Stockholm School of Economics

Can Machine Be a Good Stock Picker?:

Bridging the Gap between Fundamental Data and Machine Learning

Master Thesis

under the guidance and supervision of

Tobias Sichert

by

Tomoya Narita and Povilas Stankevicius

May 14, 2023

Abstract

We investigate the efficacy of historical accounting data and consensus forecasts for relative valuation of stocks, employing tree-based machine learning methods. We run an XGBoost model for monthly cross-sections of financial and pricing data of US equities from 1984 to 2021. We find that our model is effective for predicting pricing multiples based on non-linear relationships among various financial ratios calculated from historical financial reports, and consensus forecasts contribute to improving prediction errors of valuation. Although some predictors based on analyst forecasts score high in variable importance ranking based on SHAP, overall, they do not become consistently more important than variables based on accounting reports, when analyst forecast data is added to the models. Furthermore, when we use valuation errors as a trading signal for convergence trade, the performance is the best for the trading signals based only on historical accounting data. The convergence trade is successful for small-cap firms, earning sizable abnormal returns with limited portfolio turnover, drawdown and exposure to the Fama French 6 factors. It is suggested that the machine learning method could help to detect cheap and expensive companies within the small-cap universe while avoiding distressed firms.

Keywords: Machine Learning, XGBoost, Relative Valuation, Convergence Trade

Acknowledgement

We would like to express the gratitude to our advisor, Tobias Sichert, for his guidance and helpful feedback, and Dominykas Stankevicius for his advice on machine learing methods.

Contents

Illustrations

Figures

Tables

Chapter 1

Introduction

Machine Learning (ML) applications in Finance have been gaining popularity over the recent years. Much research has been dedicated to predicting stock returns with ML methods, due to the ability of ML to discover complex non-linear relationships in various types of data (Blitz et al., 2023). However, applying ML methods to equity financial data and performing fundamental valuation is still a topic that has not been explored deeply enough, as most researchers have been focusing on time series predictions of future returns. Hanauer et al. (2022) and Geertsema and Lu (2023) implemented a novel approach of using ML models to perform relative valuation of stocks. Inspired by this process we apply our own version of the authors' ML modeling process and explore further what drives the relative valuation of stocks, and whether resulting ML-based valuation can be used as a mispricing signal to earn superior returns for investors.

Relative valuation is a frequently used technique by practitioners to determine the value of a company, find market implied mispricing, and stock picks. The method involves comparing pricing multiples, such as the P/E ratio for a target company against that of a selection of peer companies that have similar characteristics to the target. Pricing equities this way is based on a fundamental idea in finance that two assets with the same risk characteristics should have the same price. However, every company is different, there are no perfect substitutes in the stock market, and many analysts can have different interpretations on what constitutes a good peer group for a single stock. Thus, relative valuation remains an ambiguous subject in finance and company analysis.

New and powerful algorithms in ML have demonstrated success in clustering, classifying, and comparing data according to patterns that are embedded in the data generating process. Thus, finding the right peer group, based on company financial data, appears to be a useful application of ML techniques, as machine can discover new patterns in the financial data, that is not fully obvious to a human, while taking in many different variables describing a company at once. With this motivation, we follow the process of Geertsema and Lu (2023) and attempt to predict pricing multiples for US stocks, based on their financial data.

Decision tree based models, which are easy to implement, can handle a lot of irregularities in the data, and are computationally efficient, have excelled in these applications. Geertsema and Lu (2023) have showed that Gradient Boosting Regression Trees (GBRT) implemented with Light GBM model outperforms traditional regression models in predicting pricing multiples for stocks. We show that relative valuation based on GBRT implemented with XGBoost, another popular ML model similar to Light GBM, generates similar predictions to those of Geertsema and Lu (2023).

What kind of information is used in relative valuation can be completely up to the analyst to decide in practice. Company financial data reports tell us a lot about the business and its performance but may not capture the full picture of the business's future. Therefore, we expand the original framework of ML-based relative valuation to include data from stock analyst consensus forecasts. We expect that as many efforts are put into equity research globally, consensus forecasts should offer more accurate valuation on relative basis, especially as earlier research has found links between stock prices and analyst forecasts (Beaver et al., 2008; Da & Schaumburg, 2011). Therefore, expanding the original framework by Geertsema and Lu (2023) to include consensus data in the ML models is our contribution to this topic.

We find that the inclusion of consensus data lowers out-of-sample prediction errors, although the improvement in prediction accuracy is marginal.

One application of relative valuation is to integrate the techniques into trading strategies. We use the difference between ML-suggested valuation and actual valuation observed in the financial market as a mispricing signal, and examine the efficacy of convergence trade (Liu & Timmermann, 2013; Bartram & Grinblatt, 2018; Hanauer et al., 2022; Geertsema & Lu, 2023). Following Geertsema and Lu (2023), we construct quintile portfolios and long/short portfolios based on ML-suggested valuation errors.

We find that, in line with Geertsema and Lu (2023), the efficacy of the aforementioned strategy drops as we weight the long/short portfolios by market capitalization. By grouping companies based on their market capitalization and constructing long/short portfolios within each group, we report that the long/short portfolios consisting of small companies achieve higher sharpe ratio with significantly low drawdown and smaller exposure to well-known factors. While transaction costs and liquidity consideration would be important for actual trading (Avramov et al., 2021), it is suggested that ML can contribute to detecting better trading opportunities for small companies that are less scrutinized and transparent.

Chapter 2

Related Literature

Valuation Theory and Multiples

Our study is most directly related to the literature focused on relative valuation of equity shares. Equity valuation can be divided into four categories: discounted cash flow valuation, liquidation and accounting valuation, relative valuation, and contingent claim valuation (Damodaran, 2007). While the discounted cash flow valuation attempts to estimate the present value of expected future cash flows on the firm, relative valuation estimates the value of a firm by comparing the pricing of similar firms relative to underlying factors, such as earnings, cash flows, efficiency, and financial soundness. Therefore, it is possible that a company valued fairly on a relative basis is overvalued based on absolute valuation (i.e. discounted cash flow valuation: DCF) when comparable firms are overvalued compared with their intrinsic value in such cases as a stock market bubble (Bartram & Grinblatt, 2018).

In theory, the DCF valuation can be embedded in a single pricing multiple. Justified pricing multiples is an approach, where simplified expressions for multiples are derived from the cash flows of a company. Damodaran (2007) derives expressions for price-to-book (P/B), and price-to-sales (P/S) multiples. This approach can give an insight into what kind of financial data needs to be included among the predictors to get results that are grounded in valuation theory. For P/B and P/S multiples, which will be our target variables in the ML models, ROE and net profit margin appear directly in these equations, and we will explore how ML models take these variables, among others, into account.

Agudze and Ibhagui (2020) find that valuation multiples for aggregate stock indices depend on a set of fundamental variables, such as profit margins, ROA, and ROE. However, the paper only uses 4 different financial metrics, thus it is not clear how a broader set of fundamental variables and analyst forecasts can affect relative valuation from this study.

Rabier (2018) find that growth, R&D, relative size of target companies, are significant in predicting deal pricing in acquisitions. Growth assumptions are a key ingredient in company valuation based on the model of O'Brien (2003). However, Kulshrestha and Nanda (2006) find that payout and beta are important when predicting P/E, but growth is only secondary, although still significant. Another variable, empirically important for stock returns, and consequentially, pricing is firm spending on R&D, especially when done by firms with good corporate governance (Chan et al., 2015). Booth et al. (2006) also perform an empirical study, suggesting that R&D expenses are an important value driver for stocks, arguing that innovative investments are necessary for firms have positive residual incomes, but the valuation of R&D expenses varies in different markets.

Stock analyst growth forecasts are associated with higher analyst target multiples, while higher measures of financial risk of a company are correlated with lower multiples, based on the study by Yin et al. (2014). Beaver et al. (2008) explore analyst forecast revisions and previous forecast errors and detect significance in explaining abnormal stock returns. Furthermore, analyst target price implied relative valuation multiples do carry important content for investors when valuing firms in the same industry or sector, based on the findings of Da and Schaumburg (2011). Johnston et al. (2021) explore how the relevance of analyst forecasts has evolved through time and find that although including analyst information helps to improve the explanatory power for book values and earnings, such relevance has been declining over time. Thus, based on previous research, analyst forecasts appear to have some informational content for stock prices, although to a varying degree.

Different studies have variations in determining which financial variables affect pricing multiples the most. However the common factor from most of these studies revolves around profitability, payouts, riskiness of the company, and growth being the fundamental drivers behind valuation multiples. Hence, using ML as an innovative approach on different datasets to uncover relationships between pricing and financial metrics can add to the existing understanding of relative stock valuation.

Relative Valuation and Return Predictability

The earlier works on relative valuation include Bhojraj and Lee (2002), Liu et al. (2002), Rhodes–Kropf and Viswanathan (2004), and Rhodes–Kropf et al. (2005). Bhojraj and Lee (2002) estimate firms' "warranted" multiples using each firm's fundamentals and its deviation from industry means as drivers of valuation, and show that comparable firms suggested by the warranted multiples forecast future valuation. Liu et al. (2002) investigate the relationship between stock prices and value drivers within the same industry and cross section. Rhodes–Kropf and Viswanathan (2004) and Rhodes–Kropf et al. (2005) decompose the market-to-book ratio to three components: the firm-specific pricing deviation from short-run industry pricing, sector-wide short run deviations from firms' long-run pricing, and long-run pricing to book and show that mispricing drives mergers.

More recent works by Cooper and Lambertides (2014) and Bartram and Grinblatt (2018) examine whether relative valuation can be exploited to predict stock returns. While Cooper and Lambertides (2014) find no evidence that a valuation error (mispricing) predicts a future return, Bartram and Grinblatt (2018) report that a convergence strategy constructed on relative valuation-based mispricing signals earns risk-adjusted returns ranging from 4% to 10% on annual basis.

While prior works (Bhojraj & Lee, 2002; Liu et al., 2002; Rhodes–Kropf & Viswanathan, 2004; Rhodes–Kropf et al., 2005; Cooper & Lambertides, 2014; Bartram & Grinblatt, 2018) assume a linear relationship between target and explanatory variables and employs ordinary least squares method, Hanauer et al. (2022) and Geertsema and Lu (2023) applied tree-based ML method to relative valuation. Hanauer et al. (2022) report quintile portfolios based on mispricing signals result in substantially higher risk-adjusted returns compared with that of linear regression models in European stock markets. Geertsema and Lu (2023) find that ML-based relative valuation outperforms traditional models in out of sample tests in terms of prediction accuracy in the US stock markets. The outperformance of ML-based approaches relative to traditional approaches suggests the importance of allowing for non-linearities and interconnections among underlying variables to estimate valuation multiples (Hanauer et al., 2022; Geertsema & Lu, 2023).

It should be noted that identifying comparable firms for relative valuation requires judgment in some cases. While the concept of grouping economically similar firms is simple, companies operate in multiple business areas, and the classification can be arbitrary with multiple classification standards, such as the Standard Industrial Classification (SIC), the North American Industry Classification System (NAICS), and the Global Industry Classification Scheme (GICS) (Kim & Ritter, 1999; Weiner, 2005; Lee et al., 2015). For example, Ding et al. (2019) report the effectiveness of applying ML methods for identifying peer firms. The merit of using the tree-based algorithm is that companies with similar features fall on the same leaf based on fundamental variables, therefore decreasing the reliance on a certain industry classification standard. Geertsema and Lu (2023) include industry codes from Fama and French's 48 industry classification as one of the features.

Machine Learning and Data Handling Methods

ML models, have been growing in popularity over recent years as they can handle complex data and generate accurate predictions (Blitz et al., 2023). Although regression models are widely used in financial and economic research, they rely on a specified set of assumptions, which are not always true in real-world conditions. In contrast, ML models can learn the underlying relationships in the data without many pre-set assumptions. In addition, ML models can learn complex, non-linear data relationships. The ability to handle missing data, outliers, and noise is another advantage of ML. Tree based models, which are used in our case, perform especially well with these irregularities in the data, and are well suited as an "off the shelf" application for ML (Hastie, 2009). Missing values and outliers are prevalent in our dataset, which strengthens the case for ML algorithms for our specific problem.

At the same time, there are some caveats related to ML methods. As Gu et al. (2020) and Hanauer et al. (2022) point out, the number of parameters increases dramatically when applying ML models compared with traditional linear models, possibly leading to unreliable estimated parameters in the absence of enough observations. Avramov et al. (2021) also report the profitability of ML-based trading performance while the existence of trading costs and high turnover erode the performance. Another caveat is that only specific traditional econometric methods can work well when confronted with a panel dataset. One of the main challenges in these datasets is accounting for both time-series and cross sectional dependencies, and badly specified models can make inaccurate predictions (Tayebi et al., 2021). We split the dataset up into monthly cross sections and estimate separate models for each month, which was also done by Geertsema and Lu (2023). This acts as a work-around for the temporal dependency issue.

Cross sectional splitting for stock fundamental data and relative valuation due to varying time trends in variables has been also discussed by Bartram and Grinblatt (2018), who employed cross-sectional regressions relating market values to accounting information, in a similar approach to Geertsema and Lu (2023). Based on the authors' arguments, each firm's regression peer-implied fair value evolves month to month for two reasons. Firstly, market capitalizations, the cross-sectional regression's dependent variable, change, influencing predictor coefficients. Thus, even if the financial information does not change month to month, regression coefficients and fitted values can be different. Relative pricing of market sectors can also have a similar effect on predictions. A cross-sectional approach to relative valuation can capture these effects almost in real time. Also, when new accounting information arrives, the regression equation coefficients change as well, thus new information becomes fully incorporated as soon as it is available. However, as Bartram and Grinblatt (2018) admit, that this approach does not take into account the time varying market preferences for certain stocks or the market over or under valuation as a whole. For example, a certain stock may be fairly valued relative to its peer companies, but may be expensive relative to its own history, if the price of the target company and the peer group have been rising. Therefore, such pure relative valuation may not fully capture the intrinsic value of a company. However, Aswath Damodaran (2009) suggests that comparing valuation multiples across time is difficult, as outside factors, such as interest rates can have effects on pricing multiples. Thus, to correctly include both cross sectional and time varying effects into a single econometric or ML model may require more data transformations, extra exogenous variables, and tailored model specifications.

ML models are generally used on large datasets. In our case, the monthly cross section used to train the model is a relatively small sample by ML standards. Although splitting into cross sections is done to avoid issues arising from temporal dependencies, the accuracy of XGBoost, which is our main ML model, and other models may decline with a reduction in sample size. However, tree-based models, such as XGBoost have built-in regularization terms, that assist in preventing overfitting of the data and have been found to perform better than other ML algorithms when sample size is small (Chen & Guestrin, 2016; dmlc, 2022; Zou et al., 2022).

Nevertheless, Geertsema and Lu (2023), show the success of ML applications to relative valuation, and academic evidence provides a link between stock analyst forecasts, target prices carry and returns. Therefore, our hypothesis is that including consensus forecast data in a ML approach could improve prediction accuracy and uncover new relationships in the relative valuation field.

Chapter 3

Data and Methodology

Main Database

Our dataset contains US equities from 1984 to 2021 that exist in the Financial Ratios Firm Level by WRDS. The stock universe contains companies from CRSP dataset. The financial ratios database takes company financial data from Compustat, CRSP and IBES. The dataset contains over 70 different financial ratios, from which we use valuation multiples as our target variables and the rest as predictors. The dataset is compiled on a monthly frequency where financial data for a given balance sheet date is lagged by two months to avoid a look ahead bias.[1] However, Geertsema and Lu (2023) use three months as the period before financial data becomes publicly available.

In addition, we include S&P Capital IQ industry classification, and a few other historical data variables as predictors, taken from the Compustat monthly and quarterly databases in WRDS. All variable descriptions can be found in Table 3.2 and List of Variables in Appendix. Monthly total returns, used in testing ML based trading strategies are taken from the Compustat monthly dataset as well.

Stock Universe and Variable Selection

Following the process by Geertsema and Lu (2023), we select only US stocks that have been trading on NYSE, Nasdaq, and AMEX. Furthermore, we filter out the 10% of companies with lowest assets, sales, and book equity, to mitigate the effect from micro and

1. See: WRDS Industry Financial Ratio.

cap stocks. Size filtering is done based on accounting variables, not market capitalization, to avoid the effect of market capitalization on valuation multiples.

We note that the WRDS Compustat Financial ratios database does not contain financial companies, while Geertsema and Lu (2023) have included these in their analysis. Most of the financial ratios in the WRDS do not have a solid economic meaning for financial companies (Damodaran, 2009, 2013). Thus, we choose to proceed with leaving the financial companies out, which can also test the robustness of this approach by having a different set of companies in the main dataset.

There is a range of valuation multiples in the WRDS Compustat Financial ratios database. For our analysis, we select P/B ratio and P/S ratio as the target multiples to be used in ML models. We proceed to filter out observations where both target multiples are missing, and transform these with the natural logarithm.

Practitioners generally use a range of other multiples, such as price to earnings, enterprise value to EBITDA, price to cashflow, and others. We make a choice to optimize for the multiples with the most non missing and non-negative values. Therefore, we focus solely on P/B and P/S. P/B multiple can be compared to market-to-book value ratio, used by Geertsema and Lu (2023), who also used EV/Sales and EV/Assets as their targets. We use prediction error statistics for P/B to compare our model with market-to-book value model of Geertsema and Lu (2023).

After performing data fltering described above, we end up with a dataset containing 1,297,423 observations from 1984 to 2021. This appears to be relatively in line with the Geertsema and Lu (2023) 1,811,786 observations on a comparable time period (1980 – 2019), which also included financial sector companies. We use all the financial ratios, except for the valuation multiples in the Compustat Financial Ratios dataset, and include CAPM beta and Capital IQ industry classification. Furthermore, we add total asssets, sales, and equity, last twelve-month growth in Sales, EBIT, EBITDA, Net Income, and Common Equity, together with a measure of operating leverage as predictors, which we calculated from the quarterly financial database from Compustat. After these steps, we end up with 70 predictor variables for our target multiples, while Geertsema and Lu (2023) used 97 predictors in total.

Exploratory Analysis

Summary statistics for the size of our dataset and correlation metrics for selected predictors are represented in table 3.1 and Figure 3.1, respectively. We note, that missing values are prevalent among the predictor variables in our dataset, as we only have 122,579 observations with no missing values. Tree based ML models, such as XGBoost are capable to deal with missing values on their own, emphasizing the flexibility of ML approach to our problem (Hastie, 2009; Chen & Guestrin, 2016).

Simple correlation analysis for the target multiples and predictors unveils a few key insights. Firstly, certain predictors, such as EBITDA margin, Net Profit margin, R&D to Sales ratio are strongly correlated, which means they give less distinct patterns for the ML model to recognize. In practice, this means that some highly correlated variables may be redundant, however in tree based models, dropping highly correlated variables can result in significantly decreased performance (Kuhn, Johnson, et al., 2013). We use all the variables in our dataset, following the process by Geertsema and Lu (2023).

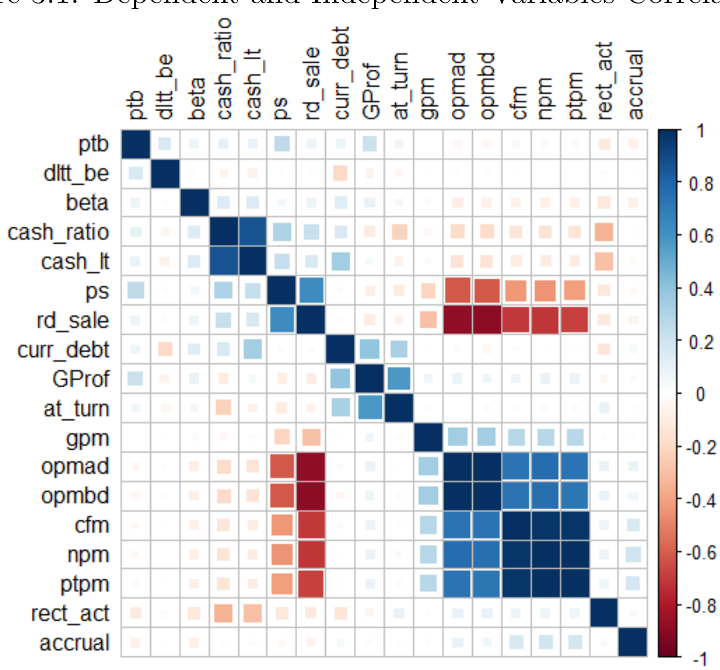
Secondly, the same highly inter-correlated variables have strong correlation with our P/S target multiple, meaning that either they carry important economic information, or company sales being in the numerator, or the denominator of these ratios cause the correlation to be high without much economic meaning. We proceed to estimate our models and re-examine this issue in the next sections.

Table 3.1: Monthly Observations

	# Observations	**# Complete cases**
Min	2103	43
Q1	2613	61
Median	2715	83
Mean	2851	269.4
Q3	3085	617.5
Max	3833	749
Total # observations	1297423	122579
Total # unique firms	10606	1646

Data is taken from WRDS Compustat Financial Ratio Suite. Observations are filtered from 1984 February to 2021 December. Complete cases describe the number of rows (observations) in the dataset, without a single missing value.

Figure 3.1: Dependent and Independent Variables Correlation Matrix



Average of monthly cross-sectional correlation over the sample period.
See List of Variables for variable definitions.

Consensus Forecast Dataset

Our extension to the approach by Geertsema and Lu (2023) is to include consensus data from stock analyst forecasts as explanatory variables when predicting our target multiples. We use S&P Capital IQ analyst consensus data for the purpose of analyzing the information value of analyst forecasts. To avoid the differences in the company fiscal years, we use the next twelve-month consensus forecasts for all financial variables in our selection. Consensus forecast variables that are used in the model, their summary statistics and correlation with the target variables are described in Table 3.2 and Figure 3.2. Missing values are prevalent in this dataset as well, but as already described, XGBoost can deal with these efficiently (Chen & Guestrin, 2016). The data from our Capital IQ sample is available starting the year 2000 for most companies, thus we have a shorter time frame to evaluate the predictive power of analyst forecasts.

From the correlation matrix of the consensus forecast dataset, we can see that variables related to profit margin are highly inter-correlated, as is the case with our main dataset. However, there is no high correlation between these and our target variables, in contrast to the main dataset. ROA, ROE, average broker recommendation, and EPS revisions have the highest correlation coefficients with our target multiples.

We also take market capitalization numbers from Capital IQ, which we use later to analyze the distribution of prediction errors and construct trading strategies. This is available from 2000s, as well as the variables from the consensus forecast dataset. Therefore, our trading strategy analysis will be done from the year 2000, to have all versions of our model on a comparable time frame, even as the main dataset starts in 1984.

Table 3.2: Consensus Forecast Descriptive Statistics

Variable	Name	Q1	Median	Mean	Q3	Missing
roe_ntm	ROE	7.40	11.70	12.40	18.30	90%
roa_ntm	ROA	1.10	4.60	4.80	8.80	92%
nd_ebitda	ND/EBITDA	1.00	2.04	3.74	3.61	53%
avg_rec	Ave. Broker Reccom.	1.75	2.14	2.16	2.64	16%
eps_revi	EPS Revisions	-3.00	0.00	0.08	3.00	39%
ep_ntm_g	EPS Growth	-0.10	21.23	58.91	53.76	20%
rev_ntm_g	Revenue Growth	2.39	7.53	13.10	16.22	26%
ni_ntm_g	Net Income Growth	-28.60	8.09	11.73	37.55	44%
ebitda_ntm_m	EBITDA margin	9.07	15.0	-24.13	24.99	45%
ebit_ntm_m	EBIT margin	5.43	11.40	-50.18	20.62	45%
fcf_ntm_m	FCF margin	2.14	6.38	-79.74	12.27	72%
ebt_ntm_m	EBT margin	4.86	11.11	-68.55	21.01	61%
ni_ntm_m	Net Income Margin	3.31	7.51	-50.50	14.41	44%

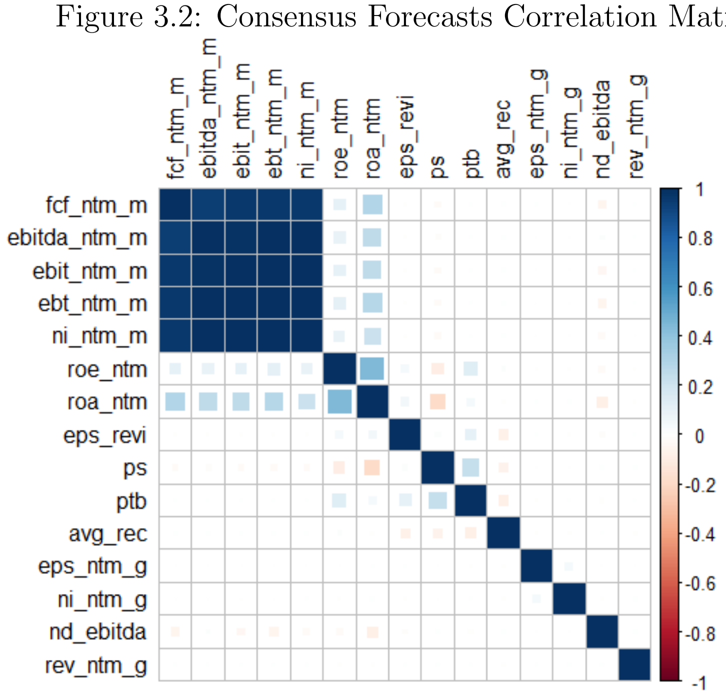
Number of observations: 730987

Average broker recommendation is defined between 0 to 5 points (the higher, the better). EPS revision is calculated as the number of upward revisions divided by the number of downward revisions per last 3 months. Other variables are in percent. The data is from August 2000 to December 2021.

For other variable definitions, see List of Variables in Appendix.

Source: S&P Capital IQ

Figure 3.2: Consensus Forecasts Correlation Matrix



Average of monthly cross-sectional correlation over the sample period.

Boosted Tree Implementation through XGBoost

Our main method for using ML for relative stock valuation closely follows the one used by Geertsema and Lu (2023). We use XGBoost[2], which is a model implementing Gradient Boosted Regression Tree framework, to generate out-of-sample estimation for target multiples for different US equities (Chen & Guestrin, 2016). The whole model is implemented in R statistical programming language using xgboost package as our main workhorse for model estimation and evaluation.

XGBoost works by building and sequentially adding decision trees to the model, with each new tree attempting to correct the errors of previous trees. The objective function is the sum of a desired loss function and a regularization term, which reduces overfitting. The model employs gradient-based optimization, which calculates the gradient of the loss function with respect to the model parameters and updates them in a direction that minimizes the loss function. In addition, XGBoost employs a voting system that combines the predictions of all trees to create one strong prediction, which is known as gradient boosting. Furthermore, the model employs feature and training data subsampling in addition to shrinkage techniques, allowing to extract signals from challenging data, while reducing overfitting and improving the generalization of the model (Chen & Guestrin, 2016; Jason Brownlee, 2021; dmlc, 2022).

2. Geertsema and Lu (2023) uses LightGBM, another implementation of decision tree models.

Main Machine Learning Model

In our procedure, following the steps of Geertsema and Lu (2023), we divide the whole dataset of monthly observations of company financials and pricing multiples into monthly cross-sections. For each month, we split the available stock database on 60-20-20% ratios for model training, validation, and testing respectively. We use 60% of the monthly cross-sectional data to train our model, then we proceed to optimize the number of trees parameter using 20% of the monthly sample as the validation set. The validation is performed using the built-in early stopping function of XGBoost, where predictions are generated for the validation set for every number of trees and the model continues to add more trees until the prediction error of the validation set stops improving. We use the number of trees chosen at the validation step and then proceed to generate out-of-sample stock multiple predictions for the test set.

This process is repeated 5 times per month where 5 monthly test sets are generated using random sampling without replacement, stratified by industry, while training and validation sets are generated from the leftover data for each choice of the test set. The monthly estimation and prediction procedure is performed separately for each month on the given cross-section. This way, we estimate a total of 2275 separate ML models for our dataset (455 months x 5 models per month). This allows us to generate out-of-sample predictions for the whole monthly sample of stock pricing multiples.

Normally, other hyperparameters for tree-based models also need to be optimised in the validation step to improve prediction accuracy. While there exist various hyperparameter tuning techniques, such as grid search and Bayesian optimization, given our case, where we estimate over 2000 different models, tuning hyperparameters becomes too computationally and time intensive. Therefore, we use the default values for other hyperparameters, except for the learning rate, which we set to 0.1, as in Geertsema and Lu (2023).

Machine Learning Model with Consensus Forecasts

ML model with consensus forecast is our extension to the main ML model, designed to include analyst forecasts as additional features. Estimation and prediction process for ML model with consensus forecast is identical to the main ML model, but we make our extension by changing the dataset. From the output of the main model, we calculate the mean absolute percentage SHAP values for each explanatory variable, averaged across all months. Percentage SHAP (SHapley Additive exPlanations) values are obtained by dividing the mean absolute SHAP value for each variable by the sum of mean absolute SHAP values. SHAP values are commonly used to explore how important a variable is in the model (den Broeck et al., 2022). Then we select variables with the mean absolute percentage SHAP values above 1% from the original dataset used in the main ML model to retain the key information from the original dataset.

Afterwards, we merge the selected variables with analyst forecast data to create a combined dataset. We do not exclude observations that contain missing feature values from any of the two datasets. We expect that model using analyst forecast data can improve prediction accuracy over the main model, if stock pricing reflects forward looking information that can be incorporated in analyst forecasts, but not necessarily be represented in historical financial data. Next, we proceed to estimate the models and make out-of-sample predictions as in the main ML model.

Chapter 4

Result

Machine Learning Model Performamce

After running the model estimation and out of sample prediction processes described above, we obtain the out of sample predicted multiples. We use Root Mean Squared Error (RMSE) and Out of Sample R Squared (ROOS) calculated from out of sample prediction errors to evaluate our models. Firstly, we compare the RMSE and ROOS for the P/B model from our main dataset with the results from Geertsema and Lu (2023). The RMSE of 0.52 is comparable to 0.56 of Geertsema and Lu (2023). Although the difference does not appear to be significant, our RMSE is slightly lower, which can be due to several factors such as:

• Differences of predictive accuracy between XGBoost and Light GBM.

• The fact that Geertsema and Lu (2023) include financial companies in their analysis, which might result in lower predictive accuracy, especially as most of the predictive variables used by in the authors' study relate to non-financial companies. Hence, having financial company multiples without the relevant predictors could lower the predictive accuracy of the Geertsema and Lu (2023) models.

• Other variations in the dataset, such as the different time lag used to deal with look ahead bias (2 months used by Compustat Financial Ratios database, compared with 3 months used by Geertsema and Lu (2023). However, given that RMSE is not dramatically lower in our model, the look-ahead bias should not be significant. Variations in stock universe selection process and some of the predictor variables can also cause a slight difference in RMSE.

Comparing the ROOS measure with that of the Geertsema and Lu (2023), results also look similar: 0.56 in our case, versus 0.54. Given that there are only slight variations in the predictive accuracy measures for the price to book multiple between our analysis and Geertsema and Lu (2023), we presume that our model reasonably tracks the original. Details on predictive accuracy for different models can be found in Table 4.1.

Table 4.1: ML-Valuation Performance

Model	**RMSE**	**ROOS**	**Sample Period**	**# Obs.**
P/B	0.52	0.56	Feb 1984 - Dec 2021	1297101
P/B (Cons. Forecast)	0.50	0.64	Aug 2000 - Dec 2021	730987
P/B (Sample)	0.53	0.58	Aug 2000 - Dec 2021	730987
P/S	0.53	0.53	Feb 1984 - Dec 2021	1297101
P/S (Cons. Forecast)	0.51	0.79	Aug 2000 - Dec 2021	730437
P/S (Sample)	0.54	0.76	Aug 2000 - Dec 2021	730437

RMSE and ROOS are defined as in Geertsema and Lu (2023). Sampled errors of main models are calculated from August 2000 for direct comparison with the consensus forecast versions.

One of the main goals of our analysis is to see whether adding consensus forecasts can improve predictive accuracy of the ML models. Based on the results displayed in Table‘4.1, we can see that there is some improvement in predictive error on fully comparable data length with the same number of observations. For P/B, RMSE drops from 0.53 to 0.50, representing a 6% improvement in RMSE. ROOS also improves from 0.58 to 0.64. For P/S, RMSE drops by 6% as well (from 0.54 to 0.51), while ROOS improves from 0.76 to 0.79.

Although the difference in predictive measures is not significant, given that there is no variation in the dataset, but only in predictors, we suggest that there is some predictive power in consensus analyst forecasts, but its contribution is marginal. Thus, analyst forecasts can assist in the relative valuation process, but based on our ML analysis, they need to be taken in the context of historical financial data as well. Because consensus forecast data offered a small improvement in our analysis, we suggest that historical financial data should still form the basis for fundamental relative valuation analysis, when used in practice.

Prediction Errors Through Time

Geertsema and Lu (2023) note that prediction errors from ML models vary through time. In their analysis, the predictive error increased during times of financial crises or market stress. We display the calculated RMSE for our models per month in Figure 4.1 and Figure 4.2. Our results follow a similar pattern to that of Geertsema and Lu (2023).

Given that prediction errors rise during times of market stress, we compare the pattern of RMSE from our P/S model to the VIX index, which is generally understood as the proxy of market stress (CBOE, 2017). From a simple time series chart in Figure 4.3, we show that the VIX index and prediction errors follow a similar pattern.

Furthermore, the time series of RMSE has been trending upwards since 2018. Given that the ML error captures some relationship between the price and the fundamental value of a stock, as discussed by Geertsema and Lu (2023), and extremely loose monetary policy has caused stock multiples to rise, our ML based valuation error increase could be a result of the recent monetary policy measures by the global central banks (McKinsey, 2014).

Figure 4.1: RMSE and ROOS for P/B Valuation

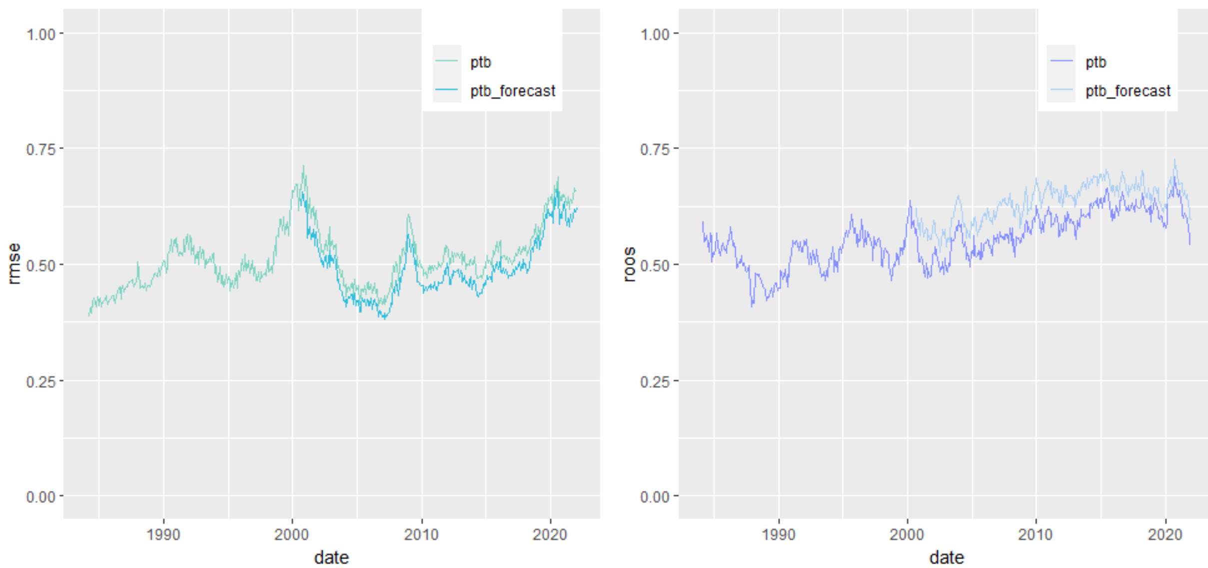
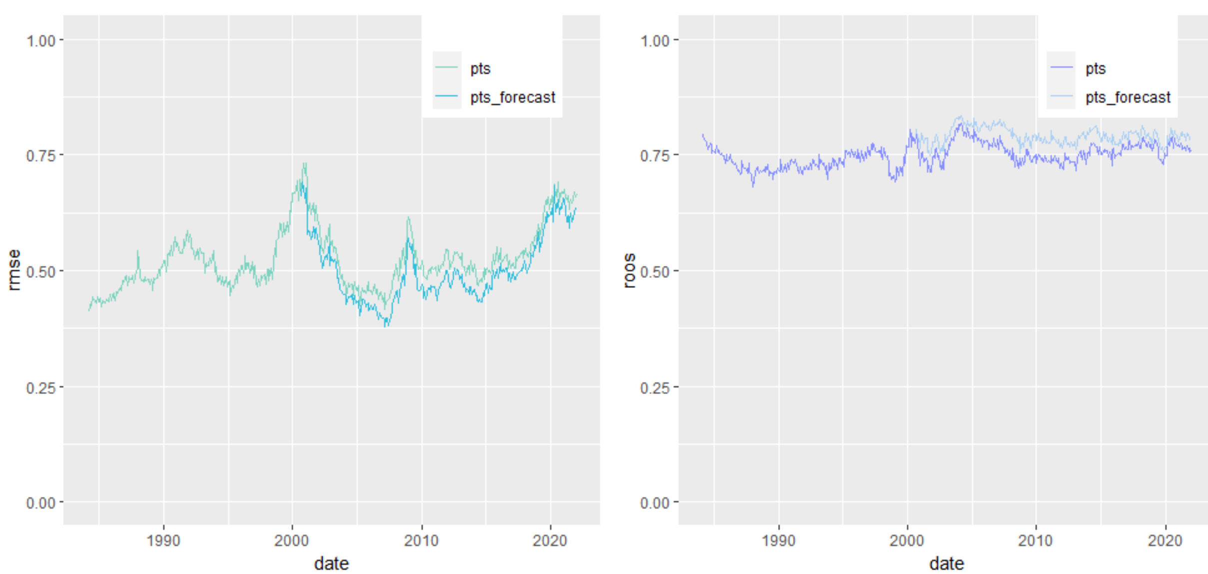
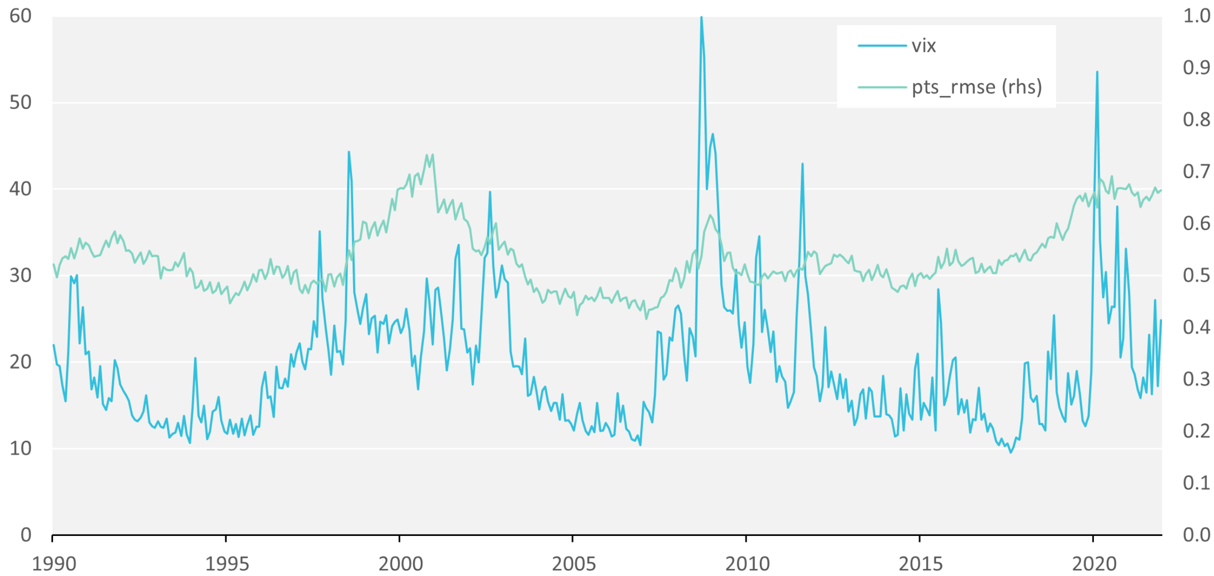


Figure 4.2: RMSE and ROOS for P/S Valuation



RMSE and ROOS measures, calculated for monthly cross sections.

Figure 4.3: Valuation Error and Market Stress



RMSE for the Price to Sales model, calculated for monthly sections and the VIX index.

Patterns in Prediction Errors

Geertsema and Lu (2023) slice the ML prediction errors based on a few variables to explore how do the ML models behave. We follow a similar process, where we calculate the RMSE based on errors divided into quantile ranges of ROE, beta, valuation multiples, and market capitalization. These are reported in Figure 4.4 and Figure 4.5.

There are some interesting results when we split up the valuation error by quantiles of underlying variables. Firstly, higher beta stocks tend to have a higher valuation error. Although high error can be caused both by high over and under valuation, this partly corresponds to the discussion of Zhang (2005), where a low multiple, or undervaluation, can go hand in hand with distress or riskiness of the stock. Therefore, it could be suggested that trading strategies based on our ML framework could fall into value traps – where a stock has a low valuation, but such valuation is justified.

However, separating errors by the multiples themselves, we can see that the highest model errors lie in the tails of the multiple distribution – both high and low valued stocks have high RMSE. If the ML valuation error contains information about the discrepancy between pricing and fundamental value, then this shows that such discrepancy is valid for both cheap and expensive stocks. As tree-based ML prediction can be calculated as a weighted average of peer multiples, as in Geertsema and Lu (2023), then low multiple stocks could get a prediction higher than the actual multiple, and vice versa for expensive stocks. Thus, trading strategies, which we explore later, could be correlated with a value factor.

Interestingly, observations with low ROE also tend to have higher valuation errors. This result is similar to Geertsema and Lu (2023). This raises a further question of value traps – high error prevalence among low ROE firms, could mean that the ML model could suggest an undervaluation that is in reality well justified due to low profitability. If this is true, then our trading strategies, which we explore later, should not perform well. On the other hand, if a trading strategy, where a buy signal is generated for ML undervalued stocks performs well, this could testify that ML framework correctly identifies fundamental value irrespective of profitability and pricing discrepancies relative to fundamentals is simply more prevalent among low profitability companies.

Figure 4.4: RMSE by Quantile for P/B Valuation

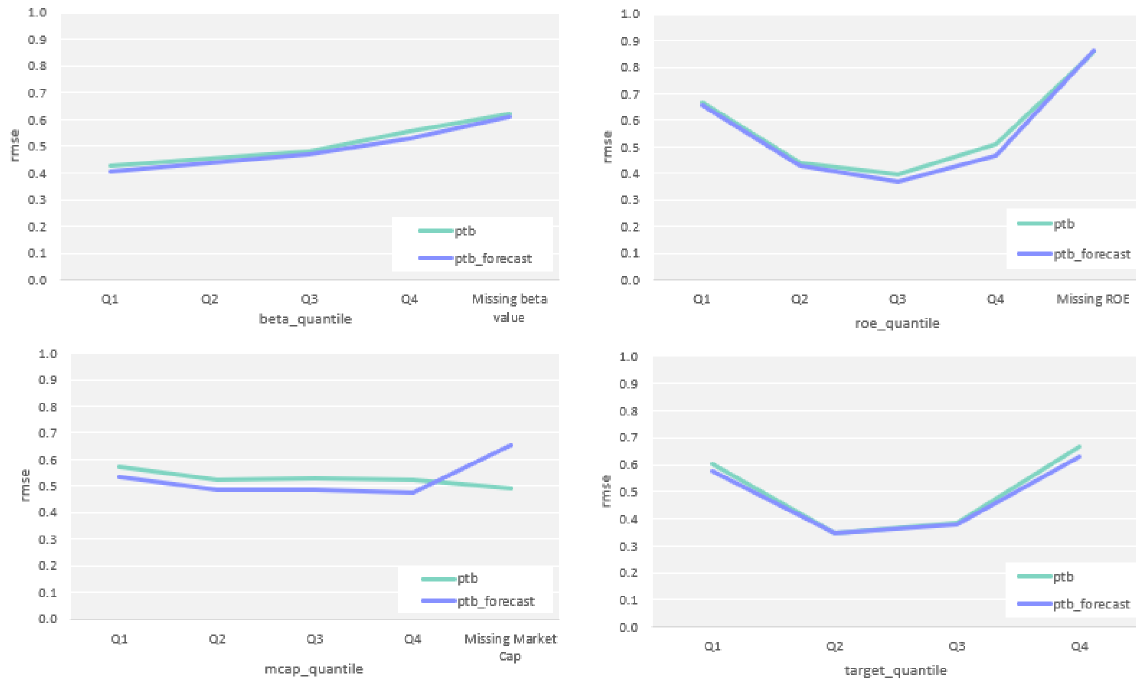
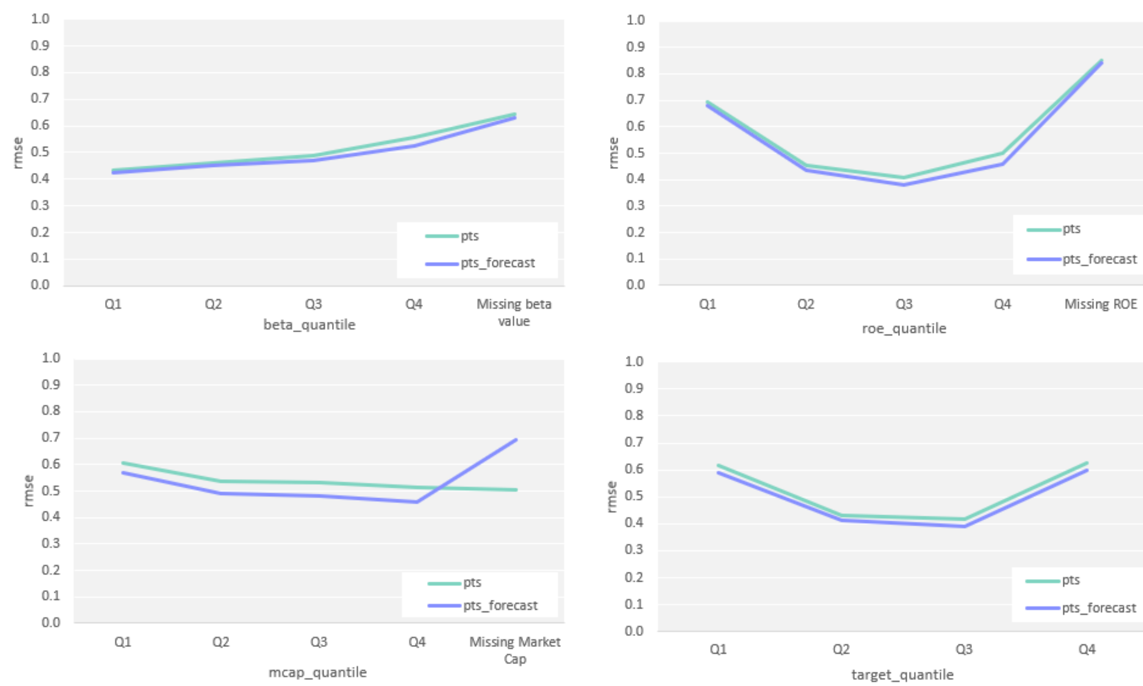


Figure 4.5: RMSE by Quantile for P/S Valuation



Target quantiles are the quantiles of the target multiples of each model. Data is split up into quantiles of underlying variables and RMSE is calculated for each quantile.

Looking at the errors in market capitalization quantiles, small cap companies tend to have higher ML error. This is consistent with the findings of Avramov et al. (2021), where ML based mispricing signals and inclusion in long portfolios based on these are more frequent for small companies. This raises a question, if ML mispricing signals are only profitable for small, illiquid stocks, hence are correlated with the size factor.

Observations with missing values for beta and ROE also tend to have higher prediction errors. A few reasons could contribute to this effect. Firstly, ROE and beta have high variable importance scores, which we discuss in the section below. If these variables are missing then the model does not have the full information needed, and the resulting prediction error is higher. Secondly, the absence of beta and ROE could indicate that other predictor variables for an observation are missing as well, suggesting that the general completeness of an information set for a given observation is important in the prediction process.

Observations where the market capitalisation is missing have higher valuation errors for models with forecast data, but not for the original dataset. This could be a result from the differences in data bases between WRDS and Capital IQ. Observations with missing market capitalisation values from Capital IQ can also have many other missing variables from this database, possibly due to identification mismatches, contributing to the lower accuracy. However, we find that that there are only 11,373 observations with missing market capitalisation out of the whole Capital IQ dataset (730,987 observations), therefore, this represents only a small part of the overall stock universe. Nevertheless, this effect could skew our trading strategy, discussed in the next chapter, towards picking the lesser known stocks, with data contribution issues in the consensus forecast version of the model.

Insights from Variable Importance

We use the mean absolute SHAP values, expressed in percentages to analyze which features are dominant in our ML models. While XGboost has a built-in function to summarize features that are used frequently to construct the boosted decision trees, SHAP values show the influence of each explanatory variable for each prediction and are commonly used to investigate the impact of predictor variables in ML models, therefore, thought to be more appropriate for our analysis (den Broeck et al., 2022). The variable importance tables containing 20 highest scoring predictor variables for our main and consensus forecast models are given in Figure 4.6 and Figure 4.7.

For P/B models, ROE is the dominant variable used in predictions, which is supported by the justified P/B multiple:

$$justified\, P/B = \frac{ROE - g}{r - g}$$

where g is the sustainable growth rate ($ROE \times Retention\, rate$) and r is the required rate of return on equity. For P/S, the most important variable is Sales/Equity ratio, followed by net profit margin, which also enters the equation for justified P/S multiple:

$$justified\, P/S = \frac{E_1/S_0(1 - b)(1 + g)}{r - g}$$

where E_0 is the net income, S_0 is the sales and b is the retention rate (Damodaran, 2007).

Hence, ML models appear to capture the main theoretical determinants of pricing multiples. R&D to Sales ratio ranks highly up in all versions of our ML models. This coincides with the findings of Geertsema and Lu (2023), and related literature describing that R&D spending can impact stock returns (Booth et al., 2006; Chan et al., 2015). Also, we can see that different versions of the model capture short-term growth through sales growth in the last twelve months, or the expected next twelve months in the case for the consensus forecast model.

In the Geertsema and Lu (2023), industry classification is ranked as the most important variable for the P/B version of the model. However, in our case, industry is only the 14th most important variable in the SHAP ranking. This could be partly due to the different classification systems between S&P Capital IQ and the Fama-French industry classification. Also, the dataset of Geertsema and Lu (2023) contains financial companies, which can make the industry distinction more important in the ML decision trees than in our dataset that is without financial companies.

A test of our consensus forecast dataset is to look if any of the new variables become highly ranked based on SHAP. Models for P/B and P/S have NTM margins of Net Income, EBITDA, and FCF, NTM growth for Revenue, Net Income, and EPS, and average broker recommendations among the 20 most important variables. In addition, P/S model with consensus forecast has NTM Net Income margin as the top 2 variable. The fact that ML models choose the variables from our consensus forecast dataset suggests that these variables carry important information for stock valuation and justifies the use of the extended dataset. However, we observe, that the top ranked variables continue to be based on historical accounting information.

Figure 4.6: SHAP for P/B and P/S Valuation

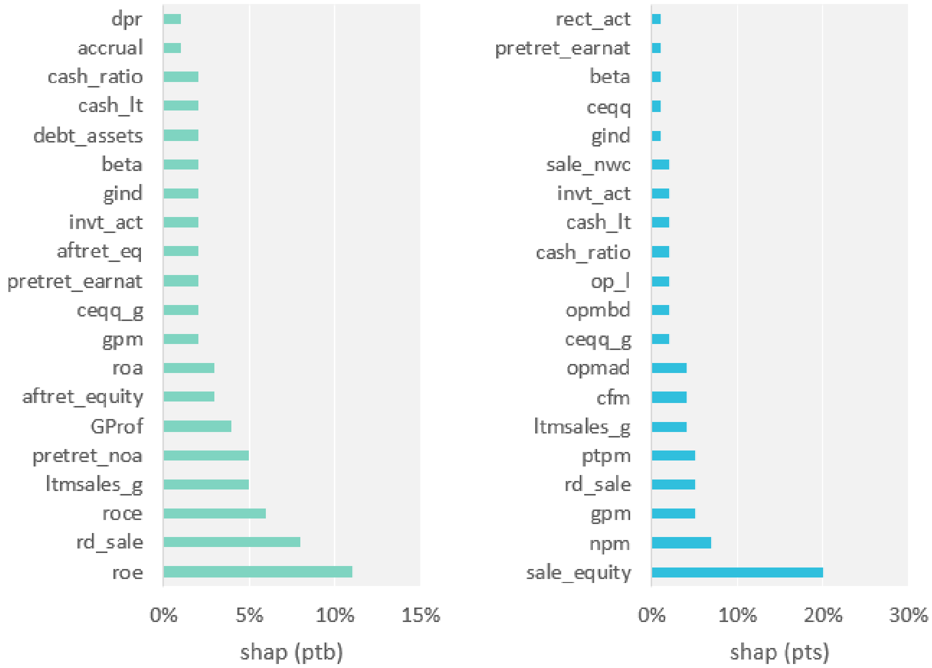
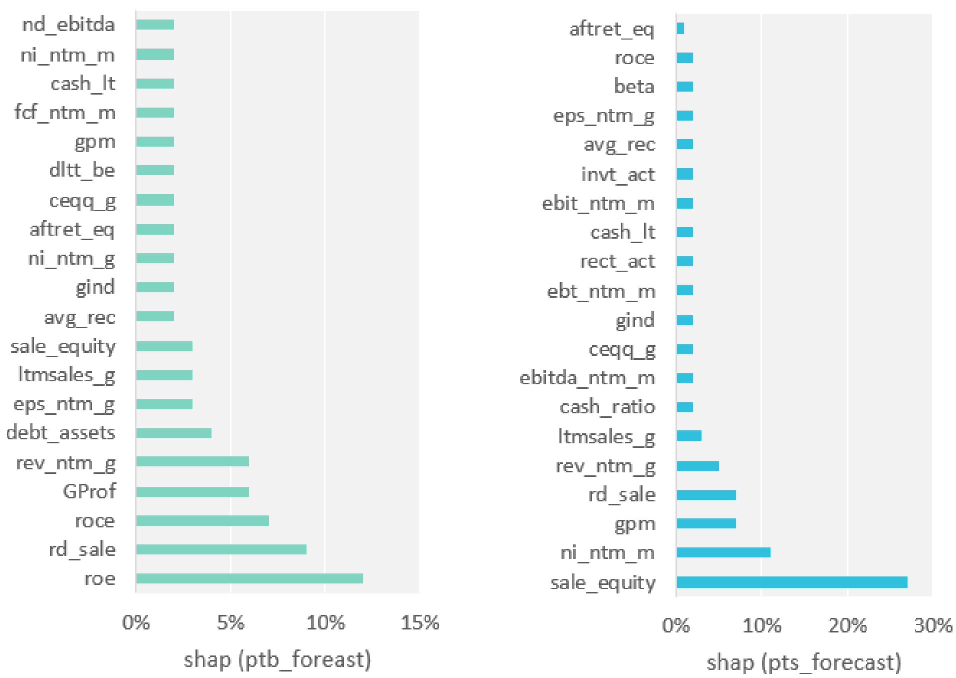


Figure 4.7: SHAP for P/B and P/S Valuation with Consensus Forecast



Mean absolute SHAP values for different predictors. SHAP Values are averaged across monthly cross sections and expressed in percentage such that SHAP values for all predictors sum up to 100%.

Variable importance can vary over time – we report SHAP values for top 5 important features by month in the Figure 4.8 and Figure 4.9. From these figures, it appears that some profitability-based measures such as ROE, Sales/Equity ratio have lost their importance over the recent years. For P/B valuation, the importance of ROE has been decreasing constantly from around 2010, while the importance of Sales/Equity has been decreasing since from around 2005 for P/S valuation. Interestingly, R&D/Sales ratio has been gaining importance, which can suggest that information in R&D spending is becoming more important for investors, and the fundamental value difference between high and low R&D spending companies could be increasing (Booth et al., 2006; Chan et al., 2015; Geertsema & Lu, 2023).

For the models with consensus forecast versions, growth in NTM revenue was among the top 5 important predictors for both P/B and P/S versions. Its importance was almost constant for the P/S version and decreased since the 2005 for the P/B version. NTM Net Income margin also has been declining in importance for the P/S forecast model. This suggests that analyst forecasts, already lagging behind historical based predictors based on SHAP, are not becoming more important through time. Thus analyst forecasts bring only marginal contribution towards relative stock pricing, even though ML prediction errors are lower.

Figure 4.8: SHAP (P/B and P/B with Consensus Forecast)

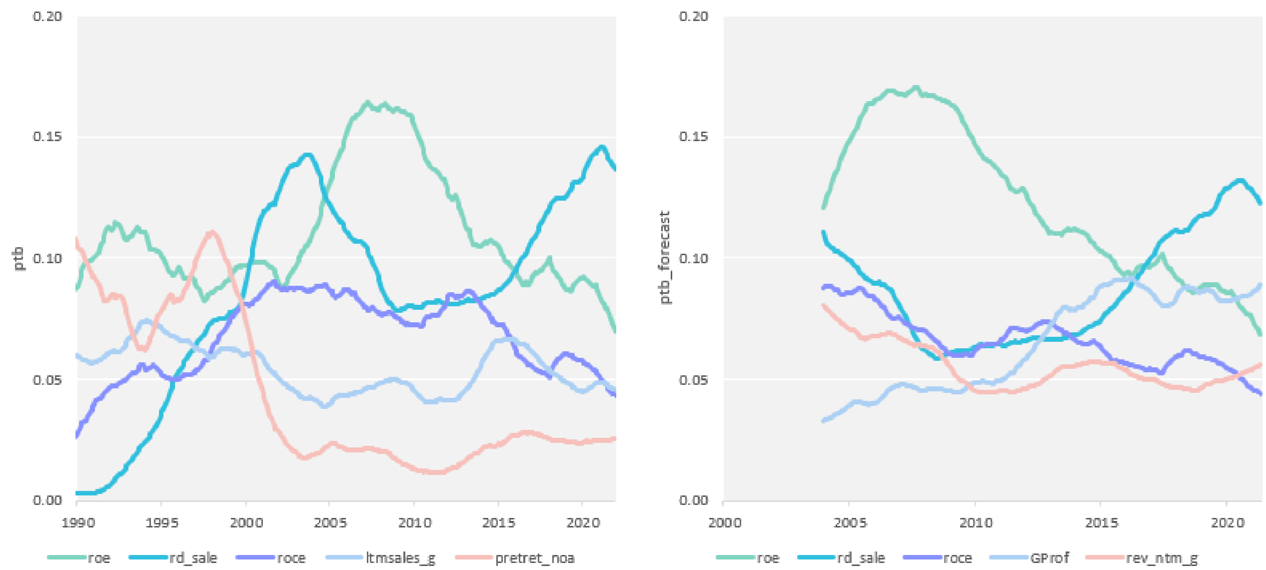
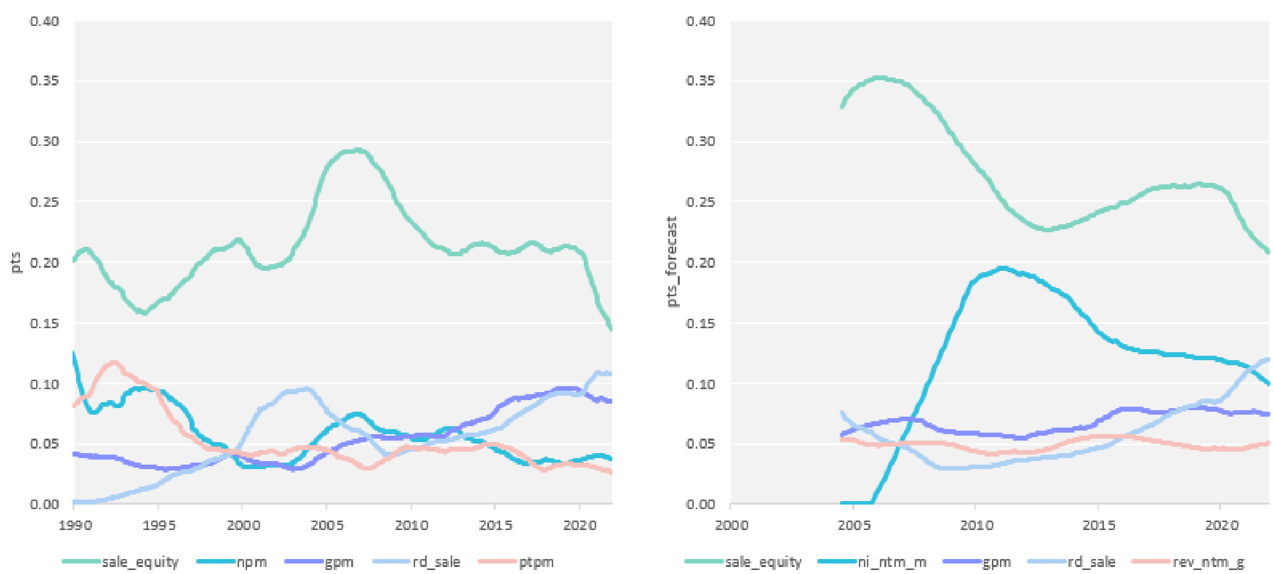


Figure 4.9: SHAP (P/S and P/S with Consensus Forecast)



Mean absolute SHAP values for different predictors. Monthly SHAP values are expressed in percentage such that SHAP values for all predictors sum up to 100% and a 4 year rolling mean is represented in the chart.

For variable definitions, see Table 3.2 and List of Variables.

Correlated Variables

Correlation matrices for our main and consensus forecast datasets showed that some variables, especially those related to profit margins are highly inter-correlated. In the variable importance section, we see, however, that many margin based predictors are ranked highly in the feature importance charts, meaning that the ML models use some of the inter-correlated predictors together, even as based on the correlation analysis, one variable should suffice.

Having highly correlated variables as explanatory variables causes multicollinearity when it comes to ordinary least squares regression. To test if these intercorrelations do not cause problems in the model, we run a P/S model version on our main dataset, where we remove some of the variables, which had the highest inter-correlation and were also correlated with the target multiple (cash flow margin, pre-tax margin, EBIT margin), leaving only net profit margin from this group. Our resulting RMSE is 0.53, which is the same as for the main P/S model. Therefore, we suggest that even as the variables are inter-correlated in our framework, which should be the same in Geertsema and Lu (2023) as many of the margin-based profitability variables are used there as well, this does not cause significant problems for our ML models.

Chapter 5

Strategy

Machine Learning-Valuation and Convergence Trade

This section investigates whether relative valuation with ML can be used to generate superior trading performance, in terms of abnormal return (alpha), risk factors, and practicality. If the ML method provides a more accurate stock valuation by considering non-linear relationships among underlying fundamental variables, one can engage in a convergence trade by going long under-valued stocks and going short over-valued stocks, since divergence from the fundamental value is expected to decrease over time (Liu & Timmermann, 2013; Bartram & Grinblatt, 2018).

Following Bartram and Grinblatt (2018), Hanauer et al. (2022), and Geertsema and Lu (2023), we sort stocks into quintiles by the difference (error) between actual market valuation (P/B and P/S ratio) and corresponding valuation suggested by ML in the previous chapter. If ML-based valuation overshoots actual market valuation, the stock is thought to be under-valued and vice versa.

We construct five long-only portfolios for each quintile and one long/short portfolio by going short Q1 (most over-valued quintile) and going long Q5 (most under-valued quintile). The long/short portfolio is expected to be less affected by the overall market direction (beta) and offer downside protection during a market downturn. The aforementioned portfolios are held for the next month and rebalanced at the month end, from January 2001 to December 2021, given the data availability for S&P Capital IQ consensus forecasts (ML valuation period from December 2000 to November 2021). Monthly total returns take dividends reinvestment into consideration, and when companies are delisted during the holding period, delisting returns acquired from CRSP are added. We use the risk-free rate from the data library website of Prof. Kenneth French, which is the 1-month US Treasury bill rate, and transaction costs are not considered.

We analyze both value-weighted (VW) and equal-weighted (EW) portfolios. VW portfolios are thought to be affected more by the movements of large-cap companies, while EW portfolios offer more diversification benefits by giving more weight to small-cap companies. EW portfolios are usually associated with higher costs from high turnover and transaction costs, but (Novy-Marx & Velikov, 2019), Qin and Singal (2022) point out trading costs of EW portfolios are limited, keeping EW portfolios superior to VW portfolios. Ilmanen (2011, p. 257–258) summarizes that VW has a tendency to overweight overpriced stocks, and EW has a contrarian bias, since the latter sells stocks whose prices have risen and buys those whose prices have declined in rebalance to keep the weight constant. Therefore, It is important to be aware of the tendencies caused by the weighting schemes.

Our stock universe used for ML valuation in the previous chapter includes approximately 2000 to 4000 stocks (also see Table 3.1), however, nano and micro-cap companies are associated with low liquidity, high volatility and high trading costs (Rabener, 2018). To control the effect of those companies and ensure the practicality of trades, smallest 20% of stocks in the investable universe are removed every month, since market capitalization grows over time and the definition of nano and micro-caps by company size in absolute term can change accordingly.

One caveat for our long/short strategy based on ML-valuation is that it might still be exposed to several risk factors, such as systematic risk (beta), size and value (Fama &

French, 1995). For example, as Figure 4.4 and Figure 4.5 suggest, small-sized companies tend to have higher valuation errors. Unless over-valued and under-valued companies are distributed evenly across company size, the implication from EW and VW portfolios might not be the same. Due to characteristic differences between Q1 (over-valued stocks) and Q5 (under-valued stocks), the resulting long/short portfolio can be still bearing certain risk factors, and this can limit the profitability and diversification benefit of the convergence trades.

Another caveat is the frequency of rebalancing. Our one-month rebalance frequency follows Bartram and Grinblatt (2018), Hanauer et al. (2022), and Geertsema and Lu (2023), but more frequent rebalancing might capture more convergence trade opportunities, while it is a trade-off with high transaction costs. It should be noted that our monthly financial ratio data from WRDS is lagged by two months to avoid a look-ahead bias, and there could be some inefficiency between the timing that the data becomes publicly available and the timing to conduct ML valuation and construct portfolios for some companies.

Table 5.1 and Table 5.2 report return excess of the risk-free rate (Ex Return), standard deviation (Std Dev) and sharpe ratio (SR) for each quintile and spread return from long/short portfolios (VW and EW). Table 5.3 and Table 5.4 show the result of the Fama-French 6 factors time-series regressions on the spread returns of the long/short portfolios (See Fama and French (2018), and note in Table 5.3). Quintile Portfolio Returns until 2019 and Spread Returns Time-series Regression until 2019 in Appendix report the result until 2019 to make the calculation period comparable with Geertsema and Lu (2023) and remove the effect of the COVID-19 shock.

Excess return and sharpe ratio are the highest for Q5 (most under-valued quintile) across valuation metrics from 0.63 to 0.72 for the VW portfolios, and 0.73 to 0.80 for the EW portfolios. There are significant differences in sharpe ratio between VW and EW long/short portfolios, suggesting the existence of heterogeneity in terms of size between Q1 and Q5. The long/short portfolios based on P/S record a higher sharpe ratio compared with portfolios based on P/B regardless of whether incorporating consensus forecasts or not. The VW long/short portfolio based on P/B and consensus forecast reports the lowest sharpe ratio at 0.04

Abnormal returns (Intercept/alpha) from the long-short portfolios ranged from 0.07% to 0.33% per month in the VW portfolios and 0.36% to 0.63% per month in the EW portfolios, while not statistically significant at 1% level for the VW portfolios. The sign of coefficients for the systematic risk (MKTRF/beta) is negative for the VW portfolios, especially 0.1% level for the P/B portfolio, but not statistically significant for the EW portfolios. The sign of regression coefficients for value factor (HML) are positive, and negative for momentum factor (UMD) with significance level at 0.1% across all the portfolios, suggesting the ML-based long/short portfolios have a tendency towards value and contrarian style. The EW portfolios are more exposed to the factors except for the systematic risk (MKTRF/beta).

Table 5.1: VW Portfolio Returns (Annualized, %)

Valuation	**Quintile**	**Ex Return**	**Std Dev**	**SR**
Price/Book	Q1	8.83	17.02	0.52
	Q2	6.83	14.67	0.47
	Q3	8.98	14.88	0.60
	Q4	9.99	15.60	0.64
	Q5	11.90	17.98	0.66
	Spread (Q5-Q1)	1.78	10.06	0.18
Price/Book (Forecast)	Q1	9.23	16.75	0.55
	Q2	8.12	15.19	0.53
	Q3	8.59	15.09	0.57
	Q4	8.63	15.30	0.56
	Q5	10.89	17.42	0.63
	Spread (Q5-Q1)	0.37	9.48	0.04
Price/Sales	Q1	8.15	16.59	0.49
	Q2	7.54	14.88	0.51
	Q3	9.51	14.88	0.64
	Q4	9.93	15.91	0.64
	Q5	12.59	18.04	0.70
	Spread (Q5-Q1)	3.15	9.90	0.32
Price/Sales (Forecast)	Q1	7.31	16.62	0.44
	Q2	8.63	15.04	0.57
	Q3	8.97	14.81	0.61
	Q4	10.28	15.71	0.65
	Q5	12.77	17.73	0.72
	Spread (Q5-Q1)	4.18	9.43	0.44

Q1 (Q5) is the most over-valued (under-valued) quintile measured by the difference between ML suggested valuation metric and actual valuation.

During the calculation period (January 2001 - December 2021), comparable excess return, standard deviation and sharpe ratio for S&P500 (VW) are 8.02%, 14.92% and 0.53, respectively.

Table 5.2: EW Portfolio Returns (Annualized, %)

Valuation	**Quintile**	**Ex Return**	**Std Dev**	**SR**
Price/Book	Q1	9.21	20.23	0.46
	Q2	11.00	17.84	0.62
	Q3	11.97	18.15	0.66
	Q4	15.23	19.59	0.78
	Q5	18.95	24.50	0.77
	Spread (Q5-Q1)	8.44	10.14	0.83
Price/Book (Forecast)	Q1	10.05	19.90	0.51
	Q2	11.38	17.92	0.63
	Q3	12.58	18.47	0.68
	Q4	14.73	19.78	0.74
	Q5	17.61	24.00	0.73
	Spread (Q5-Q1)	6.27	9.31	0.67
Price/Sales	Q1	8.36	20.08	0.42
	Q2	10.97	18.01	0.61
	Q3	12.77	18.07	0.71
	Q4	14.64	19.58	0.75
	Q5	19.65	24.59	0.80
	Spread (Q5-Q1)	9.99	10.55	0.95
Price/Sales (Forecast)	Q1	9.09	19.86	0.46
	Q2	10.99	17.91	0.61
	Q3	12.77	1834	0.70
	Q4	14.63	19.62	0.75
	Q5	18.90	24.42	0.77
	Spread (Q5-Q1)	8.52	10.02	0.85

Q1 (Q5) is the most over-valued (under-valued) quintile measured by the difference between ML suggested valuation metric and actual valuation.

During the calculation period (January 2001 - December 2021), comparable excess return, standard deviation and sharpe ratio for S&P500 (EW) are 10.50%, 17.54% and 0.59, respectively.

Table 5.3: Spread Returns Time-series Regression (VW)

	P/B	**P/B(Forecast)**	**P/S**	**P/S(Forecast)**
Intercept	0.22	0.07	0.29*	0.33*
MKTRF	-0.13***	-0.09*	-0.12**	-0.06
SMB	0.04	0.02	0.10	0.05
HML	0.35***	0.27***	0.34***	0.27***
RMW	0.08	0.08	0.10	0.12
CMA	-0.01	0.01	0.07	0.08
UMD	-0.29***	-0.21***	-0.31***	-0.21***
Adj. R^2	0.36	0.24	0.45	0.26
# Obs.	252	252	252	252

Table 5.4: Spread Returns Time-series Regression (EW)

	P/B	**P/B(Forecast)**	**P/S**	**P/S(Forecast)**
Intercept	0.53***	0.36***	0.63***	0.51***
MKTRF	-0.02	0.00	-0.01	0.02
SMB	0.26***	0.23***	0.29***	0.26***
HML	0.16***	0.14**	0.18***	0.18***
RMW	0.29***	0.26***	0.33***	0.30***
CMA	0.28***	0.23***	0.27***	0.19**
UMD	-0.42***	-0.37***	-0.44***	-0.38***
Adj. R^2	0.69	0.65	0.71	0.65
# Obs.	252	252	252	252

Fama French 5 Factors plus momentum are obtained from WRDS. Factor definitions are as follows: market return in excess of the risk-free rate (MKTRF), small minus big in capitalization (SMB), high minus low in book-to-price ratio (HML), robust minus weak in operating profitability (RMW), conservative minus aggressive in investment (CMA), up minus down in 12 months return (UMD).

*, **, *** stand for $p < 0.001$, $p < 0.01$ and $p < 0.05$, respectively.

Size $\times$ Machine Learning-Valuation Strategy

In this section, we analyze the efficacy of ML-valuation based convergence trade by size. In the previous section, VW long/short portfolios show lower sharpe ratio, and EW long/short portfolios show higher sharpe ratio while more exposed to risk factors, including size. We construct long/short portfolios by grouping stocks into three groups by market capitalization (size), and go long 1/3 of under-valued stocks and go short 1/3 of over-valued stocks based on ML-valuation within each group (S1: small-size, S2: mid-size and S3: large-size). As a result, 1/9 of total stocks go long and 1/9 of total stocks go short within each group.

Table 5.5 and Table 5.6 report return excess of the risk-free rate, standard deviation (Std Dev) and sharpe ratio of long/short portfolios for each group (VW and EW). Table 5.7 and Table 5.8 show the result of the Fama-French 6 factors time-series regressions on the spread returns of the long/short portfolios and their portfolio characteristics, including average drawdown (Ave. DD), maximum drawdown (Max. DD) and turnover for S1 (See Table 5.7 for details).

Excess return and sharpe ratio are the highest for S1 from 9.17% to 9.99% across ML-valuation metrics for the V portfolios, and highest for P/S. Sharpe ratio decreases as size grows from S1 to S3, and since heterogeneity of size decreases by market capitalization based grouping, the result is similar between VW and EW portfolios with the latter having a slightly higher sharpe ratio. P/S based long/short portfolios record sharpe ratio as high as 1.09 and 1.13 for VW and EW portfolios, respectively. Adding consensus forecast to

ML-valuation input (P/B and P/S with consensus forecast) results in lower sharpe ratio compared with ML-valuation without consensus forecast.

Abnormal returns (Intercept/alpha) from the long-short portfolios ranges from 0.48% to %0.69 per month in the VW portfolios and 0.53% to 0.75% per month in the EW portfolios, and all of them are statistically significant at 0.1% level. The sign of coefficients for the systematic risk (MKTRF/beta) is positive but small compared with other factors and the result from the previous section. The coefficients for value factor (HML) become considerably small relative to the previous section and they are not statistically significance at 0.5% level, suggesting these portfolios are less value tilted. On the other hand, the coefficient for momentum (UMD) remains negative and statistically significant at 0.1% level for both VW and EW portfolios. At the same time, the P/S based long/short portfolios (with and without consensus forecast) have a style tilt towards operating profitability (RMW) at 0.1% level for VW portfolios and 1% level for EW portfolios.

Average drawdown is from -1.50% to -2.77% for VW portfolios and -1.28% to -2.21% for EW portfolios. High portfolio turnover leads to high transaction costs, eroding the profitability of the long/short portfolios. Monthly average turnover measured by number of stocks is stable around 30% across all portfolios. The P/S based portfolios (without consensus forecast) result in superior performance relative to all other portfolios in terms of sharpe ratio and risk characteristics regardless of weighting method.

Figure 5.1, Figure 5.2, Figure 5.3 and Figure 5.4 show the hypothetical performance and risk characteristics of two selected long/short portfolios based on P/B and P/S in S1 (without consensus forecast, VW).

Table 5.5: Size × ML-Valuation VW Spread Returns
(Annualized, %)

Valuation	Size	Ex Return	Std Dev	SR
Price/Book	S1	9.17	8.96	1.02
	S2	4.67	7.68	0.61
	S3	0.38	7.88	0.05
Price/Book (Forecast)	S1	7.16	8.84	0.81
	S2	3.41	7.59	0.45
	S3	-1.19	7.33	-0.16
Price/Sales	S1	9.99	9.16	1.09
	S2	5.10	8.14	0.63
	S3	1.00	7.26	0.14
Price/Sales (Forecast)	S1	9.57	9.72	0.98
	S2	3.76	8.21	0.46
	S3	2.40	7.51	0.32

Table 5.6: Size × ML-Valuation EW Spread Returns
(Annualized, %)

Valuation	Size	Ex Return	Std Dev	SR
Price/Book	S1	9.49	9.08	1.04
	S2	5.07	7.76	0.65
	S3	2.47	7.19	0.34
Price/Book (Forecast)	S1	7.80	9.01	0.87
	S2	3.68	7.60	0.48
	S3	0.59	6.56	0.09
Price/Sales	S1	10.76	9.21	1.13
	S2	5.32	8.13	0.65
	S3	2.86	7.65	0.37
Price/Sales (Forecast)	S1	10.29	9.77	1.08
	S2	4.24	8.27	0.51
	S3	1.57	7.02	0.22

S1 (S3) is the group of smallest (largest) 1/3 of total stocks by market capitalization. Within each group, 1/3 of under-valued (over-valued) stocks based on ML valuation go long (short), and spread returns are calculated.

Table 5.7: Small Cap × ML-Valuation Spread Returns Time-series Regression (VW)

	P/B	P/B(Forecast)	P/S	P/S(Forecast)
Intercept	0.63***	0.48**	0.69***	0.63***
MKTRF	0.08*	0.08*	0.08*	0.09**
SMB	0.12*	0.14**	0.12*	0.14**
HML	0.02	0.02	0.03	0.04
RMW	0.15*	0.08	0.19***	0.24***
CMA	0.17*	0.12	0.10	-0.00
UMD	-0.35***	-0.31***	-0.37***	-0.36***
Adj. R^2	0.55	0.49	0.45	0.50
# Obs.	252	252	252	252
Ave. DD	-1.74%	-2.77%	-1.50%	-2.03%
Max. DD	-15.02%	-17.17%	-13.93%	-15.36%
Turnover	30.95%	32.98%	29.22%	31.14%

Table 5.8: Small Cap × ML-Valuation Spread Returns Time-series Regression (EW)

	P/B	P/B(Forecast)	P/S	P/S(Forecast)
Intercept	0.64***	0.53***	0.75***	0.70***
MKTRF	0.08**	0.08**	0.08**	0.09**
SMB	0.15**	0.16***	0.14**	0.18**
HML	-0.02	-0.03	-0.00	0.00
RMW	0.14*	0.07	0.18**	0.19**
CMA	0.15*	0.13	0.10	0.04
UMD	-0.37***	-0.34***	-0.38***	-0.39***
Adj. R^2	0.58	0.55	0.60	0.58
# Obs.	252	252	252	252
Ave. DD	-1.63%	-2.21%	-1.28%	-1.70%
Max. DD	-14.14%	-14.35%	-11.53%	-13.45%
Turnover	30.95%	32.98%	29.22%	31.14%

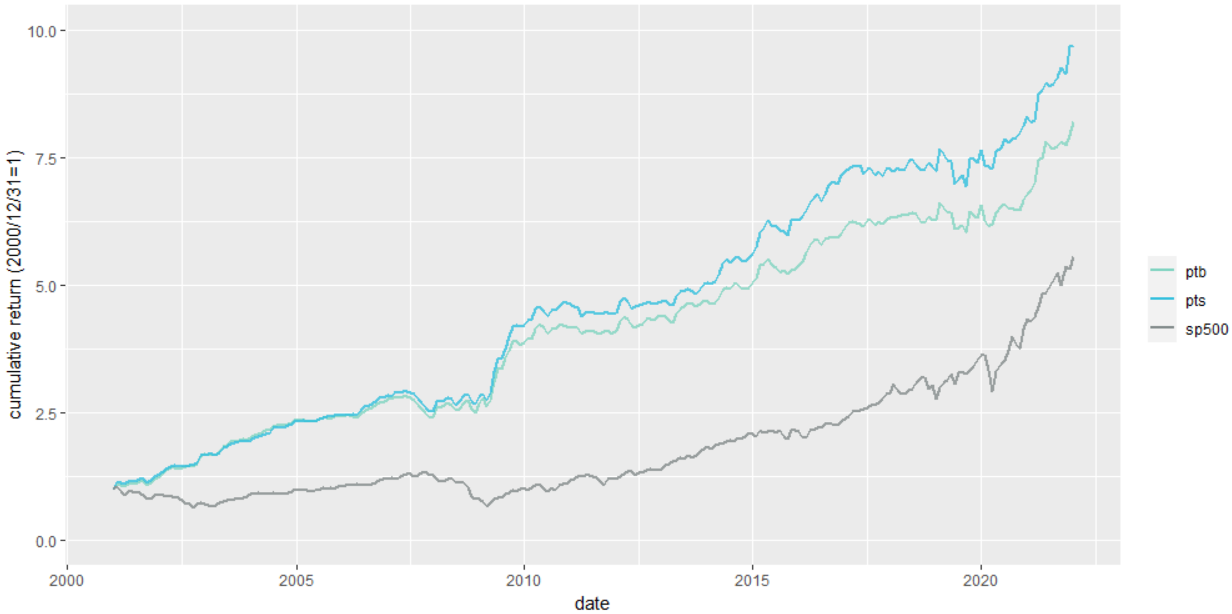
Fama French 5 Factors plus momentum are obtained from WRDS. Factor definitions are as follows: market return in excess of the risk-free rate (MKTRF), small minus big in market capitalization (SMB), high minus low in book-to-price ratio (HML), robust minus weak in operating profitability (RMW), conservative minus aggressive in investment (CMA), up minus down in 12 months return (UMD).

Draw Down (DD) is calculated as the percent difference between the highest cumulative return and the current cumulative return.

Turnover is defined as the number of different stocks from the previous month relative to the total number of stocks in the previous month (monthly average).

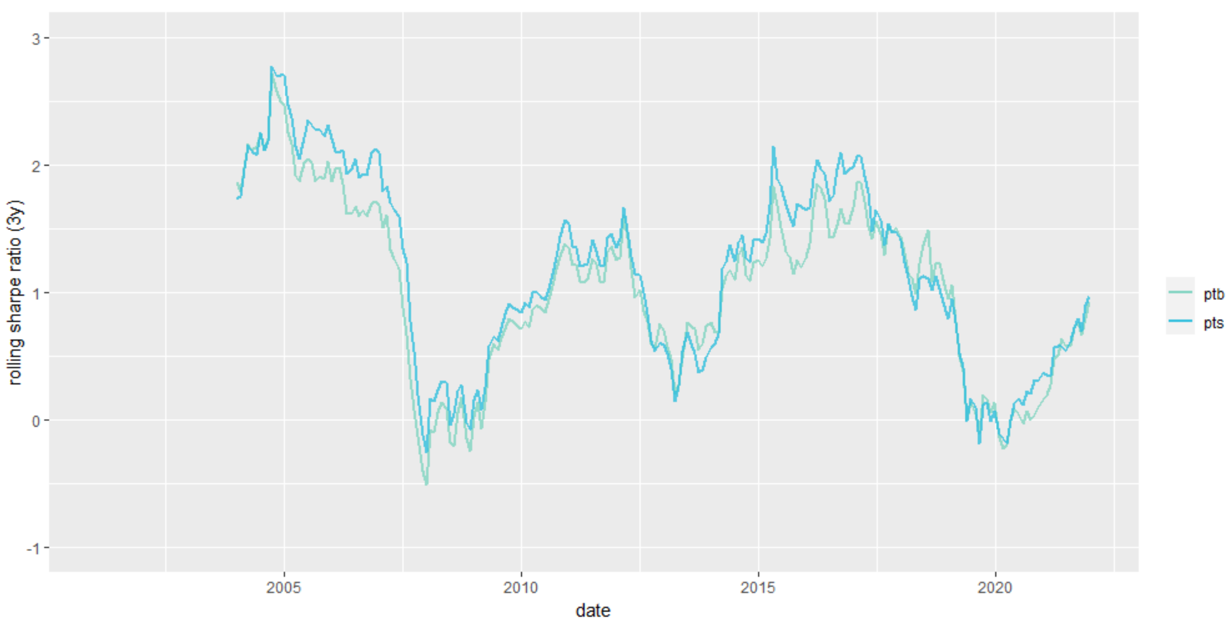
*, **, *** stand for $p < 0.001$, $p < 0.01$ and $p < 0.05$, respectively.

Figure 5.1: Cumulative Returns for Selected Portfolios



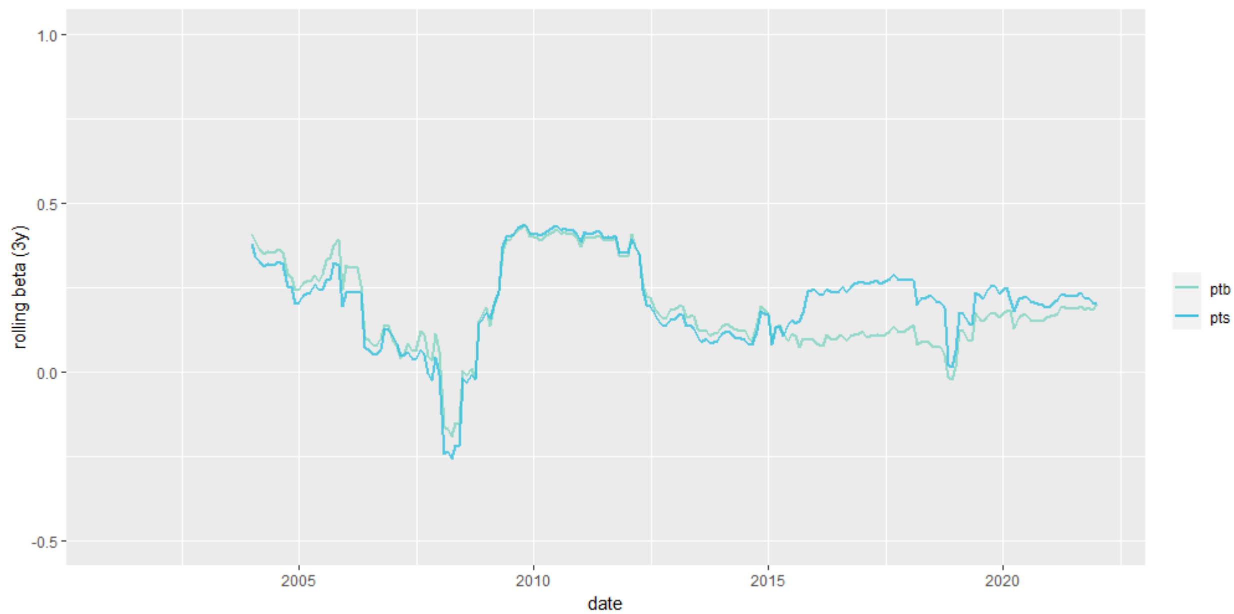
See Size × Machine Learning-Valuation Strategy for details of the strategies.
Ptb stands for P/B and pts stands for P/S.

Figure 5.2: Rolling Sharpe Ratio for Selected Portfolios



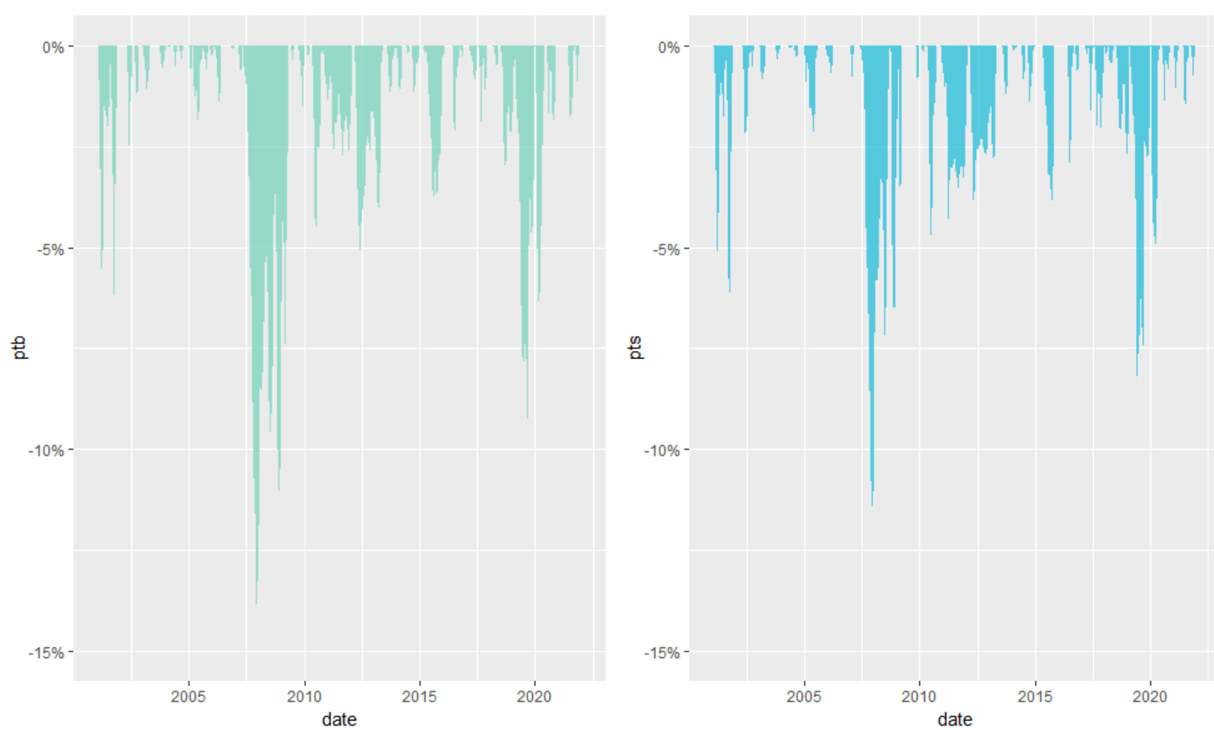
Rolling sharpe ratio is calculated as 3 years (36 months) moving window sharpe ratio (annualized).

Figure 5.3: Rolling Beta for Selected Portfolios



Rolling beta is calculated as 3 years (36 months) moving window regression coefficient on S&P500 (VW).

Figure 5.4: Draw-down for Selected Portfolios



For drow-down calculation, see Table 5.7.

Discussion

Our ML-based convergence strategies result in comparable trading performance with Geertsema and Lu (2023) when P/B ratio is used and all-sized firms are included, while abnormal returns for VW portfolios are not statistically significant. Sharpe ratio for the long/short portfolios drops for VW compared with EW portfolios, and as we include the period after the COVID-19 shock in 2020, the performance deteriorates as well. It should be noted that sharpe ratio based on P/B Geertsema and Lu (2023) report is higher than our result, possibly reflecting the period they include between 1984 and 2000, as the efficacy of strategies tends to decay over time, and advanced quantitative techniques represented by ML have become popular and replicated by many invetors (Falck et al., 2022).

The long/short portfolios for small companies result in 0.48% to 0.69% abnormal returns per month across valuation inputs on VW basis at 0.1% significant level after controlling the effect of nano and micro-cap companies. ML could provide a more accurate valuation for small companies that are not covered by analysts and scrutinized less often by investors, similar with the findings of Avramov et al. (2021). However, our portfolios are less exposed to well-known factors reported by Avramov et al. (2021), and portfolio turnover, which can significantly erode the profitability of trading strategies by incurring high transaction costs, is limited with smaller downside risk measured by drawdown. That said, further analysis on factors, such as liquidity and transaction costs, would be needed to confirm the practical applicability of our ML-based strategy, since our portfolios might still include additional risks that are unique in stocks of small companies.

The use of P/S, instead of P/B as valuation metrics leads to superior trading performance overall. While P/B and P/S include similar information, Choi et al. (2021) report that growing R&D and SG&A expenditures can add noise to book value, whereas P/S does not suffer this problem. In addition to P/B, Geertsema and Lu (2023) use enterprise value-to-assets and enterprise value-to-sales. There is a limitation in valuation metrics, such as price-to-earnings and EV/EBITDA, since they are not meaningful when earnings are negative, but the efficacy of other valuation metrics in addition to P/B and P/S should be further investigated.

On the other hand, including consensus forecast results in lower sharpe ratio compared with input only based on reported numbers in most portfolios. It could be partly attributed to model overfitting for large companies, since most consensus forecasts are only available for large companies, and by incorporating consensus forecast information, our ML algorithms might adjust parameters to fit better for large companies when making out-of-sample estimates. Another possible reason is that consensus forecasts help to reflect valuation on over-extrapolation of the past growth and optimistic assumptions (Ilmanen, 2011, p. 504).

The overall performance of our strategy could suggest that ML-based valuation successfully avoids distressed firms (value trap). At the same time, there could be a correlation between beta and valuation errors (Figure 4.4 and Figure 4.5), and in that case, our long/short portfolios might not be market neutral. Further sophistication in portfolio construction might contribute to improving trading performance.

Chapter 6

Conclusion

We analyze whether ML methods can contribute to improving the accuracy of valuation on a relative basis following Geertsema and Lu (2023), and investigate the efficacy of consensus forecast and its application to the convergence trade. We find that a tree-based ML model, employing XGboost, produces an overall accurate valuation from fundamental variables, and consensus forecasts contribute to improving the valuation accuracy. We also show that predictive errors vary through time and follow the general volatility patterns in the US stock market. Furthermore, predictive errors are not evenly distributed across fundamental variables, as high beta, low ROE, low market capitalization companies, as well those with valuation multiples that are far away from the median, tend to have higher errors. Hence, there is a pattern of a company that is expected to have a high ML prediction error, suggesting that predictions can be correlated with market factors.

While we show that ML-based relative valuation is effective, using the tree-based algorithms including XGboost, other ML algorithms, such as neural networks, might produce a better result. Analysis of textual structure of news and market sentiment that are not fully captured by our models might shed new light on the drivers of valuation.

It should be noted that while the ML models with consensus forecasts contribute to lowering prediction error of valuation metrics (P/B and P/S), trading performance is higher for basic models without consensus forecasts. This could be because models based on consensus forecasts overfit the market pricing, while models with only historical data are more general and thus capture the intrinsic value of the companies better. At the same time, it has been found that grouping companies by market capitalization is important for the ML-based convergence trade. For large-cap companies, fundamental variables and valuation metrics are more frequently scrutinized, leading to fewer opportunities to make an economically significant profit. On the other hand, ML-based valuation offers more benefits for small companies, leading to higher trading performance.

It is true that investing in small-caps is a trade-off between higher transaction costs, lower liquidity and other idiosyncratic risks, but our small-cap $\times$ ML valuation portfolio would offer considerable diversification benefits, low turnover with less exposure to the existing factors.

Appendix A

List of Variables

Category	Variable	Name	Source
Dependent Variables			
Valuation	ptb	Price/Book	WRDS
	pts	Price/Sales	
Independent Variables			
Capitalization	capital_ratio	Capitalization Ratio	WRDS
	equity_invcap	Common Equity/Invested Capital	
	debt_invcap	Long-term Debt/Invested Capital	
	totdebt_invcap	Total Debt/Invested Capital	
Efficiency	at_turn	Asset Turnover	WRDS
	inv_turn	Inventory Turnover	
	op_l	Operating Leverage	
	pay_turn	Payables Turnover	
	rect_turn	Receivables Turnover	
	sale_equity	Sales/Stockholders Equity	
	sale_invcap	Sales/Invested Capital	
	sale_nwc	Sales/Working Capital	
Financial Soundness	cash_lt	Cash Balance/Total Liabilities	WRDS
	cfm	Cash Flow Margin	
	cash_debt	Cash Flow/Total Debt	
	curr_debt	Current Liabilities/Total Liabilities	
	fcf_ocf	Free Cash Flow/Operating Cash Flow	
	int_debt	Interest/Average Long-term Debt	
	int_totdebt	Interest/Average Total Debt	
	invt_act	Inventory/Current Assets	
	dltt_be	Long-term Debt/Book Equity	
	lt_debt	Long-term Debt/Total Liabilities	
	ocf_lct	Operating CF/Current Liabilities	
	profit_lct	Profit Before Depreciation/Current Liabilities	
	rect_act	Receivables/Current Assets	
	short_debt	Short-Term Debt/Total Debt	
	debt_ebitda	Total Debt/EBITDA	
	lt_ppent	Total Liabilities/Total Tangible Assets	
Liquidity	cash_conversion	Cash Conversion Cycle (Days)	WRDS
	cash_ratio	Cash Ratio	
	curr_ratio	Current Ratio	
	quick_ratio	Quick Ratio (Acid Test)	

List of Variables - continued from previous page

Category	Variable	Name	Source
Market	beta	Beta (60 Months)	CRSP
Other Ratios	accrual	Accruals/Average Assets	WRDS
	adv_sales	Advertising Expenses/Sales	
	staff_sales	Labor Expenses/Sales	
	rd_sales	Research and Development/Sales	
Profitability	aftret_eq	After-tax Return on Average Common Equity	WRDS
	aftret_invcapx	After-tax Return on Invested Capital	
	aftret_equity	After-tax Return on Total Stockholders' Equity	
	efftax	Effective Tax Rate	
	gpm	Gross Profit Margin	
	GProf	Gross Profit/Total Assets	
	npm	Net Profit Margin	
	opmad	Operating Profit Margin After Depreciation	
	opmbd	Operating Profit Margin Before Depreciation	
	ptpm	Pre-tax Profit Margin	
	pretret_noa	Pre-tax Return on Net Operating Assets	
	pretret_earnat	Pre-tax Return on Total Earning Assets	
	roa	Return on Assets	
	roce	Return on Capital Employed	
	roe	Return on Equity	
Solvency	intcov_ratio	After-tax Interest Coverage	WRDS
	intcov	Interest Coverage Ratio	
	debt_capital	Total Debt/Capital	
	de_ratio	Total Debt/Equity	
	debt_assets	Total Debt/Total Assets	
	debt_at	Total Liabilities/Total Assets	
Growth	ceqq_g	Growth in Common Equity, last 12 months	WRDS
	ltmni_g	Net Income growth, last 12 months	
	ltmopprofit_g	EBIT growth, last 12 months	
	ltmopprofitbd_g	EBITDA growth, last 12 months	
	ltmsales_g	Sales growth, last 12 months	

WRDS stands for Financial Ratios Firm Level by WRDS with CRSP, Compustat and IBES subscriptions. All data are obtained through Wharton Research Data Services except for growth data calculated from S&P Capital IQ data.

Also see Table 3.2 for other variable definitions.

Appendix B

Quintile Portfolio Returns until 2019

Table A.1: VW Portfolio Returns until 2019 (Annualized, %)

Valuation	**Quintile**	**Ex Return**	**Std Dev**	**SR**
Price/Book	Q1	6.17	16.23	0.38
	Q2	5.43	14.05	0.39
	Q3	8.23	14.18	0.58
	Q4	8.82	14.70	0.60
	Q5	11.09	17.72	0.63
	Spread (Q5-Q1)	3.52	9.47	0.37
Price/Book (Forecast)	Q1	7.01	16.04	0.44
	Q2	6.49	14.42	0.45
	Q3	7.52	14.44	0.52
	Q4	7.77	14.31	0.54
	Q5	9.68	17.11	0.57
	Spread (Q5-Q1)	1.27	8.92	0.14
Price/Sales	Q1	6.28	15.93	0.39
	Q2	5.65	14.19	0.40
	Q3	7.85	14.15	0.55
	Q4	9.17	15.00	0.61
	Q5	11.71	17.77	0.66
	Spread (Q5-Q1)	4.02	9.59	0.42
Price/Sales (Forecast)	Q1	5.09	15.88	0.32
	Q2	6.95	14.40	0.48
	Q3	7.29	14.02	0.52
	Q4	9.46	15.07	0.63
	Q5	11.78	17.47	0.67
	Spread (Q5-Q1)	5.29	9.08	0.58

Q1 (Q5) is the most over-valued (under-valued) quintile measured by the difference between ML suggested valuation metric and actual valuation.

During the calculation period (January 2001 - December 2019), comparable excess return, standard deviation and sharpe ratio for S&P500 (value-weighted) are 6.43%, 14.33% and 0.45, respectively.

Table A.2: EW Portfolio Returns until 2019 (Annualized, %)

Valuation	**Quintile**	**Ex Return**	**Std Dev**	**SR**
Price/Book	Q1	7.31	19.32	0.39
	Q2	9.91	17.03	0.58
	Q3	11.05	17.16	0.64
	Q4	14.51	18.53	0.78
	Q5	17.84	23.61	0.76
	Spread (Q5-Q1)	8.83	10.00	0.88
Price/Book (Forecast)	Q1	8.57	19.06	0.45
	Q2	10.46	17.06	0.61
	Q3	11.54	17.40	0.66
	Q4	13.83	18.80	0.74
	Q5	16.52	23.12	0.71
	Spread (Q5-Q1)	6.55	9.09	0.72
Price/Sales	Q1	6.85	19.23	0.36
	Q2	9.81	17.09	0.57
	Q3	11.77	17.06	0.69
	Q4	13.99	18.65	0.75
	Q5	18.53	23.66	0.78
	Spread (Q5-Q1)	10.29	10.45	0.98
Price/Sales (Forecast)	Q1	7.36	18.83	0.39
	Q2	9.98	17.12	0.58
	Q3	12.17	17.46	0.70
	Q4	13.63	18.55	0.74
	Q5	17.81	23.52	0.76
	Spread (Q5-Q1)	9.06	9.89	0.92

During the calculation period (January 2001 - December 2019), comparable excess return, standard deviation and sharpe ratio for S&P500 (equal-weighted) are 9.35%, 16.79% and 0.56.

Appendix C

Spread Returns Time-series Regression until 2019

Table A.3: Spread Returns Time-series Regression until 2019 (VW)

	P/B	**P/B(Forecast)**	**P/S**	**P/S(Forecast)**
Intercept	0.25	0.06	0.28	0.34*
MKTRF	-0.06	-0.03	-0.06	-0.00
SMB	0.05	0.03	0.11	0.08
HML	0.26***	0.16*	0.30***	0.19*
RMW	0.20*	0.16	0.17	0.20*
CMA	0.05	0.08	0.09	0.12
UMD	-0.29***	-0.21***	-0.31***	-0.21***
Adj. R^2	0.32	0.18	0.41	0.22
# Obs.	228	228	228	228

Fama French 5 Factors plus momentum are obtained from WRDS. Factor definitions are as follows: market return in excess of the risk-free rate (MKTRF), small minus big in capitalization (SMB), high minus low in book-to-price ratio (HML), robust minus weak in operating profitability (RMW), conservative minus aggressive in investment (CMA), up minus down in 12 months return (UMD).

*, **, *** stand for $p < 0.001$, $p < 0.01$ and $p < 0.05$, respectively.

Table A.4: Spread Returns Time-series Regression until 2019 (EW)

	P/B	**P/B(Forecast)**	**P/S**	**P/S(Forecast)**
Intercept	0.54***	0.36**	0.63***	0.52***
MKTRF	-0.01	0.02	-0.01	0.06
SMB	0.29***	0.25***	0.31***	0.30***
HML	0.19***	0.12*	0.21***	0.17**
RMW	0.33***	0.29***	0.39***	0.38***
CMA	0.21**	0.21**	0.22**	0.16*
UMD	-0.43***	-0.38***	-0.46***	-0.40***
Adj. R^2	0.69	0.63	0.71	0.65
# Obs.	228	228	228	228

References

Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, *50*(1), 131–155. https://doi.org/https://doi.org/10.1111/j.1540-6261.1995.tb05169.x

Kim, M., & Ritter, J. R. (1999). Valuing IPOs. *Journal of Financial Economics*, *53*(3), 409–437. https://doi.org/https://doi.org/10.1016/S0304-405X(99)00027-6

Bhojraj, S., & Lee, C. M. C. (2002). Who is my peer? A valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, *40*(2), 407–439. https://doi.org/https://doi.org/10.1111/1475-679X.00054

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal of Accounting Research*, *40*(1), 135–172. https://doi.org/https://doi.org/10.1111/1475-679X.00042

O'Brien, J., Thomas. (2003). A simple and flexible DCF valuation formula. *Journal of Applied Finance*, *13*(2), 54–62. https://ssrn.com/abstract=253448

Rhodes–Kropf, M., & Viswanathan, S. (2004). Market valuation and merger waves. *Journal of Finance*, *59*(6), 2685–2718. https://doi.org/https://doi.org/10.1111/j.1540-6261.2004.00713.x

Rhodes–Kropf, M., Robinson, D. T., & Viswanathan, S. (2005). Valuation waves and merger activity: The empirical evidence. *Journal of Financial Economics*, *77*(3), 561–603. https://doi.org/https://doi.org/10.1016/j.jfineco.2004.06.015

Weiner, C. (2005). The impact of industry classification schemes on financial research. *Humboldt University of Berlin*. https://doi.org/https://dx.doi.org/10.2139/ssrn.871173 (accessed: 02.15.2023)

Zhang, L. (2005). The value premium. *Journal of Finance*, *60*(1), 67–103. https://doi.org/https://doi.org/10.1111/j.1540-6261.2005.00725.x

Booth, G. G., Junttila, J., Kallunki, J.-P., Rahiala, M., & Sahlström, P. (2006). How does the financial environment affect the stock market valuation of rd spending? *Journal of Financial Intermediation*, *15*(2), 197–214. https://doi.org/https://doi.org/10.1016/j.jfi.2005.03.003

Kulshrestha, R., & Nanda, S. (2006). Determinants of price-earning ratio. *Management Dynamics*, *6*(2), 13–25. https://doi.org/10.57198/2583-4932.1206

Damodaran, A. (2007). Valuation approaches and metrics: A survey of the theory and evidence. *Foundations and Trends in Finance*, *1*(8), 693–784. https://doi.org/http://dx.doi.org/10.1561/0500000013

Beaver, W., Cornell, B., Landsman, W. R., & Stubben, S. R. (2008). The impact of analysts' forecast errors and forecast revisions on stock prices. *Journal of Business Finance Accounting*, *35*(5/6), 709–740. https://doi.org/10.1111/j.1468-5957.2008.02079.x

Aswath Damodaran. (2009). Relative Valuation. https://pages.stern.nyu.edu/~adamodar/pdfiles/DSV2/Ch4.pdf (accessed: 05.12.2023)

Damodaran, A. (2009). Breach of trust: Valuing financial service firms in the post-crisis era. https://doi.org/http://dx.doi.org/10.2139/ssrn.1798578

Hastie, T. (2009). *The elements of statistical learning : Data mining, inference, and prediction* (2. ed). Springer.

Da, Z., & Schaumburg, E. (2011). Relative valuation and analyst target price forecasts. *Journal of Financial Markets*, *14*(1), 161–192. https://doi.org/https://doi.org/10.1016/j.finmar.2010.09.001

Ilmanen, A. (2011). *Expected returns: An investor's guide to harvesting market rewards*. John Wiley & Sons.

Damodaran, A. (2013). Valuing financial services firms. *Journal of Financial Perspectives*, *1*(1), 427–472. https://doi.org/https://doi.org/10.1016/j.jfineco.2012.05.011

Kuhn, M., Johnson, K., et al. (2013). *Applied predictive modeling* (Vol. 26). Springer.

Liu, J., & Timmermann, A. (2013). Optimal convergence trade strategies. *The Review of Financial Studies*, *26*(4), 1048–1086. http://www.jstor.org/stable/23355388

Cooper, I. A., & Lambertides, N. (2014). Is there a limit to the accuracy of equity valuation using multiples? https://doi.org/https://dx.doi.org/10.2139/ssrn.2291869

McKinsey. (2014). What effect has quantitative easing had on your share price? https://www.mckinsey.com/client_service/corporate_finance/latest_thinking/mckinsey_on_finance/~/media/5966C71286604E2DA0A2630B224E7F79.ashx (accessed: 05.04.2023)

Yin, Y., Peasnell, K., Lubberink, M., & Hunt, I., Herbert G. (2014). Determinants of analysts' target p/e multiples. *Journal of Investing*, *23*(3), 35–42. https://ez.hhs.se/login?url=https://www.proquest.com/trade-journals/determinants-analysts-target-p-e-multiples/docview/1561988031/se-2?accountid=39039

Chan, K., Chen, H.-K., Hong, L.-H., & Wang, Y. (2015). Stock market valuation of R&D expenditures—the role of corporate governance. *Pacific-Basin finance journal*, *31*, 78–93. https://doi.org/https://doi.org/10.1016/j.pacfin.2014.12.004

Lee, C. M., Ma, P., & Wang, C. C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, *116*(2), 410–431. https://doi.org/https://doi.org/10.1016/j.jfineco.2015.02.003

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. https://doi.org/https://dl.acm.org/doi/10.1145/2939672.2939785

CBOE. (2017). A Practitioner's Guide to Reading VIX. https://www.spglobal.com/spdji/en/education-a-practitioners-guide-to-reading-vix.pdf (accessed: 05.04.2023)

Bartram, S. M., & Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, *128*(1), 125–147. https://doi.org/https://dx.doi.org/10.2139/ssrn.2802478

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252. https://doi.org/https://doi.org/10.1016/j.jfineco.2018.02.012

Rabener, N. (2018). Factor Investing in Micro and Small Caps. https://blogs.cfainstitute.org/investor/2018/10/08/factor-investing-in-micro-and-small-caps/ (accessed: 04.11.2023)

Rabier, M. R. (2018). Value is in the eye of the beholder: The relative valuation roles of earnings and book value in merger pricing. *Accounting Review*, *93*(1), 335–362. https://doi.org/10.2308/accr-51785

Ding, K., Peng, X., & Wang, Y. (2019). A machine learning-based peer selection method with financial ratios. *Accounting Horizons*, *33*(3), 75–87. https://doi.org/https://doi.org/10.2308/acch-52454

Novy-Marx, R., & Velikov, M. (2019). Comparing cost-mitigation techniques. *Financial Analysts Journal*, *75*(1), 85–102. https://ssrn.com/abstract=3253359

Agudze, K., & Ibhagui, O. (2020). Do fundamentals drive relative valuation? Evidence from global stock market indices. *Journal of Financial Management, Markets and Institutions*, *08*(02), 1–42. https://doi.org/https://doi.org/10.1142/S2282717X20500024

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/https://doi.org/10.1093/rfs/hhaa009

Avramov, D., Cheng, S., & Metzker, L. (2021). Machine learning versus economic restrictions: Evidence from stock return predictability. *Management Science*. https://doi.org/http://dx.doi.org/10.2139/ssrn.3450322

Choi, K.-S., So, E. C., & Wang, C. C. Y. (2021). Going by the book: Valuation ratios and stock returns. http://dx.doi.org/10.2139/ssrn.3854022

Jason Brownlee. (2021). A Gentle Introduction to XGBoost for Applied Machine Learning. https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/ (accessed: 05.12.2023)

Johnston, J., Guidry, R. P., & Trimble, M. (2021). Temporal changes in the value relevance of analysts' forecasts. *Journal of Corporate Accounting Finance (Wiley)*, *32*(2), 7–21. https://doi.org/10.1002/jcaf.22478

Tayebi, Z., Önel, G., & Moss, C. B. (2021). Use of panel time-series data with cross-section dependence in evaluating farmland valuation: A cautionary note. *Applied Economics Letters*, *28*(6), 487–492. https://doi.org/10.1080/13504851.2020.1761527

den Broeck, G. V., Lykov, A., Schleich, M., & Suciu, D. (2022). On the tractability of shap explanations. *The Journal of artificial intelligence research*, *74*, 851–886. https://doi.org/https://doi.org/10.1613/jair.1.13283

dmlc. (2022). Introduction to Boosted Trees. https://xgboost.readthedocs.io/en/stable/tutorials/model.html (accessed: 05.12.2023)

Falck, A., Rej, A., & Thesmar, D. (2022). When do systematic strategies decay? *Quantitative Finance*, *22*(11), 1955–1969. https://doi.org/10.1080/14697688.2022.2098810

Hanauer, M. X., Kononova, M., & Rapp, M. S. (2022). Boosting agnostic fundamental analysis: Using machine learning to identify mispricing in european stock markets. *Finance Research Letters*, *49*(3), 329–376. https://doi.org/https://dx.doi.org/10.2139/ssrn.3977872

Qin, N., & Singal, V. (2022). Equal-weighting and value-weighting: Which one is better? *Review of Quantitative Finance and Accounting*, *58*, 743–768. https://doi.org/https://doi.org/10.1007/s11156-021-01008-w

Zou, M., Jiang, W.-G., Qin, Q.-H., Liu, Y.-C., & Li, M.-L. (2022). Optimized xgboost model with small dataset for predicting relative density of Ti-6AL-4V parts manufactured by selective laser melting. *Materials*, *15*(15). https://www.mdpi.com/1996-1944/15/15/5298

Blitz, D., Hoogteijling, T., Lohre, H., & Messow, P. (2023). How can machine learning advance quantitative asset management? *Journal of Portfolio Management*. https://doi.org/https://doi.org/10.3905/jpm.2023.1.460

Geertsema, P., & Lu, H. (2023). Relative valuation with machine learning. *Journal of Accounting Research*, *61*(1), 329–376. https://doi.org/https://doi.org/10.1111/1475-679X.12464