Boosting Agnostic Fundamental Analysis

Using Machine Learning to Identify Mispricing in European Stock Markets

Hanauer, Matthias X.; Kononova, Marina; Rapp, Marc Steffen

Document Version
Final published version

Published in:
Finance Research Letters

DOI:
10.1016/j.frl.2022.102856

Publication date:
2022

Citation for published version (APA):
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*, *48*, Article 102856. https://doi.org/10.1016/j.frl.2022.102856

Finance Research Letters 48 (2022) 102856

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Boosting agnostic fundamental analysis: Using machine learning to identify mispricing in European stock markets[☆]

Matthias X. Hanauer[a], Marina Kononova[b], Marc Steffen Rapp[b,1,*]

[a] Technische Universität München (TUM); Germany; Hanauer is also employed by Robeco Institutional Asset Management the Netherlands. The views expressed in this paper are not necessarily shared by Robeco Institutional Asset Management

[b] School of Business and Economics, Accounting and Finance Area, Philipps-Universität Marburg, Germany

ARTICLE INFO

JEL Classification:
C45
C53
G11
G14
G15

Keywords:
Fundamental analysis
Market efficiency
Stock return
Machine learning
Random forest
Gradient boosting
European markets

ABSTRACT

Interested in fundamental analysis and inspired by Bartram and Grinblatt (2018, 2021), we apply linear regression (LR) and tree-based machine learning (ML) methods to estimate monthly peer-implied fair values of European stocks from 21 accounting variables. Comparing LR and ML models, we document substantial heterogeneity in the importance of predictors as measured by SHAP values. Examining trading strategies based on deviations from fair values, we find ML-strategies earn substantially higher risk-adjusted returns ("alpha") than simple LR-counterparts (48–66 vs. 11–36 bp per month for value-weighted portfolios). Our findings document the importance of allowing for non-linearities and interactions in fundamental analysis.

1. Introduction

Does fundamental stock analysis work, and how should analysts derive fundamental values? While the theoretical literature has developed discounted cash flow models and other highly stylized fundamental valuation models, Bartram and Grinblatt (2018, 2021) (hereafter BG) recently suggested an agnostic approach to fundamental analysis. The authors "take the view of a statistician with little knowledge of finance" (BG, 2018, p. 125) and use linear regression analysis to proxy a firm's market value of equity as a linear function of 21 commonly reported and readily available accounting items.

BG document that deviations from their "peer-implied fair value" reliably predict future returns in the US (BG, 2018) and most regions around the world (BG, 2021). Interestingly, however, the strategy proposed by BG seems to not work in Europe, which is quite

☆ We gratefully acknowledge helpful comments and suggestions from editor Jonathan A. Batten, two anonymous referees, Hossein Asgharian, Tobias Kalsbach, and Andrew Reek. Any errors of analysis and interpretation are our own.

* Corresponding author at: Philipps-Universität Marburg, School of Business & Economics, Am Plan 1, 35032 Marburg, Germany.
E-mail address: msr@m-s-rapp.de (M.S. Rapp).

[1] Marc Steffen Rapp is affiliated with and a part of the Nordic Finance and the Good Society research network organized and hosted by the Center for Corporate Governance at Copenhagen Business School. Furthermore, he serves as an academic co-director at the Marburg center for Institutional Economics (MACIE).

https://doi.org/10.1016/j.frl.2022.102856
Received 8 December 2021; Received in revised form 28 February 2022; Accepted 3 April 2022
Available online 4 April 2022
1544-6123/

puzzling given that the literature finds ample evidence of inefficiencies in European stock markets (e.g., Walkshäusl, 2021; Abourachid et al., 2017). One might argue that this is because eventually, BG are not perfectly agnostic. Indeed, while BG are agnostic to predictors' relevance, they impose a linear relationship between market value and fundamentals. Even while conceptually sound (e.g., Ohlson, 1995), this assumption is discretionary.

A natural alternative would be to take the view of a data scientist (also with only very little finance knowledge) and allow the data to speak for itself. Our objective is to explore this data scientist view: Interested in fundamental analysis and inspired by BG (2018, 2021), we apply linear regression (LR) and machine learning (ML) methods to estimate monthly fair values for stocks from 17 European countries over the years 1993–2019 and study the return predictability of corresponding mispricing signals, i.e., the difference between the stock's model-based fair value and its actual market value. The LR approach closely follows BG (2018, 2021). Regarding ML, we apply least absolute shrinkage and selection operator (LASSO) to the 21 accounting variables and its full set of cross-products, and tree-based regression methods. Specifically, we deploy random forest and gradient boosting models, as they can deal with any form of non-linearity and interactions, handle noisy features we face, but at the same time do not require subtle tuning compared to more complex methods (see Athey and Imbens, 2019). Moreover, we consider an ML ensemble that combines random forest and gradient boosting.

Training the models with the accounting variables suggested by BG (2021), our main findings can be summarized as follows: ***First***, comparing LR and ML models, we document substantial heterogeneity in the predictors' importance measured by SHapley Additive exPlanations (SHAP) values. *Second*, examining trading strategies building on mispricing signals, we find that tree-based ML-based trading strategies earn significant risk-adjusted monthly value-weighted portfolio returns ("alpha") of 48 – 66 bp and substantially outperform their LR counterparts. These results are robust across different factor models and varying portfolio construction choices. *Third*, in cross-sectional Fama MacBeth (1973) regressions, only ML methods remain highly significant, while LR methods show little ability to predict returns. *Fourth*, while a naïve extension of the linear model for cross-products and imposing parameter parsimony by applying LASSO improves the predictive power compared to simple LR, such an approach does not result in a trading strategy that generates returns comparable to tree-based ML-based strategies. *Finally*, the performance of the three tree-based ML methods is quite similar, with a small advantage of the ML ensemble. Our findings document the importance of allowing for non-linearities and interactions in fundamental analysis, as well as substantial non-naïve market inefficiencies in European stock markets.

This study contributes to the literature in at least three ways. *First*, it adds to the discussion of whether fundamental valuation adds value, how valuation (value) strategies have performed over the last decades, and how they could be improved (e.g., Arnott et al., 2021; Blitz and Hanauer, 2021; Israel et al., 2021; Park, 2019). *Second*, it contributes to the rapidly expanding literature on machine learning methods in finance and accounting research (e.g., Bali et al., 2021; Bianchi et al., 2021; Erel et al., 2021; Ghosh et al., 2021; Bao et al., 2020; Gu et al., 2020). *Finally*, it adds to the literature on asset pricing in European markets (e.g., Drobetz et al., 2019; Walkshäusl, 2014), aiming to reduce the underrepresentation of the non-US studies in finance (Karolyi, 2016).

2. Data and methodology

2.1. Data

Our sample construction closely follows BG (2021). *First*, we start with all active and dead stocks in Refinitiv Datastream issued by firms incorporated in EU17 countries (EU15, Switzerland, and Norway) over the 1987–2019 period. *Second*, as it is common in the literature (e.g., Hanauer, 2020; Ince and Porter, 2006), we exclude non-common equity stocks, foreign listings, secondary listings, financial firms, and firms with non-positive total assets or missing industry identifiers. Also, we require non-missing values for all accounting variables used in BG (2021).[2] *Third*, in line with BG (2021), we draw returns in USD and apply dynamic screens for stock returns as recommended in the literature. *Finally*, we eliminate microcap stocks using a monthly threshold of a ten million USD market capitalization. Our final sample comprises 8,121 unique firms. Appendix A provides details on variable definitions, sourcing, and cleaning.

2.2. Linear fundamental analysis

BG (2018, 2021) propose a simple linear approach to fundamental valuation, in which a company's fundamental ("fair") equity value V is a linear function of contemporaneous and readily available accounting items $x_1, \ldots, x_N$, i.e.:

$$V_{i,t} = a_{1,t} \cdot x_{i,1,t} + \cdots + a_{N,t} \cdot x_{i,N,t}, \tag{1}$$

with i indicating the firm, t indicating time, and coefficients $a_{1,t}, \cdots, a_{N,t}$ determined by a cross-sectional OLS regression of market capitalization on contemporaneous $x_1, \ldots, x_N$. Specifically, the valuation model of BG (2021) relies on $N = 21$ accounting items from the firm's cash flow statement, income statement, and balance sheet. To replicate the analysis of BG (2021), we run monthly

[2] Our analysis consciously builds on the variable set of BG (2021), despite multicollinearity problems, noted by BG, and despite the fact that ML methods could manage much larger variable sets.

cross-sectional OLS regressions of market capitalization on winsorized versions of the 21 accounting items and deploy coefficients of these regressions to estimate a stock's fundamental equity value according to model (1).[3] We refer to this exercise as *LR*(*BG*).

2.3. Non-linear fundamental analysis

While BG (2018, 2021) are agnostic to the coefficients $a_{1,t}, \cdots, a_{N,t}$ in model (1), they still impose an important restriction: Their valuation model (1) specifies a time-specific *linear* relation between the peer-implied fair value V_t and accounting variables $(x_{1,t},\ldots,x_{N,t})$. We aim to relax this assumption and allow for time-specific, highly non-linear relationships between a firm's fundamental value V and its contemporaneous accounting items, i.e.:

$$V_{i,t} = \varphi_t\left(x_{i,1,t}, \ldots, x_{i,N,t}\right) \tag{2}$$

To specify the relation, we employ random forest (subsequently referred to as *RF*) and gradient boosting (*GBRT*) as two popular and powerful tree-based ML methods (e.g., Varian, 2014) and fit the φ_ts to minimize the degree of mispricing in the market.[4] Interested in improving prediction performance even further, we also average the predictions of both methods in the spirit of ensemble averaging (and refer to this as *Combi*).

However, allowing for non-linearity comes with costs. While a linear model asks for a single parameter for each predictor, in the case of non-linear ML models, the number of parameters to estimate rapidly expands even with a moderate number of predictors (e.g., Gu et al., 2020). As such, a cross-sectional modeling approach following BG will arguably suffer from a low observations-to-parameters ratio. Thus, we fit the non-linear valuation models φ_t leveraging information from pooled data of 48 cross-sections that were publicly known at time *t*. Moreover, we follow the standard approach of ML applications and use model (2) "out-of-sample", which means that we train and tune our valuation models φ_t on:

$$\begin{pmatrix} x_{i,1,t-1} & \cdots & x_{i,N,t-1} \\ & \ddots & \\ x_{i,1,t-48} & \cdots & x_{i,N,t-48} \end{pmatrix}. \tag{3}$$

To do so, we follow standard ML procedures and transform all accounting items by cross-sectionally ranking them in ascending order and mapping these ranks into the [−1,1] interval (e.g., Freyberger et al., 2020; Gu et al., 2020).[5] Moreover, we deflate firms' market capitalization by the total market value to alleviate the effect of changing market tastes and time-specific valuation norms. Internet Appendix IV describes our implementation choices on the hyperparameter tuning for the ML methods in detail. Similarly, Internet Appendix V provides detailed information on the sample splitting and validation design.

One might argue that allowing ML models to evaluate multiple cross-sections of data gives an advantage to ML models. Relatedly, the data transformation applied for ML methods differs from the approach used for the linear models above. To ensure that the results of ML methods are not due to the leveraging of more information or alternative data transformation, we also fit OLS on the pooled (transformed) data (3) and use the estimate for monthly linear out-of-sample predictions following (1) (*LR*(*pooled*)).

To further bridge the gap between linear regression and ML methods and to better understand the nature of potential non-linearities and interactions, we augment the set of 21 accounting items by their quadratic and interaction terms.[6] To estimate the fundamental value from the resulting 252 variables, we employ the least absolute shrinkage and selection operator (*LASSO*).[7] *LASSO* is inherently a linear model but can handle a large set of variables acting as variable selection model and thus, mitigating multicollinearity concern. For tuning and fitting *LASSO*, we apply the same sample splitting schema as for ML and LR(pooled) models.

2.4. Misprising signals and return predictability

Once we have predicted a firm's fundamental equity value for time *t* using a particular model *m*, we follow BG (2018, 2021) and calculate the corresponding mispricing signal as the percentage difference between the fundamental value $V_{i,t,m}$ and the observed market value $MV_{i,t}$ as:

$$MS_{i,t,m} = \frac{V_{i,t,m} - MV_{i,t}}{MV_{i,t}} \tag{3}$$

[3] For this, we winsorize accounting variables at the top and bottom 5% of the distribution of the normalized variable (% of total assets, to avoid size effects) with data available prior to the evaluation date *t*.

[4] The general idea behind a single tree is to split the data into subsamples consisting of observations with similar behavior. Thereby, the function is approximated with the average of the outcome variable within each split (Breiman et al., 1984). Yet, while a deep tree can fit perfectly, it is often a weak learner and prone to overfitting. One way to improve the performance is to utilize an ensemble.

[5] In untabulated results, we apply an alternative data transformation method. Instead of rank-transformation, we normalize accounting items by the min-max normalization method as a common practice in the ML literature (e.g., Al Shalabi et al., 2006). The min-max normalization maps a value of each item to the range [0, 1]. Our qualitative conclusions remain unchanged. We thank an anonymous reviewer for raising this point.

[6] We thank an anonymous reviewer for this suggestion.

[7] In untabulated results, we also fit LASSO without additional terms to address the concerns of overfitting and multicollinearity in the linear model. LASSO leads to a sparse version of the input set and thus, might improve results of the linear model. However, the results obtained from the "first-order" LASSO model are nearly identical to those obtained from LR(pooled) model.

Table 1
Summary Statistics. The table describes the averages of selected characteristics of the sample firms. In particular, Panel A reports the time-series average of the mean characteristics across all firms, the average cross-sectional correlation of the characteristic with the particular mispricing signal, and the average of the mean characteristics across quintiles of firms sorted by the mispricing signal from Q1 (low) to Q5 (high) employing breakpoints based on large firms. Large firms are defined as those which account for an aggregated share of 90% of the total market value in a particular month (Fama and French, 2017). Statistics are shown separately for the mispricing signals based on the LR model of Bartram & Grinblatt (2021), pooled LR, LASSO (augmented by squared and interaction terms), Random Forest, Gradient Boosting, and the combination of Random Forest and Gradient Boosting. Panel B reports average cross-sectional Spearman's rank-order correlation of mispricing signals in the lower diagonal and time-series correlation of spread (Q5-Q1) value-weighted portfolio (industry-adjusted) returns in the upper diagonal of the matrix for a particular method.

Panel A: Summary statistics for mispricing signals and selected firms' characteristics

			Mispricing Signal Quintiles				
	All	Correlation	Q1 (low)	Q2	Q3	Q4	Q5 (high)
LR(BG)							
Mispricing	4.29	1.000	−1.89	−0.40	−0.14	0.18	6.74
$BM_{(current)}$	0.80	0.398	0.56	0.37	0.43	0.52	1.04
Ln (Market Cap)	12.27	−0.473	13.78	14.43	14.23	13.86	11.64
Momentum	0.10	−0.146	0.17	0.20	0.17	0.14	0.05
LR (pooled)							
Mispricing	−11.02	1.000	−50.18	−0.14	0.52	1.41	16.83
$BM_{(current)}$		0.154	0.70	0.44	0.46	0.49	0.99
Ln (Market Cap)		0.329	11.38	15.06	14.46	13.94	12.29
Momentum		0.042	0.10	0.18	0.17	0.16	0.07
LASSO							
Mispricing	0.76	1.000	−24.13	0.02	0.71	1.66	36.96
$BM_{(current)}$		0.024	0.81	0.44	0.48	0.54	0.95
Ln(Market Cap)		−0.053	11.84	14.74	14.38	13.93	11.80
Momentum		−0.010	0.11	0.17	0.16	0.14	0.06
RF							
Mispricing	1.63	1.000	−0.54	−0.25	−0.05	0.19	2.89
$BM_{(current)}$		0.605	0.21	0.32	0.41	0.52	1.10
Ln(Market Cap)		−0.448	13.79	13.78	13.70	13.27	11.32
Momentum		−0.185	0.34	0.22	0.17	0.13	0.02
GBRT							
Mispricing	0.65	1.000	−0.50	−0.13	−0.03	0.08	1.58
$BM_{(current)}$		0.399	0.41	0.47	0.52	0.58	1.13
Ln(Market Cap)		−0.275	12.52	13.53	13.84	13.55	11.49
Momentum		−0.176	0.27	0.18	0.14	0.11	0.00
Combi							
Mispricing	1.14	1.000	−0.45	−0.18	−0.04	0.13	2.10
$BM_{(current)}$		0.572	0.23	0.35	0.43	0.52	1.11
Ln(Market Cap)		−0.420	13.39	13.74	13.78	13.35	11.35
Momentum		−0.198	0.34	0.21	0.17	0.13	0.01

Panel B: Correlation analysis of mispricing signals and of the corresponding spread returns

	LR(BG)	LR(pooled)	LASSO	RF	GBRT	Combi
LR(BG)	1.000	0.453	0.552	0.708	0.504	0.674
LR(Pooled)	−0.050	1.000	0.958	0.496	0.226	0.447
LASSO	−0.067	−0.289	1.000	0.596	0.336	0.553
RF	0.622	−0.117	0.033	1.000	0.743	0.942
GBRT	0.459	0.073	−0.006	0.671	1.000	0.815
Combi	0.605	−0.073	0.030	0.966	0.809	1.000

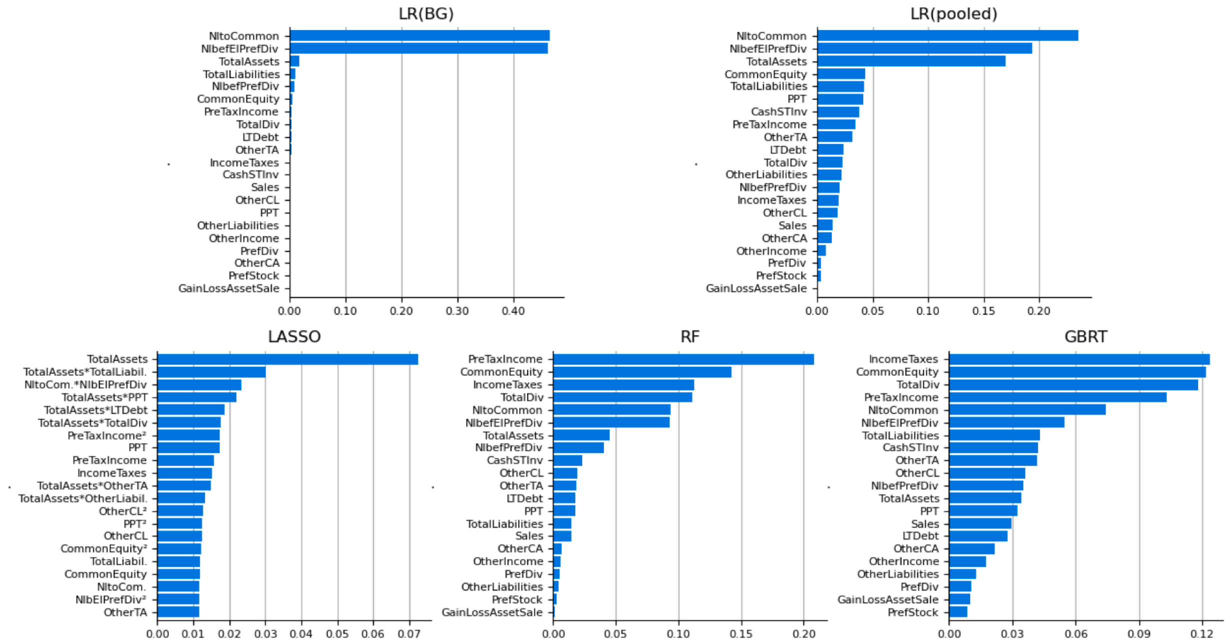

Fig. 1. SHAP values to illustrate predictors' importance. The figure plots variable importance for the LR model of Bartram & Grinblatt (2021), pooled LR, LASSO (augmented by squared and interaction terms), Random Forest, and Gradient Boosting. Variable importance for each model is defined based on SHAP (SHapley Additive exPlanations) values as a time-series average of mean predictor importance across all firms in a test sample. The SHAP values are normalized to sum to one, enabling the interpretation of relative importance in a particular method. For the LASSO method, we illustrate the first 21 most important variables from 252 variables used in the LASSO estimator.

We then evaluate the return predictability of these mispricing signals following BG (2021). Thereby, we use the LR results as a benchmark since BG have demonstrated their performance in regions all over the globe.

3. Results

We proceed in five steps to derive our empirical results. *First*, we determine fundamental stock values using the six valuation approaches, i.e., *LR*(*BG*), *LR*(*pooled*), *LASSO*, *RF, GBRT*, and *Combi*, and calculate corresponding mispricing signals. We do this for the last day of the month over the 01/1993 – 11/2019 period.[8]

Second, for *LR*(*BG*), *LR*(*pooled*), *LASSO*, *RF*, and *GBRT*, we compute SHAP values for all covariates to shed some light on the question which variables contribute most strongly to a model's fair value estimate.[9] Fig. 1 shows the average SHAP value of the 21 accounting items over the sample period. LR models seem to draw information mainly from two or three variables, i.e., *Net Income Available to Common* (Equity), *Net Income before Extraordinary Items/Preferred Dividends*, and *Total Assets*. While *LASSO* identifies *Total Assets* as most important, it indicates the relevance of both linear effects of used items and their interactions (e.g., *Total Assets* and *Total Liabilities*) or their quadratic terms (e.g., *Pre-tax Income*). In contrast, the ML models seem to draw information more uniformly along covariates, with *Pre-tax Income/Income Taxes, Common Equity*, and *Dividends* being among the most relevant ones. It is interesting to see that the data confirms the considerations proposed in the theoretical literature. For instance, Ohlson (1995) relates prices to earnings, equity book value, and dividends.

[8] With Refinitiv Datastream providing reliable data for European stock markets from 1987 onwards and training our ML models on a four-year period, our prediction period starts in 1993.

[9] Pioneered by Lundberg and Lee (2017), SHAP is a game-theoretic approach to estimate the extent to which a particular predictor contributes to pushing the model's output away from the unconditional expectation. Thereby, it is important to bear in mind, that while SHAP values make the prediction generation process somewhat transparent, they do not indicate any causality. For RF and GBRT, we also calculate impurity-based importance (Breiman et al., 1984) and obtain very similar results.

Table 2

Portfolio Sorts. The table reports the results from a quintile portfolio analysis based on mispricing signals obtained from different models. The breakpoints for sorting stocks in portfolios are calculated based on large firms. Large firms are defined as those which account for an aggregated share of 90% of the total market value in a particular month (Fama and French, 2017). Panel A shows the average monthly industry-adjusted returns of the value-weighted quintile portfolios as well as the spread portfolios (Q5-Q1), based on the LR model of Bartram & Grinblatt (2021), pooled LR, LASSO (augmented by squared and interaction terms), Random Forest, Gradient Boosting, and the method ensemble, respectively. Panel B shows results from a factor model time-series regressions of the following form: $r_{q,t+1} = a_q + \sum_{l=1}^{L} \beta_{q,l} F_{l,t+1} + \varepsilon_{q,t+1}$, where $r_{q,t+1}$ is the month t+1 industry-adjusted returns of the value-weighted quintile portfolios, a_q is the intercept in the time-series regression, and $F_{l,t+1}$ is the return of the lth factor portfolio. For brevity, Panel B reports the results for the difference in the alphas of Q5 and Q1 (Q5-Q1) only. We test whether the industry-adjusted value-weighted Q5-Q1 portfolio returns can be explained by the following asset pricing models: The Fama-French 5 factor model plus momentum (FF5FM+MOM); the Fama-French 5 factor model plus momentum plus short-term and long-term reversals (FF5FM+MOM+LT&ST Rev.); the modified Fama-French 5 factor model, using HML based on the current market value of equity (as by Asness and Frazzini, 2013) plus momentum ($FF5FM_{HMLcurrent}$+MOM); the modified Fama-French 5 factor model, using HML based on the current market value of equity (as by Asness and Frazzini, 2013) plus momentum plus short-term and long-term reversals ($FF5FM_{HMLcurrent}$+MOM+LT&ST Rev.). Numbers in parentheses are t-statistics. *, **, *** correspond to statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Industry-adjusted returns						
	LR(BG)	LR(pooled)	LASSO	RF	GBRT	Combi
1 (low)	−0.24*	−0.25*	−0.24*	−0.35**	−0.25**	−0.37***
	(−1.93)	(−1.85)	(−1.84)	(−2.57)	(−2.27)	(−2.77)
2	−0.16	−0.07	−0.11	−0.23*	−0.54***	−0.33**
	(−1.22)	(−0.68)	(−0.98)	(−1.83)	(−4.10)	(−2.49)
3	−0.19	−0.02	−0.06	−0.25*	−0.27**	−0.26*
	(−1.44)	(−0.24)	(−0.62)	(−1.82)	(−2.07)	(−1.95)
4	−0.16	−0.03	−0.02	−0.02	0.08	0.05
	(−1.38)	(−0.39)	(−0.28)	(−0.18)	(0.60)	(0.36)
5 (high)	0.12	0.05	0.12*	0.22**	0.24**	0.23**
	(1.17)	(1.01)	(1.77)	(2.29)	(2.12)	(2.23)
Q5-Q1 (Spread)	0.36***	0.30**	0.37**	0.57***	0.49***	0.60***
	(2.65)	(2.18)	(2.55)	(3.36)	(3.11)	(3.46)
Panel B: Alphas and t-values from assets pricing tests						
FF5FM+MOM	0.23*	0.22***	0.27***	0.60***	0.66***	0.64***
	(1.79)	(2.82)	(2.83)	(4.99)	(4.76)	(4.86)
FF5FM+MOM+LT&ST Rev.	0.22*	0.21***	0.26***	0.57***	0.56***	0.59***
	(1.68)	(2.71)	(2.72)	(5.05)	(4.73)	(4.96)
$FF5FM_{HMLcurrent}$+MOM	0.11	0.15**	0.18**	0.48***	0.54***	0.50***
	(1.04)	(2.18)	(2.21)	(4.77)	(4.46)	(4.57)
$FF5FM_{HMLcurrent}$+MOM+LT&ST Rev.	0.11	0.13**	0.17**	0.48***	0.49***	0.49***
	(1.04)	(2.04)	(2.12)	(4.96)	(4.51)	(4.75)

Third, we follow BG and sort stocks into five quintiles based on the corresponding mispricing signal.[10] Table 1, Panel A,[11] reports time-series averages of mispricing signals and selected firm characteristics for the quintile portfolios. All methods show large negative (positive) mispricing signals for the first (fifth) quintile. Thereby, ML and *LASSO* signals are considerably smaller because of the non-linearity of these valuation models and thus their ability to better fit the data. Further, Panel B shows the correlation between mispricing signals as well as between the corresponding spread returns. Interestingly, correlations between LR, LASSO, and ML models are moderate, which can be attributed to the heterogeneity in the relevance of accounting variables to the valuation models, as shown in Fig. 1. Furthermore, the correlation between mispricing signals from *RF* and *GBRT* is also limited, suggesting that ensemble averaging (*Combi*) may be valuable by "averaging out" noise.

Fourth, we calculate value-weighted and industry-adjusted monthly portfolio returns and perform a time-series analysis to study the relation between mispricing signals and subsequent month's returns.[12] Table 2, Panel A, shows that ML investment strategies earn statistically and economically significant industry-adjusted returns spreads, profiting quite uniformly from the long and short positions. While LR and *LASSO* quintile spreads are significant, their economic relevance is substantially weaker, with a higher portion of

[10] In line with the current standards (e.g., Hou et al., 2020; Fama and French, 2017, 2012), we calculate breakpoints using large stocks to avoid small stocks dominating our results. Large stocks are defined as the largest stocks in the sample that add up to 90% of the aggregate market capitalization. We also experiment with intra-country breakpoints as in BG and sample breakpoints. As expected, after these changes our trading strategies deliver even higher risk-adjusted returns. Results of these exercises are documented in the Internet Appendix.

[11] See Internet Appendix I for the extensive summary statistics.

[12] We follow BG and adjust returns by average industry portfolio, using Fama-French 38 industry classifications.

Table 3
Fama-MacBeth Regressions. The table shows results from Fama and MacBeth (1973) regressions. In Panel A, the next-month return is regressed on the mispricing signal based on a particular model: Bartram & Grinblatt (2021)-based LR model in Column 1, pooled LR in Column 2, LASSO (augmented by squared and interaction terms) in Column 3, Random Forest in Column 4, Gradient Boosting in Column 5, and the method ensemble in Column 6. Each specification controls for market beta, book-to-market, using the current market value of equity as in Asness and Frazzini (2013), the natural logarithm of market capitalization, momentum, short-term reversal, long-term reversal, accruals, gross profitability, earnings yield, and earnings surprise as well as for country and industry (38 Fama-French industries) fixed effects. Regressions employ quintile dummies for all covariates (except for fixed effects dummies). All quintiles are calculated with breakpoints based on large firms. Large firms are defined as those which account for an aggregated share of 90% of the total market value in a particular month (Fama and French, 2017). The regressions include dummy variables for quintiles 2, 3, 4, and 5 of each signal or firm characteristic. We display the coefficients of the quintile 5 dummy (Q5) for brevity. In Panel B, Columns 1–4 (5–8), the specification from Panel A, Column 1 (2), is extended by *LASSO* mispricing quintiles in Column 1 (5), by *RF* mispricing quintiles in Column 2 (6), by *GBRT* mispricing quintiles in Column 3 (7), and by *Combi* mispricing quintiles in Column 4 (8). Numbers in parentheses are t-statistics. *, **, *** correspond to statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Regression for a single method

Model	LR(BG)	LR(pooled)	LASSO	RF	GBRT	Combi
Dependent	Return($t+1$)					
	(1)	(2)	(3)	(4)	(5)	(6)
Mispricing (Q5)	0.16*	0.02	0.15***	0.33***	0.26***	0.32***
	(1.69)	(0.26)	(3.52)	(3.11)	(3.89)	(3.29)
Beta (Q5)	−0.08	−0.08	−0.09	−0.08	−0.09	−0.08
	(−0.54)	(−0.58)	(−0.65)	(−0.55)	(−0.66)	(−0.58)
$BM_{current}$ (Q5)	0.17*	0.22**	0.23**	0.14	0.15	0.13
	(1.70)	(2.17)	(2.17)	(1.38)	(1.47)	(1.29)
Ln(MarketCap) (Q5)	−0.14	−0.17	−0.10	−0.07	−0.14	−0.10
	(−1.04)	(−1.15)	(−0.73)	(−0.52)	(−1.03)	(−0.77)
Momentum (Q5)	0.94***	0.93***	0.93***	0.96***	0.97***	0.97***
	(6.74)	(6.68)	(6.69)	(7.06)	(7.09)	(7.16)
ST reversals (Q5)	−0.67***	−0.67***	−0.67***	−0.66***	−0.64***	−0.65***
	(−5.36)	(−5.37)	(−5.31)	(−5.27)	(−5.20)	(−5.23)
LT reversals (Q5)	−0.29***	−0.29***	−0.29***	−0.27***	−0.28***	−0.27***
	(−3.98)	(−4.02)	(−3.94)	(−3.76)	(−3.83)	(−3.76)
Accruals (Q5)	−0.28***	−0.28***	−0.27***	−0.26***	−0.27***	−0.26***
	(−5.21)	(−5.16)	(−5.01)	(−4.87)	(−4.97)	(−4.89)
Gross profitability (Q5)	0.31***	0.30***	0.31***	0.32***	0.31***	0.33***
	(5.78)	(5.63)	(5.79)	(6.12)	(5.80)	(6.11)
Earnings yield (Q5)	0.32***	0.37***	0.38***	0.31***	0.31***	0.30***
	(5.70)	(6.59)	(6.75)	(5.48)	(5.25)	(5.26)
SUE (Q5)	−0.00	−0.01	−0.04	−0.02	−0.02	−0.02
	(−0.03)	(−0.21)	(−0.75)	(−0.28)	(−0.27)	(−0.30)
Constant	2.43**	1.30	0.73	0.21	0.83	1.63
	(2.25)	(1.22)	(1.33)	(0.42)	(1.57)	(1.52)
Industry FE (38 FF)	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Aver. observations	2742	2742	2742	2742	2742	2742
Aver. Adj. R2	0.081	0.081	0.080	0.081	0.081	0.081

Panel B: Marginal effect of the LR mispricing signals

	LR(BG) vs:				LR(pooled) vs:			
Model	LASSO	RF	GBRT	Combi	LASSO	RF	GBRT	Combi
Dependent	Return($t+1$)				Return($t+1$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mispricing (Q5)	0.15***	0.29***	0.23***	0.28***	0.15***	0.34***	0.26***	0.32***
	(3.35)	(2.87)	(3.62)	(3.03)	(3.42)	(3.07)	(3.99)	(3.22)
Mispricing $(Q5)_{OLS}$	0.13	0.11	0.12	0.10	0.04	0.05	0.01	0.05
	(1.35)	(1.24)	(1.31)	(1.09)	(0.63)	(0.83)	(0.21)	(0.72)
Constant	1.24**	0.79	0.48	0.89*	1.18**	0.68	0.90	0.96*
	(2.05)	(1.30)	(0.80)	(1.68)	(2.19)	(0.65)	(1.54)	(1.75)
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (38 FF)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Aver. observations	2742	2742	2742	2742	2742	2742	2742	2742
Aver. Adj. R2	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081

alpha's profitability coming from the short leg.

In Panel B, we verify our results employing four different factor models, with factors constructed following the approach of Fama and French (2012, 2017) for international data.[13] Controlling for these common factors largely explains the returns of the *LR*(*BG*) strategy and renders its alpha insignificant when using the monthly updated HML factor (third and fourth model). Similarly, the alphas for *LR*(*pooled*) decrease for all models, although they remain significant (t-value between 2.04 and 2.82). In contrast, the alphas for the ML models remain similar or become even stronger across all factor models. Their annualized returns amount to 6% to 8%, depending on the method and factor model.[14] Interestingly, the alphas for *LASSO* are only slightly higher than for *LR*(*pooled*), indicating that trivial non-linear forms are not sufficient to improve the performance of LR compared to those obtained from ML. As such, ML methods seem to detect hidden non-linearities that are important in predicting stocks' fundamental value.

Finally, we dig deeper into whether using ML methods adds value from a portfolio selection perspective. Indeed, although ML seems to dominate LR in economic terms (Table 2), we still find positive and significant returns for LR. To analyze if LR is priced after controlling for ML (and other variables), i.e., measuring the marginal effect of the LR mispricing signals, we conduct Fama and MacBeth (1973) cross-sectional regressions. We include the same set of controls as BG and closely follow their variable definition. As in BG, we run regressions with quintile dummies, calculated as described above, for easier interpretation of coefficient estimates. Table 3 confirms our findings, showing statistically and economically significant coefficient estimates on the Q5 for ML-strategies (Panel A, Columns 4–6), while Q5 returns for LR-strategies (Panel A, Columns 1–2) are statistically and economically neglectable. Even more important, when including ML methods and *LR*(*BG*) (*LR*(*pooled*)) jointly in Panel B, Columns 2–4 (6–8), we find that only ML methods remain significant, while LR methods have little marginal ability to predict returns. Finally, Q5 returns for the *LASSO*-strategy are statistically significant in single specifications (Panel A, Column 3) and remain significant in joint specifications with LR methods (Panel B, Columns 1 and 5). Although the significance of *LASSO* in cross-sectional regressions is comparable to ML methods, the point estimate is still substantially lower than that of ML strategies. This finding underlines the importance of interaction and non-linearities for stocks' value predictions, on the one hand, and the superiority of ML methods in their detection, on the other hand.

4. Conclusion

Consistent with the intuition that ML may discover additional structure in the data, we document that portfolio spreads based on BG-inspired ML mispricing signals can earn large and significant alphas and outperform corresponding LR mispricing signals. These findings suggest that it is important to allow for non-linearities and interactions in fundamental analysis. Also, these findings suggest substantial non-naïve market inefficiencies. Future research may test the performance of other ML methods, such as multilayer perceptron or recurrent neural networks, and a larger array of predictors, probably outside simple accounting items.

CRediT authorship contribution statement

Matthias X. Hanauer: Conceptualization, Methodology, Validation, Supervision, Writing – original draft, Writing – review & editing. **Marina Kononova:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Marc Steffen Rapp:** Conceptualization, Data curation, Methodology, Software, Validation, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.frl.2022.102856.

Appendix

A. Variables

In our analysis, we use accounting and return data. We source accounting data from Refinitiv Worldscope Fundamentals. In contrast to BG, we do not use point-in-time but annual accounting data. However, we are very conservative when feeding our models with accounting information. Specifically, we allow our models to learn from recent accounting information only six months after the calendar year-end in order to ensure that the information is indeed in the public domain (e.g., Walkshäusl, 2014; Fama and French, 1992). Studying European stock markets, we draw accounting data in EUR.

We source monthly return data from Refinitiv Datastream. Following BG (2021), we calculate monthly returns from the total return index in USD. However, as described by Ince and Porter (2006), raw return data from Datastream may not be error-free. To ensure data quality, we follow Ince and Porter (2006) and Hanauer (2020) and apply dynamic screens. To eliminate delisted firms, we delete all zero returns (in local currency) from the end of the time series to the first non-zero return. In addition, we remove returns above 300%

[13] Next to the standard HML factor, we also use a monthly updated value factor that uses the most recent market capitalization ($HML_{(current)}$) as the latter is seen as superior in combination with momentum (cf., Barillas et al., 2020; Hanauer, 2020; Hanauer and Lauterbach, 2019; Asness and Frazzini, 2013). Appendix A describes all factors used in the analysis.

[14] We also experiment with equal-weighted portfolio returns. Our qualitative conclusions remain unchanged. Results of these exercises are documented in the Internet Appendix.

reversing within one month. Finally, following Jacobs (2016), we cross-sectionally winsorize all returns at the 0.1th and 99.9th percentiles.

All variables used are reported and defined in Table A.

Table A
Variable definition. The table presents the names, definitions, or construction for all variables used in the paper. The table also specifies the Refinitv code for accounting data items. All factor model variables and variables for cross-sectional regression are based on the Datastream data and are self-constructed.

Name	*Definition*	*Refinitiv Code*
Accounting items		
TotalAssets	Total Assets	WC02999
NItoCommon	Net Income Available to Common	WC01751
NIbefEIPrefDiv	Net Income before Extraordinary Items/Preferred Dividends	WC01551
PrefDiv	Preferred Dividends Requirements	WC01701
NIbefPrefDiv	Net Income before Preferred Dividends	WC01651
Sales	Net Sales or Revenues	WC01001
GainLossAssetSale	Extraordinary Items & Gain/Loss Sale of Assets	WC01601
PPT	Property, Plant and Equipment - Net	WC02501
LTDebt	Long Term Debt	WC03251
CommonEquity	Common Equity	WC03501
PrefStock	Preferred Stock	WC03451
OtherIncome	Other Income/Expense - Net	WC01262
TotalLiabilities	Total Liabilities	WC03351
PreTaxIncome	Pretax Income	WC01401
IncomeTaxes	Income Taxes	WC01451
OtherTA	Other Assets Total	WC02652
OtherLiabilities	Other Liabilities	WC03273
CashSTInv	Cash & Short-Term Investments	WC02001
OtherCA	Total Current Assets	WC02201
OtherCL	Total Current Liabilities	WC03101
TotalDiv	Cash Dividends Paid - Total	WC04551
MV	Market Capitalization	MV
Factor model variables		
Mkt	Value-weighted monthly market return net of risk-free rate (one-month Treasury bill rate, obtained from Kenneth French's website)	
SMB	Monthly small minus big size portfolio return, using market capitalization (MV) to construct size	
HML	Monthly high minus low book-to-market portfolio return with latest available book value of common equity, divided by the market capitalization at the end of December of the same year as book equity	
$HML_{(current)}$	Monthly high minus low book-to-market portfolio return with latest available book value of common equity, divided by the most recent monthly market capitalization ($BM_{current}$)	
CMA	Monthly conservative minus aggressive investment portfolio return, with investment factor based on asset growth	
RMW	Monthly robust minus weak profitability portfolio return, with profitability factor defined as gross income plus depreciation and depletion, scaled by total assets	
WML	Monthly winners minus losers portfolio return with prior (2–12) returns to construct momentum	
ST reversals	Monthly low prior portfolio returns minus high prior portfolio returns with prior (1–1) returns to construct short-term reversals.	
LT reversals	Monthly low prior portfolio returns minus high prior portfolio returns with prior (13–60) returns to construct long-term reversals	
Variables for cross-sectional regression		
Beta	Monthly market beta with regards to the European market estimated over prior 60 months, using value-weighted returns	
$BM_{current}$	Latest available book value of common equity, divided by the most recent market capitalization updated each month	
MarketCap	Natural logarithm of stock market capitalization (in USD)	
Momentum	Return in prior year excluding prior month	
ST reversals	Return in prior month	
LT reversals	Return in prior five years excluding prior year	
Accruals	The percentage difference in net operating assets in the current and previous year. Net operating assets is defined as operating assets minus operating liabilities. Operating assets is defined as total assets less cash and short-term investments. Operating liabilities is defined as total assets less total debt less book value of total common and preferred equity less minority interest	
Gross profitability	Gross income plus depreciation and depletion, scaled by total assets	
Earnings yield	Latest available net income after preferred dividends, divided by the market capitalization	
SUE	Yearly earnings surprise based on a rolling random walk model	

References

Abourachid, H., Kubo, A., Orbach, S., 2017. Momentum strategies in European equity markets. Finance Res. Lett. 23, 147–151.
Al Shalabi, L., Shaaban, Z., Kasasbeh, B., 2006. Data mining: a preprocessing engine. J. Comput. Sci. 2, 735–739.
Arnott, R.D., Harvey, C.R., Kalesnik, V., Linnainmaa, J.T., 2021. Reports of value's death may be greatly exaggerated. Financial Anal. J. 77, 44–67.
Asness, C., Frazzini, A., 2013. The devil in HML's details. J. Portf. Manag. 39, 49–68.
Athey, S., Imbens, G.W., 2019. Machine learning methods economists should know about. Annu. Rev. Econom. 11, 685–725.
Bali, T.G., Goyal, A., Huang, D., Jiang, F., Wen, Q., 2021. Different strokes: return predictability across stocks and bonds with machine learning and big data. Research Paper Series 20–110.
Bao, Y., Ke, B., Li, B., Yu, Y.J., Zhang, J., 2020. Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. J. Account. Res. 58, 199–235.
Barillas, F., Kan, R., Robotti, C., Shanken, J., 2020. Model comparison with Sharpe ratios. J. Financ. Quant. Anal. 55, 1840–1874.
Bartram, S.M., Grinblatt, M., 2021. Global market inefficiencies. J. Financ. Econ. 139, 234–259.
Bartram, S.M., Grinblatt, M., 2018. Agnostic fundamental analysis works. J. Financ. Econ. 128, 125–147.
Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Rev. Financ. Stud. 34, 1046–1089.
Blitz, D., Hanauer, M.X., 2021. Resurrecting the value premium. J. Portf. Manag. 47, 63–81.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth and Brooks/Cole, Monterey.
Drobetz, W., Haller, R., Jasperneite, C., Otto, T., 2019. Predictability and the cross section of expected returns: evidence from the European stock market. J. Asset Manag. 20, 508–533.
Erel, I., Stern, L.H., Tan, C., Weisbach, M.S., 2021. Selecting directors using machine learning. Rev. Financ. Stud. 34, 3226–3264.
Fama, E.F., French, K.R., 2017. International tests of a five-factor asset pricing model. J. Financ. Econ. 123, 441–463.
Fama, E.F., French, K.R., 2012. Size, value, and momentum in international stock returns. J. Financ. Econ. 105, 457–472.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Finance 47, 427–465.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81, 607–636.
Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Rev. Financ. Stud. 33, 2326–2377.
Ghosh, P., Neufeld, A., Sahoo, J.K., 2021. Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. Finance Res. Lett., 102280
Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33, 2223–2273.
Hanauer, M.X., 2020. A comparison of global factor models. SSRN Working Paper no. 3546295, doi:10.2139/ssrn.3546295.
Hanauer, M.X., Lauterbach, J.G., 2019. The cross-section of emerging market stock returns. Emerg. Mark. Rev. 38, 265–286.
Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. Rev. Financ. Stud. 33, 2019–2133.
Ince, O.S., Porter, R.B., 2006. Individual equity return data from Thomson Datastream: handle with care! J. Financ. Res. XXIX, 463–479.
Israel, R., Laursen, K., Richardson, S., 2021. Is (systematic) value investing dead? J. Portf. Manag. 47, 38–62.
Jacobs, H., 2016. Market maturity and mispricing. J. Financ. Econ. 122, 270–287.
Karolyi, A.G., 2016. Home bias, an academic puzzle. Rev. Financ 20, 2049–2078.
Lundberg, S., Lee, S., 2017. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 47, 68–77.
Ohlson, J.A., 1995. Earnings, book values, and dividends in equity valuation. Contemp. Account. Res. 11, 661–687.
Park, H., 2019. An intangible-adjusted book-to-market ratio still predicts stock returns. Critic. Finance Rev. 25, 207–236.
Varian, H.R., 2014. Big data: new tricks for econometrics. J. Econ. Perspect. 28, 3–28.
Walkshäusl, C., 2021. Predicting stock returns from the pricing and mispricing of accounting fundamentals. Q. Rev. Econ. Finance 81, 253–260.
Walkshäusl, C., 2014. The MAX effect: European evidence. J. Bank. Financ. 42, 1–10.