the distribution expected if their relative performances each year were purely random. Average accuracy is estimated across individuals, and the observed distribution of analyst forecast accuracies is compared with the expected distribution for purely random relative performances. The forecast accuracy metric used is simply the average absolute forecast (estimation) error. Absolute forecast error is defined as

$$E_{a,s,t} = |A_{s,t} - F_{a,s,t}|\,, \tag{1}$$

where $A_{s,t}$ denotes actual EPS for stock s in year t, and $F_{a,s,t}$ denotes the forecast from analyst a.

ALGORITHM A

Absolute forecast error

```
double[][][] absolute_forecast_error() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double E[][][] = double[analysts][stocks][days]; //absolute forecast error
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  for (a = 0; a < analysts; a++) {
```

```
    for (s = 0; s < stocks; s++) {
      for (t = 0; t < days; t++) {
        if (cover[a][s][t])
          E[a][s][t] = abs(A[s][t] - F[a][s][t]);
      }
    }
  }
  return E;
}
```

O'Brien also points out that average squared forecast error is another commonly used accuracy criterion. However, using squared forecast error can result in skewed and fat-tailed residual distributions, so it is often less than ideal as a test statistic.

Stickel (1992) studies the relation between equity analysts' reputation and estimation skill using three criteria of evaluation: forecast (estimation) accuracy, frequency of forecast issuance, and impact of forecast revisions on equity prices. The accuracy measure is identical to that of O'Brien, but Stickel also reports absolute scaled forecast error, where the actual reported EPS is used in the denominator:

$$ASE_{a,s,t} = \frac{| A_{s,t} - F_{a,s,t} |}{| A_{s,t} |}, \tag{2}$$

ALGORITHM B

Absolute scaled forecast error

```
double[][][] absolute_scaled_forecast_error() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double ASE[][][] = double[analysts][stocks][days]; //absolute scaled forecast error
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  for (a = 0; a < analysts; a++) {
    for (s = 0; s < stocks; s++) {
      for (t = 0; t < days; t++) {
        if (cover[a][s][t]) {
          if (A[s][t] <> 0)
            ASE[a][s][t] = abs(A[s][t] - F[a][s][t])/A[s][t];
        }
      }
    }
  }
  return ASE;
}
```

Mikhail et al (1999) investigate if earnings forecast accuracy matters to equity analysts by examining its relation to analyst turnover. Two measures of forecasting accuracy are used, one absolute metric which measures proximity of the analyst forecast to actual earnings, and one relative measure which measures proximity of the forecast to the actual earnings relative to peer analysts. The absolute measure is calculated as follows. First, the absolute percentage error is calculated as

$$APE_{a,s,t} = \frac{| A_{s,t} - F_{a,s,t} |}{P_{s,t}}, \tag{3}$$

where $A_{s,t}$ and $F_{a,s,t}$ are defined as before, and $P_{a,s,t}$ is the stock price at the beginning of the period. The stock price, which should be of similar magnitude to EPS, is used as a deflator[1] instead of actual reported EPS to avoid some potential statistical problems with (2). The absolute metric used is then calculated as the average absolute percentage error across all firms in an analyst's coverage universe (usually an industry), multiplied by minus one (-1) so that high (low) levels correspond to more (less) accurate analysts.

ALGORITHM C

Absolute percentage error metric

```
double[][] absolute_percentage_error_metric() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double P[][] = double[stocks][days];
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  int coveredStocks[][] = int[analysts][days];
  double metric[][] = double[analysts][days];
  for (a = 0; a < analysts; a++) {
    for (t = 0; t < days; s++) {
      for (s = 0; s < stocks; t++) {
        if (cover[a][s][t]) {
          metric[a][t] += (abs(A[s][t] - F[a][s][t]) / P[s][t]);
          coveredStocks[a][t]++;
        }
      }
      metric[a][t] = metric[a][t] / coveredStocks[a][t] * -1;
    }
  }
  return metric;
}
```

The relative measure is instructive as an example of how one can scale ranks to be able to relate and compare several ranking measures to each other. It is computed based on the absolute measure by ranking an analyst's $APE_{a,s,t}$ as in (3) relative to that of all other analysts with the same primary industry following the same stock. The rank is then divided by the number of analysts issuing forecasts for that stock and year. This measure ranges from $1/n$ to 1 with high levels corresponding to relatively more accurate analysts.

[1] By deflator here is meant the denominator in a ratio calculation, which is used to "deflate" the nominator, to allow comparison between stocks.

$$score_{a,s,t} = \frac{rank_{a,s,t}}{number\ of\ analysts_{s,t}} \tag{4}$$

Finally, the relative metric used in the study is the average of this rank accuracy for all firms in an analyst's primary industry.

ALGORITHM D

Absolute percentage error score

```
double[][][] absolute_percentage_error_score() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double P[][] = double[stocks][days];
  double APE[][][] = double[analysts][stocks][days]; //absolute percentage error
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  int coveringAnalysts[][] = int[stocks][days];
  double metric[][][] = double[analysts][stocks][days];
  double score[][][] = double[analysts][stocks][days];
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              APE[a][s][t] = (abs(A[s][t] - F[a][s][t]) / P[s][t]);
              coveringAnalysts[s][t]++;
           }
        }
     }
  }
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t])
              score[a][s][t] = rank(APE[a][s][t],APE[][s][t]) / coveringAnalysts[s][t];
        }
     }
  }
  return score;
}
```

A similar ranking approach for forecast accuracy is used by Hong et al. (2000). The starting point here is absolute forecast error as in (1), and the analysts who cover a firm in one year are then sorted and ranked based on these forecast errors. Instead of using an average, a scaled score measure is used as follows:

$$score_{a,s,t} = 100 - \left(\frac{rank_{a,s,t} - 1}{number\ of\ analysts_{s,t} - 1} \right) \times 100 \tag{5}$$

With this procedure, an analyst with the rank of one receives a score of 100; an analyst who is the least accurate receives a score of zero. Finally, the accuracy metric used is the average scores for all of the analyst's covered firms in year t and the preceding two years. Hong et al. argue that by using three-year averages they get a less noisy proxy for the true forecasting ability.

ALGORITHM E

Absolute forecast error score

```
double[][][] absolute_forecast_error_score() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double E[][][] = double[analysts][stocks][days]; //absolute forecast error
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  int coveringAnalysts[][] = int[stocks][days];
  double metric[][][] = double[analysts][stocks][days];
  double score[][][] = double[analysts][stocks][days];
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              E[a][s][t] = abs(A[s][t] - F[a][s][t]);
              coveringAnalysts[s][t]++;
           }
        }
     }
  }
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t])
              score[a][s][t] = 100 - (rank(E[a][s][t], E[][s][t]) - 1) _
              / (coveringAnalysts[s][t] - 1) * 100;
        }
     }
  }
  return score;
}
```

In a study by Loh & Mian (2006), a measure of forecasting accuracy relative to other analysts is constructed. The metric, proportional mean absolute forecast error, is defined as follows

$$PMAFE_{a,s,t} = \frac{|E_{a,s,t}| - |\overline{E}_{s,t}|}{|\overline{E}_{s,t}|}, \tag{6}$$

where $\overline{E}_{s,t}$ is the mean absolute forecast error of all analysts (the consensus error).

The metric can be interpreted as analyst a's fractional forecast error relative to the consensus error for stock s in year t. Negative (positive) values of $PMAFE_{a,f,t}$ represent above (below) average accuracy. The rationale behind subtracting the consensus mean from the analyst's absolute forecast error is to control for stock-year effects. Stock-year effects result from stock- or year-specific factors that make certain stocks' earnings harder or easier to forecast in certain years, for instance macro-economic shocks. Scaling the numerator by the consensus error controls for heteroscedasticity[2] of forecast error distributions across firms, which can be important for example if the metric is to be used as a variable in a linear regression analysis.

ALGORITHM F

Proportional mean absolute forecast error

```
double[][][] proportional_mean_absolute_forecast_error () {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double E[][][] = double[analysts][stocks][days]; //absolute forecast error
  double E_bar[][] = double[stocks][days]; //consensus absolute forecast error
  double PMAFE[][][] = double[analysts][stocks][days];
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  int coveringAnalysts[][] = int[stocks][days];
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              E[a][s][t] = abs(A[s][t] - F[a][s][t]);
              coveringAnalysts[s][t]++;
           }
        }
     }
  }
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t])
              E_bar[s][t] += E[a][s][t] / (coveringAnalysts[s][t];
        }
     }
  }
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t])
              PMAFE[a][s][t] = (E[a][s][t] - E_bar[s][t])/ E[a][s][t];
        }
     }
  }
  return PMAFE;
}
```

[2] Heteroscedasticity is a statistical concept, which means in principle that the variance of the data is inconsistent in magnitude over one or more of the independent variables (often time in time-series data). The presence of heteroscedasticity is a concern as it can affect the validity of statistical regression significance tests, i.e.. we cannot be certain the results of a regression are reliable under heteroscedasticity. For a thorough discussion see Gujarati (2003) pp. 387-428.

2.1.2 Decreasing uncertainty

In the O'Brien (1990) study, forecasts are only included in the sample if they are made at least 120 trading days prior to the annual earnings announcement. This minimum horizon is devised to provide comparability, because forecast accuracy generally improves as the horizon decreases. Stickel (1992) tries to mitigate the problem with decreasing uncertainty as the announcement date approaches by dividing the yearly data into monthly sub-periods so that only forecasts with equal horizons are compared (those forecasts that are issued in the same sub-period).

Cooper et al. (2000) develop procedures for ranking the performance of analysts based on three criteria: forecast accuracy, abnormal trading volume associated with these forecasts, and "timeliness" of earnings forecasts. The first criterion, forecast accuracy, is measured exactly as in (2). However, to control for the bias related to decreasing uncertainty as the forecast horizon becomes shorter, the absolute scaled forecast error from (2) is regressed by linear regression on the length of time from the forecast release date to the annual earnings announcement by the following model:

$$ASE_{a,s,t} = b_0 + b_1 T_{s,t} + \varepsilon_{a,s,t} \;, \tag{7}$$

where $T_{s,t}$ is the number of days at time t from the forecast release date until the earnings announcement date for stock s, b_0 and b_1 are the intercept and the slope coefficient respectively and $\varepsilon_{a,s,t}$ are the residuals. Since the residuals are free of bias related to the length of the forecast horizon, the average of the absolute value of the residuals over analysts' coverage universes can be used as an unbiased measure to rank the analysts' relative accuracy. Moreover, the signs and relative magnitudes of the slope and the intercept can be used to draw conclusions about whether analysts were initially too optimistic (positive slope) or pessimistic (negative slope) and also estimate at what point in time they changed their sentiment.

ALGORITHM G

Absolute scaled forecast error with time regression

```
double[][] absolute_scaled_forecast_error_time_regression_metrics() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double ASE[][][] = double[analysts][stocks][days]; //absolute scaled forecast error
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  int daysUntilReport[][][] = int[analysts][stocks][days]: //days until EPS report
  double epsilon[][] = double[analysts][stocks]; //residuals
  double intercept[][] = double[analysts][stocks]; //intercept
  double metrics[] = double[analysts][3];  //residuals, intercept and slope metrics
  double averageCoveredStocks[] = double[analysts]; //average covered stocks over time
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
```

```
        for (t = 0; t < days; t++) {
          if (cover[a][s][t]) {
            if (A[s][t] <> 0) {
              ASE[a][s][t] = abs(A[s][t] - F[a][s][t])/A[s][t];
              averageCoveredStocks[a]++;
            }
          }
        }
      }
      averageCoveredStocks[a] /= days;
    }
    for (a = 0; a < analysts; a++) {
      for (s = 0; s < stocks; s++) {
        Do regression with ASE as dependent variable and daysUntilReport as _
        explanatory variable, with intercept.
        Save residuals in epsilon[a][s]
        Save intercept in intercept[a][s]
        Save slope in slope[a][s]
      }
    }
    for (a = 0; a < analysts; a++) {
      for (s = 0; s < stocks; s++) {
        metric[a][0] += abs(epsilon[a][s])/averageCoveredStocks[a];
        metric[a][1] += intercept[a][s]/averageCoveredStocks[a];
        metric[a][2] += epsilon[a][s]/averageCoveredStocks[a];
      }
    }
    return metric;
  }
```

2.1.3 Leading/herding

Hong et al. (2000) investigate the relation between analysts' career concerns and herding of earnings forecasts. Herding is when analysts copy the action of others, changing their estimates to follow the majority. The opposite of herding is called leading. In this context, leading means that one analyst changes his/her estimates and then the majority of analysts follow suit (with a time lag). One possible explanation for this is information free-riding where herding analysts simply delay their revisions of estimates until a leading analyst produces new information which they subsequently use in their own forecasts. Generally, all other things equal, a leading analyst behavior is in most circumstances preferable to a herding behavior. However, it should be pointed out that being bold (leading) and bad (having low accuracy) is certainly not a desirable combination, so herding/leading should never be used as the sole criteria. Hong et al. measure leading (or forecast boldness as they call it) with a metric defined as follows

$$\textit{deviation from consensus}_{a,s,t} = |F_{a,s,t} - \overline{F}_{s,t}|, \tag{8}$$

where $F_{a,s,t}$ is defined as in (1) and A is the set of all analysts who issue an earnings estimate for stock s in year t, so that $\overline{F}_{s,t}$ is a measure of the consensus forecast. Starting with this measure, the same ranking methodology as previously described for forecast accuracy is used to construct a score for leading/herding similar to that in (5). Higher (lower) values of the metric correspond to a more leading (more herding) analyst behavior.

ALGORITHM H

Forecast boldness score

```
double[][][] forecast_boldness_score () {
 double A[][] = double[stocks][days]; //actual EPS reported by the company
 double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
 double F_bar[][][] = double[stocks][days]; //consensus forecast
 double deviation[][][] = double[stocks][days]; //deviation from forecast
 double score[][][] = double[analysts][stocks][days];
 bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
 int coveringAnalysts[][] = int[stocks][days];
 for (a = 0; a < analysts; a++) {
   for (s = 0; s < stocks; s++) {
     for (t = 0; t < days; t++) {
       if (cover[a][s][t]) {
         coveringAnalysts[s][t]++;
       }
     }
   }
 }
 for (a = 0; a < analysts; a++) {
   for (s = 0; s < stocks; s++) {
     for (t = 0; t < days; t++) {
       if (cover[a][s][t])
         F_bar[s][t] += F[a][s][t] / (coveringAnalysts[s][t];
     }
   }
 }
 for (a = 0; a < analysts; a++) {
   for (s = 0; s < stocks; s++) {
     for (t = 0; t < days; t++) {
       if (cover[a][s][t])
         deviation[a][s][t] = abs(F[a][s][t] - F_bar[s][t]);
     }
   }
 }
 for (a = 0; a < analysts; a++) {
   for (s = 0; s < stocks; s++) {
     for (t = 0; t < days; t++) {
       score[a][s][t] = 100 - (rank(deviation[a][s][t], deviation[][s][t]) - 1) _
       / (coveringAnalysts[s][t] - 1) * 100;
     }
   }
 }
 return score;
}
```

The third criterion used in the study by Cooper et al. (2000), “timeliness”, is an attempt to incorporate a leading/herding measure into his analysis by quantifying to what extent an analyst is a leader or a follower. Assuming the information free-riding scenario outlined earlier, forecast revisions by a leading analyst should be followed closely by forecast revisions of other analysts. The idea of "timeliness" is illustrated in Figures 1 and 2 below.

FIGURE 1

Expected pattern of forecast revision dates surrounding the forecast revision of a lead analyst. The timeline shows forecast revision dates for analyst L and the two most recent forecast revisions before (C and D) and after (X and Y) L's revision. The *LFR* metric for L = (10 + 9) /(1 + 2) = 6 1/3 > 1. (after Cooper et al. (2000) p. 394.)

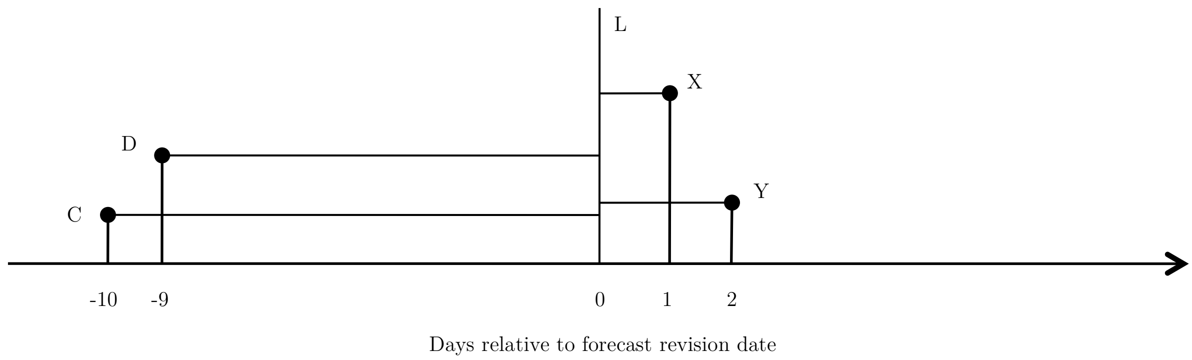
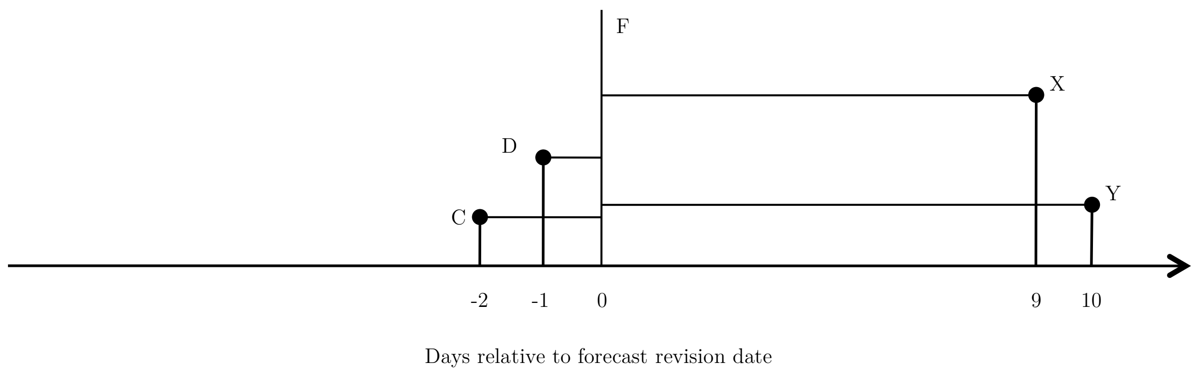


FIGURE 2

Expected pattern of forecast revision dates surrounding the forecast revision of a following analyst. The timeline shows forecast revision dates for analyst F and the two most recent forecast revisions before (C and D) and after (X and Y) F's revision. The *LFR* metric for F = (2 + 1) /(9 + 10) = 3/19 < 1. (after Cooper et al. (2000) p. 394.)



Conditional on the release of a leading analyst estimate, we assume that the times until release of revised forecasts by follower analysts have independent exponential distributions

$$\frac{1}{\theta_1} e^{-t/\theta_1}, \tag{9}$$

where θ_1 is the expected time until the next forecast release by another analyst, which is assumed to be the same for each follower analyst. Similarly, conditional on the release of a follower analyst forecast revision, the times until the next forecast release have independent exponential distributions with expected time until next release given by θ_0. Herding followers will quickly update their forecasts after an earnings forecast release by a leading analyst. However, they have no incentive to revise their forecasts in response to forecast revisions by other followers. As a consequence of this logic, θ_0 must be greater than θ_1.

Next, the cumulative analyst days required to generate N forecasts by competing analysts preceding and following each of the K forecasts by an analyst is computed. Let $t^0{}_{n,k}$ and $t^1{}_{n,k}$ denote the number of days by which forecast n either precedes or follows the k^{th} forecast by an analyst. The cumulative lead-time for the K forecasts is then

$$T_0 = \sum_{k=1}^{K} \sum_{n=1}^{N} t_{n,k}^{0} \tag{10}$$

Similarly, the cumulative follow-time for these K forecasts is

$$T_1 = \sum_{k=1}^{K} \sum_{n=1}^{N} t_{n,k}^{1} \tag{11}$$

The maximum likelihood estimators[3] of the expected forecast arrival times during pre- and post-release periods are T_0/N and T_1/N respectively. Since $2T_0/\theta_0$ and $2T_0/\theta_0$ are distributed as $\chi^2_{(2KN)}$, Cooper et al. (2000) can form the test statistic

$$LFR = \frac{2T_0 / \theta_0}{2T_1 / \theta_1}, \tag{12}$$

which is distributed as $F_{(2KN,\ 2KN)}$. Since θ_0 and θ_1 are assumed to be constants we can simplify and calculate the pleasingly parsimonious metric

$$LFR = \frac{T_0}{T_1}, \tag{13}$$

which we call the leader-follower ratio. If leading is defined as systematically releasing forecast revisions before other analysts, leading analysts are those who have an *LFR* metric greater than 1 and conversely herding analysts have an *LFR* metric less than 1. Cooper et al. (2000) suggest calculating firm-specific *LFR* statistics by computing lead and follow times across all forecast

[3] Gujarati (2003)

revisions for a given analyst on a firm-by-firm basis. As an alternative, an industry-specific *LFR* can be calculated by accumulating across all forecasts for the firms that an analyst follows.

ALGORITHM I

Leader-follower ratio

```
double[][][] leader-follower_ratio () {
  int F_dates[][][] = int[analysts][stocks][];//dates with a forecast revision
  int F_revisions[][] = int[analysts][stocks]; //number of forecast revisions
  double LFR[] = double[analysts]; //leader-follower metric
  bool cover[][][] = bool[analysts][stocks]: //stock coverage matrix
  int T0; int T1;
  int currentDate; //the date for the forecast being evaluated
  int previousDate; //the preceding forecast date by the same analyst
  int nextDate; //the next forecast date by the same analyst
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        if (cover[a][s]) { //if covers stock
           previousDate = null;
           T0 = 0;
           T1 = 0;
           for (d = 0; d < F_revisions[a][s]; d++) {
              currentDate = F_dates[a][s][d];
              if (d < F_revisions[a][s])
                 nextDate = F_dates[a][s][d+1];
              else
                 nextDate = null;
              for (aa = 0; aa < analysts; aa++) { //if analyst cover stock
                 if (aa <> a && cover[aa][s]) { //if not the same analyst and covers
                    for (dd = 0; dd < F_revisions[aa][s]; d++) {
                    if (F_dates[aa][s][dd] < currentDate && _
                          (F_dates[aa][s][dd] > previousDate || previousDate == null))
                          T0 += F_dates[aa][s][dd] - currentDate;
                       else {
                          if (F_dates[aa][s][dd] > currentDate && _
                             (F_dates[aa][s][dd]  < nextDate || nextDate == null))
                             T1 += currentDate - F_dates[aa][s][dd];
                       }
                    }
                 }
              }
           previousDate = currentDate;
           }
        }
     }
  LFR[a] = T0/T1;
  }
  return LFR;
}
```

2.2 Methods for evaluating recommendations

From a client perspective, analysts' recommendations are merely a means to an end: generating adequate profits on investments. Therefore recommendation profitability or performance is undeniably the single most important metric for evaluating equity analysts. In financial literature, researchers have interested themselves in this subject mainly from an efficient market hypothesis-point of view. More specifically whether equity analysts' recommendations have investment value by consistently generating abnormal returns. We will discuss abnormal returns in much more detail shortly, but to be able do so we must first understand the mechanics of how we move from an analyst's set of recommendations on the stocks that he or she covers, to something that we can measure.

2.2.1 Portfolio formation

In principle, the ideal single metric can be thought to embody the portfolio that an analyst would run, if analysts actually ran portfolios. Thus, performance evaluators invent ways to create such a synthetic portfolio by weighting the stocks covered by the analyst (the coverage universe) consistently with his or her recommendations. The simplest stock rating system consists of the ratings ”Buy”, ”Hold” and ”Sell”. Most brokerage firms, however, use expanded forms adding such ratings as “Overweight” and “Underweight” or “Outperform” and “Underperform”. In practice most analyst recommendations are submitted electronically to database providers such as Reuters, Zacks, and First Call (Thomson). These providers standardize the ratings by converting them to a numerical scale (usually 5-point). To exemplify a common technique of weighting the returns from recommended stocks we can look at for instance at Loh & Mian (2006). They employ a five-point system: 1 = “Strong Buy”, 2 = “Buy”, 3 = “Hold”, 4 = “Underperform”, and 5 = “Sell”. These ratings translate into weightings as follows: “Buy” equals a long position in the stock by 100%, so the position simply earns the same return as the stock. A “Strong buy” gets a weighting of 200%, so the position earns twice the return from the stock, In reality this could be achieved by taking a leveraged position (borrowing and investing) in the stock. “Holds” are treated as a special case. It has been observed in many studies that “Sell” recommendations are underrepresented relative to positive recommendations, which has led to the belief among researchers that some “Sells” are actually hidden behind the euphemism of “Hold” due to conflicts of interest such as existing, and potential, investment banking relationships with the companies. To counter this bias, it has become a relative common practice in studies to equate a “Hold” with a “Sell”, which is the approach adopted by Loh & Mian. Furthermore, “Sells” receive a weighting of -100%, thus creating a position earning the opposite of what the stock returns. This could be interpreted as investors short-selling the stock (in principle borrowing a stock, selling it in the markets with the aim to buy it again at a lower price before returning the stock to its owner). For “Underperform” and “Sell” they use a weighting of -200%, implying a leveraged short-sell in the stock equivalent to twice the amount of a “Hold” recommendation.

ALGORITHM J

Raw recommendation portfolio returns

```
double[] recommendations_portfolio_return() {
  double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
  double stockReturn[][] = double[stocks][days]; //daily stock returns
  bool cover[][][] = bool[analysts][stocks]: //stock coverage matrix
  double weights[] = double[5];
  weights[0] = 1; //weight for a ‘Buy’ recommendation
  weights[1] = 0.5; //weight for an ’Outperform’ recommendation
  weights[2] = 0; //weight for a ‘Hold’ recommendation
  weights[3] = -0.5; //weight for a ’Underperform’ recommendation
  weights[4] = -1; //weight for a ‘Sell’ recommendation
  double portfolioReturn[] = double[analysts];
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (d = 0; d < days; d++) {
           if (cover[a][s][t])
              portfolioReturn[a] += weight[recs[a][s][d]]*stockReturn[s][d]];
     }
  }
  return portfolioReturn;
}
```

The question of how to weight a “Hold” recommendation and the extent of the bias towards positive recommendations perhaps warrants some more attention. The results are a bit mixed. For example, a study by Francis & Soffer (1997, pp. 199-200) using a sample of 1483 U.S. stock recommendations during 1988-1991 and a three-point scale cites 46% “Buy”, 43% “Hold” and about 10% “Sells”. On the other hand, a study by Asquith et al (2003, p.10) cites the following proportions in their sample of 1126 U.S. stock recommendations during 1997-1999: 30.8% “Strong buy”, 40.0% “Buy” 28.7% “Hold” and merely 0.5% “Sell/Strong Sell”. Jegadeesh et al (2004), relying on a large international sample with data for all G7 countries for the years 1993-2002, cite the following average for all countries and years: 24.0% “Strong Buy”, 25.1% “Buy”, 37.3% “Hold” and 13.6% “Sell/Strong Sell”. The authors observe that there are substantial differences between the U.S. and other countries: “The frequency of sell recommendations is the lowest in the U.S. In fact, during our sample period, sell recommendations are about four to five times as frequent in the other countries as in the U.S. These results support the general notion that the analysts in the U.S. face the largest conflicts of interest. Therefore, if conflicts of interests were a dominant factor in determining the value of analysts’ forecasts, then we would expect the value of analyst recommendation to be the lowest in the U.S.“ (p. 2)

Researchers frequently use techniques to create relative categories such as “Downgrades” and “Upgrades” with the aim to better capture the value-creating information content in newly published recommendations. For example, Womack (1996) creates event categories by using changes to and from the extremes: either stocks added to or removed from the most attractive ratings (“added-to-buy” and “removed-from-buy”) or stocks added to or removed from the least attractive ratings (“added-to-sell” and “removed-from-sell”). Creating events like this can be useful when you are interested in measuring the impact of recommendations on share prices. Barber et al (2001) takes this approach even further and builds up a whole 5 by 5 matrix to capture all possible changes between recommendations, not only from the extremes. Another technique is to treat upgrades a bit differently from downgrades, for example Green (2006, p. 5) classify recommendation changes as upgrades only when they are shifts to “Strong Buy” or “Buy”, but all downgrades are included regardless of levels. Moreover, to ensure that a recommendation represents a shift in opinion, Green only considers those recommendations, which are not reiterations of the same recommendation or new initiations.

Jegadeesh & Kim (2006) make a strong case for using relative changes in recommendations as a metric. They have found that in a regression model setting with up to 12 other predictive variables, relative *changes* have larger predictive power over future performance of a recommendation than the actual *level* of the change. Further analysis shows that the superior performance of recommendation changes is due largely to the fact that recommendation changes are less affected by the growth bias that afflicts the level variable. The explanation for this is that the level measure suffers more from an analyst bias towards making more positive recommendations for high growth ‘glamour’ stocks as opposed to ‘value’ stocks. “Stocks that receive higher recommendations (as well as more favorable recommendation revisions) tend to have positive momentum (both price and earnings) and high trading volume (as measured by their turnover ratio). They exhibit greater past sales growth, and are expected to grow their earnings faster in the future. /…/ Our results indicate that the economic consequences of sell-side incentives that impair analyst objectivity can also extend to the type of the stocks they choose to recommend. /…/ Growth firms, and firms with higher trading activity, make for more attractive investment banking clients. These firms also tend to be widely held by the institutional clients that place trades with the brokerage houses. Thus, sellside analysts have significant economic incentives to publicly endorse high growth stocks with glamour characteristics. These incentives may cause analysts to, knowingly or otherwise, tilt their attention and recommendations in favor of growth stocks.” (Jegadeesh & Kim, 2006, pp. 1084-1085)

2.2.2 Relative return and risk adjustment

When equity analysts publish recommendations on stocks, they usually restrict their opinion to the industry that they are specialized, i.e. the recommendation is valid relative to its peer stocks in the same industry. The opposite, an absolute recommendation, is of course also possible but is not feasible in practice because it means that the analyst must incorporate his or her opinions about all possible external factors that could affect the return on the stock, and this is usually not within their field of expertise. Therefore, most stock recommendations are effectively relative to other stocks in the same industry classification. The industry scope also makes sense from our analyst evaluation standpoint because it would be undesirable to allow differences between industries to affect the relative evaluation of analysts, for example one industry being more difficult to analyze (e.g. the banking sector where the regulatory framework is currently being completely reworked versus for example utilities which are generally very stable businesses) or one industry simply having a particularly bad or good year.

One way to compare analysts on an equal basis is to simply restrict the comparisons that you make to within an industry. For example, Mikhail et al. (1999) rank analysts within industries by their raw returns as a proxy for analysts' skills in making profitable recommendations. Alternatively, one must find a way to calculate comparable returns. This could be done by simply subtracting the return of a broad industry index from the raw return of the stock, or, all covered stocks could be categorized by their industry sectors and the average return within a sector can be subtracted from the raw returns. This is a very common practice in academic papers (for example Womack, 1996) and the result is called abnormal returns. One question which is sometimes discussed in this context is whether such a comparison index should be equally weighted or weighted by the market capitalization of each stock. A study by Barber et al. (2001) cites two reasons for value-weighting the returns. First, an equal weighting of daily returns is said to lead to portfolio returns that are severely overstated due to the cycling over time of a firm's closing price between its bid and ask (commonly referred to as the bid-ask bounce). Second, a value-weighting is better at capturing the economic significance of the results, since larger and more important firms will be more heavily represented in an aggregated return than those of the smaller firms.

ALGORITHM K

Simple market index risk-adjusted recommendation portfolio returns

```
double[] recommendations_market_adjusted_portfolio_return() {
  double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
  double stockReturn[][] = double[stocks][days]; //daily stock returns
  double marketReturn[] = double[days]; //daily market returns
  bool cover[][][] = bool[analysts][stocks][days]; //stock coverage matrix
  double weights[] = double[5];
  weights[0] = 1; //weight for a ‘Buy’ recommendation
  weights[1] = 0.5; //weight for an ’Outperform’ recommendation
  weights[2] = 0; //weight for a ‘Hold’ recommendation
  weights[3] = -0.5; //weight for a ’Underperform’ recommendation
  weights[4] = -1; //weight for a ‘Sell’ recommendation
  double sum_AR[] = double[analysts]; //sum abnormal returns
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (d = 0; d < days; d++) {
           if (cover[a][s][d])
              sum_AR[a] += weight[recs[a][s][d]]*stockReturn[s][d]] - _
              marketReturn[d];
        }
     }
  }
```

```
  return sum_AR;
}
```

Another related issue is that comparison of returns should reflect in some way the risk (price volatility) associated with the stock. The reason for this is in principle that it matters not only how high the return on a stock is, but also which path the stock price followed to reach that return. This is because it is assumed that investors are risk averse, i.e. if two stocks have the same expected return, investors will prefer the stock with the lowest risk. Hence we need some way to adjust returns for the fact that stocks have different volatility. One straightforward way to achieve this risk-adjustment is the technique used by Mastrapasqua and Bolten (1973, p. 708) and calculate the so called Sharpe-ratio by taking ratio of the abnormal return to the stock volatility as measured by the standard deviation of historical stock prices. For a portfolio p of stocks:

$$AR_p^{adj} = \frac{AR_p}{\sigma_{AR}} = \frac{r_p - r_i}{\sqrt{\operatorname{var}(r_p - r_i)}}, \tag{14}$$

where AR_p^{adj} is the risk-adjusted abnormal return, AR_s is the abnormal return to portfolio p, σ_{AR} is the standard deviations of these abnormal returns, r_p is the raw portfolio return and r_i is the average raw return for industry i.

ALGORITHM L

Sharpe-ratio for recommendations portfolio

```
double[] recommendations_portfolio_Sharpe() {
  double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
  double stockReturn[][] = double[stocks][days]; //daily stock returns
  double marketReturn[] = double[days]; //daily market returns
  bool cover[][][] = bool[analysts][stocks][days]; //stock coverage matrix
  double weights[] = double[5];
  weights[0] = 1; //weight for a ‘Buy’ recommendation
  weights[1] = 0.5; //weight for an ’Outperform’ recommendation
  weights[2] = 0; //weight for a ‘Hold’ recommendation
  weights[3] = -0.5; //weight for a ’Underperform’ recommendation
  weights[4] = -1; //weight for a ‘Sell’ recommendation
  double AR[][] = double[analysts][days]; //abnormal returns
  double sum_AR[][] = double[analysts]; //sum abnormal returns
  double vol[] = double[analysts]; //portfolio abnormal returns volatility
  double Sharpe[] = double[analysts]; //portfolio sharpe ratio
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              AR[a][t] += weight[recs[a][s][t]]*stockReturn[s][t]] - _
              marketReturn[t];
              sum_AR[a] += AR[a][t];
           }
```

```
      }
    }
    Sharpe[a] = sum_AR[a]/stdev(AR[a][]);
  }
  return Sharpe;
}
```

A slightly different way of achieving a similar result would be to follow the example of Francis & Soffer (1997), who adjust for risk based on forming portfolios of companies with equal standard deviation (within the same industry).

Another common way to adjust for risk is to employ the capital asset pricing model (CAPM) as in for example Bjerring et al (1983). This model postulates that a stock's expected return above the risk-free rate (excess returns) should be proportional to its systematic risk, measured as the ratio of the stock's covariance with the market index to the variance of the market index, better known as the stock's beta (β). CAPM can equivalently be expressed as linear time-series OLS regression of portfolio p excess return on market excess return as follows:

$$\varepsilon_{p,t} = (r_{p,t} - r_{f,t}) - (\alpha_p + \beta_p(r_{m,t} - r_{f,t})) , \tag{15}$$

where $r_{p,t}$ is the raw return to portfolio p, $r_{f,t}$ is the risk-free rate, $r_{m,t}$ is the market index raw return, α_p is the estimated intercept (Jensen's alpha), β_p is the estimated portfolio beta and $\varepsilon_{p,t}$ is the regression error term. The sum of the error terms can be used as a measure of the abnormal return over the theoretical expected return according to the CAPM.

ALGORITHM M

Risk-adjusted recommendations portfolio returns using CAPM

```
double[] recommendations_CAPM_adjusted_portfolio_return() {
  double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
  double stockReturn[][] = double[stocks][days]; //daily stock returns
  double marketReturn[] = double[analysts][days]; //daily market returns
  double riskFreeRate[] = double[days]; //risk-free rate
  bool cover[][][] = bool[analysts][stocks][days]; //stock coverage matrix
  double weights[] = double[5];
  weights[0] = 1; //weight for a ‘Buy’ recommendation
  weights[1] = 0.5; //weight for an ’Outperform’ recommendation
  weights[2] = 0; //weight for a ‘Hold’ recommendation
  weights[3] = -0.5; //weight for a ’Underperform’ recommendation
  weights[4] = -1; //weight for a ‘Sell’ recommendation
  double returns[][] = double[analysts][days]; //portfolio returns
  double epsilon[][] = double[analysts][days]; //CAPM regression residuals
  double sum_res[] = double[analysts]; //sum portfolio residuals
  for (a = 0; a < analysts; a++) {
    for (s = 0; s < stocks; s++) {
      for (t = 0; t < days; t++) {
        if (cover[a][s][t]) {
```

```
            returns[a][t] += weight[recs[a][s][t]]*stockReturn[s][t]] - _
            riskfreeRate[t];
         }
      }
    }
  }
  for (a = 0; a < analysts; a++) {
    Do OLS regression with returns as dependent variable and marketReturn -
    riskFreeRate as explanatory variable
    Save residuals in epsilon[a][]
  }
  for (a = 0; a < analysts; a++) {
    for (t = 0; t < days; t++) {
      sum_res[a] += epsilon[a][t];
    }
  }
  return sum_res;
}
```

Another way of adjusting for risk is based on the premise that companies of similar size should have similar risk characteristics. Barber et al. (2001) compares returns within decile portfolios formed by ranking stocks by their company market cap, and Green (2006) forms portfolios both for size and industry. Womack (1996) uses a more complex way of taking size into account, which is based on the Fama-French three-factor model. This model is based on the idea that uses the same market excess return factor as the CAPM and two additional factors: the first to approximate excess return for smaller companies (small market cap) over big companies and secondly excess return for companies with a low book-to-market value (low valuation) over companies with a high book-to-market value.

$$\varepsilon_{p,t} = (r_{p,t} - r_{f,t}) - (\alpha_p + \beta_{m,p}(r_{m,t} - r_{f,t}) + \beta_{s,p} SMB + \beta_{u,p} HML)\,, \tag{16}$$

where $\beta_{m,p}$ is the three-factor model factor on the market excess return (which is not the same as the CAPM beta as there are additional regression factors), $\beta_{s,p}$ is the loading for the factor capturing excess returns of small caps over big caps (SMB) and $\beta_{u,p}$ is the loading for the factor capturing excess returns of value stocks over growth stocks (HML).[4]

ALGORITHM N

Risk-adjusted recommendations portfolio returns using the Fama-French three-factor model

```
double[] recommendations_Fama-French_3Factor_adjusted_portfolio_return() {
  double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
  double stockReturn[][] = double[stocks][days]; //daily stock returns
  double marketReturn[][] = double[days]; //daily market returns
  double SMB[] = double[days]; //small-minus-big market capitalisation factor
```

[4] See Appendix B for an explanation on how these factors are calculated.

```
  double HML[] = double[days]; //high-minus-low book-to-market factor
  double riskFreeRate[] = double[days]; //risk-free rate
  bool cover[][][] = bool[analysts][stocks][days]; //stock coverage matrix
  double weights[] = double[5];
  weights[0] = 1; //weight for a ‘Buy’ recommendation
  weights[1] = 0.5; //weight for an ’Outperform’ recommendation
  weights[2] = 0; //weight for a ‘Hold’ recommendation
  weights[3] = -0.5; //weight for a ’Underperform’ recommendation
  weights[4] = -1; //weight for a ‘Sell’ recommendation
  double returns[][] = double[analysts][days]; //portfolio returns
  double epsilon[][] = double[analysts][days]; //CAPM regression residuals
  double sum_res[] = double[analysts]; //sum portfolio residuals
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              returns[a][t] += weight[recs[a][s][t]]*stockReturn[s][t]] - _
              riskFreeRate[t];
           }
        }
     }
  }
  for (a = 0; a < analysts; a++) {
     Do regression with returns as dependent variable and marketReturn - riskFreeRate,
     SMB and HML as explanatory variables
     Save residuals in epsilon[a][]
  }
  for (a = 0; a < analysts; a++) {
     for (t = 0; t < days; t++) {
        sum_res[a] += epsilon[a][t];
     }
  }
  return sum_res;
}
```

Finally, Barber (2001) takes the complexity one step further and uses an extended Fama-French three-factor model with an additional price momentum factor. The rationale for using price momentum comes from Jagadeesh and Titman (1993) who show that the strategy of buying stocks that have performed well in the recent past and selling those that have performed poorly generates significant positive returns over 3- to 12-month holding periods.

$$\varepsilon_{p,t} = (r_{p,t} - r_{f,t}) - (\alpha_p + \beta_{m,p}(r_{m,t} - r_{f,t}) + \beta_{s,p}SMB_t + \beta_{u,p}HML_t + \beta_{mom,t}MOM_t), \tag{17}$$

where $\beta_{mom,t}$ is the loading for the factor capturing excess returns of stocks with high positive momentum (a clear positive price trend) over stocks with high negative momentum (a clear negative trend).

ALGORITHM O

Risk-adjusted recommendations portfolio returns using the Fama-French three-factor model, with a fourth momentum factor

```
double[] recommendations_Fama-French_Momentum_adjusted_portfolio_return() {
  double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
  double stockReturn[][] = double[stocks][days]; //daily stock returns
  double marketReturn[] = double[days]; //daily market returns
  double SMB[] = double[days]; //small-minus-big market capitalisation factor
  double HML[] = double[days]; //high-minus-low book-to-market factor
  double Momentum[] = double[days]; //high-minus-low momentum factor
  double riskFreeRate[] = double[days]; //risk-free rate
  bool cover[][][] = bool[analysts][stocks][days]; //stock coverage matrix
  double weights[] = double[5];
  weights[0] = 1; //weight for a ‘Buy’ recommendation
  weights[1] = 0.5; //weight for an ’Outperform’ recommendation
  weights[2] = 0; //weight for a ‘Hold’ recommendation
  weights[3] = -0.5; //weight for a ’Underperform’ recommendation
  weights[4] = -1; //weight for a ‘Sell’ recommendation
  double returns[][] = double[analysts][days]; //portfolio returns
  double epsilon[][] = double[analysts][days]; //CAPM regression residuals
  double sum_res[] = double[analysts]; //sum portfolio residuals
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              returns[a][t] += weight[recs[a][s][t]]*stockReturn[s][t]] - _
              riskFreeRate[t];
           }
        }
     }
  }
  for (a = 0; a < analysts; a++) {
     Do regression with returns as dependent variable and marketReturn - riskFreeRate,
     SMB, HML and Momentum as explanatory variables
     Save residuals in epsilon[a][]
  }
  for (a = 0; a < analysts; a++) {
     for (t = 0; t < days; t++) {
        sum_res[a] += epsilon[a][t];
     }
  }
  return sum_res;
}
```

2.2.3 Portfolio effects

One important aspect we have not discussed yet is our goal to find a method of evaluation that is fair in the sense that it should reward skill but not luck. In a portfolio context, this is a related but different issue to the problem of risk adjusting returns, which we have just explored. For example, if an analyst's whole sector has had a positive return on average, then, even after calculating abnormal returns within the sector, an analyst could potentially get the recommendations completely wrong on some of the stocks, and still earn a high portfolio return as long as he or she got the recommendation right for some of the best performing stocks they are covering.

TABLE 1

Example of portfolio effects. Analyst A's portfolio is more profitable than analyst B's portfolio, while Analyst B's assessments of individual stock are clearly superior to those of Analyst A.

Stock	Analyst A	Analyst B	Return
1	Buy	Buy	20%
2	Hold	Hold	15%
3	Buy	Hold	10%
4	Hold	Hold	5%
5	Buy	Sell	-1%

This is easier to see if we illustrate this with an example, consider two analysts, which cover the same four stocks. To simplify matters, assume that the analysts don't change their ratings during the year. Consider that the stocks perform as in Table 1. One straightforward portfolio formation approach (which we will adopt here) is for the analyst's portfolio to hold a double weighting in each of his Buy ratings, a single weighting in each of his Hold ratings, and zero weighting in any of his Sell ratings. We then calculate a form of relative return by subtracting the return of the equally weighted analyst's coverage universe from the hypothetical portfolio's return.

Using this methodology, the portfolio return for analyst A is 15.6%. The return of the coverage universe is 9.8%. So analyst A's value added is 5.8%. Analyst B, on the other hand, earned a portfolio return of only 14% with the same 9.8% coverage universe return. Therefore analyst B's added value is only 4.2%. So measured by this metric we would conclude that analyst A is the better analyst.

However, analyst A's three Buy-rated stocks had an average return of 9.7%, while his two Holds had an average return of 10%. So one could argue that analyst A's stock selection is fundamentally flawed, despite his apparent high portfolio return. Analyst B, on the other hand, had one Buy rating which returned 20%, three Holds which averaged 10%, and one Sell which returned -1%. Analyst B's stock selection is close to perfect, despite an unimpressive portfolio return. This conundrum arises because the return measure is asked to measure two things with one number: predicting the general market direction and stock selection. These two are quite different aspects of analyst skill.

Mastrapasqua and Bolten (1973) recognize that these issues need to be addressed: "Another difficulty is the failure [of traditional techniques] to disentangle general market movements from the analyst's performance and to consider his forecasts [forecasts here refers to returns, not company earnings]. Frequently, an analyst may demonstrate forecasting accuracy in spite of poor selection methods and incorrect forecasting simply because of a rising market. A more appropriate measure should distinguish the analyst's record from market influences as well as consider the analyst's pre-disposition to that market." (p. 708). They then propose a method to evaluate analyst recommendations based on probabilities and a clever use of Bayes' theorem as follows. Starting with the probability for abnormal returns

$$\mathrm{P}(r_a \geq r_e)\,, \tag{18}$$

where r_a is the actual return to a portfolio of covered stocks and r_e is expected risk-adjusted return for the recommended portfolio. (18) is an objective or ex-ante probability of recommendation profitability. Next, the analyst's recommendations are incorporated into (18) to derive the conditional probability, i.e. the probability of the portfolio return exceeding or equaling the market return given an increase in the value of the market portfolio was predicted by the analyst:

$$\mathrm{P}(r_a \geq r_e \mid U_m^*)\,, \tag{19}$$

where U_m^* is the analyst's prediction of an increase in the value of the suggested portfolio for each period. Then, using Bayes' theorem

$$\mathrm{P}(r_a \geq r_e \mid U_m^*) = \frac{\mathrm{P}(r_a \geq r_e)\,\mathrm{P}(U_m^* \mid r_a \geq r_e)}{\mathrm{P}(r_a \geq r_e)\,\mathrm{P}(U_m^* \mid r_a \geq r_e) + \mathrm{P}(r_a < r_e)\,\mathrm{P}(U_m^* \mid r_a < r_e)}, \tag{20}$$

where $\mathrm{P}(U_m^* \mid r_a \geq r_e)$ is the ex-post (revised) probability that given a portfolio return greater than or equal to the expected risk adjusted portfolio return, the analyst had predicted such an increase. Finally, the revised probability is compared to the prior probability to determine how effective the recommendations were:

$$\mathrm{P}(r_a \geq r_e \mid U_m^*) \,/\, \mathrm{P}(r_a \geq r_e) \tag{21}$$

When this ratio is 1, the analyst's recommendations did not add any value, because the performance of the portfolio is exactly what would have been expected just by holding all the stocks under the analyst's coverage without taking into account the recommendations on them. However, as the ratio becomes larger (smaller) than 1, following the recommendations on these stocks adds (subtracts) more value for an investor holding the portfolio.

ALGORITHM P

Portfolio effects ratio

```
double[] portfolio_effects_ratio() {
  double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
  double portfolio_AR[][] = double[analysts][days]; //abnormal portfolio returns
  bool cover[][][] = bool[analysts][stocks][days]; //stock coverage matrix
  double weights[] = double[5];
  weights[0] = 1; //weight for a ‘Buy’ recommendation
  weights[1] = 0.5; //weight for an ’Outperform’ recommendation
  weights[2] = 0; //weight for a ‘Hold’ recommendation
  weights[3] = -0.5; //weight for a ’Underperform’ recommendation
  weights[4] = -1; //weight for a ‘Sell’ recommendation
  int pos_AR[] = int[analysts] //count for positive abnormal returns
  double P_pos_AR[] = double[analysts] //corresponding probability
  int pos_rec[] = int[analysts] //count for positive average recommendation
  int pos_AR_|_pos_rec[] = int[analysts];     //count positive abnormal
                                              //returns conditional on positive
                                              //average rec
  double P_pos_AR_|_pos_rec[] = double[analysts]: //corresponding probability
  int pos_rec_|_pos_AR[] = int[analysts];        //count positive abnormal
                                                 //rec when there was
                                                 //positive AR
  double P_pos_rec_|_pos_AR[] = double[analysts]; //corresponding probability
  int neg_AR_|_pos_rec[] = int[analysts];        //count negative abnormal
                                                 //rec conditional on negative
                                                 //average rec
  double P_neg_AR_|_pos_rec[] = double[analysts]; //corresponding probability
  int cases[] = int[analysts]; //total number of evaluated cases per analyst
  int coveredStocks[][] = int[analysts][days]; //covered stocks over time
  double avg_rec[][] = double[analysts][days];
  double metric[] = double[analysts];
  for (a = 0; a < analysts; a++) {
     for (t = 0; t < days; t++) {
        for (s = 0; s < stocks; s++) {
           if (cover[a][s][t]) {
              coveredStocks[a][t]++;
              cases[a]++;
              avg_rec[a][t] += recs[a][s][t];
        }
        avg_rec[a][t] /= coveredStocks[a][t]; //average portfolio rec at time t
        if (avg_rec[a][t] >= 0) {
           pos_rec[a]++;
           if (portfolio_AR[a][t] >= 0)
              pos_AR_|_pos_rec[a]++;
        }
        if (portfolio_AR[a][t] >= 0) {
           pos_AR[a]++;
           if (avg_rec[a][t] >= 0)
              pos_rec_|_pos_AR[a]++;
        }
        else {
           if (avg_rec[a][t] >= 0)
              neg_AR_|_pos_rec++;
        }
     }
```

```
  }
  for (a = 0; a < analysts; a++) {
     P_pos_AR[a] = pos_AR[a]/cases;
     P_pos_AR_|_pos_rec[a] = pos_AR_|_pos_rec[a]/pos_rec[a];
     P_pos_rec_|_pos_AR[a] = pos_rec_|_pos_AR[a]/pos_AR[a];
     P_neg_AR_|_pos_rec[a] = neg_AR_|_pos_rec[a]/pos_rec[a];
     metric[a] = P_pos_rec_|_pos_AR[a] / (P_pos_AR[a] * P_pos_rec_|_pos_AR[a] + _
                  (1-P_pos_AR[a])* P_neg_AR_|_pos_rec[a]);
  }
  return metric;
}
```

In a similar way we can also calculate whether the analyst's downside recommendations add any value.

$$\mathrm{P}(r_a \geq r_e \mid D_m^*), \tag{22}$$

where D_m^* is the analyst's prediction of a decrease in the value of the suggested portfolio for each period under consideration. Calculating both these measures allows us to determine an analyst's bias as follows. If he or she issues more profitable Buy recommendations than Sell recommendations we will find that

$$\mathrm{P}(r_a \geq r_e \mid U_m^*) > \mathrm{P}(r_a \geq r_e \mid D_m^*), \tag{23}$$

and the opposite will be true for an analyst whose Sell recommendations are likely to be more profitable than his or her Buy recommendations.

3 Equity analyst evaluation in industry practice

This section aims to provide the reader with an idea about common industry practices in evaluation of equity analysts. Although all investment banks no doubt have developed their own proprietary evaluation procedures in place for their equity research departments, obtaining such information and analyzing it, for reasons of brevity, falls outside the scope of this thesis. We limit ourselves here to describing the most important publicly available equity analyst rankings and one example from one bank.

3.1 Institutional Investor Research Team Rankings

The success of equity analysts is ultimately depending on whether they bring in business to their firms, and by extension whether clients understand their advice, find it useful and are prepared to act on it. The importance of clients' opinions is expressed nowhere else more comprehensively than in the yearly rankings of equity research teams published by the periodical Institutional Investor. The rankings are based on qualitative surveys sent out to investment managers at institutions worldwide and there are several rankings, each focusing on a specific geographical region (All-America, All-Asia, All-Europe, and so on).

The surveys are carried out as follows. The respondents can score individual analysts and/or research teams on a scale 1 to 10 on four evaluation criteria: estimation accuracy, stock-picking ability, quality of written reports and overall service. They are free to do this for as many of about 50 industry sectors, and country sectors as they see fit. There are also categories for "Economics & Strategy" but those categories fall slightly out-of-scope for our purposes since they are not relevant to the "classical" equity analyst, but rather for macroeconomists or derivative strategists. Each vote is weighted based on the respondent's equity assets under management in the relevant stock universe (e.g. European equities) and points are given to individuals and/or entire teams the respondents have ranked in first, second, third or fourth place. Results are reviewed by an independent auditor. Firms are then ranked within sectors and regions as well as on the total score over all sectors and regions. (Institutional Investor, 2013)

The rankings have been conducted since 1985 and are widely regarded as one of the most important evaluation measures within the equity research industry. This is understandable since it reflects the actual opinions of the people responsible for managing a sizeable part of the equity capital in the world. For example for the 2012 All-Europe Equity Research survey, Institutional Investor consulted some 2200 money managers from 760 institutions, managing total assets of $5.7 trillion in European equities, which represents about 84 percent of the MSCI Europe index's market capitalisation of $6.8 trillion. An example which illustrates the ranking's importance in relation to other factors is Stickel (1992), who cites a Wall Street Journal article: "At *most* firms, the important factors affecting pay [for equity analysts] are an evaluation of the analyst by the brokerage sales force, standing in the Institutional Investor poll, and job offers from competitors. A smaller set of firms expand the set of factors to include investment banking business generated, trading volume in recommended stocks, and the success of buy and sell recommendations. Accuracy of earnings forecasts is rarely an explicit factor, but is subsumed in the other factors. As one analyst put it, 'If your estimates aren't accurate, nobody's going to buy your stocks'." (p. 1811)

Loh & Mian (2006) point out that it is usually assumed that superior analysts score high on all four of the Institutional Investor rankings four evaluation criteria, implying a correlation between them. Obviously, the criteria "Quality of written reports" and "Overall service" are of a subjective nature and thus cannot be measured without great difficulty. To add to the list of immeasurable items, well-honed communication skills, an extensive network of personal contacts with clients as well as such esoteric qualities as charisma can also be of great importance. The fact that several of the most important aspects of an analyst's work cannot be measured and quantified could go a long way in explaining the relative scarcity of research focused explicitly on the subject of analyst evaluation. Nevertheless, a structured method to evaluate those criteria which can indeed be quantified can be a useful component in a comprehensive evaluation procedure. We will now turn our attention to an example of such a quantitative method.

3.2 Financial Times/Starmine Global Analyst Awards

The Financial Times in cooperation with the research analytics company StarMine publishes an annual ranking based on quantitative data for the top equity analysts and brokerages in the US, Europe and Asia, respectively. For each of these regions, awards are presented to the top three stock pickers and earnings estimators in each industry, to the top 10 stock pickers and earnings estimators overall, and to the 10 brokerage firms that have won the most individual analyst awards. Stock picking in this context refers to recommendation performance. There is also an award for the most Productive Broker which goes to the firm which has the highest number of individual awards in relation to the number of analysts and a Top Global Broker award for the firm with the best cumulative results for America, Asia and Europe combined. StarMine has published analyst rankings for a number of years and their methodology in quantitative measurement of research performance has become widely respected. Although their methodology is proprietary, some details about how the rankings are compiled were described following the 2012 awards.

3.2.1 Estimates

The earnings estimation ranking is based on StarMine's proprietary metric Single-stock Estimate Score (SES). The measure can range from 0 to 100, with 50 representing the average analyst. To receive a score higher than 50, an analyst must make estimates that are both significantly different from and more accurate than other analysts' estimates. SES reportedly incorporates several factors: the analyst's absolute forecast error, the relative error compared with other analysts, the variance of the analysts' errors, the timing of the estimates and the absolute value of the actual earnings for the stock. SES is computed on a daily basis and aggregated to provide scores on individual stocks, industries, and the analyst overall.

3.2.2 Recommendations

For the industry stock picking awards, the method is as follows. Analysts are ranked according to their Industry Excess Return, computed from a portfolio simulation technique that measures each analyst's recommendations performances relative to a market capitalization weighted portfolio of all the stocks in a given industry and region. StarMine builds portfolios based on each analyst's recommendations. The portfolio is rebalanced each month and whenever the analyst adds coverage, drops coverage or changes a rating. For each "Buy" recommendation, the portfolio is one unit long the stock and one unit short the benchmark. This means that the analyst's portfolio return will increase (decrease) by the amount the stock outperforms (underperforms) the benchmark. "Strong buys" get a larger investment of two units long the stock and two units short the benchmark. "Holds" invest one unit in the benchmark. "Sells" are the reverse: long one unit in the benchmark and short one unit in the stock. "Strong sells" get two units long the benchmark and short one unit in the stock. The portfolio is then opportunity-adjusted to account for differences between analysts' coverage. Exactly how this adjustment is done is not explained. The adjusted return is the analyst's Industry Excess Return. This metric is then weighted by the number of stocks the analyst covers in each industry and aggregated into overall excess return (for analysts that cover more than one industry).

3.3 Existing techniques used at one bank

The following describes the existing techniques for evaluating equity analysts at one bank.

3.3.1 Estimates

For estimates the bank relies on three metrics. The first one is called "Average score" and is defined as the average error by one analyst divided by average consensus error on the stock. A score below 1.0 indicates better estimates than consensus.

$$\overline{Z}_a^{avg} = \sum_{s=1}^{N} \frac{\overline{e}_s^{a} \;/\; \overline{e}_s^{consensus}}{N} \,, \tag{24}$$

where e is defined as ASE in (2) and $\overline{e}_s^{a}$ is the average error by analyst a for stock s.
One obvious problem with this metric is that it weights equally all errors, not taking into account the decreasing uncertainty of estimates. It does not indicate whether estimates have improved over the reporting year.

ALGORITHM Q

"Average score" used by one bank to evaluate forecasts

```
double[] average_score() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double F_cons[][][] = double[stocks][days]; //consensus forecasted EPS
  double ASE_bar[][] = double[analysts][stocks]; //analyst average ASE
  double ASE_cons_bar[] = double[stocks]; //consensus average ASE
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  double Z[] = double[analysts];
  int coveredStocksInPeriod[] = int[analysts]; //count covered stocks over t
  bool covered;
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        covered = false;
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              if (!covered) //just include in consensus average once
                 ASE_cons_bar[s] += abs(A[s][t] - F_cons[s][t])/abs(A[s][t]);
              covered = true;
              ASE_bar[a][s] += abs(A[s][t] - F[a][s][t])/abs(A[s][t]);
           }
        }
        if (covered)
           coveredStocksInPeriod[a]++;
        Z[a] += ASE_bar[a][s]/ASE_cons_bar[s]; //same denominator so it disappears
     }
     Z[a] /= coveredStocksInPeriod[a];
  }
  return Z;
}
```

The second metric is called "Stdev score" and is calculated as the standard deviation of the analyst's errors divided by the standard deviation of the consensus errors. A score below 1.0 indicates less fluctuating estimates than consensus, i.e. higher level of stability during the year

$$\overline{Z}_a^{stdev} = \sum_{s=1}^{N} \frac{\text{stdev}(e_s^a) \,/\, \text{stdev}(e_s^{consensus})}{N} \tag{25}$$

ALGORITHM R

"Stdev score" used by one bank to evaluate forecasts

```
double[] stdev_score() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double F_cons[][][] = double[stocks][days]; //consensus forecasted EPS
  double ASE[][][] = double[analysts][stocks][days]; //analyst ASE
  double ASE_stdev[][] = double[analysts][stocks]; //stdev for analyst ASE
  double ASE_cons[][] = double[stocks][days]; // consensus ASE
```

```
  double ASE_cons_stdev[][] = double[stocks][days]; // stdev for consensus ASE
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  double Z[] = double[analysts];
  int coveredStocksInPeriod[] = int[analysts]; //count covered stocks over t
  bool covered;
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        covered = false;
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              covered = true;
              ASE[a][s][t] = abs(A[s][t] - F[a][s][t])/abs(A[s][t]);
              ASE_cons[s][t] = abs(A[s][t] - F_cons[s][t])/abs(A[s][t]);
           }
        }
        if (covered)
           coveredStocksInPeriod[a]++;
        Z[a] = stdev(ASE[a][s][])/stdev(ASE_cons[s][]);
     }
     Z[a] /= coveredStocksInPeriod[a];
  }
  return Z;
}
```

The third metric is called "Sum of errors score" and simply sums the forecast errors by the analyst and divides by the sum of consensus errors. The bank would also use this metric on a stock-by-stock basis and let a ratio over 1.0 constitute a "win" over consensus, and then simply count the number of "wins" and calculate each analyst's ratio of "wins".

$$Z_a^{sum} = \frac{\sum_{n=1}^{N} e_s^a}{\sum_{n=1}^{N} e_s^{consesnsus}} \tag{26}$$

ALGORITHM S

"Sum of errors score" used by one bank to evaluate forecasts

```
double[] sum_errors_score() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double F_cons[][][] = double[stocks][days]; //consensus forecasted EPS
  double ASE_sum[][] = double[analysts][stocks]; //analyst average ASE
  double ASE_cons_sum[] = double[stocks]; //consensus average ASE
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  double Z[] = double[analysts];
  bool covered;
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        covered = false;
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
```

```
            if (!covered) //just include in consensus sum once
               ASE_cons_sum[s] += abs(A[s][t] - F_cons[s][t])/abs(A[s][t]);
            covered = true;
            ASE_sum[a][s] += abs(A[s][t] - F[a][s][t])/abs(A[s][t]);
         }
      }
      if(covered)
         Z[a] += ASE_sum[a][s]/ASE_cons_sum[s];
   }
 }
 return Z;
}
```

3.3.2 Recommendations

The technique employed by the bank to evaluate recommendations again is based on comparison with consensus and the basic long-only strategy. The bank uses only four levels of recommendations: “Buy”, “Accumulate”, “Reduce” and “Sell” and thus there is no “Hold” recommendation. The bank weights the recommendations as follows: “Buy” = 1 unit long, “Accumulate” = 0.5 unit long, “Reduce” = 0.5 unit short and “Sell” = 1 unit short. To be able to compare with consensus, which has five recommendation levels, the bank decided to count a consensus “Hold” as corresponding to its “Accumulate” recommendation. Recommended portfolios are formed based on these weightings, and raw (non risk-adjusted) portfolio returns are calculated and compared to raw returns of consensus portfolios and a long-only strategy portfolio where all stocks are given a weighting of 1 unit long. The bank considers that beating the consensus portfolio with more than 1% constitutes a “win” and then simply count the number of “wins” and calculate each analyst’s ratio of “wins”. The same approach is then repeated for comparison with the long-only alternative.

ALGORITHM T

Wins over consensus as used by one bank to evaluate recommendations

```
double[] recommendations_consensus_comparison() {
 double recs[][][] = double[analysts][stocks][days]; //recommendations indexed 0 to 4
                                                          //except no 2 (Hold)
 double recs_cons[][] = double[stocks][days];         //consensus recommendations
                                                     //indexed 0 to 4
 double stockReturn[][] = double[stocks][days]; //daily stock returns
 bool cover[][][] = bool[analysts][stocks][days]; //stock coverage matrix
 double weights[] = double[5];
 weights[0] = 1; //weight for a ‘Buy’ recommendation
 weights[1] = 0.5; //weight for an ’Accumulate’ or ‘Outperform’ recommendation
 weights[2] = 0.5 //weight for a ‘Hold’ recommendation
 weights[2] = -0.5; //weight for a ’Reduce’ or ‘Underperform’ recommendation
 weights[3] = -1; //weight for a ‘Sell’ recommendation
 double AR_sum[][] = double[analysts][stocks]; //sum analyst portfolio returns
 double AR_cons_sum[][] = double[analysts][stocks]; //sum consensus portfolio returns
 bool covered;
 int coveredStocks[a] = int[analysts]; //covered stocks at any time in period
 double winRatio[] = double[analysts];
```

```
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        covered = false;
        for (t = 0; t < days; d++) {
           if (cover[a][s][t])
              if (!covered) {
                 coveredStocks[a]++;
                 AR_cons_sum[a][s] += getConsensusWeight(recs_cons[s][t]) * _
                 * stockReturn[s][t]]; //consensus weight must be averaged from 2
                                         //weight points
              }
              AR_sum[a][s] += weight[recs[a][s][t]]*stockReturn[s][t]];
        }
        if(covered)
        if (AR_sum[a][s] - AR_cons_sum[a][s] ~ 0.01)
           winRatio[a]++;
     }
     winRatio[a] /= coveredStocks[a];
  }
  return winRatio;
}
```

4 Proposed solution

In this section we will endeavor to suggest new improved algorithms to evaluate equity analysts based on the theoretical techniques described in section 2 and the industry practices described in section 3. We do this from the perspective of a bank, with the aim that the algorithms should be useful for comparing analysts to their competitors covering the same industries at other banks, but also between analysts at the same bank, covering different industries.

4.1 Estimates

We want to suggest that the main aspects of evaluating estimates are related to the following areas:

- Precision compared to relevant peers
- Consistency of precision - difficulty of achieving precision changes over time as the uncertainty in the remaining unpublished company information for the accounting year decreases
- Interpreting new information - timing of estimate revisions can reveal which analysts are good at translating new information into better precision (leading/herding)

For evaluating precision as such it seems clear to us that a very good candidate is Algorithm F, which has several attractive traits. It incorporates the consensus, which lets us compare individual analysts to their competitors, without the need for data on all analysts, and it does not build on a ranking. Ranking, although useful in some ways, does not tell us how much better one higher ranked individual is over another lower ranked individual, just that he or she is better. So with ranking there is some loss of information. Also, Algorithm F automatically adjusts for stock-year effects, which is an attractive feature.

We suggest that the problem of taking consistency of estimates into account may be addressed in conjuncture with the problem of reduced uncertainty, all in one fell swoop. Clearly an analyst who estimates correctly the annual earnings per share already in the first quarter deserves higher praise than an analyst who manages this only in the fourth quarter. One way to do this would be through Algorithm G, which uses a regression on time as an explanatory variable to obtain residuals, which are uncorrelated to the amount of time left to the reporting date (and thus uncertainty) in the data. By utilizing Algorithm G, if the fit of the regression is good, this should indicate that the estimates have not fluctuated much over the year. Therefore, in connection with "good" b_0 and b_1-coefficients, a higher R-square measure for the regression should indicate a higher consistency of accuracy. We can interpret the size of the intercept as follows: Assuming that the accuracy of estimates increases over time (as more information becomes available), an imaginary regression line should cross the Y-axis close to zero, i.e. the closer to zero the intercept a is, the better the estimate. The regression sloop can be interpreted as follows: If the accuracy of the estimate is high already at the start of the year, a lower slope should denote higher accuracy. We propose a small variation to Algorithm G, replacing the *ASE* metric by *APE*, to avoid problems with inflated error metrics in cases where actual earnings per share are close to zero. Adjusted algorithm below.

ALGORITHM U

Adjusted absolute scaled forecast error with time regression using APE from equation (3) instead of ASE from equation (2).

```
double[][] absolute_percentage_error_time_regression_metrics() {
  double A[][] = double[stocks][days]; //actual EPS reported by the company
  double F[][][] = double[analysts][stocks][days]; //forecasted EPS by analysts
  double APE[][][] = double[analysts][stocks][days]; //absolute percentage error
  double P[][] = double[stocks][days]; //stock price
  bool cover[][][] = bool[analysts][stocks][days]: //stock coverage matrix
  int daysUntilReport[][][] = int[analysts][stocks][days]: //days until EPS report
  double epsilon[][] = double[analysts][stocks]; //residuals
  double intercept[][] = double[analysts][stocks]; //intercept
  double metrics[] = double[analysts][3];  //residuals, intercept and slope metrics
  double averageCoveredStocks[] = double[analysts]; //average covered stocks over time
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        for (t = 0; t < days; t++) {
           if (cover[a][s][t]) {
              APE[a][s][t] = abs(A[s][t] - F[a][s][t])/P[s][t];
              averageCoveredStocks[a]++;
           }
        }
     }
     averageCoveredStocks[a] /= days;
  }
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        Do regression with APE as dependent variable and daysUntilReport as _
        explanatory variable, with intercept.
        Save residuals in epsilon[a][s]
        Save intercept in intercept[a][s]
        Save slope in slope[a][s]
     }
  }
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        metric[a][0] += abs(epsilon[a][s])/averageCoveredStocks[a];
        metric[a][1] += intercept[a][s]/averageCoveredStocks[a];
        metric[a][2] += epsilon[a][s]/averageCoveredStocks[a];
     }
  }
  return metric;
}
```

Finally, to address the problem of leading/herding, it is straightforward to see that Algorithm I and the *LFR* metric can be calculated. We suggest that, if detailed data on forecast revisions by competing analysts is not available, that changes to the consensus can be used as a proxy for individual revisions by other analysts.

ALGORITHM V

Leader-follower ratio with consensus revisions instead of individual analyst forecast revisions

```
double[][][] leader-follower_ratio () {
  int F_dates[][][] = int[analysts][stocks][]; //dates with a forecast revision
  int F_cons_revisions[] = int[stocks]; //number of consensus forecast revisions
  int F_cons_dates[][] = int[stocks][] //dates with consensus forecast revision
  double LFR[] = double[analysts]; //leader-follower metric
  bool cover[][][] = bool[analysts][stocks]: //stock coverage matrix
  int T0; int T1;
  int currentDate; //the date for the forecast being evaluated
  int previousDate; //the preceding forecast date by the same analyst
  int nextDate; //the next forecast date by the same analyst
  for (a = 0; a < analysts; a++) {
     for (s = 0; s < stocks; s++) {
        if (cover[a][s]) { //if covers stock
           previousDate = null;
           T0 = 0;
           T1 = 0;
           for (d = 0; d < F_revisions[a][s]; d++) {
              currentDate = F_dates[a][s][d];
              if (d < F_revisions[a][s])
                 nextDate = F_dates[a][s][d+1];
              else
                 nextDate = null;
                 for (dd = 0; dd < F_cons_revisions[s]; d++) {
                    if (F_cons_dates[s][dd] < currentDate && _
                           (F_cons_dates[s][dd] > previousDate || previousDate == null))
                           T0 += F_cons_dates[s][dd] - currentDate;
                    else {
                       if (F_cons_dates[s][dd] > currentDate && _
                          (dd < nextDate || nextDate == null))
                          T1 += currentDate - F_cons_dates[s][dd];
                    }
                 }
              }
           }
           previousDate = currentDate;
           }
        }
     }
  LFR[a] = T0/T1;
  }
  return LFR;
}
```

We do not see an absolute need to integrate or weight together the separate metrics produced by these three algorithms (unless the goal of the evaluation is to rank the analysts). A different approach could be to look at all the metrics in conjunction as a type of "scorecard" and this should give a well balanced view on which aspects of forecasting the analyst is good at, and where efforts to improve can be directed.

4.2 Recommendations

For recommendations, we suggest that the following aspects of evaluation are important and should be considered

- Profitability of recommendations compared to relevant peers
- Risk adjustment of returns – stripping out returns due to known risk factors
- Taking into account portfolio effects – separating stock picking ability from ability to predict general market direction

The basic methodology of portfolio formation seems to be a standard so we will adopt this approach, with the standard scale of five levels and apply the weights as follows: “Buy” 1.0, “Outperform” 0.5, “Hold” 0, “Underperform” -0.5 and “Sell” -1.0. As with estimates, we suggest that recommendation performance should be evaluated against a relevant benchmark on a stock-by-stock basis. This can be achieved by benchmarking against the consensus recommendation, by calculating a return differential. On an aggregate level this differential will contain more information than a ranking or a count of “wins”. The differential should be taken at the level being compared, so if we are evaluating on the single stock level, then the differential should be return if one were to invest according to the analyst’s recommendations on that stock minus the return if one were to invest according to the consensus recommendation on that stock. If evaluating on the portfolio level then the differential would be the return of the portfolio of stocks weighted by the analyst’s recommendations minus the return of the portfolio of stocks weighted by the consensus recommendations. As with estimates, the advantage of using the consensus is that we do not need to have data for all competing analysts.

To properly risk adjust the recommendation returns we propose to use the 3-factor Fama-French model in Algorithm N. We suggest leaving the momentum factor out from the abnormal return regression, as we do not think it is appropriate in the context of stock recommendations to consider returns due to momentum to be factored out.

On a portfolio level comparison, the portfolio effects metric in Algorithm P should also be calculated to control for portfolio effects in the analysts recommendations. We will also add the calculation of bias in (23), which will give valuable new insights on how analyst add value for clients.

5 Implementation and results

In this section, we will implement the proposed algorithms using the bank as an example. We will compare the output data from the new algorithms to the existing to show how the new methods can unlock new insights from the data.

5.1 Data

The data set used for this implementation was provided by the bank and contains estimates and recommendations for 34 equity analysts covering 240 stocks. Two analysts (Analystid 6 and 28) were excluded from the data as they were junior analysts who were not yet lead analysts on any stocks. The data set contains a total of 3852 estimations for the fiscal year 2006/2007 and 7874 recommendations from these analysts.

The data set also contains consensus estimations (12205 data points) and consensus recommendations (7046 data points) over the same time period, as well as reported earnings per share. Data on stock prices, index values, rates and data necessary for the calculation of the Fama-French factor model such as stocks market cap and book-to-market values were all collected from Bloomberg. The data was organized into a object-relational database, and the algorithms were implemented as user-defined-functions and stored procedures using T-SQL on a Microsoft SQL server. See Appendix A for a database diagram.

5.2 Estimates

5.2.1 Existing algorithms

TABLE 2

"Average score". The null values for analysts 4, 32 and 34 are due to some missing data for their stocks.

AnalystId	Average score	AnalystId	Average score	AnalystId	Average score
1	0.843	13	11.065	24	0.887
2	5.204	14	0.802	25	1.056
3	0.600	15	1.004	26	0.711
4	NULL	16	0.767	27	0.865
5	1.083	17	2.857	29	1.068
7	14.013	18	0.891	30	13.710
8	6.112	19	1.213	31	1.707
9	1.200	20	0.733	32	NULL
10	2.916	21	3.317	33	0.790
11	3.826	22	1.835	34	NULL
12	0.774	23	1.633		

TABLE 3

"Stdev score". The null values for analysts 4, 32 and 34 are due to some missing data for their stocks.

AnalystId	Stdev score	AnalystId	Stdev score	AnalystId	Stdev score
1	0.823	13	3.213	24	1.539
2	1.597	14	0.526	25	1.887
3	1.899	15	0.700	26	1.870
4	NULL	16	1.265	27	0.282
5	1.946	17	1.221	29	2.342
7	2.014	18	5.481	30	1.664
8	1.225	19	0.740	31	4.989
9	2.222	20	1.081	32	NULL
10	3.591	21	3.710	33	1.379
11	15.578	22	10.928	34	NULL
12	0.770	23	1.757		

TABLE 4

"Sum of errors" The null values for analysts 4, 32 and 34 are due to some missing data for their stocks.

AnalystId	Score	Wins	Win ratio	AnalystId	Score	Wins	Win ratio
1	0.777	2	0.400	18	0.536	10	0.833
2	0.601	7	0.700	19	0.927	1	0.500
3	0.591	2	1.000	20	0.831	3	1.000
4	NULL	NULL	NULL	21	2.181	2	0.400
5	0.592	3	0.600	22	1.464	4	0.444
7	1.643	4	0.571	23	0.956	4	0.571
8	4.193	4	0.667	24	0.747	6	0.857
9	0.952	2	0.667	25	0.869	4	0.667
10	1.091	4	0.800	26	0.512	7	1.000
11	0.815	5	0.833	27	0.516	1	1.000
12	0.540	5	0.714	29	0.740	7	0.700
13	0.936	9	0.692	30	4.910	5	0.455
14	0.400	2	1.000	31	1.547	2	0.222
15	0.842	6	0.857	32	NULL	NULL	NULL
16	0.705	3	0.600	33	0.643	3	0.600
17	4.033	2	0.333	34	NULL	NULL	NULL

5.2.2 New algorithms

TABLE 5

PMAFE metric. The null values for analysts 4, 32 and 34 are due to some missing data for their stocks.

AnalystId	PMAFE	AnalystId	PMAFE	AnalystId	PMAFE
1	-0.108	13	22.476	24	0.018
2	27.493	14	0.183	25	0.320
3	-0.372	15	0.305	26	0.200

4	NULL	16	-0.219	27	5.012
5	0.336	17	3.616	29	4.179
7	57.416	18	1.646	30	28.673
8	6.714	19	7.687	31	2.122
9	1.667	20	-0.313	32	NULL
10	5.347	21	6.487	33	-0.214
11	3.372	22	1.619	34	NULL
12	0.426	23	1.381		

TABLE 6

APE time regression, averages over covered stocks. The null values for analysts 4 and 32 are due to some missing data for their stocks.

AnalystId	Sum(Residuals)	Sum(Abs(Residuals))	Slope	Intercept	R^2
1	-0.005	0.495	3.17E-05	0.009	0.701
2	-0.024	0.824	2.54E-05	0.010	0.655
3	-0.053	13.698	4.67E-04	0.022	0.356
4	NULL	NULL	NULL	NULL	NULL
5	-0.013	2.067	1.17E-04	0.002	0.525
7	-0.002	0.640	1.88E-05	0.010	0.441
8	-0.004	1.312	1.96E-05	0.012	0.444
9	0.000	0.186	5.59E-06	0.001	0.298
10	-0.002	0.477	5.39E-05	0.000	0.519
11	-0.001	2.968	1.42E-04	-0.009	0.541
12	-0.010	1.259	7.35E-05	0.010	0.641
13	-0.008	5.211	2.24E-04	0.010	0.433
14	0.000	0.006	1.89E-06	0.001	0.613
15	0.000	0.608	2.52E-05	0.010	0.534
16	-0.003	0.725	2.60E-05	0.003	0.518
17	0.000	0.953	2.05E-05	0.031	0.525
18	-0.005	2.546	2.56E-04	0.047	0.518
19	-0.063	9.692	4.08E-04	0.259	0.419
20	-0.002	1.053	9.68E-06	0.006	0.553
21	-0.003	1.274	3.74E-05	0.007	0.554
22	-0.040	8.061	3.44E-04	0.005	0.571
23	-0.007	1.223	4.42E-05	0.007	0.399
24	-0.015	3.676	2.94E-04	0.068	0.550
25	0.006	1.697	1.04E-04	0.003	0.469
26	0.023	0.873	5.74E-05	-0.001	0.539
27	0.000	0.163	3.56E-06	0.016	0.310
29	-0.007	1.750	9.95E-05	0.015	0.561
30	-0.001	1.340	5.65E-05	0.006	0.494
31	-0.010	6.804	5.50E-05	0.049	0.361
32	NULL	NULL	NULL	NULL	NULL
33	-0.005	0.704	4.43E-05	0.002	0.740
34	-0.001	0.153	4.71E-06	0.000	0.303

TABLE 7

Leader/Follower Ratio (LFR)

AnalystId	LFR	AnalystId	LFR	AnalystId	LFR
1	0.978	13	1.219	24	1.221
2	1.187	14	0.462	25	1.007
3	0.794	15	0.719	26	1.016
4	0.931	16	0.961	27	0.899
5	0.718	17	1.451	29	0.797
7	0.828	18	1.085	30	0.921
8	0.941	19	0.953	31	0.783
9	0.543	20	1.204	32	1.062
10	0.963	21	1.616	33	1.221
11	1.187	22	0.587	34	1.007
12	0.978	23	1.108		

5.3 Recommendations

5.3.1 Existing algorithms

TABLE 8

Recommendation wins over consensus

AnalystId	Wins	Stocks	Win ratio	Average abnormal return	AnalystId	Wins	Stocks	Win ratio	Average abnormal return
1	2	5	0.400	-0.020	18	4	12	0.333	-0.053
2	3	11	0.273	-0.123	19	0	2	0.000	-0.108
3	3	8	0.375	-0.058	20	1	5	0.200	-0.057
4	1	9	0.111	-0.206	21	1	7	0.143	-0.154
5	2	5	0.400	-0.031	22	4	7	0.571	0.059
7	4	7	0.571	-0.048	23	3	8	0.375	-0.142
8	2	5	0.400	0.106	24	4	7	0.571	0.081
9	1	5	0.200	-0.067	25	3	6	0.500	0.098
10	2	7	0.286	-0.132	26	3	6	0.500	-0.025
11	2	6	0.333	-0.124	27	0	3	0.000	-0.068
12	3	8	0.375	-0.005	29	3	10	0.300	-0.043
13	7	13	0.538	0.010	30	4	10	0.400	-0.072
14	1	2	0.500	0.139	31	5	7	0.714	-0.003
15	4	7	0.571	-0.006	32	1	1	1.000	0.162
16	4	6	0.667	-0.047	33	2	6	0.333	-0.049
17	2	7	0.286	-0.150	18	4	12	0.333	-0.053

5.3.2 New algorithms

TABLE 9

Risk-adjusted recommendation wins over consensus

AnalystId	Wins	Stocks	Win ratio	Average abnormal return	AnalystId	Wins	Stocks	Win ratio	Average abnormal return
1	4	5	0.800	0.025	18	6	12	0.500	0.019
2	4	11	0.364	-0.027	19	0	2	0.000	-0.125
3	3	8	0.375	-0.035	20	1	5	0.200	-0.046
4	4	9	0.444	0.004	21	3	7	0.429	-0.003
5	1	5	0.200	-0.061	22	1	7	0.143	-0.109
7	2	7	0.286	-0.059	23	1	8	0.125	-0.072
8	1	5	0.200	-0.024	24	4	7	0.571	-0.021
9	2	5	0.400	-0.039	25	2	6	0.333	0.065
10	3	7	0.429	-0.046	26	2	6	0.333	-0.015
11	2	6	0.333	-0.036	27	2	3	0.667	0.006
12	2	8	0.250	-0.009	29	3	10	0.300	-0.034
13	7	13	0.538	-0.021	30	4	10	0.400	0.020
14	2	2	1.000	0.073	31	3	7	0.429	-0.099
15	3	7	0.429	0.007	32	0	1	0.000	-0.049
16	5	6	0.833	0.030	33	3	6	0.500	-0.003
17	1	7	0.143	-0.049	18	6	12	0.500	0.019

TABLE 10

Recommendations, effectiveness and portfolio effects. The null values for analysts 32 and 34 are due to some missing data for their stocks.

AnalystId	$P(r_a \geq r_e \mid U_m)$	Positive recommendation effectiveness	$P(r_a \geq r_e \mid D_m)$	Negative recommendation effectiveness	Bias
1	0.421	0.951	0.500	1.130	-0.079
2	0.483	1.368	0.470	1.329	0.014
3	0.470	1.001	0.498	1.060	-0.027
4	0.557	1.052	0.506	0.956	0.051
5	0.527	1.093	0.426	0.884	0.101
7	0.458	1.116	0.480	1.171	-0.022
8	0.567	1.084	0.433	0.828	0.134
9	0.375	0.898	0.669	1.601	-0.294
10	0.508	1.236	0.348	0.848	0.160
11	0.539	1.058	0.499	0.979	0.040
12	0.449	0.987	0.500	1.098	-0.051
13	0.449	1.030	0.488	1.121	-0.040
14	0.632	1.074	0.488	0.829	0.144
15	0.536	1.217	0.000	0.000	0.536
16	0.520	1.004	0.547	1.054	-0.026

17	0.527	1.118	0.267	0.567	0.260
18	0.501	1.077	0.516	1.108	-0.014
19	0.524	1.100	0.000	0.000	0.524
20	0.625	1.077	0.316	0.545	0.309
21	0.379	0.895	0.513	1.212	-0.134
22	0.502	1.042	0.369	0.766	0.133
23	0.469	1.084	0.491	1.133	-0.021
24	0.490	1.004	0.484	0.992	0.006
25	0.458	1.834	0.559	2.238	-0.101
26	0.521	1.188	0.457	1.044	0.063
27	0.560	1.023	0.000	0.000	0.560
29	0.449	0.960	0.563	1.206	-0.115
30	0.421	0.997	0.626	1.481	-0.204
31	0.439	1.018	0.437	1.015	0.001
32	0.435	1.000	NULL	NULL	NULL
33	0.510	1.020	0.499	0.999	0.011
34	0.348	1.087	NULL	NULL	NULL

5.4 Example ranking of analysts

We will exemplify how a bank could use the new algorithms to rank analysts, using the results above. The assignment of weights to the different metrics is a subjective decision by the evaluator. However, to keep this example simple we will give the three precision metrics “Average score”, “Sum of errors score”, “Win ratio” and the consistence metric “Stdev score” 25% weight each. This would give us the following ranking of the analysts:

TABLE 11

Estimation ranking based on existing algorithms

Rank	AnalystId	Rank	AnalystId	Rank	AnalystId	Rank	AnalystId
1	14	9	16	16	25	25	7
2	27	10	33	18	9	25	30
3	26	11	18	19	11	27	22
4	3	12	1	19	23	28	31
4	12	13	2	21	8	29	21
4	20	14	5	21	10		
7	15	14	29	23	13		
8	24	16	19	24	17		

Using the new algorithms we will assign equal weights (33% each) to precision (PMAFE), consistency (sum of absolute residuals) and leadership (LFR). For simplicity consistency will be represented by sum of absolute residuals. For PMAFE we use the absolute values for the ranking. This would give us the following ranking of the analysts:

TABLE 12

Estimation ranking based on new algorithms

Rank	AnalystId	Rank	AnalystId	Rank	AnalystId	Rank	AnalystId
1	1	8	17	17	18	25	3
2	33	10	23	18	5	26	30
3	20	11	9	19	10	27	31
4	26	12	15	20	11	28	22
5	24	13	21	20	13	29	19
6	16	13	25	22	7		
7	14	15	27	22	8		
8	12	16	2	22	29		

For ranking recommendations based on the existing algorithms, we give equal weight (50% each) to the two metrics "Win ratio" and "Average abnormal return" and we end up with the following ranking of the analysts based on recommendations:

TABLE 13

Recommendation performance ranking based on existing algorithms

Rank	AnalystId	Rank	AnalystId	Rank	AnalystId	Rank	AnalystId
1	32	8	13	17	3	25	2
2	24	10	16	17	29	25	10
3	22	11	7	17	33	27	17
3	31	12	26	20	18	27	27
5	14	13	1	21	23	29	19
6	25	14	5	22	11	30	21
7	15	14	12	22	20	31	4
8	8	16	30	24	9		

Using the new algorithms for evaluating recommendations we give equal weight (25%) to the four metrics "Win ratio", "Average abnormal return", "Positive recommendation effectiveness" and "Negative recommendation effectiveness", giving the analysts the following ranking:

TABLE 14

Recommendation performance ranking based on new algorithms

Rank	AnalystId	Rank	AnalystId	Rank	AnalystId	Rank	AnalystId
1	25	9	5	17	22	25	30
2	2	10	23	18	13	26	12
3	10	11	8	19	27	27	29
4	15	12	18	20	33	28	1
5	26	13	20	21	31	29	9
6	17	14	14	22	24	30	21
7	7	15	11	23	16	31	
8	19	16	4	24	3		

Below is a table listing the outcome using existing and new algorithms and the difference in ranking between the two.

TABLE 15

Example ranking comparison. The missing values for analysts 4, 32 and 34 are due to some missing data for their stocks.

AnalystId	Estimations			Recommendations		
	Existing ranking	New ranking	Difference	Existing ranking	New ranking	Difference
1	12	1	-11	13	29	16
2	13	16	3	25	2	-23
3	4	25	21	17	24	7
4				31	16	-15
5	14	18	4	14	9	-5
7	25	22	-3	11	7	-4
8	21	22	1	8	11	3
9	18	11	-7	24	30	6
10	21	19	-2	25	3	-22
11	19	20	1	22	15	-7
12	4	8	4	14	27	13
13	23	20	-3	8	18	10
14	1	7	6	5	14	9
15	7	12	5	7	4	-3
16	9	6	-3	10	23	13
17	24	8	-16	27	6	-21
18	11	17	6	20	12	-8
19	16	29	13	29	8	-21
20	4	3	-1	22	13	-9
21	29	13	-16	30	31	1
22	27	28	1	3	17	14
23	19	10	-9	21	10	-11
24	8	5	-3	2	22	20
25	16	13	-3	6	1	-5
26	3	4	1	12	5	-7
27	2	15	13	27	19	-8
29	14	22	8	17	28	11
30	25	26	1	16	26	10
31	28	27	-1	3	21	18
32				1		
33	10	2	-8	17	20	3
34						

Looking at the results for individual analysts in detail, we note that for estimations, analyst 3 have lost some 21 ranking positions by going from scoring well on all measures by the existing algorithms to an average score on precision (rank 14) and scoring quite poorly on consistency (rank 23) and rock-bottom on leadership (rank 30). Looking at recommendations, we note that analyst 2 have gained 23 ranking positions mainly due to scoring exceptionally well on recommendation effectiveness, both for positive and negative recommendations. With the existing algorithms the high effectiveness in the recommendation was hidden, and analyst 2 scored well below average with a (rank 25) for recommendations.

6 Conclusions

We have proposed new algorithms to improve the evaluation of equity analysts. We have found these new techniques by identifying weaknesses in the existing techniques and exploring academic literature to come up with alternative ideas and adjustments. We have described how these new algorithms take into consideration new aspects into the evaluation, such as leadership in estimation revision, proper risk adjustment of returns and much more. Based on this, we think that the new proposed algorithms are improvements on the existing ones, as they give a more unbiased assessment. So we feel confident that we have achieved the goal of this thesis in that respect.

It is difficult to quantify how much “improvement” the new algorithms offer over the existing ones because there are more than one aspect/metric to consider and the weighting of these is basically up to each evaluator’s priorities. But what we can demonstrate is whether the new algorithms give different results when applied and, if so how much this could affect the evaluation results. If the new algorithms have added much new information content to the evaluation, then one would expect large differences in the evaluation. In the previous section we have done just that by showing an example ranking of analysts. As we have seen the new algorithms yielded a marked difference in ranking for many of the analysts, which were revealed to be better/worse analysts on estimation or recommendation performance than previously indicated by the existing algorithms. Thus, based on this example, it would seem that the new algorithms have managed to capture new, important aspects in the evaluation.

So far we have not discussed the choice between qualitative techniques, such as the Institutional Investor rankings described in section 3.1, and the more quantitative techniques, which we have described at length throughout this thesis. In conclusion there is clearly a strong case for the latter as they are free from any human biases that may have affected the subjective surveys. However, it seems equally clear that ideally a good analyst should score highly on both subjective and quantitative measures, and in a relationship industry such as banking the importance of favorable reviews from clients cannot be overestimated. Important aspects of the equity analyst job include for example creativity in new idea generation, successful marketing of investment ideas, personal industry contacts etc., which are all hard or impossible to incorporate into a quantitative framework. We would therefore suggest that subjective measures are at least considered as a part in a comprehensive evaluation of equity analysts. For example one of the leading industry surveys, such as the Institutional Investor rankings, could be used, or alternatively it could be worthwhile conducting client surveys, depending on budget and time constraints.

7 Suggestions for further research

It would be interesting to explore further any link between analysts forecasting abilities and their abilities to issue profitable recommendations, and how this could potentially be used in the context of evaluating analysts. In the academic literature, there is some evidence on the link between forecast accuracy and recommendation profitability. Mikhail et al (1999) found evidence that an analyst is more likely to change jobs if his or her forecast accuracy is lower than for peer analysts. Interestingly, they found no statistical relation between absolute or relative profitability of an analyst's stock recommendations and the probability of a job change. This result held regardless of the number of times the analyst has changed jobs. Thus it would seem that most investment banks indeed incorporate forecast accuracy in the evaluation of equity analysts, assuming that there is a connection between evaluation and the changing of jobs. One possible explanation for this result is that if analyst turnover is dependent on assessments of an analyst's unobservable effort, relative forecast accuracy may be more revealing of effort than the short-run profitability of stock recommendations.

The study also makes a very strong case for the use of evaluation by relative rather than absolute measures: "The use of relative performance evaluation appears ideal in our setting because analysts face common uncertainty, their work output is a noisy measure of their effort, and unambiguous benchmarks (e.g., actual earnings per share) and reference groups (e.g., other analysts providing forecasts for the same firm and time period) exist to evaluate their performance. Holmstrom (1982) demonstrates that in settings such as ours, relative performance evaluation results in improved risk sharing because the agent is held accountable only for those outcomes over which he has control." (Mikhail et al, 1999, p. 193)

Another study is Loh & Mian (2006), where a measure of recommendation profitability is constructed which corresponds to their earnings forecasting ability measure described earlier. Analysts are sorted into three groups according to forecast accuracy and a statistically significant difference is found between profitability between best third and worst third forecasters. This average difference in profitability is around 10% although there are large differences in magnitude over industries. Thus there seems to exist a subset of analysts with superior stock picking skills, which are not by chance since they also have earnings forecasting skills. This finding provides some support to the semi-strong form of the efficient market hypothesis regarding informational efficiency. The fact that an analyst with superior expectations data is able to transform that into value-creating stock recommendations shows that market prices do not always fully reflect all available information, otherwise information gatherers like equity analysts would not be rewarded for their efforts.

References

Books

Brealey, R. & S. Myers, 2000, *Principles of Corporate Finance*, 6th ed., New York: Irwin/McGraw-Hill.

Hawawini, G. & D.B. Klein, 1994, “On the Predictability of Common Stock Returns: Worldwide Evidence”, in R.A. Jarow, V. Maksinovic & W.T. Ziembas (eds), *Finance*, Amsterdam: North-Holland.

Gujarati, D. N., 2003, *Basic Econometrics*, 4th ed., New York: McGraw-Hill.

Journals

Barber, B., R. Lehavy, M. McNichols & B. Trueman, 2001, “Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns”, *The Journal of Finance*, Vol. 56, pp 531-563.

Bjerring, J. H., J. Lakonishok & T. Vermaelen, 1983, “Stock Prices and Financial Analysts’ Recommednations”, *The Journal of Finance*, Vol. 38, pp. 187-204.

Cooper, R. A., T. E. Day & C. M. Lewis, 2001, “Following the Leader: A Study of Individual Analysts’ Earnings Forecasts”, *Journal of Financial Economics*, Vol. 61, pp 383-416.

Francis J. & L. Soffer, 1997, “The Relative Informativeness of Analysts’ Stock Recommendations and Earnings forecast Reviews”, *Journal of Accounting Research,* Vol. 35, pp. 193-211.

Green, C. T., 2006, “The Value of Client Access to Analyst Recommendations”, *Journal of Financial & Quantitative Analysis*, Vol. 41, pp. 1-24.

Hawawini, G. & D. Keim, 1995, “On the predictability of common stock returns: World-wide evidence”, *Handbooks in Operations Research and Management Science*, Vol. 9, pp 497-544.

Hong, H., J. D. Kubik & A. Solomon, 2000, ”Security Analysts’ Career Concerns and Herding of Earnings Forecasts”, *The RAND Journal of Economics*, Vol. 31, pp 121-144.

Jegadeesh, N. & W. Kim, 2006, “Value of Analyst Recommendations: International Evidence”, *Journal of Financial Markets*, Vol. 9, pp 274-309.

Jegadeesh, N., J. Kim, S. D. Krische & C. M. C. Lee, 2004, ”Analyzing the Analysts: When Do Recommendations Add Value?”, *The Journal of Finance*, Vol. 59, pp. 1083-1124.

Jegadeesh, N. & S. Titman, 1993, “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency”, *The Journal of Finance,* Vol. 48, pp. 65-91.

Loh, R. K. & G. M. Mian, 2006, “Do Accurate Earnings forecasts Facilitate Superior Investment Recommendations?”, *Journal of Financial Economics*, Vol. 80, pp 455-483.

Mastrapasqua F. & S. Bolten, 1973, “A Note on Financial Analyst Evaluation”, *The Journal of Finance*, Vol. 28, pp. 707-712.

Mikhail, M.B., B.R. Walther & R.H. Willis, 1999, “Does Forecast Accuracy Matter to Security Analysts?”, *The Accounting Review,* Vol. 74, pp 185-200.

O’Brien, P. C., 1990, “Forecast Accuracy of Individual Analysts in Nine Industries”, *Journal of Accounting Research,* Vol. 28, pp 286-304.

Stickel, S. E., 1992, “Reputation and Performance Among Security Analysts”, *The Journal of Finance,* Vol. 47, pp. 1811-1836.

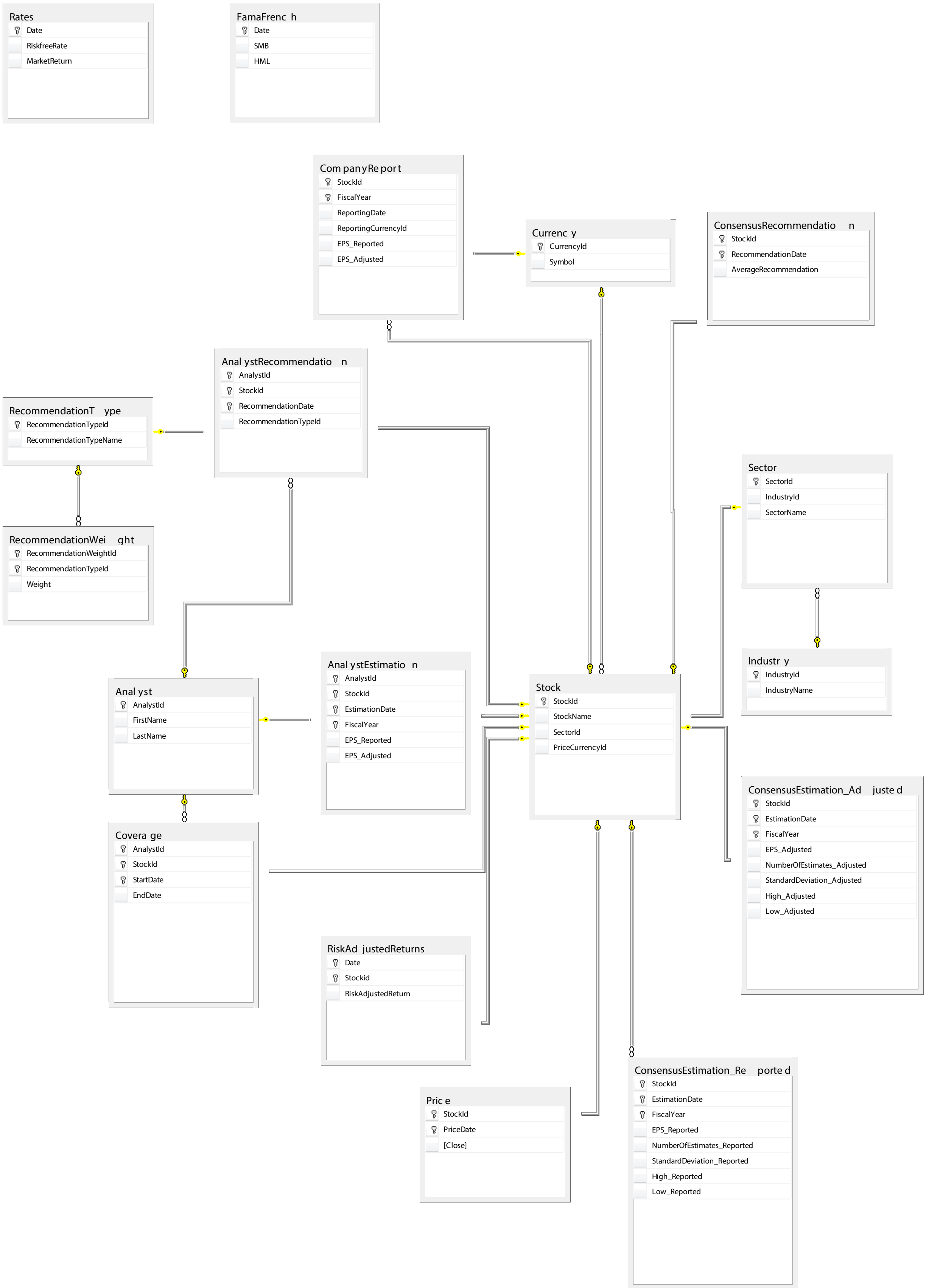
Womack, K. L., 1996, “Do Brokerage Analysts’ Recommendations Have Investment Value?”, *The Journal of Finance*, Vol. 51., pp 137-167.

Internet resources

“Methodology”, *Institutional Investor*, retrieved 2012-12-31 from [http://www.institutionalinvestor.com/Research/3582/Methodology.html]

“Methodology: How StarMine compiled the 2012 awards”, *Financial Times*, retrieved 2012-12-31 from [http://www.ft.com/cms/s/0/a9f095ea-a8ae-11e1-be59-00144feabdc0.html]

Appendix A: Database diagram



Appendix B: Fama-French three-factor model

The 3-factor model was originally developed by Fama and French (1993), who used six portfolios formed from sorts of stocks on market value of equity (*ME*) and book-to-market equity, which is the ratio of book value of equity to market value of equity (*BE*/*ME*). The portfolios are intended as mimics of underlying risk factors in returns related to size and differences between growth stocks (low *BE*/*ME*) and value stocks (high *BE*/*ME*). Each month all stocks are ranked on *ME*. The median size is then used to split the stocks into two groups, small and big (*S* and *B*). Each month stocks are also ranked on *BE*/*ME* and three groups of stocks are formed based on the breakpoints for the bottom 30% (*L*), medium 40% (*M*) and top 30% (*H*) of the ranked values of *BE*/*ME*.

Six portfolios (*SL*, *SM*, *SH*, *BL*, *BM*, *BH*) are then created from the intersections of the two *ME* and the three *BE*/*ME* groups. Daily value-weighted returns on the six portfolios are calculated, and portfolios are reformed monthly. The size factor-mimicking portfolio *SMB* (small minus big), is calculated as the difference, each day, between the average return on the three small stock portfolios (*SL*, *SM* and *SH*) and the average return on the three big-stock portfolios (*BL*, *BM*, and *BH*). The value/growth factor-mimicking portfolio *HML* (high minus low), is calculated as the difference, each day, between the average return on the two value-stock portfolios (*SH* and *BH*) and the average return on the two growth-stock portfolios (*SL* and *BL*).

For anyone interested in the Fama-French factor model, Kenneth French has published an abundance of data with ready calculated factors for different countries on his excellent website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html