The potential for machine-learning based investment analysis: analyst versus machine-learning based risk assessments

William Rees* and Alistair Haig*

Abstract

We examine the potential for machine learning in investment research by comparing analyst and machine-learning risk assessments made available by an independent investment research firm. In our early assessment, both indicators can benefit by incorporating information from the other and both are more accurate in countries considered to have superior informational environments. Analysts seem to under-estimate risk for stocks they classify as "buy", but we find no equivalent mis-calibration in machine-learning predictions. Our results confirm both the value of analysts' outputs and considerable potential for machine learning in financial analysis. We anticipate rapid developments in machine-learning analysis.

*University of Edinburgh Business School. Correspondence: alistair.haig@ed.ac.uk

Analyst versus machine-learning based risk assessments

"No human can beat a computer at chess. And no computer is better than at chess than a human supported by a computer."

Lopez de Prado (2018, p15)

Machine learning (hereafter ML) algorithms could become ubiquitous in investment analysis. Certain tasks performed by analysts, such as risk assessments, could be systematically re-estimated and calibrated on a continuous basis. If successful, investment research firms could substantially extend coverage, reduce costs, mitigate biases found in human decision making, and, possibly, improve the quality of predictions. Early signs of adoption, which could change the role of both analyst and investment manager, are already appearing. Leading buy-side and sell-side firms have recently unveiled prodigious capabilities in ML (Dunn, 2018). But the potential for change surely hinges on whether machine learning can match or even surpass analysts in predictive tasks. We contrast the ability of analysts and ML in quantifying the distribution of possible outcomes.

The prospect of wider coverage with frequent, cheaper and less biased valuations makes ML appealing. Yet, while ML has advantages in speed, scale and objectivity, human fundamental analysts, powered by expertise and intuition, can process unstructured data to identify and evaluate news, e.g., managerial changes, contracts won, or controversies exposed. A systematic approach may struggle to capture subtle contextual information. The potential for ML therefore depends upon its effectiveness when contrasted with alternative approaches, most obviously existing human approaches.

Our setting allows us to make such a comparison. We use risk measures derived from data provided by Morningstar Inc., an independent research organization with over 100 analysts covering some 1,500 firms. Morningstar compliments its analysts' research with ML output including, trading recommendations, valuations and risk assessments for more than 20,000 firms.

We use subsequent return volatility to measure the effectiveness of risk assessments. The prior evidence suggests that financial analysts can assess risk with substantially more success than predicting returns (Joos et al, 2016). In an efficient market, only exceptional individuals will consistently beat the market. There is no equivalent competitive element to risk assessment. For most stocks, risk in the next period is strongly associated with risk in the last. Thus, whether analysts or machine learning can provide forecasts that have predictive ability is not contentious: it would be surprising if they did not. We therefore use risk assessment as an achievable benchmark, and use this compare analysts to ML in order to gauge the prospects for ML.

Analyst Risk Assessment (ARisk). Morningstar analysts allocate firms to a low, medium, high or very-high risk category. Each category is associated is mapped to the distribution around the analyst's fair value estimate. Morningstar define a standard range, or spread, between the boundaries for each category. The range between the upper and lower quartile, divided by the fair value estimate, for each category is 44, 63, 88 and 111 percent spread, with a mean of 76% for our sample. Approximately 5% of firms are designated low, 47% medium, 38% high and 10% very high. This risk assessment is coupled with the analysts' fair value to price computation to determine the investment recommendations: the riskier the stock the more extreme the fair value to price required to justify a particular star rating (Morningstar, 2015). Thus, analyst uncertainty rating can be expressed as either a category (e.g. Low, Medium, High or Very High), as in Liu et al. (2007), or a score (e.g. 0.44, 0.63, 0.88 or 1.11) as in Joos et al. (2016).

Machine Learning Risk Assessment (MRisk)
Morningstar's machine learning approach is based on the random forest method (Das et al., CFA 2018 p77). The intention is that the ML models will provide substitutes for the analysts' output. ML produces 500 valuation estimates, each an iteration of the model estimated on a random subsample of the available data. The output is then summarized as a mean, the ML valuation, and MRisk, which describes the interquartile range of the fair value estimates. Morningstar describe this as "comparable to the [Analyst Risk Assessment] for the analyst Fair Value Estimate" (Morningstar 2013, p2). For our sample MRisk has a mean of 14% and lower and upper quartiles of 9% and 17% respectively. ARisk is therefore more than 5 times higher in mean score (0.76) than the quantitative uncertainty (0.14). The former is intended as an indicator of the potential for change in share prices over the coming three years, Morningstar's nominal investment window, whereas MRisk is derived from uncertainty about the current valuation.

The relative distribution (not shown) of the four analyst risk assessment (ARisk) categories and ML risk assessment (MRisk) deciles reveals a measure of agreement but also some disagreement; specifically, the presence of a few high MRisk cases in low ARisk codes and low MRisk in high uncertainty. Although not tabulated here, we found the distributions of US and non-US subsamples to be similar.

Future volatility.
Our risk measure is return volatility for each security, measured as the annualized standard deviation of weekly returns for the forthcoming quarter. Liu et al (2007) also examine volatility, in their case using the log of daily volatility, and present sensitivity analysis to demonstrate that their results are not affected by the precise calculation of volatility. Joos et al. (2016) examine the valuation error (the difference between analysts' fair value estimate and one year ahead share price). The different risk measures are strongly correlated, and our sensitivity analysis suggests that our results are robust to different definitions of share price volatility. We use weekly readings over one quarter which reduces the impact of thin trading and avoids the noise inherent in using only a few readings for each year.

The fact that our two risk assessments, quantitative uncertainty and analyst uncertainty, and the outcome, volatility, are measured differently could lead to confusion. The mean ARisk and MRisk for our main sample is 0.142 and 0.767 respectively whereas the outcome variable being predicted, volatility, has a mean of 0.240. This means that the relationship between the two risk assessments and volatility will have very different coefficients. We reduce this problem by using Fama MacBeth regressions to derive a forecast of volatility from the data at each quarter and then assessing the explanatory power of the derived forecast in the following quarter. Our three derived forecasts, one based on ARisk measure, one on the MRisk, and a third estimated using the financial characteristics of the firms, all have the same mean, 0.239, close to that of the outcome volatility.

Derived Forecasts
We estimate the relationship between the risk indicators at time t-1 and volatility at t and then use the estimated relationship to forecast volatility at t+1 based on the risk indicators at t. Using Fama-MacBeth cross-sectional regressions we evaluate the effectiveness of the forecasts by estimating the explanatory power of the forecast in each quarter and the average across the 27 quarters in our seven year sample period. (One quarter is necessarily dropped as the risk measures at the end of 2018 would need to be assessed against volatility in the first quarter of 2019). We have named the forecasts the analyst derived forecast (ADFor) and the machine derived forecast (MDFor). The third forecast we use follows a parallel approach save that we based the forecast on the financial characteristics of the firms: the last quarter volatility, the last year volatility, the market capitalization of the firm, the firm's beta and the dispersion of analyst valuations available on I/B/E/S. This forecast is a simple model of the forecast a user could access should they not have access to the analyst or machine risk measures and has been named the financial derived forecast (FDFor). Using derived forecasts generated at each cross-section differs from previous studies which have typically estimated the relationship across a panel of firms and time. We have tested the sensitivity of our results by replicating the panel data approach, and found consistent results with those that we report, but using quarterly forecasts more closely models the situation that a potential user would experience.

We find that output from both analyst and ML techniques can be used as risk indicators. Both provide information beyond that contained in a set of control variables and each can be improved by using information contained in the other. These results hold broadly across time and countries. However, the different constituents of the sample present different challenges. To examine this, we split the sample into the US, other common law countries, where we argue that the information environment will be relatively benign, and code law countries, where the information environment may be less complete. We also divide the sample into cases where I/B/E/S collates fair values and forecasts and those without I/B/E/S coverage. The presence of an I/B/E/S following indicates that at least one analyst is monitoring the firm and disseminating information to investors.

Analyst and machine-derived forecasts are both less effective indicators of volatility in code law countries than where common law is applied. The decline in analysts' predictive ability across this divide is such that the ML predictions marginally outperform those from analysts for the code law sample. This indicates a potential advantage beyond the obvious ability of machine-based techniques to produce more frequent, cost effective and comprehensive coverage than is feasible with analysts: for our sample, the ML approach handles unfamiliarity and limited informational environments better than analysts. However, if this were a general result we would expect analysts to cope less well for firms not covered by other analysts (i.e., with no I/B/E/S following). Our results show no appreciable difference between firms with and without I/B/E/S coverage.

As we study analysts from an independent research provider, the adverse incentives faced by sell-side analysts are removed (Allee et al. 2017). This bias is most obviously demonstrated by the fair value to price assessment. For example, Joos and Piotroski (2017) report a mean fair value to price of 1.159 for their (2007-2012) Morgan Stanley sample with only 22 of cases percent below one. The mean fair value to price for our sample is 1.055 with 45 percent below one. Yet, despite the absence of typical sell-side incentives our tests still demonstrate mild bias. We show that analyst risk assessments are relatively optimistic for companies for which they issue buy recommendations. We find no equivalent bias within the ML or financials-derived forecasts.

We propose that the main contribution of this paper is to provide early evidence of the prospects for ML by evaluating the effectiveness of ML in performing tasks performed by experts. The experts in our study are analysts at an independent research firm; biases in their outputs cannot be the result of sell-side incentives and therefore can be attributed to the human condition. Rather than relying on stylized models, our data comes from a well-established, commercially available source. Additionally, our work is related to recent research on the second moment of analysts' predictions and the expanding independent category of investment analysts. Finally, we present global results and therefore build on the minority of studies featuring global or international findings an analysts' work.

The study has a number of limitations. As with most studies of risk assessment we use a single-firm setting. We present evidence that the analyst and ML outputs are credible and widely available. Our sample firm uses supervised ML to mimic analyst outputs from a narrow range of input variables. These variables may become more or less important over time. Other ML algorithms might use wider sets of input data. Other supervised learning approaches such as neural networks might also be used and may be more or less effective (Kelly, 2018). Unsupervised learning approaches are also popular and available to discretionary investment managers, i.e., where the ultimate investment decision is subjective rather than systematic (Lopez de Prado, 2016). Examples include algorithms which establish sentiment scores from unstructured text. We expect a surge in the study of ML in investment management and research.

Prior Research and Hypothesis Development

Prior Studies

The valuations in our sample data serve as target prices, which have become ubiquitous in analyst reports over recent decades. Target prices provide a more informative summary than recommendations or earnings forecasts (Bradshaw et al., 2013; Bonini et al., 2010 and Bilinski et al., 2012). A recent, systematic literature review of the analyst literature in accounting, finance and management journals (Spence et al., 2018) reveals that two thirds of papers study earnings or recommendations compared to only 10% examining other forecasts, of which target prices are only one type.

More recently, analysts at some firms have included summary information about the distribution around their target price. These risk assessments, sometimes known as bull-bear analysis, usually take the form of alternative target prices for positive and negative scenarios. Several recent papers relate these risk assessments to outcome variables which represent fundamental risk. Liu et al. (2007) focus on future share price volatility, whilst Joos et al. (2016) concentrate on absolute pricing error. Further, Liu et al. (2012) report that the publication of analysts' risk assessments affect share prices and investment returns, and Joos and Piotroski (2017) show how the relationship between upper and lower bounds of risk and the analysts' target price can be used to discriminate between more and less accurate investment predictions. Hashim and Strong (2015) also find that target prices accompanied by risk assessments out-perform those without. Hashim and Strong (2015) use a wide-ranging sample of reports on US firms while Liu et al. (2012) and Joos and Piotroski (2017) concentrate on data from one brokerage firm; all use risk assessments provided by sell-side analysts, with their well-established propensity to biases (Bradshaw 2011), and all concentrate on US firms. To our knowledge, technological innovation has not previously been investigated in this strand of the literature.

The inclusion of explicit risk assessment in analysts' output is still growing. Hashim and Strong (2015) take a snap shot of practice in 2008 and 2009 and in analyzing a large cross-section of sell-side analyst reports they find close to one-third include bull-bear analysis (BBA). The Joos et al. (2016) and Joos and Piotroski (2017) papers also examine BBA and conveniently include an example from their data provider, Morgan Stanley. This shows that the analysts have considered and valued explicit up- and down-side scenarios. An alternative practice is to allocate a risk description to each stock. For example, Liu et al. (2007 & 2012) use risk categories from Citi (branded at that time as Salamon Smith Barney). Their first sample (2007) is taken in each January and February from 1997 to 2002 and the second (2012) from various sources from 1999 to 2006. In this coverage the analysts categorize stocks as being suitable for different types of investor ranging from conservative through aggressive to sophisticated (Liu et al. 2007). Few stocks fall into the extreme categories. Overall, the picture from recent research shows a strong presence of risk assessment in analysts' output. Hashim and Strong's (2015) work would suggest that in the US, shortly after the 2008 crash, it was not uncommon but also not typical.

The Global Settlement and associated legislation placed restrictions on investment bank analysts and prompted a flurry of studies which compare investment bank and non-investment bank analysts. Prior studies tend to group independent analysts with "unaffiliated" peers, who work at non-investment bank brokers) and refer to this category as "independent" (Allee et al., 2017, is a rare exception). Haig and Rees (2017) provide a discussion of the regulatory changes between the Global Settlement and Current European regulation (MiFIDII); they document emerging interest in independent investment research. MiFIDII addresses the conflicts faced by brokers regardless of their affiliation to investment banks. Truly independent firms such as Morningstar do not earn dealing commissions and therefore are treated separately. Studies of the characteristics of independent analysts is therefore likely to be valuable.

Hypothesis Development

Experts are subject to human error and make mistakes. The possibility of using systematic processes to augment or replace expert decision making is therefore of interest. Simple static models have been shown to be more accurate than experts in domains such as psychological diagnosis (Meehl, 1956), political forecasting (Tetlock, 2005) and systematic, quantitative investment (Fabozzi et al., 2009). Systematic investment models have tended to be static and to require analyst skill and experience to determine when and how it should be re-estimated in order to deal with living, evolving financial markets. Models which not only update data but also re-estimate the relationship between variables frequently, e.g. daily or intraday, are likely to be of interest. Developments in ML have led to the introduction of dynamic models which learn and update without human intervention.

Risk assessment differs from volatility prediction. The former as a means for the analyst to convey their expectations regarding the distribution around the target price; it is an indication of risk rather than a precise estimate of ex ante return variance (Joos et al, 2016). The latter, volatility prediction, has various purposes including investment and risk management. The substantial literature on volatility prediction, reviewed by Poon and Granger (2003; 2005) "reflects the importance of volatility in investment, security valuation, risk management, and monetary policy making." Poon and Granger (2003, p478). Implied volatility offers an appealing alternative estimate of risk; a forward-looking consensus estimate derived from option prices.

Recent papers such as Baltussen et al. (2012) and Szado (2018) review the potential for implied volatility to be incorporated in stock selection strategies. All these studies use US implied volatility as a predictor. OptionMetrics provide US data on several thousand US companies (Szado, 2018). Even so, studies such as Szado (2018) restrict their analysis to S&P500 companies due to liquidity constraints. Outside the US, single stock options tend to be available only on a narrow selection of the largest listed companies. In the UK, exchange traded options exist for around 120 companies; in France and Switzerland options are available only on the largest companies. Additionally, implied volatility is not reliable for illiquid options. Our direct interactions with practitioners, and our own desk work, confirm that it is impractical to use implied volatility as in input to broad market equity strategies in non-US markets. It is only available where liquid single stock options are traded which makes it much less applicable for mid-cap companies which feature in our risk assessment sample.

Since historic volatility is generally available we include it as an explanatory variable in order tounderstand whether analysts or ML can provide incremental information in their risk assessments.

Research Question 1: Can ML mimic analyst risk assessments?

The ML algorithm learns from and attempts to mimic analysts' target prices. If successful, we expect analyst and ML risk assessments to be related:

H1: Analyst risk assessments vary with ML risk assessments.

We can test this hypothesis by comparing the determinants of analysts' risk assessments.

Research Question 2: Are analyst or ML risk assessments more accurate?

Investors will surely be interested not only in the extent to which ML can mimic analyst risk assessments but also the relative predictive accuracy.

Prior studies of investors' geographic proximity to investee companies have often found a home bias to exist. This is usually explained either by the familiarity hypothesis or the information asymmetry hypothesis. Familiarity may be due to cultural, linguistic and emotional factors (Grinblatt and Keloharju, 2000) or behavioural biases such as overconfidence and self-attribution of knowledge (Barber and Odean, 2003). Analysts might therefore be likely to place too narrow a range around their target prices for companies with which they are familiar, such as those in their home market.

Information asymmetry explanations for proximity effects are based on the idea that analysts who can make "house calls rather than conference calls" (Malloy 2005, p1) are in a better position to find and interpret news which could affect local companies. Investment managers have been shown to generate higher returns on local rather than distant firms. Coval and Moskowitz (2001) attribute this to the value of local contextual information, such as face-to-face meetings with management. Similarly, analysts meet the CEOs face-to face and survey the firm's operations directly and use this private information to make more accurate predictions (Malloy 2005).

Since Morningstar has a longer history of covering US companies, and the majority of analysts are based in the US or other common law countries, we can expect analysts to be more accurate in calibrating valuations for equities listed in the US and other common law countries, where analysts tend to be based, than those in code law countries.

H2a: Analyst risk assessments will be most informative for US stocks.

H2b: Analyst risk assessments will be most informative for common law stocks.

Research Question 3: Do independent analysts produce unbiased forecasts of risk?

Analysts have been found to exhibit optimism bias due to behaviour and incentives. Our sample is unlike that of prior studies which have merged independent analysts with their peers at non-investment brokerages. In aggregate, Morningstar analyst recommendations display bell-shaped distribution rather than the typical sell-side research firm which tends to have more buys than sells. We can therefore determine whether independent analysts suffer from the behavioural biases found in other samples which should be free from adverse incentives such as buy-side analysts (Groysberg, et al 2008)

If independent analysts place a narrower range around buy-rated stocks than sell-rated stocks this would be evidence of self-attribution of knowledge, i.e., overconfidence in having picked a winner. We expect this to be the case.

If analysts forecast a narrower spread for US stocks than international stocks, after adjusting for risk, this would be evidence of familiarity bias. We expect this to be the case. If analysts do not suffer from self-attribution bias we expect analyst and ML spread to be similar for buy/neutral/sell-rated stocks.

> ***H3a: Analysts risk assessments underestimate risk for buy-rated stocks (self-attribution bias).***

> ***H3b: Machine learning risk assessments do not vary between (implied) recommendation category.***

Research Question 4: How can investors can best use analyst and ML risk assessments?

This paper could be framed as a contest between human and artificial intelligence. It is, however, possible that the best risk predictions are made by a combination of mind and machine. In the absence of perfect correspondence between the two sources, it may still be possible for investors to combine the two valuations to their advantage. Taking stock of recent evidence on the use of ML to compliment analysts' earnings forecasts (Ball and Ghysels, 2017), we have at least some basis to expect that a hybrid approach may be optimal. Our final hypothesis is:

> ***H4: A combination of ML and analyst makes superior predictions (than either source in isolation).***

Our large sample of ML-only predictions allows us to perform further tests on this hypothesis.

Setting, data and method

Our data comes from Morningstar, a relatively large independent provider of equity research. With around 100 analysts covering securities issued by 1,500 companies (Morningstar, 2015, 2017) its scale is comparable to a large brokerage firm (although smaller than the very largest global investment banks) but without the incentives faced by sell-side analysts (Barber et al. 2007). Analyst data is available from 2002 and over time the coverage has expanded from the original US sample and includes all sectors. Each analyst covers, on average, 15 firms and typically hold similar credentials to sell-side analysts (Kang et al. 2018). Morningstar's equity research reports and outputs resemble those of a brokerage firm and they provide two risk assessments. The first is an analyst risk classification into low, medium, high and very high (plus extreme but we have not identified any such case in our sample), where each category is also allocated a score reflecting the interquartile range of expected investment outcomes. The second, ML derived, is the interquartile range of valuations derived from 500 iterations of their random forest model.

We also analyse a sample of firms for which Morningstar provide ML data without accompanying analyst output. This is a much bigger sample which reflects the wider coverage Morningstar are able to provide using the ML approach. The firms in this extension sample are often drawn from less-developed or developing economies where the information environment might be expected to be shallower than in countries with long-established stock markets. When we match ML and analyst data in our comparison sample we include 969 target firms. When we analyse firms with ML data, but no matched analyst data, our extension sample, we have 11,164 firms.

Although the machine-learning approach used by Morningstar is proprietary, the firm does provide documentation and their research team assisted us in answering further questions. The input factors resemble those used in established stock selection tools. Morningstar make their outputs widely available. Morningstar's annual report discloses the number of licences and the typical price per individual firm; from this we estimate that over 3,000 firms have access to the analyst and ML data which features in our analysis; this estimate was confirmed by industry specialists. The credibility of

Morningstar as a provider of equity analysis was reinforced through informal discussions with former employees and consultants who help investment managers to source and evaluate investment research. Morningstar's machine-learning product has been in widely available since 2012. It is reasonable to assume that other firms have developed comparable and possibly superior models, these are much less widely available. Our sample firm therefore provides a unique opportunity to evaluate the effectiveness of a working ML process rather than a prototype model.

Research Method

Liu et al. (2007) and Joos et al. (2016) first model the determinants of the analysts' and ML uncertainty measures and future volatility. Although this analysis is descriptive, if future volatility is associated with a particular variable, and that variable is not associated with the risk measure, it would suggest that the risk measure neglects relevant information. To ensure consistency with previous studies we have reported determinants. We include controls for past volatility, the log of market capitalization, beta and the dispersion of I/B/E/S target prices. Apart from the last, these control variables have been used in previous studies.

We also considered including option implied volatility as a viable forecast of volatility. Using implied volatility would be feasible for most US stocks, and a sub-set of the larger non-US firms, and we might expect analysts to refer to option prices in forming their expectations. However, deriving viable implied volatilities is not straightforward when the liquidity, moneyness and horizon of options differ. Poon and Granger (2005) point out that implied volatility forecasts work well if calibrated by historical volatility, and that historical volatility is also an effective model when used alone. Our controls include historical volatility, and we experimented with the volatility index (US VIX and CBOE's ex-US equivalent) to incorporate changing market-wide expectations, but this, whilst statistically significant, had little impact on our research question. We have also included the dispersion (coefficient of variation) of target prices available on I/B/E/S as a benchmark for valuation uncertainty. Our goal is not to develop the best possible risk prediction model but to contrast the effectiveness of analysts' and ML risk metrics. The control variables are used to ensure that the risk metrics provide information beyond what is readily available to investors.

Previous papers have typically also included a set of firm-specific accounting-based variables. However, these variables add little when the other control variables are included, and, for the sake of brevity, we omit the accounting variables from our reported results.

The model is estimated separately for the US, other common law countries and code law countries. Our assumption is that the difficulty of forecasting risk will depend, in part, on the informational environment and that this will vary with legal system and the development of the economy. We also view the comparison of three large but distinct samples as a useful triangulation test for our results. If our results are reliable they should appear across our different samples.

Information content of analyst and ML risk assessments.

We estimate the information content of the risk assessments with a model similar to that used in previous research. Unlike previous studies we have transformed our risk indicators into explicit forecasts of volatility using only data that was available at or before the forecast date. We have named these "derived" forecasts. To identify the relative information content of the two risk assessments we estimate the results using the analyst (ADFor), machine learning (MDFor) and financials (FDFor) derived forecasts separately and then together, with and without the control variables. **Equation 1**, estimated using Fama-MacBeth estimation, is:

$$\sigma_{i,qt+1} = \alpha_0 + \beta_1 ADFor\ or\ MDFor\ or\ FDFor_{i,t} + \beta_2\sigma_{i,qt} + \beta_3\sigma_{i,yt} + \beta_4 Log(MktCap)_{i,t} + \beta_5 Beta_{i.t} + \beta_6 ibes\_tp\_cv_{i.t} + \varepsilon_{i.t}$$

$\sigma_{i,qt+1}$ is volatility for the forthcoming quarter, $ADFor_{i,t}$, $MDFor_{i,t}$, $FDFor_{i,t}$, are respectively the derived analysts, machine-leaning and financial-based forecasts of forthcoming volatility based on variables at time *t* and the relationship between those variables at *t-1* and volatility at t, $\sigma_{i,qt}$ and $\sigma_{i,yt}$ are the last quarter and last year volatility $Log(MktCap)_{i,t}$ is the log of market capitalization in US dollars, and $Beta_{i,t}$ is the Bayesian-adjusted beta estimated on the prior 36 months using either the US market index for US firms or the FTSE world (ex US) index for non-US firms, and $ibes\_tp\_cv_{i,t}$ is the coefficient of variation of the I/B/E/S reported fair values. Definitions of all variables are provided in Appendix 1.

Equation 1 is similar to approaches used in previous studies save that we omit accounting-based variables, which can be statistically significant, but do not affect our conclusions. We also have included the coefficient of variation of I/B/E/S fair values as an independent surrogate for the ML valuation uncertainty which is also based on the distribution of valuation estimates. We find this variable to be an effective indicator of future volatility. Finally, we have chosen to use Fama-MacBeth cross-sectional estimation. Previous research has typically used panel data estimation and we find the results are similar whether we used panel or Fama-MacBeth estimation. In practice a user of analyst output will have to predict next period's volatility using currently available indicators and experience of the relationship between those indicators and volatility in prior periods. By including forecasts derived from the previous quarter's relationship between the indicators and volatility, and evaluating the performance of those derived forecasts quarter by quarter, we closely simulate the decision making circumstances of a user of investment analysis – whether performed by analysts or ML. The significance tests are adjusted for autocorrelation using the Newey-West approach (Newey and West, 1987).

Data

In our dataset, we have two different indicators of risk from the same supplier. The first is a traditional risk classification into low, medium, high and very high (plus extreme but we have not identified any such case in our sample). This is notionally built on a bull and bear analysis with 25 percent probabilities assigned to the chance of exceeding each boundary, and the spread between the two can be interpreted as the interquartile range. This range is used to classify each firm into one of five categories, with 5 percent classified as low risk, 48 percent medium, 38 percent high and 9 percent very high. The spread between bear, designated "consider buy" and bull, "consider sell", divided by the average of the two, is 44%, 63%, 88% and 111% percent, respectively. We use these to create ARisk in the results presented. As a robustness test we replace these with an ordered classification with no appreciable difference in our results. Morningstar have chosen the blunt simplicity of ordinal categories rather than allowing analysts to assign ranges directly.

The second Morningstar variable is a measure of dispersion within the valuation model calculated as the interquartile range from 500 iterations of the model. It therefore reflects uncertainty but is not explicitly a prediction of expected risk. In essence a model is estimated which relates the predicted variable to a subset of available explanatory variables. The full set of explanatory variables is not used as this is likely to result in over-fitting. Each iteration uses a random sub-sample of the available data and the results consolidated to a point estimate.. Appendix 2 provides a short summary of Morningstar (2013).

The analysts' prediction is based on risk categories identified by analysts and we can utilize the categories as a continuous variable without impact on our results (as per Liu et al., 2012). The categories identified by the analyst are intended to indicate the expected inter-quartile range of share price changes – which is closely related to the expected share price volatility. Conversely, the machine learning estimates are based on 500 cross sectional iterations. Both are designed to calibrate valuations and are likely to reflect share price volatility. Volatility is itself just one way to measure risk.

As with previous papers we use indicators to predict a different outcome. Prior literature on analysts risk assessments uses disparate indicators to forecast risk, an outcome which is measured in different ways). Neither ARisk nor MRisk are direct volatility forecasts. Rather, they are used as proxies to help forecast volatility, each with its own scale. The comparison of proxy variable is not uncommon in finance. An analogy might be contrasting the informational value of fair value-to-price ratios and buy-sell-hold recommendations, by relating them both to realized stock returns. Neither piece of information is a returns variable, and they differ from each other, but it would not be unreasonable to evaluate their respective worth as investment advice.

To address potential issues regarding the comparability of ARisk and MRisk we regress each on share price volatility estimated over the following quarter and predict the value of ARisk and MRisk; ADFor and MUfor respectively. This common estimation process allows us to make comparisons based on forecasts expressed on the same scale. Interestingly the correlation between ADFor and MDFor is 0.59, higher than that for the underlying ARisk and MRisk (0.39).

Results

Research Question 1: Can ML algorithm mimic analyst risk assessments?

The mean ARisk is 0.767 and MRisk is 0.142, and the mean outcome (next-quarter) volatility is 0.239. Analysis of the joint distribution of analyst and ML risk assessments (not tabulated) shows a considerable degree of agreement in that there are few ML high-risk assessments in the analysts low-risk categories, the reverse also holds true, and yet there are also cases of disagreement. Both analysts' and ML output contain information about subsequent volatility and the information from one can be used to improve the predictive ability of the other.

This variation in the calculation of risk measures, as with prior work, implies that we cannot directly compare the coefficient transforming different types of predictions to outcomes of risk. To overcome this issue, we focus on our standardized measures: AQFor and MQFor.

In **appendix 4** we report models of the determinants of volatility and risk assessments. The first three columns show determinants of volatility and the raw risk assessments provided by Morningstar. Forecasts derived from our Fama MacBeth estimation are shown on the rightmost columns. In all models, realised volatility over the preceding quarter, beta and target price dispersion are significantly positively related to the outcome variable. Company size is inversely related to risk in models of volatility and analyst risk assessments. ML risk assessments are not related to company size. Analysis of US and non-US samples (not tabulated) confirms only subtle differences in the significance of variables.

Since the factors affecting raw risk assessments and derived forecasts are consistent, we proceed with the latter. In general, we conclude that the risk assessments tend to be influenced by the same characteristics as volatility, but to different extents. There is scope to improve both sets of risk assessments; for example the analysts might pay more attention to 12-month realised volatility and ML might be trained on log market cap rather than the level of this variable.

We considered modelling option implied volatility. The inclusion of VIX, a market-level option-implied volatility measure led to no significant improvement in explanatory power. The Fama MacBeth approach relies on cross-sectional estimation and since implied volatility data is sparse outside the US and certain large-cap stocks in other markets, we drop implied volatility from our analysis.

We evaluate the explanatory power of the variables and the models as a whole by comparing the t-statistics and R-squared. The model of next quarter volatility has a higher explanatory variable than analyst or ML outputs, either in raw or standardised form, and this can most likely be explained by differences in the loading on realised volatility. This does not necessarily imply that risk assessments are inferior, as un-modelled elements may be related to future volatility. For example, the analyst may incorporate contextual information regarding the sector outlook. The analysis of determinants does, however, suggest that risk assessments under emphasise factors which are known to predict future volatility.

In panel analysis not reported here we considered the importance of accounting characteristics, specifically, leverage, book to market, return on investment, negative earnings, negative equity and combinations of these variables (as per Joos et al. 2016). In many instances these variables are statistically significant, but their contribution to the explanatory power of the models is slight, none are robustly significant across all models and sometimes the sign switches despite being statistically significant. This lack of stability within the accounting variables may be caused by the relatively high correlation between them. This is typical for such ratios. However, the significance of non-accounting variables, reported in appendix 5 (determinants table) is unchanged (domnated) by the inclusion of accounting characteristics. Previous research also found their results to be unaffected by the inclusion or exclusion of these accounting controls. For the sake of brevity we exclude the accounting variables and retain realised volatility, beta, target price dispersion and company size.

In summary, our results for the determinants of volatility and analyst risk assessments echo those of previous researchers.

The model is estimated for our entire comparison sample and separately for code and common law countries, and also for the large US component. We could estimate these jointly, but we want to contrast the results for a consistent and homogenous sample with a changing and heterogeneous one.

Research Question 2: Are analyst or ML risk assessments more accurate?

The workhorse for our analysis is the Fama Macbeth procedure (Fama and Macbeth, 1973). The results allow an examination of the stability of the relationship and eliminates forward-looking bias. We test the influence of firm specific financial variables but find them to be unimportant and drop them from the reported tables for the sake of brevity. The analyst spread variable is firstly included as risk categories, as in Liu et al. (2007). We also test the continuous version as a robustness test.

To identify the relative information content of the two spread measures we estimate the results incorporating analyst and ML spread separately and then together, with control variables included in both cases.

The variable definitions are the same as for the previous model. In this instance we do not incude the accounting variables. They are often individually significant, especially in the absence of prior volatility, market capitalization and beta variables, but make no substantive difference to the results of interest. Again, standard errors are clustered by firm and quarter.

Research Question 3: Do independent analysts produce unbiased forecasts of risk?

Our sample comes from independent analysts and should not display the strong biases typically found amongst sell-side analysts (Barber et al. 2007). Even so, there is no reason to suppose that our analysts will not be subject to the general behavioral biases which typically affect decision makers. For example, Groysberg et al. (2008) find strong optimism bias among buy-side analysts. As the analysts have to attribute a recommendation from one to five to each stock, we can identify firms that the analyst looks upon favorably. If that benign attitude to the investment influences the assessment of spread, we might expect analysts to bias their spread measure down. To investigate this, we estimate the following relationship where the spread is conditioned by buy, hold and sell categories, and RiskFor represents $ADFor_{i,t}$, $MDFor_{i,t}$, $FDFor_{i,t}$, i.e., the derived analysts, machine-leaning and financial-based forecasts of forthcoming volatility based on variables at time t.

$$\sigma_{i,t+1} = \alpha_0 + \beta_1 Sell_{i,t} * RiskFor_{i,t} + \beta_2 Hold_{i,t} * RiskFor_{i,t} + \beta_3 Buy_{i,t} * RiskFor_{i,t} + \varepsilon_{i,t}$$

If the analysts are unbiased, we would not expect β_1, β_2 and β_3 to be significantly different. A positive bias towards stocks categorized as buy would increase β_3, whereas a negative attitude towards sell stocks would decrease β_1. We also run the model in the absence of fixed effects to ensure that the relationship between spread and volatility is a direct test and the results are robust. Here the variables are as before plus $Sell_{i,t}$, $Hold_{i,t}$ and $Buy_{i,t}$ represent the analysts' or ML recommendations. For the analysts, we have merged strong sell with sell, and strong buy with buy, as there are few cases in the extreme categories. This is also consistent with the ML version which provides three categories: fairly, over and under-valued. Buy, hold and sell are distributed 27, 50 and 23 percent by the analysts and 32, 38 and 30 percent by the ML system.

Information content of risk assessments

To examine the ability of analyst and ML to make informative risk assessments, we run the model for each sample using the different spread indicators separately, together, and with and without the control variables (prior volatility, company size, target price dispersion and beta).

We use Fama-MacBeth estimation procedures, where a separate cross-sectional regression is conducted for each quarter and for which we have excluded industry (and country) dummies. This conforms more closely to the investor's circumstances – at each point in time the risk metrics are used to predict the next period's volatility. The procedure allows us to check consistency through time by reviewing the relative explanatory power for each quarter and contrast the relative effectiveness of our two measures quarter by quarter.

In each case the results, shown in **table 2,** are as expected. Volatility in the subsequent quarter is significantly and positively associated with both analyst and ML risk assessment, current volatility, target price uncertainty, beta and negatively related to size. The explanatory power of the models, measured as the average of cross-sectional coefficient of determination (average R-squared), varies from 0.215 with MDFor alone to 0.583 with both risk assessments plus control variables.

These results show that a) analyst and ML risk assessments each contain information about future volatility, but b) both risk assessments omit relevant information. Since the control variables retain significance in the full model (columns 4, 6, 7 and 8), risk assessments are in one sense inefficient because they do not incorporate this information. Moreover, they do not fully incorporate the information included in the rival measure. (Morningstar could point out that predicting volatility is not the function, or at least not the main function, of either analysts' risk categories or ML valuation variability.) Even so, both analysts and ML measures are statistically significant in the presence of the other and inclusion of either or both risk assessments leads to greater explanatory power. Both sources are useful in assessing risk.

Research Question 4: How can investors can best use analyst and ML risk assessments?

It is apparent that when modelled together, analyst and ML risk assessments are both statistically significant and produce higher explanatory power than either alone. A possible reason is differential effectiveness between US and non-US samples. The US sample is stable and relatively homogenous and should be easier to analyze using either traditional or machine-learning methods. In **table 3** we do indeed find that explanatory power, i.e., average R-squared, tends to be greater in the US sample.

We investigated this further by dividing the non-US sample into common law and code law countries. For Australia, Canada, Hong Kong, New Zealand, South Africa and the United Kingdom the results follow the US model. The predictive ability of analyst and ML risk assessments in common law countries is similar when modelled separately. When used together ADFor is marginally more significant and therefore marginally more effective in predicting future volatility than the MDFor. This would not surprise practitioners who view the operating practices of common law capital markets as having much in common. However, for the sample of code law countries, dominated by Europe plus

Japan, we find that ADfor and MDfor to have similar significance, and when used jointly, MDFor is marginally more significant than ADFor. We notice a deterioration in the value of analysts' risk assessments while the significance of MDFor remains fairly constant. If the work of investment analysts is more difficult in code law countries, they may struggle to match the insights of analysts based in common law countries. The systematic nature of ML might be immune to such difficulties.

Morningstar use ML to extend their stock coverage. The extension sample provides a further opportunity to explore the potential for ML. The ML model is trained to predict analyst fair value to price. For companies outside the Morningstar analyst coverage universe the model is trained on the valuation of peer companies as defined (although not disclosed) by the model provider. Does predictive ability decline for stocks not covered by Morningstar analysts? By comparing R-squared of ML in our comparison sample (Table 2, column 2) with the extension sample (Table 4, columns 1 and 4) we can see that the coefficient on MDFor remains significant. There is, however a drop in overall explanatory power. The comparison sample R-squared drops from 0.215 (table 3) to 0.12 and 0.18 in Table 4; the more marked decline in R-squared is evident for companies with no I/B/E/S coverage (column 1) than those with I/B/E/S data (column 4). The extension sample shows ML to perform better in the presence of sell-side analysts, i.e., where more extensive investor information is likely to be available.

Conclusion

We contrast the effectiveness of risk assessments derived from traditional financial analysis and machine learning for a large international sample drawn from 2012-2018. We find that both assessments are good predictors and, whilst we find a small statistical advantage to the analyst assessments in common law countries and to the ML assessments in code law countries, there is little practical difference. Our evidence suggest that an investor may not be able to distinguish ML-based versus analyst -based forecasts. We also show, unsurprisingly, that the analysts' assessments retain an element of behavioral bias not found in the ML assessments.

Our results have two implications for the work of financial analysts. Firstly, combining the assessments from the ML process with that produced by analysts clearly improves the information content of the analysts' work. On the assumption that risk assessment is useful, and we note that the buy/sell/hold recommendations made by our data provider are a function of the risk assessment, and hence presumably important, analysts may benefit from combining the insights from traditional investment analysis with that from ML. Our understanding is that analysts frequently have internal access to ML output to assist them with their work. Secondly, ML learning appears to closely match the effectiveness of traditional analysis in producing informative risk assessments and its cost effectiveness leads to substantial increases in coverage and much faster updating. Particularly in the market for investment research, where regulatory changes have put considerable pressures on costs structures, cost effective ML techniques may become widely adopted.

Research into ML in investment research is relatively new. We believe our study is the first to conduct a head-to-head comparison of ML and analyst based financial analysis. Our results suggest that ML is effective but for decisions where unstructured and unquantified information play a larger role ML may find it more difficult to match the insights of analysts. Preliminary evidence on fair value to price estimation suggest that this is the case (Haig & Rees, 2018). As ML is already established, and we anticipate its use will grow rapidly, so more research concerning its effectiveness would be welcome.

References

Allee, K. D., D. Erickson, A. M. Esplin and S. Larocque (2017). "Independent Analysts' Estimates of Firm Value."

Ball, R. T. and E. Ghysels (2017). "Automated Earnings Forecasts: Beat Analysts or Combine and Conquer?" Management Science.

Barber, B., R. Lehavy, M. McNichols and B. Trueman (2001). "Can investors profit from the prophets? Security analyst recommendations and stock returns." The journal of finance 56(2): 531-563.

Barber, B. M., R. Lehavy, M. McNichols and B. Trueman (2006). "Buys, holds, and sells: The distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendations." Journal of accounting and economics 41(1-2): 87-117.

Barber, B. M., R. Lehavy and B. Trueman (2007). "Comparing the stock recommendation performance of investment banks and independent research firms." Journal of Financial Economics 85(2): 490-517.

Barberis, N. and R. Thaler (2003). "A survey of behavioral finance." Handbook of the economics of finance 1: 1053-1128.

Bilinski, P., D. Lyssimachou and M. Walker (2012). "Target price accuracy: International evidence." The Accounting Review 88(3): 825-851.

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

Bradshaw, M., Y. Ertimur and P. O'Brien (2017). "Financial Analysts and Their Contribution to Well-Functioning Capital Markets." Foundations and Trends® in Accounting 11(3): 119-191.

Bradshaw, M. T. (2011). "Analysts' forecasts: what do we know after decades of work?" Working Paper.

Bradshaw, M. T., L. D. Brown and K. Huang (2013). "Do sell-side analysts exhibit differential target price forecasting ability?" Review of Accounting Studies 18(4): 930-955.

Bradshaw, M. T., A. G. Huang and H. Tan (2014). "Analyst target price optimism around the world." Unpublished Working Paper.

Brav, A. and R. Lehavy (2003). "An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics." The Journal of Finance 58(5): 1933-1967.

Breiman, L. (2001). "Random forests." Machine learning 45(1): 5-32.

Cowen, A., B. Groysberg and P. Healy (2006). "Which types of analyst firms are more optimistic?" Journal of accounting and economics 41(1): 119-146.

Da, Z., K. P. Hong and S. Lee (2016). "What Drives Target Price Forecasts and Their Investment Value?" Journal of business finance & accounting 43(3-4): 487-510.

Das, S. R., DeFusco, Richard A., , D. W. McLeavey, Pinto Jerald E., and D. E. Runkle (2018). Multiple Regression and Machine Learning. CFA refresher readings, CFA Institure. 2019.

Fama, E. F. and J. D. MacBeth (1973). "Risk, return, and equilibrium: Empirical tests." Journal of political economy 81(3): 607-636.

Grinblatt, M. and M. Keloharju (2000). "The investment behavior and performance of various investor types: a study of Finland's unique data set." Journal of financial economics 55(1): 43-67.

Groysberg, B., P. Healy and C. Chapman (2008). "Buy-side vs. sell-side analysts' earnings forecasts." Financial analysts journal 64(4): 25-39.

Groysberg, B. and P. M. Healy (2013). Wall Street research: Past, present, and future, Stanford University Press.

Groysberg, B., P. M. Healy and D. A. Maber (2011). "What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?" Journal of accounting research 49(4): 969-1000.

Haig, A. and B. Rees (2016). "The changing market for investment research: the impact of innovations in the payment mechanism on market structure and information flows."
Joos, P., J. D. Piotroski and S. Srinivasan (2016). "Can analysts assess fundamental risk and valuation uncertainty? An empirical analysis of scenario-based value estimates." Journal of Financial Economics 121(3): 645-663.
Joos, P. R. and J. D. Piotroski (2017). "The best of all possible worlds: unraveling target price optimism using analysts' scenario-based valuations." Review of accounting studies 22(4): 1492-1540.
Kang, Q., X. Li and T. Su (2018). "Sell-Side Financial Analysts and the CFA® Program." Financial analysts journal 74(2): 70-83.
Malloy, C. J. (2005). "The geography of equity analysis." The Journal of Finance 60(2): 719-755.
Meehl, P. E. (1957). "When shall we use our heads instead of the formula?" Journal of counseling psychology 4(4): 268.
Miller, E. M. (1977). "Risk, uncertainty, and divergence of opinion." The Journal of finance 32(4): 1151-1168.
Morningstar. (2013). "Morningstar's quantitative equity &
credit ratings."
Morningstar. (2015). "Equity Research Methodology. Morningstar Inc.".
Morningstar (2017). Morningstar Quantitative Research. Swiss CFA Society.
Pinto, J. E., T. R. Robinson and J. D. Stowe (2015). "Equity valuation: a survey of professional practice."
Poon, S.-H. and C. Granger (2005). "Practical issues in forecasting volatility." Financial analysts journal: 45-56.
Poon, S.-H. and C. W. Granger (2003). "Forecasting volatility in financial markets: A review." Journal of economic literature 41(2): 478-539.
Ramnath, S., S. Rock and P. Shane (2008). "The financial analyst forecasting literature: A taxonomy with suggestions for further research." International Journal of Forecasting 24(1): 34-75.
Shane, P. B., S. Rock and S. Ramnath (2006). "Financial Analysts' Forecasts and Stock Recommendations: A Review of the Research." Foundations and Trends® in Finance 2(4): 311-421.
Spence, C., M. Aleksanyan, Y. Millo, S. Imam and S. Abhayawansa "Earning the 'write to speak': sell-side analysts and their struggle to be heard." Contemporary accounting research.
Szado, E., H. Kazemi and T. Schneeweis (2018). "Option Informed Stock Picking." The Journal of Alternative Investments 21(1): 48-66.
Tetlock, P. E. (2017). Expert political judgment: How good is it? How can we know?, Princeton University Press.

Appendix 1. Variable definitions.

Name	Definition	Source
Last Qtr Vol *Last Year Vol*	The standard deviation of 13 (52) trailing weekly returns, including dividends, in local currency and annualized by multiplying by the square root of 52. Estimates below 0.1 or above 0.75 are set to missing.	Datastream
Next Qtr Vol	The standard deviation of 13 leading weekly returns, including dividends, in local currency and annualized by multiplying by the square root of 52. Estimates below 0.1 or above 0.75 are set to missing.	Datastream
AU	Analysts' uncertainty is the difference between analyst's 'consider sell' price, which is the price at which a one-star (sell) recommendation would be triggered and 'consider buy' price, which is the price at which a 5-star (buy) recommendation would be triggered, (scaled by analysts' fair value estimate).	Morningstar
QU	Machine learning uncertainty is the interquartile range of 500 quantitative valuations generated by Morningstar quant's machine learning algorithm scaled by the machine learning fair value estimate. Estimates above 0.75 are set to missing.	Morningstar
ABFor	Forecast volatility in the subsequent quarter based on latest analyst risk assessment (AU).	
MBFor	Forecast volatility in the subsequent quarter based on latest ML risk assessment (QU).	
FBFor	Forecast volatility in the subsequent quarter based on controls.	
Analyst Buy, Analyst Hold, Analyst Sell	Analysts' recommendation is a set of indicator variables where A_Buy = 1 if analyst has assigned a 4 or 5 star recommendation, A_Hold =1 if analysts has assigned 3-star, and A_Sell if analyst assigned 4 or 5 star and zero otherwise.	Morningstar
Beta	Bayesian adjusted beta estimate based on 5 prior years of monthly data where Bayesian beta = 0.4 + .6(estimated coefficient). Estimates below zero or above 2.25 set to missing.	Datastream
Log(MktCap)	Firm size measured by US$ market capitalization of equity.	Datastream
TP CV	Target price coefficient of variation measured by the standard deviation of target price estimates divided by the mean target price. Estimates greater than 0.4 are set to missing.	I/B/E/S
Industry	GICS level 2 industry classification	Datastream
Country	Country of listing	Morningstar

Appendix 2 Sample by Legal System

	Comparison Sample (with both analyst and ML risk indicators).				Extension Sample (with ML risk indicators only).			
	Common	Code	US	All	Common	Code	US	All
2012q2	111	89	412	612	561	4,527	34	5,122
2012q3	121	97	412	630	765	5,373	32	6,170
2012q4	138	86	419	643	988	5,571	31	6,590
2013q1	163	102	400	665	994	5,704	35	6,733
2013q2	224	110	393	727	664	5,819	34	6,517
2013q3	218	113	415	746	754	6,093	35	6,882
2013q4	225	109	415	749	1,032	6,205	37	7,274
2014q1	219	115	421	755	814	6,276	37	7,127
2014q2	240	116	427	783	1,211	6,840	38	8,089
2014q3	244	119	428	791	1,173	6,493	43	7,709
2014q4	253	122	430	805	1,212	5,861	44	7,117
2015q1	253	121	433	807	1,238	6,395	47	7,680
2015q2	261	123	440	824	1,264	5,428	47	6,739
2015q3	250	122	432	804	1,223	5,204	46	6,473
2015q4	258	131	437	826	1,317	5,843	42	7,202
2016q1	252	149	426	827	1,368	6,603	47	8,018
2016q2	259	156	452	867	1,385	7,569	55	9,009
2016q3	262	165	466	893	1,396	7,854	59	9,309
2016q4	275	182	473	930	1,428	7,889	60	9,377
2017q1	282	181	482	945	1,452	7,907	52	9,411
2017q2	281	179	469	929	1,436	7,294	53	8,783
2017q3	271	180	476	927	1,418	6,867	49	8,334
2017q4	264	177	473	914	1,396	6,267	51	7,714
2018q1	268	185	427	880	1,403	7,475	50	8,928
2018q2	280	186	486	952	1,536	8,253	60	9,849
2018q3	279	182	482	943	1,519	8,612	66	10,197
Total	6,151	3,597	11,426	21,174	30,947	170,222	1,184	202353

Appendix 3. Descriptive Statistics

Panel A: Sample for which both analysts' and machine learning based risk assessments are available					
	N	Mean	Std. Dev.	Min	Max
Next Qtr Vol	21,174	0.240	0.099	0.071	0.749
Last Year Vol	21,174	0.238	0.099	0.067	1.179
Last Qtr Vol	21,174	0.239	0.099	0.071	0.749
AU	21,174	0.767	0.181	0.439	1.765
QU	21,174	0.142	0.078	0.038	0.747
Beta	21,174	0.986	0.363	0.002	2.248
Log(MktCap)	21,174	16.366	1.333	12.397	20.810
TP CV	21,174	0.109	0.051	0.000	0.400
ABFor	21,174	0.239	0.057	0.121	0.497
MBFor	21,174	0.239	0.055	0.148	0.761
FBFor	21,174	0.239	0.086	0.046	0.748

Panel B: Sample for which only machine learning based risk assessments and I/B/E/S target prices are not available.					
	N	Mean	Std. Dev	Min	Max
Next Qtr Vol	112,793	0.316	0.133	0.070	0.750
QU	112,793	0.285	0.114	0.050	0.750
MBFor	112,793	0.316	0.055	0.144	0.623
FBFor	112,793	0.316	0.096	0.086	0.831
Last Year Vol	112,793	0.341	0.137	0.070	1.000
Last Qtr Vol	112,793	0.318	0.136	0.070	0.750
Log(MktCap)	112,793	13.020	1.172	6.596	19.225
Beta	112,793	0.864	0.434	0.000	2.250

Panel C: Sample for which only machine learning based risk assessments and I/B/E/S target prices are available.

	N	Mean	Std. Dev	Min	Max
Next Qtr Vol	89,560	0.299	0.113	0.070	0.749
QU	89,560	0.244	0.110	0.050	0.750
MBFor	89,560	0.298	0.054	0.144	0.623
FBFor	89,560	0.297	0.084	0.073	0.785
Last Year Vol	89,560	0.313	0.109	0.071	0.995
Last Qtr Vol	89,560	0.298	0.114	0.070	0.750
Log(MktCap)	89,560	14.297	1.229	9.182	19.499
Beta	89,560	0.963	0.419	0.000	2.250
TP CV	89,560	0.141	0.102	0.000	3.016

Next Qtr Vol., Last Year Vol and Last Qtr Vol are the standard deviation of weekly returns, annualised, for the preceding and following year or quarter respectively. AU and QU are the analysts' and machine learning risk assessments; Beta is the estimated sensitivity of share prices to market movements based on the preceding three years with Bayesian adjustment; Log(MktCap) is the log of market capital measured in US$; TP CV is the coefficient of variation of target prices recorded by I/B/E/S; and ABFor, MBFor and FBFor are the forecasts of volatility based on analysts, machine learning and financial data respectively.

Appendix 4. Correlation Matrix

Panel A: Sample for which both analysts' and machine learning based risk assessments are available. (N=21,174)

	Next Qtr Vol	Last Year Vol	Last Qtr Vol	AU	QU	Beta	Log (MktCap)	TP CV	ABFor	MBFor
Last Year Vol	0.674									
Last Qtr Vol	0.680	0.983								
AU	0.461	0.472	0.475							
QU	0.448	0.444	0.444	0.383						
Beta	0.252	0.268	0.274	0.285	0.214					
Log(MktCap)	-0.284	-0.281	-0.284	-0.216	-0.136	-0.069				
TP CV	0.489	0.538	0.539	0.440	0.402	0.243	-0.137			
ABFor	0.474	0.588	0.591	0.801	0.341	0.208	-0.191	0.385		
MBFor	0.454	0.545	0.544	0.297	0.804	0.137	-0.126	0.340	0.592	
FBFor	0.637	0.912	0.918	0.442	0.406	0.290	-0.329	0.588	0.669	0.617

Panel B: Sample for which machine learning, but not analysts', based risk assessments are available and for which I/B/E/S target prices are not available (N=112,793 including 243 US cases not included in regressions).

	Next Qtr Vol	QU	MBFor	FBFor	Last Year Vol	Last Qtr Vol	Log (MktCap)
QU	0.338						
MBFor	0.346	0.862					
FBFor	0.618	0.421	0.548				
Last Year Vol	0.543	0.414	0.409	0.822			
Last Qtr Vol	0.605	0.381	0.434	0.920	0.750		
Log(MktCap)	-0.031	-0.286	-0.268	-0.058	0.005	-0.009	
Beta	0.137	0.172	0.175	0.245	0.362	0.242	0.044

Panel C: Sample for which machine learning, but not analysts', based risk assessments are available and for which I/B/E/S target prices are available (N=89,560 including 941 US cases not included in regressions).

	Next Qtr Vol	QU	MBFor	FBFor	Last Year Vol	Last Qtr Vol	Log (MktCap)	Beta
QU	0.436							
MBFor	0.445	0.852						
FBFor	0.634	0.464	0.613					
Last Year Vol	0.578	0.500	0.486	0.807				
Last Qtr Vol	0.624	0.457	0.512	0.914	0.773			
Log(MktCap)	-0.194	-0.271	-0.245	-0.225	-0.213	-0.191		
Beta	0.132	0.201	0.158	0.173	0.276	0.196	-0.008	
TP CV	0.264	0.266	0.240	0.299	0.347	0.310	-0.036	0.115

Table 1 Panel A Sample Derivation

Comparison Sample: Cases with both analyst and machine learning data. Extension Sample: Cases with machine learning but not analyst data.	Comparison Sample		Extension Sample	
Firm quarters with US$ market value plus both analysts and ML risk assessments		25,670		253,578
Missing lagged or leading volatility	1,492		12,511	
Missing beta	2,022		847	
Missing I/B/E/S target price	60		na	
Missing generated forecasts	346	3,920	346	13,704
		21,750		239,874
Outliers eliminated		576		37,521
		21,174		202,353
US		11,426		1,184
Code Law		3,597		170,222
Common Law (XUS)		6,151		30,947
With I/B/E/S target price available		21,174		89,560
Without I/B/E/S target price available		na		112,793

Table 1 Panel B Sample Distribution by Industry

	Extension Sample		Comparison Sample	
	Number	Percent	Number	Percent
Auto & Parts	6,793	3.36	482	2.29
Banks	9,249	4.57	1,105	5.24
Basic Resources	9,872	4.88	858	4.07
Chem. &	11,519	5.69	548	2.6
Construct.	12,364	6.11	609	2.89
Financials	8,392	4.15	1,174	5.57
Food & Beverage	11,217	5.54	720	3.42
Healthcare	12,714	6.28	2,047	9.71
Ind. Goods	39,323	19.43	2,867	13.61
Insurance	2,977	1.47	767	3.64
Media	4,285	2.12	480	2.28
Oil & Gas	4,535	2.24	1,521	7.22
Personal & House	11,696	5.78	1,108	5.26
Real Estate	13,482	6.66	1,111	5.27
Retail	10,202	5.04	1,613	7.65
Technology	17,131	8.47	1,539	7.3
Telecom	2,358	1.17	601	2.85
Travel & Leisure	7,922	3.91	665	3.16
Unclassified	65	0.03	102	0.60
Utilities	6,257	3.09	1,257	5.97
Total	202,353	100	21,174	100

Table 1 Panel C Sample distribution by Country

Extension Sample			Comparison Sample		
Country	Number	Percent	Country	Number	Percent
China	44,557	22.02	United States	12,221	57.72
Japan	41,355	20.44	Australia	2,889	13.64
Taiwan	14,910	7.37	Canada	1,041	4.92
India	12,103	5.98	UK	944	4.46
Korea	8,281	4.09	France	604	2.85
UK	7,650	3.78	Japan	591	2.79
Germany	4,531	2.24	New Zealand	488	2.3
Sweden	4,413	2.18	Germany	430	2.03
France	4,150	2.05	Switzerland	338	1.6
South Africa	3,466	1.71	Netherlands	281	1.33
Indonesia	3,458	1.71	Singapore	256	1.21
Brazil	3,070	1.52	Italy	218	1.03
Switzerland	2,902	1.43	Denmark	164	0.77
Malaysia	2,802	1.38	Spain	158	0.75
Australia	2,701	1.33	Sweden	143	0.68
Singapore	2,700	1.33	Belgium	124	0.59
Thailand	2,520	1.25	Norway	63	0.3
Italy	2,171	1.07	Finland	60	0.28
Turkey	2,130	1.05	Korea	54	0.26
Poland	2,060	1.02	Taiwan	44	0.21
Mexico	1,941	0.96	Portugal	26	0.12
Norway	1,899	0.94	South Africa	25	0.12
Saudi Arabia	1,860	0.92	Mexico	12	0.06
Israel	1,816	0.9			
Chile	1,808	0.89			
Russia	1,572	0.78			
Spain	1,429	0.71			
Belgium	1,385	0.68			
Denmark	1,254	0.62			
Finland	1,238	0.61			
United States	1,184	0.59			
	189,316	93.55			
48 other countries	23,037	6.45			
Total	202,353	100.00	Total	21,174	100

Table 2 Predictions of Next Quarter Volatility (Comparison Sample)

	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol
ABFor	1.011^{**}				0.754^{**}	0.257^{**}		0.217^{**}	0.245^{**}
	(38.94)				(28.24)	(12.19)		(11.87)	(11.02)
MBFor		1.044^{**}			0.712^{**}		0.296^{**}	0.258^{**}	0.271^{**}
		(24.72)			(17.16)		(9.18)	(8.09)	(7.84)
FBFor			0.988^{**}						0.819^{**}
			(46.55)						(28.89)
Last Year Vol				0.159^{**}		0.150^{**}	0.137^{**}	0.133^{**}	
				(3.57)		(3.70)	(3.14)	(3.29)	
Last Qtr Vol				0.415^{**}		0.385^{**}	0.393^{**}	0.370^{**}	
				(9.22)		(8.72)	(8.08)	(7.82)	
Log(MktCap)				-0.00690^{**}		-0.00615^{**}	-0.00659^{**}	-0.00596^{**}	
				(-10.13)		(-8.50)	(-7.98)	(-7.12)	
TP CV				0.322^{**}		0.275^{**}	0.272^{**}	0.238^{**}	
				(24.84)		(20.26)	(16.94)	(15.80)	
Beta				0.0174^{**}		0.0123^{**}	0.0149^{**}	0.0108^{**}	
				(5.05)		(3.98)	(4.72)	(3.66)	
Intercept	-0.000400	-0.00916	0.00868	0.164^{**}	-0.109^{**}	0.110^{**}	0.106^{**}	0.0675^{**}	-0.0750^{**}
	(-0.08)	(-1.01)	(2.01)	(13.47)	(-11.16)	(7.16)	(5.98)	(3.41)	(-7.98)
N	21174	21174	21174	21174	21174	21174	21174	21174	21174
ave. R^2	0.244	0.215	0.528	0.558	0.330	0.570	0.574	0.583	0.556

t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$

Table 3. Difference in Predictive Ability Across Legal Systems (Comparison Sample)

	Common (XUS) Next Qtr Vol	Code Next Qtr Vol	Common (US) Next Qtr Vol	Common (XUS) Next Qtr Vol	Code Next Qtr Vol	Common (US) Next Qtr Vol	Common (XUS) Next Qtr Vol	Code Next Qtr Vol	Common (US) Next Qtr Vol
ABFor	1.062**	0.534**	1.138**				0.756**	0.372**	0.869**
	(33.05)	(10.34)	(28.44)				(17.48)	(8.68)	(23.27)
MBFor				1.063**	0.620**	1.281**	0.737**	0.499**	0.848**
				(18.73)	(12.54)	(20.44)	(10.87)	(9.58)	(20.16)
Intercept	-0.0185*	0.108**	-0.0241*	-0.0198	0.0893**	-0.0587**	-0.124**	0.0255	-0.159**
	(-2.63)	(10.11)	(-2.75)	(-1.45)	(7.31)	(-4.81)	(-9.05)	(2.04)	(-13.21)
N	6151	3597	11426	6151	3597	11426	6151	3597	11426
ave. R^2	0.243	0.114	0.299	0.234	0.165	0.234	0.348	0.214	0.392

t statistics in parentheses, $^{*}p < 0.05$, $^{**}p < 0.01$

Table 4 Predictions of Next Quarter Volatility (Extension Sample)

	Cases without I/B/E/S cover			Cases with I/B/E/S cover			
	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol	Next Qtr Vol
MBFor	0.995**		0.211**	1.106**		0.320**	0.303**
	(22.82)		(6.02)	(19.57)		(11.02)	(10.91)
FBFor		0.974**	0.923**		1.003**	0.905**	0.886**
		(44.82)	(43.84)		(38.58)	(32.89)	(32.85)
TP CV							0.0755**
							(11.38)
Intercept	0.00266	0.0113**	-0.0384**	-0.0321	0.00383	-0.0641**	-0.0606**
	(0.19)	(2.96)	(-4.22)	(-1.74)	(0.68)	(-6.86)	(-6.71)
N	112793	112793	112793	89560	89560	89560	89560
ave. R^2	0.118	0.412	0.418	0.187	0.441	0.457	0.459

t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$

Table 5. Difference in Coefficients Across Trading Recommendations

	Next Qtr Vol ABFor	Next Qtr Vol ABFor	Next Qtr Vol MBFor	Next Qtr Vol MBFor	Next Qtr Vol FBFor	Next Qtr Vol FBFor
Analyst Sell	0.923**		0.882**		0.865**	
	(18.72)		(15.51)		(67.60)	
Analyst Hold	0.904**		0.922**		0.865**	
	(19.66)		(15.62)		(70.29)	
Analyst Buy	1.004**		0.901**		0.903**	
	(21.19)		(16.15)		(72.15)	
ML Sell		0.853**		0.797**		0.844**
		(19.38)		(14.65)		(61.46)
ML Hold		0.874**		0.826**		0.859**
		(18.26)		(14.72)		(64.56)
ML Buy		0.986**		0.874**		0.900**
		(20.95)		(15.96)		(74.66)
Intercept	0.0157	0.0253*	0.0233	0.0415**	0.0302**	0.0327**
	(1.44)	(2.41)	(1.73)	(3.23)	(10.73)	(11.12)
N	21072	20755	21072	20755	21072	20755
ave. R^2	0.389	0.396	0.357	0.378	0.573	0.574
F-test	45.23**	53.58**	7.13**	22.40**	26.15**	50.69**

t statistics in parentheses, $^{*}\,p < 0.05$, $^{**}\,p < 0.01$

Table 5 shows results from equation 2:

$$\sigma_{i,t+1} = \alpha_0 + \beta_1 Sell_{i,t} * (ABFor\ or\ MBFor\ or\ FBFor_{i,t}) + \beta_2 Hold_{i,t} * (ABFor\ or\ MBFor\ or\ FBFor_{i,t}) + \beta_3 Buy_{i,t} * (ABFor\ or\ MBFor\ or\ FBFor_{i,t}) + \varepsilon_{i,t}$$

Dependent variable, Next Qtr Vol is the same in each column.

Appendix 2: A brief introduction to the machine learning algorithm

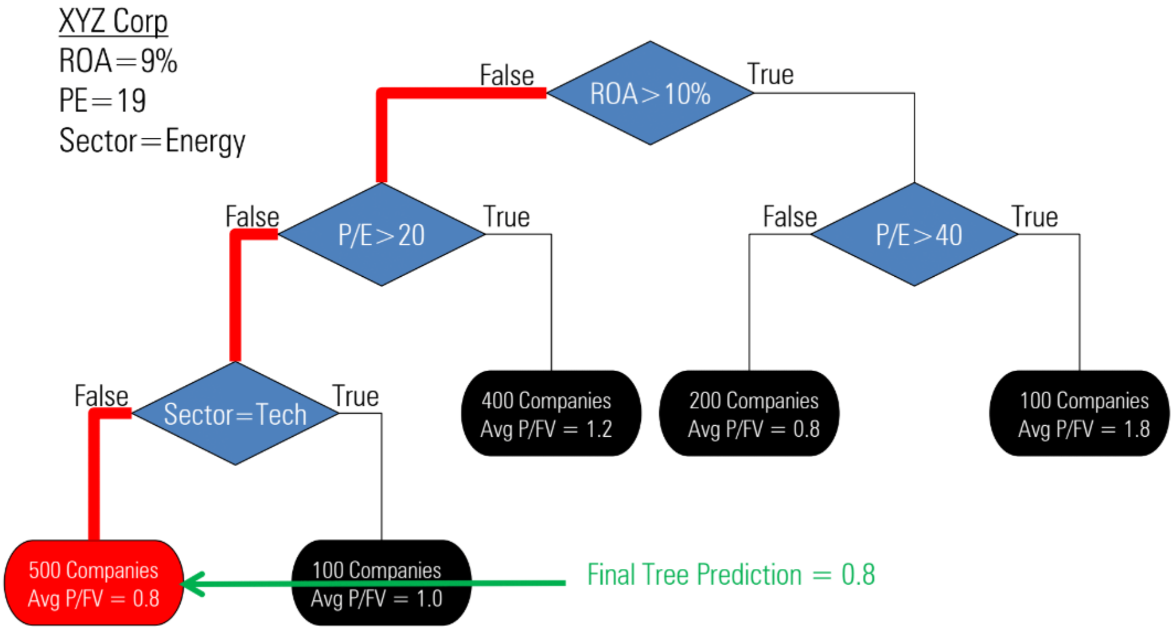
The objective is to establish a relationship between input variables and analyst valuation estimates. While this could be achieved with standard regression methods, machine learning (ML) algorithms are dynamic and learn from experience without relying in the researcher to estimate the relationship. The machine trains, i.e., estimates and updates, an algorithm to improve the predictive accuracy through an iterative process. Since the input variables are structured we refer to this a form of supervised learning (in contrast, unsupervised machine learning which can be used to analyse unstructured data such as assorted text documents.) The input variables are listed in table A1:

Table A1
ROA
Earnings to price
Sales to price
Book value to price
Equity (share price) volatility
Maximum drawdown
Total revenue
Market capitalistation
Enterprise value
Average daily volume
Enterprise value to market value
Sector

Morningstar use regression trees which split data based on statistical analysis of input variables. The learning process continuously improves their ability to perform specified tasks. In our setting, the task is to establish a relationship between twelve input variables and analyst FV/P estimates. ML approach trains, i.e., estimates and updates, an algorithm to improve the predictive accuracy of estimates through an iterative process which learns from experience. Since the input variables are structured we refer to this a form of supervised learning.

A stylised example of a regression tree, mapped out as a flowchart where we start from the top of the page and work towards the bottom, is shown in the Figure A2. The algorithm identifies a condition for predicting fair value to price that a company's return on assets (ROA) is greater than 10%. Companies on each side of the threshold are less scattered (less heterogenous) than we would find in the overall data. This criterion becomes the top node in the tree and two branches, true or false, emerge. Each branch will lead to a second node, for example, determining whether or not the company is in a specific sector.

Figure A2



Source: Morningstar presentation to the Swiss CFA Society, 2017.

Eventually the algorithm stops splitting into further branches because no further improvement in predictive power can be attained. Practical constraints can also be applied, such as a minimum number of companies in each end node. The end nodes at the foot of the tree group observations from the input data. Each node contains observations that are similar to each other based on the splits that have been made in the tree. Although it differs from conventional linear regression, a tree which estimates the value of a continuous variable is known as a regression tree.

Some trees have positive incremental ability to predict analyst FV/P. Rather than relying on a single tree to estimate FV/P, Morningstar uses an ensemble, i.e., many trees, each "grown" from a random subsample. Each tree is constructed from a different subsample of security-level estimates. Morningstar generate 500 random trees each trading day. The use of a large sample of regression trees each grown from a random subsample reduces the signal-to-noise ratio and therefore presents desirable properties. Following Breiman (2001) the approach is referred to as a random forest approach.

This is based on available in Morningstar (2013, 2017). Machine learning has been added to the 2019 CFA curriculum. Members can access a general introduction to this topic at the refresher readings page of the website.