Exploring the factor zoo with a machine-learning portfolio

Halis Sak[1], Tao Huang[2] and Michael T. Chng[3]

Abstract

With the growing reliance on machine-learning (ML) methods in finance, an understanding of their long-term efficacy and underlying mechanism is needed. We document the time-varying importance of different stock characteristics over an 18-year (1998-2016) out-of-sample period to determine whether ML models, when trained on a large set of firm and trading characteristics, can consistently outperform factor models. Utilizing a combination of linear and nonlinear models, we form a ML portfolio that consistently generates a significant alpha against factor models, ranging from 2.14 to 2.74% per month. We uncover patterns in characteristic dominance that alternates between arbitrage and financial constraint features. The variation correlates with the US credit cycle, and highlights a fundamental economic mechanism underlying the ML portfolio's performance. The study's impact extends to both academics and practitioners, providing insights into the economic drivers of stock returns and the practical implementation of ML methods in portfolio construction.

JEL classification: G12, G32.

[1]Shenzhen Audencia Financial Technology Institute, Shenzhen University, and Department of Finance, Hong Kong University of Science and Technology; Email: halis@szu.edu.cn.

[2]Faculty of Business Management, Beijing Normal University-Hong Kong Baptist University United International College; Email: taohuang@uic.edu.cn.

[3]Corresponding author: International Business School Suzhou (IBSS), Xi'an Jiaotong-Liverpool University (XJTLU) China; Email: Michael.Chng@xjtlu.edu.cn; An earlier draft is circulated under the title 'Rise of the machine-learning portfolio' http://ssrn.com/abstract=3202277. We thank Jari Kappi, Kim Sawyer, Sean Anthonisz, Kevin Zhu, Stephen Gong, Takeshi Yamada, Lin William Cong, Yi-Li Chien, Andrew Ellul and seminar participants at ANU, University of Adelaide, Deakin University, University of Sydney, University of Wollongong, Fudan University and XJTLU. We thank discussants at the 2019 Fudan-Stanford FinTech conference, 2019 SFM-Journal of Financial Markets conference, 2019 SUFE/Edinburgh-Review of Corporate Finance Studies FinTech conference. The corresponding author is grateful to Yuelan Chen for many discussions during the write-up of the paper. Halis would like to acknowledge financial support from Shenzhen Humanities & Social Sciences Key Research Bases. Tao would like to acknowledge financial support from the Key Platforms and Scientific Research Projects of Universities in Guangdong Province (2020ZDZX3067). We retain full property rights to all errors.

Keywords: Factor model; Firm characteristic; Return predictability.

"We also thought that the cross-section of expected returns came from the CAPM. Now we have a zoo of new factors."

John Cochrane, Presidential Address
2011 American Finance Association Meeting

1. Introduction

Harvey et al. (2015) noted that when Fama and French (1992, 1993) published their influential papers, they identified around 40 anomalies (i.e., irregularities in stock returns) that the capital asset pricing model (CAPM) could not explain. By 2003, the number of these anomalies had doubled to 84. Fast-forward to 2012, and the total had surged to 240. Cochrane (2011) referred to the continuing discovery of anomalies as a "factor zoo." The increasing computational ease of data mining has contributed to the identification of many stock-market anomalies. For instance, a significant stock characteristic might be explained by theories related to risk (e.g., a company's financial distress or quality), mispricing (e.g., limits to arbitrage opportunities), information (e.g., metrics derived from options), or behavioral factors (e.g., herding or anchoring behaviors). Conversely, if a characteristic is found to be insignificant, it might be dismissed as a result of data snooping or market inefficiency. It is common to find correlations among subsets of anomalies, such as those related to illiquidity (Amihud, 2002; Pástor & Stambaugh, 2003; Liu, 2006), idiosyncratic volatility (Ang et al., 2006; Fu, 2009), or coskewness (Kraus & Litzenberger, 1976; Harvey & Sidique, 2000; Conrad et al., 2013). Harvey et al. (2015) suggested that many published findings in Financial Economics are probably false.

Enormous research effort is spent discovering new anomalies and debunking extant ones. Hou et al. (2020) reported that 60% of published anomalies cannot be replicated based on a standard t-stat hurdle of 1.96. After adjusting for data snooping using a t-stat of 2.78, the proportion of insignificant anomalies increased to 80%. However, the literature hardly pays any attention to the factor-zoo characteristics that persist, such as the 20% of anomalies in Hou et al. (2020) that have been replicated.

Our motivation and contribution are to identify the characteristics that remain significant and to ascertain a possible economic mechanism behind them that is dynamic over time. Mclean and Pontiff (2016) explained a once-off rise and fall in a published

anomaly. However, it is not clear how one would explain the recurring significance of certain characteristics over a long sample period. To fill in this knowledge gap, we conducted a comprehensive out-of-sample factor analysis over a long period, without imposing any assumptions on the underlying factor structure. Our factor-zoo analysis spanned time $t = 1, ..., T$ across firms $i = 1, .., N$ for $k = 1, ..., K$ characteristics. Standard econometric tools can handle a large panel of N firms over time T, but they accommodate only a small variable choice set. This poses a problem when K is large. The estimation is normally in-sample and assumes a linear factor structure with some added non-linearity (e.g., interacting and/or squared variables). Functionally, the number and types of variables underlying the return-generating process structure are unobservable. Moreover, a given factor structure may fit some characteristic subsets but not others. Additionally, a nonlinear factor structure becomes less stable when taken out-of-sample, which is why standard factor models follow a parsimonious linear structure, with three to five characteristics between them.

Therefore, we used machine-learning (ML) models in this paper. These models specialize in prediction tasks that facilitate out-of-sample analyses. In stark contrast, overfitted econometric models degrade rapidly when applied out-of-sample. Using an objective function to maximize stock-return forecast accuracy over a 1980-1998 sample, we trained different ML models to choose from $K = 106$ firm and trading characteristics to estimate the factor structures that maximize the objective function. We applied the trained ML models on a 1998-2016 out-of-sample testing period to generate monthly return forecasts and form a portfolio of predicted winner (PW) and predicted loser (PL) stocks. The ML portfolio generated a significant alpha (α_{MLA}) against entrenched factor models, including those of Fama and French (1993; i.e., FF3, 2015 FF5, and 2018 FF6), Carhart (1997; i.e., C4), Hou et al. (2015; i.e., Q4), and Hou et al. (2021; i.e., Q5).

Using the ranked difference in normalized characteristic value between the PW and PL stocks, we identified the ML portfolio's dominant characteristics over the 18-year out-of-sample period. Given the significant α_{MLA}, long-surviving characteristic anomalies were identified. The ML portfolio analysis uncovered a noteworthy pattern in the rise and fall of dominant characteristics in the factor zoo. Specially, we found that only two small subsets of three or four characteristics play an *alternating* dominant role in generating the ML portfolio return. In the literature, these subsets are generally viewed as attributes of

investor-level arbitrage constraints (i.e., Ivol [Ang et al., 2006] and max and min effects [Bali et al., 2011]) or firm-level financial constraints (i.e., cash flow risk [Da & Warachka, 2009], growth in external financing [Bradshaw et al., 2006], sale of common preferred stock [Pontiff & Woodgate, 2008], and gross profitability [Novy-Marx, 2013]).

Given that all the characteristics in K were published by 2016, it is unlikely that any dominant characteristic could generate a significant α_{MLA} against newer factor models (e.g., FF5/FF6 and Q4/Q5). This suggests that a potential source of α_{MLA} could stem from the ML portfolio's time-varying exposure to dominant characteristics over the testing period. Using a conceptual argument alongside empirical results, we uncover the alternating importance of arbitrage and financial constraint characteristics, showing that they coincide with different stages of the credit cycle. The finding offers fundamental insight into a longer-horizon explanation of cross-sectional stock returns.

Our paper complements several recent studies. That the ML portfolio loaded on a small set of characteristics, is consistent with Mclean and Pontiff (2016) and Hou et al. (2020), who suggested that many published anomalies could not be replicated. Our emphasis is to better understand the source of a pervasively significant α_{MLA} against factor models, which we associate with two distinct alternating characteristic subsets. Avramov et al. (2023) attributed the α_{MLA} to microcap and distress stocks, arguing that an unconstrained α_{MLA} would be driven by mispricing, owing to the limits to arbitrage, rather than abnormal returns being realized. Leippold et al. (2022) built ML portfolios using Chinese stocks, confirming that illiquidity characteristics dominate in markets crowded with retail investors. Both studies examined potential sources of the α_{MLA} to better understand ML portfolio risks and rewards, but they did not focus on the dynamic characteristic exposure to out-of-sample data. Our ML portfolio not only exhibits time-varying exposure, but the timing of the exposure on arbitrage and financial constraint characteristics align with the contraction and expansion stages of the US credit cycle.

Our paper proceeds as follows. The next section reviews recent studies that employ ML in finance realms and examines different aspects of the factor zoo. Section 3 outlines our ML training procedure, Section 4 presents our ML portfolio analysis, and Section 5 provides conclusions.

2. Literature review on ML and the factor zoo

Researchers have utilized ML methods as prediction tools for a variety of issues, ranging from financial crises to energy prices (Samitas et al., 2020; Polyzos et al., 2021; Alshater et al., 2022)). For asset pricing, ML methods have been used to predict expected returns, analyze factors, assess risk exposure, determine risk premia, and calculate stochastic discount factors. These models are also used in model comparisons and trading strategy evaluations. Giglio et al. (2022) provided a comprehensive survey of the latest methodological advancements and MLAs, highlighting advancements in econometrics problems with large characteristic domains K. There are four main problems currently pursued in this field. Notably, researchers are trying to raise the acceptance hurdle to expand the factor zoo, reduce the dimensionality of the factor zoo, apply principal component and/or factor analysis to extract common latent factors, and create factor-zoo portfolio-analysis applications.

In terms of raising the acceptance hurdle, recent studies have voiced data-mining concerns in empirical asset pricing. For example, Foster et al. (1997) addressed a related problem associated caused by the increased availability of data, proposing a simple procedure to adjust the critical maximal R^2 value to account for variable snooping. Hou et al. (2020) reported that an adjusted t-stat of 2.78 at a 5% significance could reject around 80% of the published anomalies. Harvey et al (2015) documented the proliferation of anomalies, and introduced a new multiple testing framework that infers historical acceptance thresholds from past studies. They proposed that a new factor must exhibit a t-stat greater than 3.0. Chordia et al. (2019) implemented a data-mining approach that generates over two-million trading strategies. Using multiple hypothesis testing to account for covariance in trading signals and returns, they controlled for the proportion of false rejections by proposing that the 5% significance t-stat threshold should be closer to 4.0.

Regarding reducing the dimensionality of the factor zoo, Freyberger et al. (2017) used an adaptive group–based least absolute shrinkage and selection operator (LASSO) to select characteristics that provide independent information. To address model-selection bias, Feng et al. (2018) combined LASSO with a double least absolute shrinkage method that used a two-pass regression procedure (i.e., Fama–MacBeth) to identify an appropriate finite set of control variables to evaluate a new candidate factor. Giglio and Xiu

(2016) and Kelly et al. (2019) applied dimension reduction methods to estimate and test factor-pricing models. Feng et al. (2020) applied ML models to evaluate the explanatory power of any new factor over a high-dimensional set of existing factors. They found that a small set of characteristics provide statistically significant explanatory power incrementally over hundreds of known factors in the literature.

Regarding principle component and factor analyses, Kelly et al. (2017) used characteristics as instruments to extract principal components to analyze time-varying factor loadings, and Light et al. (2017) applied partial least-square estimation to extract a finite set of common latent factors from characteristics. They reported that their latent factor–sorted portfolios produced a larger spread in returns than individual characteristic portfolios. Kozak et al. (2020) used shrinkage and selection methods to estimate a stochastic discount factor that explains returns for a large number of stocks, and Freyberger et al. (2017) use similar methods to approximate a nonlinear factor structure of expected returns.

For factor-zoo applications, Moritz and Zimmermann (2016) applied tree-based models to sort portfolios, and Gu et al. (GKX, 2020) trained 13 ML models on a factor zoo of 94 characteristics, identifying a set of important characteristics (i.e., variants of momentum, liquidity, and volatility) that were common across models. GKX (2020) attributed the outperformance of boosted trees and neural networks to their ability to allow nonlinear interactions among characteristics. Avramov et al. (2023) documented a significant α_{MLA} from a portfolio formed using ML models trained on an unconstrained factor zoo. However, they found that the alpha's significance was sensitive to economic restrictions. Specifically, their unconstrained ML portfolio loads heavily on microcaps and financially distressed stocks. There are recent deep-learning finance applications that utilize multisequence techniques to capture correlations among firm characteristics over extended periods (Feng et al., 2018). Recent studies have also explored the use of deep reinforcement learning to directly enhance portfolio performance (Cong et al., 2021). Bayesian frameworks have been used for model selection as well (Barillas & Shanken, 2018; Bryzgalova et al., 2023).

Our paper complements both GKX (2020) and Avramov et al. (2023) in several aspects. Most discoveries presented in GKX (2020) involved in-sample analyses of multiyear information acquired by individual ML models during repeated training iterations;

the primary emphasis was on stock returns. In contrast, our ML portfolio is generated from an ensemble forecast from different ML models trained only once on the first half of the data to provide out-of-sample analyses. This method is less dependent on specific ML models and more equitable to firm characteristics with varying update frequencies. Our approach allows us to reverse-engineer out-of-sample patterns from the characteristic behaviors found in the ML portfolio. Lastly, the dominant characteristics reported by GKX (2020) were variant measures of volatility, illiquidity, and momentum. These are all trading characteristics that correlate to an extent with dominant arbitrage constraint characteristics (i.e., Ivol and Max/Min) in our ML portfolio. We also find that their dominance alternates with a contrasting set of firm-level financial constraint characteristics.

Avramov et al. (2023) reported a significant α_{MLA} from an unconstrained factor zoo, attributing it to difficult-to-arbitrage stocks (i.e., microcaps) and financial distress indicators (e.g., no rating or downgrades). Our ML portfolio is also generated from an unconstrained factor zoo; however, instead of restricting our test to a single source of the alpha, we dissect the ML portfolio to uncover patterns in the dominant characteristics over time. Like Avramov et al. (2023), we find that proxy arbitrage constraint characteristics are important. However, we also discover that their importance alternates with the firm-level financial constraint characteristic proxy.

Many research papers focus on ML model–specific risk premia. Alternatively, our approach incorporates asset pricing models into the aggregation of return predictions from different ML models. Few studies have attempted to directly interpret output from multiple models; however, Cong et al. (2021) did so by employing a characteristic importance method using gradients to identify the main characteristics driving their ML predictions. In contrast, we use portfolio analysis to identify the dominant characteristics afterward.

3. MLA Methodology

Our MLA inputs the firm sample, N, which is an average of 2,500 stocks listed on NYSE, NASDAQ, and AMEX) with $K = 106$ firm and trading characteristics. Different ML models are trained using 1980–1998 data, and a stock-return forecast is provided for each month, combined from the multiple trained ML model predictions. Our MLA then uses these to sort firms and identify PWs and PLs. This process provides the long and

short legs of the portfolio, which are then rebalanced monthly over the 1998–2016 testing period.

One of our research objectives is to determine whether there are any characteristics in K that survive over the long out-of-sample period. For this, we ensured that our $K = 106$ characteristics were comparable with well-cited papers. Notably, Gu et al. (2020) set their K to 94, Green et al. (2017) to 94, Mclean and Pontiff (2016) to 97, Kozak et al. (2020) to 80, and Feng et al. (2020) to 150. The consensus is that most characteristics in a large K are redundant. For example, Hou et al. (2020) found that nearly 82% of their 450 characteristics were anomalies, which rendered the results less significant in later sample periods. We confirmed that only a small subset of K was dominant in our ML portfolio during the out-of-sample period.

Our 1980–2016 full-sample period is comparable with the 1976–2017 period applied by Feng et al. (2020), which is one of the first papers to apply ML methods in asset pricing. We want to test dominant characteristics against economic states, which requires a long out-of-sample period. Hence, this period also covers normal and crisis trading conditions so that our findings are relevant to both states. The training sample includes the crash of October 1987, and the testing period includes the 2000 DotCom crash, the September 11th, 2001 terrorist attack, the 2008 global financial crisis, and the 2015–2016 Wall Street sell-off. An unbiased training/testing partition is applied by simply dividing the full 1980–2016 sample period evenly into a 1980–1998 training period and a 1998–2016 testing period. This forms the basis for our main empirical analysis.

3.1. Constructing the ML portfolio

Figure 1 provides a flowchart illustrating the key stages of constructing our ML portfolio. Starting with $K = 106$ firm and trading characteristics, we single-sort the firms on the level and change of each $k = 1, 2, ..., K$ characteristic, which yields 212 spread portfolios. In Step 1, the MLA trains different ML models on these 212 characteristic portfolios, which are then shortlisted into the model set, M, based on in-sample return forecast accuracy. To obtain an ensemble forecast, we apply stacking to generate a conditional probability distribution over trained models in M per Wolpert (1992), Ho and Hull (1994), and Kittler et al. (1998). Effective stacking requires a shortlist of important characteristics (i.e., feature selection), and numerous selection methods are available

(Chandrashekar & Sahin [2014]); however, their performance is entirely data-specific.

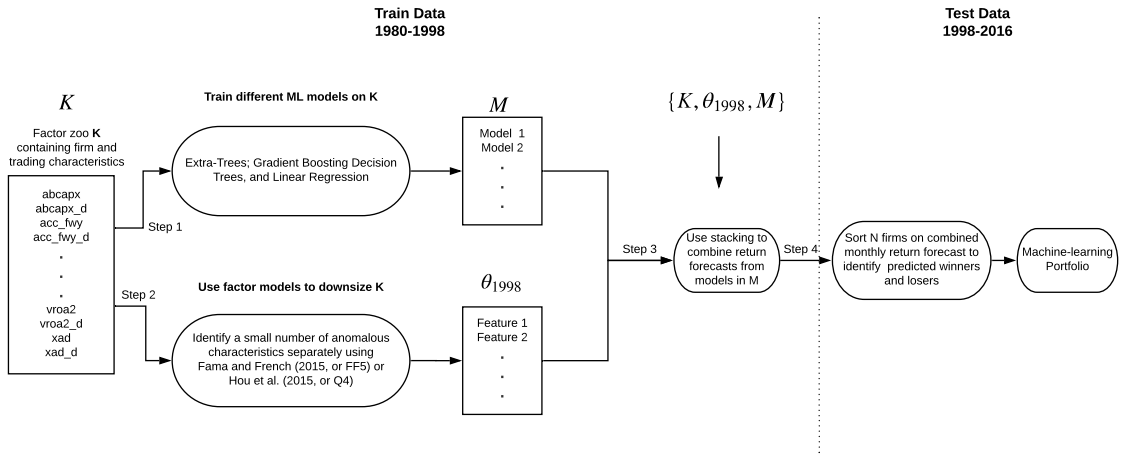


Figure 1: Blueprint for constructing the ML portfolio.
The MLA trains different ML models on K characteristics in Step 1, after which the trained models are shortlisted into model set M, based on in-sample return forecast accuracy. In Step 2, the MLA separately utilize FF5 and Q4 factors to identify important features θ_{1998}. In Step 3, the MLA uses θ_{1998} to implement stacking, which generates a probability distribution over M, after which it computes a probability-weighted return forecast for each stock. Lastly, in Step 4, firms are decile-sorted on forecast return to form PW and PL portfolios, which become a monthly rebalanced long–short ML portfolio over the testing period.

In Step 2, we address the feature selection problem by using factor models to identify training-sample anomalies. We run separate regressions of each characteristic portfolio return against FF5 and Q4 factors and rank them on α to identify a small subset, $\theta_{1998} \in K$, that is anomalous to each factor model.

In Step 3, the MLA implements stacking by using θ_{1998} to generate a probability distribution over model set M, after which it computes a probability-weighted return forecast for each stock, $\hat{r}_{i,t+1}$. Lastly, in Step 4, firms are sorted on $\hat{r}_{i,t+1}$ to form PW and PL decile portfolios, which become the long and short legs of the monthly rebalanced ML portfolio over the testing period. As a comparison, GKX (2020) trains different ML models (Figure 1 Step 1), and uses each model's predicted returns directly to sort firms into PWs and PLs (Figure 1 Step 4). Afterward, the analysis focuses on a small set of common dominant trading characteristics identified by the different trained ML models.

To complement Figure 1, Algorithm 1 outlines the key details of the ML portfolio construction process. In Stage 1, we evenly divide the 36-years full-sample period into a training sample (1980–1998) and testing sample (1998–2016). Using the training sample, we estimate different ML and linear regression models and identify training-sample

anomalies in θ_{1998}. The training procedure strictly ends in June 1998, after which there is no conditional updating of model parameters or θ_{1998}. As a robustness check, we consider a subsample partition based on the 1980–1994 training sample and the 1994–2006 testing sample. The main findings are similar; hence, we focus our discussion on the full-sample partition.

In Stage 2, the MLA trains three types of models: extra-trees (ET; Geurts et al., 2006), gradient boosting decision tree (GBDT; Friedman (2001)), and linear regression. ET and GBDT are ML variants that are trained on the entire factor zoo, K, considering a range of model configurations. The next section of this paper contains a brief technical description of the ET and GBDT models. Note that GKX (2020) provides a comprehensive and succinct overview of different ML models for readers with no computer science background. For linear regression models, the MLA estimates a two-factor specification from an exhaustive pairwise combination of θ_{1998} characteristics.

Each month during the training period, the MLA evaluates each model's in-sample ability to predict stock returns for all firms $i = 1, ..., N$. We measure forecast accuracy based on the R-value in Equation (1), where r_i is the realized return of stock i, $\hat{r}_i$ is the predicted return, and μ is the mean return for all firms. The R-value normalizes the sum-of-squared forecast error, $\sum_i^N (r_i - \hat{r}_i)^2$, across N firms using the same month's cross-sectional return variance, $\sum_i^N (r_i - \mu)^2$. Each month, the trained model with the highest R-value is shortlisted into the model set, M. Note that a trained model that generates poor return forecasts could produce a negative R^2. Hence, to calculate a valid R-value, we take the square-root of the absolute value of R^2 and add a negative sign to indicate that it is a low score. Whereas R^2 and the R-value give similar model rankings, the latter is more commonly used in the ML.

$$R^2 = 1 - \frac{\sum_i^N (r_i - \hat{r}_i)^2}{\sum_i^N (r_i - \mu)^2}$$

$$R\text{-value} = sign(R^2)\sqrt{|R^2|} \tag{1}$$

The above shortlisting procedure provides a model set, M, that includes all trained models that are ranked best in predicting stock returns at least once during the training period. The number of models in M is less than the total number of trained models

from ET, GBDT, and linear regressions. Hence, a trained model may exhibit the highest R-value over several months or even years. Another scenario involves a trained model that never produces the highest R-value throughout the training sample. As such, it is unlikely that any given model will dominate all others trained on the same sample. However, our approach accommodates possible variations in model prediction power based on the timeframe, which could be associated with its functional form or the relevance of characteristics in focus. It is important to note that the identification of M is strictly in-sample.

In Stage 3, the MLA performs stacking to combine stock-return forecasts from different trained models in M. The MLA generates a probability distribution over M, conditional on θ_{1998}. For each firm i, the MLA computes a conditional probability-weighted monthly return forecast. Intuitively, a heavier weight is assigned to trained models in which θ_{1998} characteristics are important. The point is that θ_{1998}, which contain training-sample characteristics that generate significant α against either FF5 or Q4, are more important than other characteristics in explaining stock returns over time.

Here, Stage 3 represents a technical contribution to the application of trained ML models in portfolio allocation. We use factor models to address the feature selection problem by stacking returns. It is inappropriate to weigh return predictions using the R-value of different trained models because it is measured from different parts of the training sample. It is also suboptimal to utilize R-values based on the entire training sample as weights. These R-values indicate each model's average return prediction performance over the training sample, but they would not capture the likely time-varying importance of different characteristics. Notably, the probabilities over M are conditional on θ_{1998}; however, the ML models themselves are trained on the factor zoo, K. We can confirm that the trained ET and GBDT models in M are assigned an average 85% probability weighting.

Lastly, in Stage 4, the MLA generates a monthly stock-return forecast, r_{it+1}, from each trained model in M. Using the conditional probability distribution over M, the MLA computes a probability-weighted monthly stock-return forecast, $\hat{r}_{it+1}$. Firms are sorted on $\hat{r}_{it+1}$ to form an ML portfolio that buys top decile-PWs, and short-sells the bottom-decile PLs. Hence, we have two ML portfolios that come from FF5 and Q4 to identify θ_{1998}, which the MLA apply to combine model forecasts. More than half of the

Algorithm 1: Key stages to constructing the ML portfolio.

- **Stage 1: Training and testing-sample partitions**
  1. Divide the 36-year full sample into training (1980–1998) and testing (1998–2016) sets.
  2. Use the training sample to identify model set M.
  3. Identify training-sample anomalies $\theta_{1998} \in K$ (see Section 3.2 and Algorithm 2 for details).
  4. Trained ML models are not retrained during the testing period (i.e., no dynamic updating).

- **Stage 2: The MLA training procedure**
  1. Using 1980–1998 data on all K=106 characteristics, the MLA estimates ET and GBDT models with different parameter settings. For linear regression, the MLA estimates an exhaustive pairwise combination of characteristics in θ_{1998}.
  2. For each month, t, in the training sample, the MLA computes each model's R-value to evaluate among the trained models.
  3. Each month, the model with the highest R-value is shortlisted into model set M.

- **Stage 3: Generate the ensemble forecast by stacking on θ_{1998}**
  1. The MLA trains another decision tree to generate a probability distribution over M, conditional on the training-sample anomalies, θ_{1998}.
  2. Models in M, for which one (or more) characteristic in θ_{1998} is important, receive a larger probability weight.

- **Stage 4: Generate return predictions to form the ML portfolio**
  1. For each month, t, in the testing sample, the MLA
     (a) uses each model in M to predict the next-month's return, r_{it+1}, for each firm, i.
     (b) uses the decision tree from Stage 3 to compute the ensemble forecast, $\hat{r}_{it+1}$, as the probability-weighted stock-return forecast from the models in M.
     (c) repeats this procedure on a monthly basis until the end of the testing period.
  2. decile-sorts each firm at each month t on $\hat{r}_{it+1}$ to form a long–short ML portfolio of PWs (top decile) and PLs (top decile).

characteristics in θ_{1998} are the same between FF5 and Q4, such that the main findings are consistent between the two portfolios.

3.1.1. Brief outline of the ET and GBDT models

The estimation procedures for ET and GBDT are complex, as they combine predictions from numerous decision trees. Here, a decision tree refers to a nonparametric model that resembles a tree-like structure. GKX (2020) Section 1.6, Figure 1 provides a good comparison between a decision tree on size and value, and the equivalent table for a two-way sort. It demonstrates a decision tree's ability to handle a large number of characteristics, unlike conventional sorting. The estimation procedure for the decision tree is based on squared residuals, including some regularization terms to penalize for complexity (e.g., depth of the tree).

For a random forest algorithm, the estimation procedure separately estimates multiple decision trees on random sampling from the same training sample. A prespecified weighting-scheme is used to combine predictions from different decision trees. An ET algorithm uses the full training sample to estimate decision trees. However, the decision boundaries of numerical input features are set randomly instead of being optimized. A gradient boosting method updates a model with the gradient of the loss function in an iterative fashion. The GBDT extends this idea to decision trees. The algorithm grows successive trees based on the residuals of the preceding tree. In this paper, we use LightGBM Python package (Ke et al., 2017) to estimate ET and GBDT models. This approach speeds up the training procedure of conventional GBDT using gradient-based one-sided sampling and exclusive feature bundling.

3.2. Identify training-sample anomalies for stacking

To address the feature selection problem in Stage 3, we separately use FF5 and Q4 to identify a small number of anomalous characteristics, $\theta_{1998} \in K$. As the MLA progressively moves through the 1998–2016 testing period, more observations on θ_{1998} and other characteristics become available. These are sequentially fed into the MLA to generate a monthly time-series of the next-month's stock-return forecasts. Note that θ_{1998} were identified by June 1998; they were not updated during the testing period.

3.2.1. Ranking approaches

We identify θ_{1998} based on each characteristic's ability to generate α against FF5 and Q4 factors. Following Fama and French (1996), we decile-sort firms on the level and change in each of the K =106 firms and their trading characteristics, generating 212 characteristic portfolios. The firm characteristic portfolios are rebalanced at the end of every June, whereas trading-characteristic portfolios are rebalanced monthly. Characteristics are ranked on the magnitude of their portfolio, α, from separate regressions against FF5 and Q4 factors. As FF5 and Q4 consider a similar set of factors, it is unsurprising that their corresponding θ_{1998} contain similar characteristics, albeit with slightly different rankings.

1. The Fama and French (2015) method of regressing characteristic portfolio returns using FF5 factors.

   $$r_{kt}-r_{ft} = \alpha_{Fk}+b_k(r_{mt}-r_{ft})+s_k(SMB_t)+h_k(HML_t)+i_k(CMA_t)+p_k(RMW_t)+\varepsilon_{kt},$$

   where r_{mt} is the value-weighted (VW) return of the market portfolio, r_{ft} is the risk-free return, SMB is the spread return between a portfolio of small and large firms sorted by market capitalization, HML_t is the spread return between a portfolio of high and low book-to-market ratio firms, CMA_t is the spread return between a portfolio of firms with conservative and aggressive investment strategies, and RMW_t is the spread return between a portfolio of firms with robust and weak profitability. $\varepsilon_{kt} \sim (0, \sigma_\varepsilon^2)$ is a zero-mean residual, and coefficients $\{b_k, s_k, h_k, i_k, p_k\}$ are the FF5 factor loadings ranked on $|\alpha_{Fk}|$.

2. The Hou et al. (2015) method of regressing characteristic portfolio excess returns on Q4 factors.

   $$r_{kt} - r_{ft} = \alpha_{Qk} + b_k(r_{mt} - r_{ft}) + s_k(SMB_t) + i_k(I2A_t) + p_k(ROE_t) + \varepsilon_{kt},$$

   where $I2A_t$ is the spread return between a portfolio of low and high investment stocks, ROE_t is the spread return between a portfolio of high and low profitability (i.e., return on equity) firms. The coefficients $\{b_k, s_k, i_k, p_k\}$ are loadings on Q4 factors, whose characteristics are ranked on $|\alpha_{Qk}|$.

In the preliminary analysis, we considered additional feature selection approaches

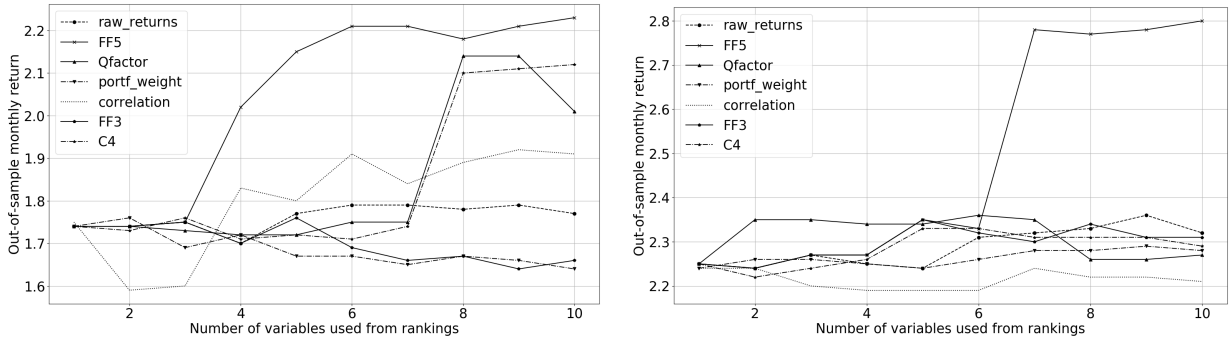
based on industry norms (e.g., spread return, correlation, and mean-variance efficient). These approaches aim to identify characteristics from K that are likely to be associated with large (positive or negative) returns over time. In the next section, we provide some in-sample results that show that these industry approaches are inferior to FF5 and Q4. Furthermore, our paper does not aim to evaluate competing feature selection methods for combining ML model forecasts. Hence, we do not consider them in the out-of-sample analysis.

3.2.2. Optimal number of anomalous characteristics in θ_{1998}

We can use FF5 and Q4 factors to identify training-sample anomalies in K. However, the optimal number of anomalies to include in θ_{1998} is an empirical choice.

We considered between 7 and 10 characteristics based on the following justifications. First, if θ_{1998} contain too many characteristics, it could introduce more noise than information when combining return forecasts across trained ML models. Second, because all characteristics in K were published by 2016, there should be only a small number of anomalies (if any) against FF5 or Q4 factors. Third, Stambaugh and Yuan (2017) highlighted that the number of characteristics that can parsimoniously explain stock returns is typically small. Entrenched factor models (e.g., FF3/FF5/FF6, C4, and Q4/Q5) have between three and six variables. Even the early generation factor models of Chen et al. (1986) and Chan and Chen (1991) had two-to-four variables.

In Figure 2a, we plotted the ML portfolio average monthly return, $\bar{r}_{ML,t}$, against the number of characteristics in θ_{1998} for the various identification approaches. The graphs confirm that ML portfolios that relied on FF5 or Q4 to identify that θ_{1998} generate greater $\bar{r}_{ML,t}$ compared with using industry norms. The FF5 approach yielded the highest average monthly return, which stabilized at 2.2% for 6 to 10 characteristics in θ_{1998}. Next is Q4, with a peak $\bar{r}_{ML,t}$ of 2.15% corresponding to eight characteristics in θ_{1998}. Third is C4, which stabilized at eight characteristics at $\bar{r}_{ML,t} = 2.1\%$. A robustness check based on subsample partitioning (Figure 2b) confirms that $\bar{r}_{ML,t}$ did not increase from expanding θ_{1998} to beyond eight characteristics. We used seven characteristics for all ranking approaches for the remaining numerical results in this paper.



(a) **Training: 1980–1998, Testing: 1998–2016** (b) **Training: 1980–1994, Testing: 1994–2006**

Figure 2: ML VW portfolio out-of-sample average monthly return against the number of anomalies in θ_{1998}.

We plotted the out-of-sample ML portfolio average monthly returns against the number of characteristics in θ_{1998} for various identification approaches: Raw returns, FF5, Q4, Portfolio weight, Correlation, FF3, and C4. The characteristics were identified based on training samples (a) 1980–1998 and (b) 1980–1994.

3.3. Comparison of ML models

We evaluated the out-of-sample monthly stock-return prediction performance results of various ML models using linear regression, ET, LightGBM, and MLA. To better compare existing studies, we report the prediction performance of the NN3 model in GKX (2020), which is an ensemble of 10 neural network models. For all models, the return prediction follows the procedure outlined in Algorithm 1. In Table 1, we report the monthly out-of-sample percentage, R^2 or R^2_{oos}, for various trained ML models. The "All" row refers to all stocks; the rows "Top 1,000" and "Bottom 1000" report the R^2_{oos} performance on the long and short sides of the characteristic portfolios. In contrast, GKX (2020) reports R^2_{oos} for the top and bottom 1,000 stocks by market value.

Table 1: Monthly out-of-sample stock-return prediction performance (percentage R^2_{oos}).

We report monthly R^2_{oos} for stock-return predictions using linear regression, ET, LightGBM, NN3, a variant of NN3 (NN3-FD), and MLA. To show model performance on the long and short sides of the characteristic portfolios, we also separately report the R^2_{oos} for the forecasted top-1,000 stocks and bottom-1,000 stocks by various ML models.

	LR	ET	LightGBM	NN3	NN3-FD	MLA
All	0.21	0.18	0.55	0.38	0.35	0.55
Top 1,000	0.83	0.44	1.02	0.87	0.88	1.03
Bottom 1,000	0.11	0.13	0.47	0.22	0.21	0.47

We used TensorFlow 2.12 and Python 3.11 for training the neural network models. We

fixed some of the hyperparameters of the LightGBM model, whereas the rest were fine-tuned using the validation data, which consisted of 20% of the training set. Specifically, we focused on the following key hyperparameters: "num_leaf," representing the maximum number of leaves in one decision tree, "feature_fraction," involving the random selection of a subset of features on each decision tree, and "bagging_fraction," which pertains to the random selection of a specific part of data without resampling. Notably, "num_leaf" was set to 70, "bagging_fraction" was fixed at 0.7, and we experimented with different values (0.2, 0.6, and 0.8) for "feature_fraction" (refer to Stage 2 of Algorithm 1). For ET, we fixed "n_estimators" at 40 and set "max_depth" to four. There were no hyperparameters for the linear regression models, as they consider only two characteristics.

For NN3, we used the hyperparameters provided in the Internet Appendix (Table A.5) of GKX (2020), with the exception of replacing $l1$-regularization with dropout regularization as in Chen et al. (2024), who reported that dropout regularization has better performance compared to conventional $l1/l2$-regularization. The difference between NN3 and NN3-FD columns in Table 1 pertains to our search method for optimal dropout rates for three hidden layers using TensorFlow's Hyperband tuner (Li & Jamieson, 2018) for each testing year separately for NN3. We also used a fixed dropout rate of 0.05 and a learning rate of 0.001 following Chen et al. (2024) for NN3-FD. A dropout rate of 0.05 in TensorFlow 2.12 is equivalent to the suggested dropout retention probability of 0.95 by Chen et al. (2024) in TensorFlow 1.1. As it easily takes weeks to search for optimal dropout rates for three NN3 hidden layers, we used a tick size of 0.01 for the dropout rate domain search. The performance between NN3 and NN3-FD was very small, so we used NN3-FD for further analysis instead of NN3.

There was a noticeable performance difference between long and short sides of all ML portfolios, as listed in Table 1. This aligns with Table 1 of GKX (2020) and the observations of Avramov et al. (2023), where long positions generate a significant economically larger payoff than the short position. Table 1 shows that LightGBM and MLA performed similarly and better than other ML models when predicting individual stock returns. As we argued in Section 2.2, it is inappropriate to choose ML models based on the R-value of different trained models on training and testing data, as it would introduce a look-ahead bias.

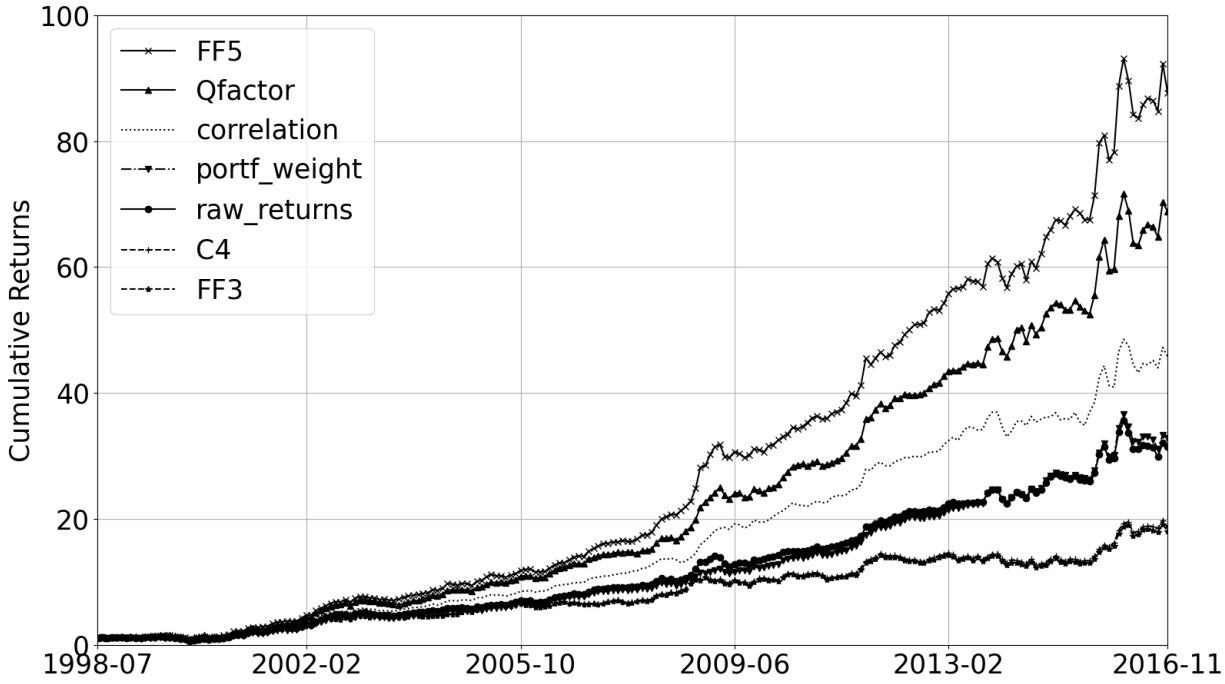


Figure 3: Out-of-sample cumulative returns for ML portfolios based on θ_{1998} identified by various ranking approaches.

Over the 1998–2016 testing period, we plotted the VW cumulative returns for ML portfolios formed based on θ_{1998} identified by various ranking approaches.

3.4. Out-of-sample long-short portfolio cumulative returns for different ranking approaches

Figure 3 plots the ML portfolio cumulative returns over the testing period for the different approaches to identify θ_{1998}. The plots confirm that using FF5 and Q4 to identify θ_{1998} lead to higher cumulative returns than industry-norm approaches. This is consistent with our finding in the Online Appendix Table A3, describing the ML portfolio average returns of FF5 and Q4. In summary, our analysis shows that factor models help to identify a small set of important features for stacking. Additionally, adding features from an exhaustive search of K yields only marginal improvements in $\bar{r}_{ML,t}$.

3.4.1. Do training-sample anomalies persist during the testing period?

To prevent the look-ahead bias, we identified θ_{1998} using the training period ending in June 1998. Based on the confirmed stylized fact, we included in θ_{1998} seven-to-eight top ranked characteristics with the largest $|\alpha|$ against FF5 or Q4 factors. Table 2 lists the training-sample anomalies in θ_{1998} against FF5 and Q4 factors. Apart from momentum (mom12) and change in daily average turnover volume (turnover-d) from FF5 and book-to-market ratio (beme) from Q4, all other characteristics had significantly negative αs.

The identified θ_{1998} were robust to subsample partitioning, albeit producing a slightly different ranking order. Five characteristics were anomalous to both FF5 and Q4 factors, including idiosyncratic volatility (ivol; Ang et al. (2006)), maximum daily return per month (max; Bali et al. (2011)), change in illiquidity (amihud-d; Amihud (2002)), month-end closing price (price), and previous months return (reversal).

The test column reports each characteristic's α and t-stat based on the 1998–2016 testing sample. It shows that many training-sample anomalies subsequently lost their significant α during the testing period. This is consistent with Mclean and Pontiff (2016), who documented a postpublication decline in characteristic spread returns. Indeed, most K characteristics in our paper were published after 1998. For the FF5 approach, four of seven characteristics in θ_{1998} (i.e., ivol, max, price, and reversal) lost their significant α for the testing period. For the other three, the magnitude of t-stat changed substantially from 6.04 to 1.78 for mom12, -7.40 to -3.92 for amihud-d, and -5.77 to -3.88 for turnover-d. The Q4 anomalies were similarly described, with six of eight characteristics losing their significant α in the testing sample. The BM-ratio and Amihud illiquidity both retained a significant α, but with lower t-stats of 1.67 and -3.92, respectively. The subsample partition revealed a similar decline in training-sample anomalies over the testing period.

The last row of Table 2 reports the ML portfolio's α_{MLA} against FF5 and Q4 for both sampling partitions. Across the four estimations, α_{MLA} ranged from 2.14 to 2.74% per month, with t-stats between 4.12$\sim$5.46. The smallest α_{MLA} of 2.14% was larger than the momentum portfolio's α=1.88%. The latter is the global maximal α from 848 estimated αs associated with Table 2 results. We single-sorted on the level and change in $K = 106$ characteristics to form 212 spread portfolios. Each characteristic portfolio was regressed against FF5 and Q4 factors, which produced a total of 424 estimated α. Because we considered both full-sample and subsample partitions, there were 848 estimated α in total, of which momentum's $\alpha = 1.88$ was the global maximum. The results show that although the ML portfolio construction was based on a progressively outdated θ_{1998} over the 18-year period, the α_{MLA} remained significantly positive against FF5 and Q4 factors in both sampling partitions. This suggests that θ_{1998} are not important sources of α_{MLA}.

Table 2: Training-sample anomalies in θ_{1998}, identified using FF5 and Q4 factors based on full- and subsample partitioning.

The table reports the estimated α (t-stat) of the top seven-to-eight anomalies to the FF5 and Q4 models based on the training and testing-sample periods. The anomalies were identified and ranked based on the training sample, after which we conducted the same estimation over the testing period. Comparing the two sets of results allowed us to gauge whether the training-sample anomalies remained anomalous during the testing period.

	Training: 1980–1998, Testing: 1998–2016					**Training: 1980–1994, Testing: 1994–2006**				
Apr.	**List**	**Train**		**Test**		**List**	**Train**		**Test**	
		Alpha	***t*-stat**	**Alpha**	***t*-stat**		**Alpha**	***t*-stat**	**Alpha**	***t*-stat**
FF5	mom12	1.88***	(6.04)	0.95*	(1.78)	mom12	1.60***	(4.61)	2.17***	(3.38)
	ivol	-1.08***	(-6.26)	-0.57	(-1.02)	ivol	-1.32***	(-6.77)	-0.73	(-0.91)
	max	-1.03***	(-5.72)	-0.60	(-1.12)	max	-1.27***	(-6.07)	-0.90	(-1.18)
	amihud_d	-0.99***	(-7.40)	-0.75***	(-3.92)	dltis	-1.12***	(-2.98)	-0.22	(-0.50)
	price_d	-0.94***	(-4.67)	-0.36	(-0.98)	price_d	-1.09***	(-4.55)	-0.63	(-1.25)
	turnover_d	0.86***	(5.77)	0.84***	(3.88)	reversal	-1.06***	(-3.66)	-0.58	(-1.08)
	reversal	-0.85***	(-3.36)	-0.34	(-0.88)	amihud_d	-0.80***	(-4.95)	-1.26***	(-5.00)
	ML Portfolio			2.21***	(5.46)	ML Portfolio			2.74***	(4.53)
Q4	reversal	-1.42***	(-4.62)	-0.34	(-0.88)	reversal	-1.73***	(-4.89)	-0.58	(-1.08)
	price_d	-1.34***	(-5.26)	-0.36	(-0.98)	price_d	-1.54***	(-5.17)	-0.63	(-1.25)
	beme	1.29***	(6.33)	0.69*	(1.67)	max	-1.47***	(-5.82)	-0.90	(-1.18)
	max	-1.22***	(-5.50)	-0.60	(-1.12)	ivol	-1.28***	(-5.08)	-0.73	(-0.91)
	ivol	-1.08***	(-4.93)	-0.57	(-1.02)	beme	1.19***	(5.06)	1.29**	(2.24)
	m2b	-1.07***	(-4.66)	-0.68	(-1.49)	m2b	-1.00***	(-3.53)	-1.24*	(-1.91)
	tobin_q	-1.04***	(-4.77)	-0.71	(-1.51)	tobin_q	-0.97***	(-3.73)	-1.28*	(-1.90)
	amihud_d	-1.03***	(-6.78)	-0.75***	(-3.92)	me_d	-0.96***	(-2.88)	0.12	(0.24)
	ML Portfolio			2.14***	(4.93)	ML Portfolio			2.26***	(4.12)

4. ML Portfolio Analysis

In this section, we answer three questions about the ML portfolio analysis: "Is the α_{MLA} significant against entrenched factor models?" "What is the source of α_{MLA}?" and "What are the dominant characteristics in the ML portfolio over the testing period?" Recall that we have two ML portfolios corresponding to FF5 and Q4 approaches to identify θ_{1998}. However, because the main findings were similar, we focus our discussion on the FF5 approach.

4.1. Significant α_{MLA} everywhere

We evaluated the ML portfolio against entrenched factor models (i.e., FF3, C4, FF5, and Q4). For additional insights, in Appendix A4, we report the regression results of the ML portfolio excess returns against FF6 (2018) and Q5 (2021) factors. The α_{MLA} remained significantly positive against factor models published after the end of the testing period. The interaction among four factor models, two sets of θ_{1998} and equal/VW portfolio returns $r_{ml,t}$, generated a total of 16 portfolio regression results. Tables 3 and 4 report the estimated factor loading and t-stat for decile portfolios based on FF5- and Q4-identified θ_{1998}. Each table has two panels corresponding to equally- and VW $r_{ml,t}$. Across both tables, the main findings are consistent between the two panels; hence, we discuss mainly Panel A results.

The portfolio analysis revealed significant α_{MLA} everywhere. All 16 ML portfolio regressions produced α_{MLA} that were highly significantly positive. The magnitude of annualized α_{MLA} ranged from 17 to 29%, with t-stats ranging from 2.97 to 9.72. On the long side, all 16 PW portfolios displayed positive α at 95% significance. On the short side, 13 PL portfolio α values were significantly negative at the 95% level, with another two significantly negative at the 90% level. Most importantly, both the magnitude and t-stat of α increased monotonically from the PL to PW portfolio. This shows clear evidence that the sorting criterion (i.e., model-combined stock-return forecast) had a strong anomalous return pattern versus all factor models.

In the first row of Table 3, the average return and t-stat both increased from the PL to PW portfolio. The PW portfolio had a significant average monthly return of 1.85%, and the PL portfolio had an average monthly return of -0.36%, which is insignificant. This is likely due to short-sale constraints on PL stocks. In the last column, the ML

portfolio had the same spread return (t-stat) of 2.21% (5.46) as reported in Table 2. The α_{MLA} was significantly positive against all entrenched factor models, ranging from 1.86% for FF5, 2.03% for Q4, and 2.48% for FF3 and C4. The t-stats ranged between 6.88 and 9.72. For all models, the magnitude and t-stat of α_{MLA} both increased monotonically from the PL to PW portfolio.

Against FF3 and C4 factors, $r_{ml,t}$ loaded negatively on MKT and SMB (i.e., the ML portfolio took negative bets on beta and size). Both PL and PW portfolios loaded positively on the market and size premia. However, because the loading on the PL exceeded that of the PW, the net loading was negative. The ML portfolio loaded positively on HML, which came from the PW portfolio chasing value stocks and the PL portfolio loading-up on growth stocks. Against C4, MOM was not significant in the ML portfolio. Both PL and PW portfolios loaded negatively on MOM. In fact, all decile portfolios loaded negatively on MOM.

Against Q4 and FF5 factors, the ML portfolio loadings on MKT and SMB were similar to those for FF3 and C4, albeit less pronounced. The loadings on MKT were insignificant (i.e., market-neutral). The loadings on SMB remained significantly negative, albeit at a smaller magnitude. In FF5, HML was no longer significant. Instead, the ML portfolio loaded positively on both profitability (RMW) and investment (CMA) factors. It also loaded positively on the corresponding ROE and I2A factors in Q4. All four loadings were highly significant. Estimates from PL and PW showed that the loadings of PW were insignificant on RMW, CMA, and I2A factors. The PW loading on ROE (-0.18) was nearly significant at the 5% level, with a t-stat of -1.97. The significance of the ML portfolio loadings on investment and profitability factors stemmed from the PL portfolio, which loaded negatively on RMW, CMA, I2A, and ROE. This suggests that the MLA can predict low returns for firms with weak profitability and/or aggressive investment policies. However, the PW portfolio is either insignificant, or it also loads negatively on profitability and investment factors.

Table 4 evaluates $r_{ml,t}$ based on θ_{1998} identified by Q4 factors. As with Table 3, the main results were consistent across equal- and VW $r_{ml,t}$. Hence, we focus our discussion on Panel A. The average return and t-stat also increased monotonically from the PL to the PW portfolio. Like the FF5 approach, the PW portfolio had a significant average return of 1.86%, but the PL return of -0.28% was insignificant. The ML portfolio had a

significant spread return (t-stat) of 2.14% (4.93), as well as a significantly positive α_{MLA} against all factor models, ranging from 1.71% for FF5, 1.84% for Q4, and 2.37% for both FF3 and C4. The t-stats ranged from 5.86 to 8.73. Across the four models, the magnitude and t-stat of α both increased monotonically from the PL to PW portfolio.

Against FF3 and C4 factors, $r_{ml,t}$ loaded negatively on MKT and SMB (i.e., the ML portfolio took negative bets on both beta and size). Both PL and PW portfolios loaded positively on the market and size premia. However, as the loading on the PL portfolio those of the PW portfolio, the net loading of the ML portfolio was negative. The ML portfolio loaded positively on HML, which came from the PW portfolio chasing value stocks, whereas the PL portfolio loaded-up on growth stocks. Against C4, MOM was not significant in the ML portfolio, and both PL and PW portfolios loaded negatively on MOM.

Against the FF5 and Q4 factors, the ML portfolio loading on MKT and SMB were similarly described for FF3 and C4, albeit less pronounced. The loading on MKT was insignificant (i.e., market-neutral). The loading on SMB remained significantly negative, but with a smaller magnitude. In FF5, the ML portfolio loaded positively on HML, RMW (profitability), and CMA (investment) factors. It also loaded positively on the corresponding ROE and I2A factors of Q4. All these loadings were highly significant. Separate results based on PL and PW showed that the significance was driven mainly by the PL portfolio, which exhibited significantly negative loading on HML, RMW, CMA, for FF5, and on I2A and ROE for Q4. In contrast, the PW portfolio loaded significantly only on RMW. This suggests that the MLA is good at predicting lower return for firms with weak profitability and/or aggressive investment policies. However, it is less able to predict higher returns for firms with strong profitability and/or conservative investment policies.

4.2. Potential source(s) of α_{MLA}

Given that the ML portfolio was generated from ML models trained on a large K of published characteristics, it is not surprising for the ML portfolio to outperform entrenched factor models. Because our aim is to ascertain the source of significant α_{MLA} everywhere, we conjecture that the ML portfolio is likely to exhibit a time-varying characteristic exposure during the testing period. By definition, a given factor model is static

in the characteristic domain, which makes it awkward to track and explain return over time in an ML portfolio with changing characteristics. To test this conjecture, we considered two alternative benchmark zoo factors ($K1$ and $K2$) that were designed to beat the ML portfolio.

Table 3: ML portfolio loading on factor models (i.e., FF3, C4, FF5, and Q4) over the 1998–2016 testing period.

At the end of each month, t, we decile-sort stocks on their month t+1 predicted return from the MLA, based on θ_{1998} identified using FF5 model. The table reports equally (Panel a) and VW (Panel b) average returns for the decile portfolios, and the ML portfolio that is long in the PW and short in the PL. Portfolio returns were regressed against Fama–French-3 factors, Carhart-4 factors, Fama–French-5 factors, and Q4 factors. The estimated factor loadings are reported with Newey–West-adjusted $t-$stat in parentheses. *, **, and *** reflect significance at the 10, 5, and 1% levels, respectively.

(a) Equal-weighted

Model	Factor	PL	2	3	4	5	6	7	8	9	PW	ML portfolio
Average ret	Ret	-0.36	0.40	0.55	0.80	0.85	0.98	1.07	1.25	1.38	1.85	2.21***
		(-0.52)	(0.75)	(1.22)	(2.05)	(2.29)	(2.81)	(3.10)	(3.54)	(3.79)	(4.29)	(5.46)
FF3	Alpha	-1.50	-0.58	-0.35	-0.03	0.07	0.22	0.31	0.48	0.59	0.98	2.48***
		(-8.14)	(-4.61)	(-3.19)	(-0.33)	(0.70)	(2.31)	(2.95)	(4.22)	(5.52)	(6.61)	(9.27)
	MKT	1.45	1.23	1.10	1.01	0.95	0.92	0.91	0.92	0.93	1.05	-0.40***
		(24.26)	(27.45)	(31.98)	(31.40)	(35.02)	(33.39)	(28.33)	(26.69)	(27.85)	(26.67)	(-4.71)
	SMB	1.35	0.97	0.79	0.59	0.47	0.40	0.41	0.43	0.49	0.68	-0.67***
		(12.93)	(22.29)	(16.00)	(9.74)	(7.03)	(4.75)	(4.63)	(4.38)	(5.16)	(6.25)	(-3.38)
	HML	-0.35	-0.07	0.09	0.21	0.31	0.36	0.37	0.37	0.33	0.20	0.55***
		(-4.04)	(-1.06)	(1.77)	(4.36)	(6.34)	(6.85)	(5.86)	(5.42)	(5.37)	(3.25)	(4.21)
C4	Alpha	-1.43	-0.50	-0.30	0.00	0.07	0.24	0.33	0.50	0.62	1.05	2.48***
		(-7.52)	(-4.09)	(-2.70)	(0.05)	(0.74)	(2.57)	(3.05)	(4.43)	(5.74)	(6.93)	(8.97)
	MKT	1.39	1.16	1.05	0.98	0.94	0.90	0.89	0.90	0.91	0.99	-0.40***
		(23.55)	(25.66)	(26.46)	(26.40)	(31.97)	(30.37)	(26.54)	(24.43)	(26.38)	(22.97)	(-4.59)
	SMB	1.38	1.01	0.81	0.61	0.47	0.41	0.42	0.45	0.50	0.71	-0.67***
		(12.61)	(22.28)	(17.73)	(11.07)	(7.31)	(5.31)	(5.15)	(4.92)	(5.66)	(7.15)	(-3.44)
	HML	-0.40	-0.12	0.05	0.19	0.31	0.35	0.36	0.35	0.31	0.15	0.55***
		(-4.95)	(-2.71)	(1.19)	(4.07)	(6.64)	(6.85)	(5.59)	(5.13)	(5.23)	(2.52)	(4.27)
	MOM	-0.13	-0.15	-0.11	-0.07	-0.01	-0.04	-0.04	-0.05	-0.05	-0.13	0.00
		(-3.24)	(-4.63)	(-4.27)	(-3.43)	(-0.35)	(-1.70)	(-1.19)	(-1.52)	(-1.64)	(-3.14)	(0.04)
FF5	Alpha	-0.98	-0.34	-0.27	-0.11	-0.05	0.06	0.15	0.29	0.44	0.88	1.86***
		(-7.66)	(-2.92)	(-2.18)	(-1.16)	(-0.56)	(0.75)	(1.71)	(3.25)	(4.55)	(6.44)	(9.72)
	MKT	1.16	1.09	1.05	1.04	1.00	0.99	0.98	1.01	1.01	1.09	-0.07
		(28.95)	(30.47)	(30.75)	(32.67)	(35.40)	(41.31)	(36.32)	(32.98)	(31.61)	(23.48)	(-1.04)
	SMB	1.07	0.87	0.77	0.67	0.55	0.53	0.55	0.57	0.63	0.74	-0.33***
		(18.79)	(14.72)	(14.92)	(14.66)	(10.70)	(10.14)	(10.02)	(9.29)	(11.05)	(8.52)	(-3.03)
	HML	-0.13	-0.02	0.03	0.04	0.12	0.16	0.18	0.13	0.11	-0.01	0.12
		(-2.27)	(-0.35)	(0.53)	(0.69)	(2.09)	(3.04)	(2.76)	(1.80)	(1.89)	(-0.18)	(1.15)
	RMW	-0.94	-0.40	-0.14	0.14	0.19	0.29	0.30	0.32	0.29	0.12	1.06***
		(-12.60)	(-5.96)	(-2.23)	(2.77)	(4.36)	(7.25)	(5.87)	(6.45)	(5.20)	(1.60)	(8.91)
	CMA	-0.36	-0.24	-0.12	0.00	0.05	0.03	-0.00	0.09	0.02	0.08	0.44***
		(-3.75)	(-2.68)	(-1.24)	(0.07)	(0.85)	(0.58)	(-0.01)	(1.53)	(0.27)	(0.62)	(2.67)
Q4	Alpha	-0.84	-0.20	-0.14	0.02	0.02	0.13	0.22	0.37	0.56	1.19	2.03***
		(-4.34)	(-1.48)	(-1.04)	(0.18)	(0.17)	(1.07)	(1.85)	(3.01)	(4.53)	(6.95)	(6.88)
	MKT	1.05	0.99	0.98	0.98	0.98	0.97	0.96	0.98	0.97	1.01	-0.04
		(16.56)	(23.00)	(22.07)	(20.38)	(25.58)	(21.60)	(19.72)	(18.91)	(19.27)	(16.62)	(-0.35)
	SMB	1.12	0.85	0.72	0.56	0.47	0.40	0.43	0.44	0.48	0.60	-0.52**
		(10.94)	(19.43)	(14.58)	(8.58)	(6.26)	(4.08)	(4.16)	(4.06)	(4.61)	(5.21)	(-2.49)
	I2A	-0.77	-0.37	-0.14	0.10	0.24	0.31	0.30	0.35	0.26	0.12	0.89***
		(-8.08)	(-4.64)	(-1.83)	(1.73)	(4.25)	(5.02)	(4.20)	(5.44)	(4.18)	(1.48)	(7.43)
	ROE	-0.88	-0.52	-0.29	-0.07	0.07	0.10	0.13	0.11	0.07	-0.18	0.69***
		(-8.45)	(-8.54)	(-5.52)	(-1.36)	(1.38)	(1.69)	(1.63)	(1.74)	(1.06)	(-2.06)	(4.41)

Table 3: (continued)

b) VW

Model	Factor	PL	2	3	4	5	6	7	8	9	PW	ML Portfolio
Average ret	Ret	-0.23	0.39	0.24	0.40	0.44	0.51	0.70	0.88	1.07	1.57	1.80***
		(-0.32)	(0.72)	(0.50)	(1.06)	(1.33)	(1.78)	(2.40)	(2.94)	(3.26)	(3.69)	(3.70)
FF3	Alpha	-1.20	-0.38	-0.44	-0.19	-0.14	-0.01	0.19	0.35	0.48	0.89	2.08***
		(-5.00)	(-2.20)	(-2.52)	(-1.28)	(-1.11)	(-0.10)	(1.60)	(3.24)	(3.14)	(3.58)	(5.70)
	MKT	1.64	1.28	1.17	1.03	0.96	0.88	0.81	0.88	0.97	1.13	-0.51***
		(19.13)	(18.99)	(22.82)	(31.40)	(29.94)	(44.66)	(22.85)	(25.54)	(19.13)	(17.15)	(-4.09)
	SMB	0.70	0.44	0.28	-0.02	-0.05	-0.17	-0.10	-0.06	0.08	0.25	-0.45
		(4.61)	(5.12)	(4.55)	(-0.47)	(-1.33)	(-4.27)	(-3.14)	(-1.36)	(1.02)	(1.55)	(-1.56)
	HML	-0.69	-0.47	-0.40	-0.13	0.05	0.10	0.07	-0.03	-0.11	-0.25	0.44**
		(-5.73)	(-3.93)	(-3.77)	(-1.64)	(1.20)	(2.71)	(1.05)	(-0.46)	(-1.76)	(-2.52)	(2.37)
C4	Alpha	-1.12	-0.28	-0.41	-0.16	-0.15	-0.01	0.17	0.34	0.47	0.96	2.08***
		(-4.67)	(-1.66)	(-2.30)	(-1.13)	(-1.21)	(-0.06)	(1.32)	(2.90)	(2.85)	(3.75)	(5.33)
	MKT	1.58	1.19	1.14	1.01	0.98	0.88	0.83	0.89	0.97	1.06	-0.51***
		(18.05)	(19.95)	(20.54)	(27.37)	(27.22)	(38.81)	(20.75)	(28.06)	(16.53)	(13.34)	(-3.57)
	SMB	0.73	0.49	0.29	-0.01	-0.06	-0.17	-0.11	-0.07	0.08	0.28	-0.45
		(4.62)	(5.17)	(4.64)	(-0.19)	(-1.38)	(-4.45)	(-3.48)	(-1.46)	(0.85)	(1.85)	(-1.54)
	HML	-0.74	-0.54	-0.42	-0.15	0.06	0.10	0.08	-0.02	-0.10	-0.31	0.44**
		(-6.64)	(-5.45)	(-4.12)	(-2.03)	(1.44)	(2.58)	(1.32)	(-0.27)	(-1.91)	(-2.84)	(2.35)
	MOM	-0.14	-0.19	-0.07	-0.05	0.03	-0.01	0.04	0.03	0.02	-0.14	-0.00
		(-2.20)	(-3.74)	(-1.56)	(-1.67)	(1.06)	(-0.31)	(1.88)	(0.63)	(0.25)	(-1.64)	(-0.03)
FF5	Alpha	-0.66	-0.07	-0.23	-0.25	-0.21	-0.15	0.04	0.23	0.41	0.91	1.57***
		(-3.00)	(-0.42)	(-1.31)	(-1.77)	(-1.55)	(-1.43)	(0.32)	(2.22)	(2.69)	(3.38)	(4.03)
	MKT	1.36	1.12	1.06	1.06	1.00	0.96	0.89	0.94	1.00	1.12	-0.24*
		(16.93)	(19.36)	(21.55)	(26.39)	(29.81)	(34.77)	(30.80)	(30.36)	(21.60)	(15.20)	(-1.83)
	SMB	0.43	0.29	0.16	-0.04	-0.00	-0.10	-0.02	-0.02	0.18	0.31	-0.13
		(3.78)	(3.21)	(2.12)	(-0.72)	(-0.10)	(-2.69)	(-0.48)	(-0.46)	(2.29)	(2.02)	(-0.53)
	HML	-0.32	-0.27	-0.27	-0.21	-0.00	0.01	-0.06	-0.14	-0.14	-0.24	0.08
		(-2.84)	(-2.49)	(-3.14)	(-2.21)	(-0.03)	(0.11)	(-1.51)	(-2.23)	(-2.13)	(-2.10)	(0.46)
	RMW	-0.87	-0.50	-0.36	0.01	0.13	0.23	0.24	0.15	0.23	0.08	0.95***
		(-7.84)	(-5.12)	(-3.64)	(0.16)	(2.30)	(5.01)	(2.97)	(2.33)	(2.78)	(0.54)	(4.73)
	CMA	-0.46	-0.27	-0.13	0.21	0.03	0.11	0.15	0.18	-0.17	-0.23	0.23
		(-2.54)	(-1.89)	(-1.01)	(1.70)	(0.46)	(1.49)	(1.93)	(1.88)	(-1.63)	(-0.99)	(0.67)
Q4	Alpha	-0.56	0.07	-0.08	-0.14	-0.13	-0.12	0.07	0.29	0.51	1.26	1.82***
		(-1.83)	(0.34)	(-0.37)	(-0.83)	(-0.83)	(-1.09)	(0.55)	(2.16)	(2.58)	(4.03)	(3.76)
	MKT	1.31	1.05	0.98	1.02	1.00	0.95	0.90	0.91	0.96	1.02	-0.29
		(10.41)	(14.40)	(18.43)	(23.81)	(23.63)	(36.32)	(29.05)	(29.17)	(18.25)	(12.23)	(-1.60)
	SMB	0.53	0.37	0.18	-0.06	-0.03	-0.16	-0.03	-0.04	0.13	0.13	-0.40
		(3.05)	(3.46)	(2.68)	(-1.05)	(-0.73)	(-3.30)	(-0.66)	(-0.75)	(1.63)	(0.85)	(-1.32)
	I2A	-1.12	-0.79	-0.63	-0.09	0.05	0.20	0.14	0.08	-0.27	-0.37	0.75***
		(-7.00)	(-4.39)	(-4.28)	(-0.77)	(0.92)	(2.99)	(1.96)	(0.94)	(-2.61)	(-2.25)	(2.93)
	ROE	-0.67	-0.42	-0.39	-0.06	0.11	0.13	0.22	0.06	0.09	-0.26	0.40*
		(-4.24)	(-4.43)	(-4.52)	(-0.84)	(1.93)	(2.37)	(4.10)	(0.83)	(0.85)	(-2.07)	(1.69)

Table 4: ML portfolio loading on factor models (i.e., FF3, C4, FF5, and Q4) over the 1998–2016 testing-sample period.

At the end of each month, t, we decile-sorted stocks on their month, t+1, predicted returns from the MLA based on the θ_{1998} identified using the Q4 model. The table reports equally (Panel a) and VW (Panel b) average returns for the decile portfolios, and the ML portfolio that is long in the PW and short in the PL. Portfolio returns were regressed against Fama–French-3 factors, Carhart-4 factors, Fama–French-5 factors, and Q4 factors. The estimated factor loadings are reported with Newey–West-adjusted $t-$stat in parentheses. *, **, and *** reflect significance at the 10, 5, and 1% levels, respectively.

(a) Q4 anomalies and equally-weighted portfolio returns.

Model	Factor	PL	2	3	4	5	6	7	8	9	PW	ML Portfolio
Average ret	Ret	-0.28	0.35	0.72	0.77	0.83	0.94	1.03	1.19	1.36	1.86	2.14***
		(-0.41)	(0.62)	(1.56)	(2.00)	(2.31)	(2.70)	(2.95)	(3.38)	(3.59)	(4.38)	(4.93)
FF3	Alpha	-1.39	-0.64	-0.17	-0.04	0.05	0.17	0.28	0.40	0.54	0.98	2.37***
		(-7.23)	(-5.26)	(-1.56)	(-0.41)	(0.56)	(1.72)	(2.48)	(3.76)	(4.25)	(6.39)	(8.73)
	MKT	1.41	1.27	1.12	1.00	0.95	0.92	0.89	0.93	0.94	1.04	-0.37***
		(21.67)	(28.61)	(30.16)	(28.68)	(32.92)	(29.29)	(27.24)	(28.20)	(26.33)	(26.23)	(-4.14)
	SMB	1.37	0.96	0.77	0.58	0.47	0.41	0.38	0.44	0.53	0.67	-0.71***
		(11.87)	(24.92)	(13.56)	(9.64)	(7.37)	(5.08)	(4.16)	(5.28)	(4.81)	(5.91)	(-3.32)
	HML	-0.46	-0.11	0.04	0.18	0.31	0.36	0.38	0.44	0.42	0.29	0.75***
		(-4.84)	(-1.74)	(0.82)	(3.91)	(6.66)	(6.51)	(5.71)	(6.16)	(5.67)	(4.18)	(5.29)
C4	Alpha	-1.31	-0.59	-0.12	-0.01	0.07	0.18	0.30	0.42	0.59	1.06	2.37***
		(-6.65)	(-4.61)	(-1.16)	(-0.13)	(0.72)	(1.75)	(2.69)	(3.89)	(4.72)	(6.95)	(8.49)
	MKT	1.34	1.23	1.07	0.97	0.93	0.91	0.87	0.91	0.90	0.96	-0.37***
		(21.90)	(26.38)	(25.04)	(24.77)	(30.06)	(28.08)	(24.72)	(25.98)	(23.05)	(22.76)	(-4.16)
	SMB	1.42	0.98	0.79	0.59	0.48	0.41	0.39	0.45	0.55	0.71	-0.71***
		(11.56)	(23.99)	(14.96)	(10.66)	(8.00)	(5.37)	(4.63)	(5.92)	(5.60)	(7.11)	(-3.38)
	HML	-0.52	-0.14	0.00	0.16	0.29	0.35	0.36	0.42	0.39	0.23	0.75***
		(-6.17)	(-2.63)	(0.03)	(3.77)	(6.46)	(6.93)	(5.55)	(5.89)	(5.22)	(3.37)	(5.41)
	MOM	-0.16	-0.09	-0.10	-0.05	-0.03	-0.02	-0.04	-0.04	-0.09	-0.16	-0.00
		(-3.55)	(-2.82)	(-4.66)	(-2.38)	(-1.25)	(-0.64)	(-1.46)	(-1.09)	(-2.97)	(-3.86)	(-0.01)
FF5	Alpha	-0.83	-0.40	-0.09	-0.08	-0.03	0.00	0.06	0.22	0.34	0.87	1.71***
		(-6.24)	(-3.61)	(-0.77)	(-0.82)	(-0.34)	(0.05)	(0.64)	(2.75)	(3.23)	(5.85)	(8.32)
	MKT	1.11	1.14	1.07	1.01	0.98	1.00	1.00	1.01	1.03	1.08	-0.03
		(24.99)	(30.76)	(29.61)	(32.61)	(37.20)	(35.72)	(36.85)	(40.84)	(31.48)	(22.56)	(-0.41)
	SMB	1.07	0.86	0.76	0.64	0.56	0.53	0.52	0.58	0.68	0.76	-0.31***
		(18.73)	(16.00)	(14.32)	(13.15)	(11.93)	(10.55)	(9.14)	(11.23)	(10.42)	(8.61)	(-2.84)
	HML	-0.23	-0.06	-0.01	0.05	0.16	0.14	0.11	0.20	0.16	0.09	0.32***
		(-4.25)	(-1.08)	(-0.19)	(1.00)	(2.92)	(2.50)	(1.67)	(2.65)	(2.05)	(1.05)	(2.96)
	RMW	-1.01	-0.41	-0.11	0.08	0.16	0.29	0.35	0.32	0.34	0.17	1.18***
		(-12.01)	(-6.06)	(-2.03)	(1.68)	(3.06)	(7.96)	(7.65)	(7.51)	(5.25)	(2.12)	(8.78)
	CMA	-0.34	-0.23	-0.17	-0.06	-0.03	0.07	0.15	0.07	0.07	0.02	0.36**
		(-3.48)	(-2.82)	(-2.45)	(-0.82)	(-0.55)	(1.21)	(2.83)	(1.19)	(0.92)	(0.14)	(2.19)
Q4	Alpha	-0.67	-0.23	0.05	-0.01	-0.03	0.09	0.16	0.29	0.51	1.17	1.84***
		(-3.33)	(-1.89)	(0.41)	(-0.07)	(-0.27)	(0.75)	(1.23)	(2.48)	(3.45)	(6.33)	(5.86)
	MKT	0.98	1.04	1.00	0.97	0.97	0.98	0.97	0.99	0.98	1.00	0.02
		(12.95)	(26.12)	(19.77)	(22.30)	(21.98)	(22.83)	(19.76)	(22.49)	(16.43)	(14.83)	(0.16)
	SMB	1.13	0.84	0.70	0.57	0.48	0.41	0.39	0.44	0.50	0.59	-0.53**
		(9.05)	(21.23)	(13.10)	(8.71)	(6.18)	(4.67)	(3.79)	(4.47)	(3.93)	(4.70)	(-2.21)
	I2A	-0.86	-0.41	-0.24	0.01	0.21	0.28	0.39	0.40	0.40	0.21	1.06***
		(-7.10)	(-5.81)	(-4.33)	(0.14)	(3.51)	(4.84)	(6.47)	(5.86)	(5.55)	(2.41)	(6.90)
	ROE	-0.96	-0.52	-0.24	-0.03	0.05	0.14	0.13	0.12	0.03	-0.18	0.78***
		(-8.48)	(-9.82)	(-4.81)	(-0.59)	(0.79)	(2.29)	(1.94)	(1.75)	(0.38)	(-1.80)	(4.31)

Table 4: (continued)

b) Q4 anomalies and VW portfolio returns.

Model	Factor	PL	2	3	4	5	6	7	8	9	PW	ML Portfolio
Average ret	Ret	-0.18	0.29	0.49	0.44	0.38	0.63	0.75	0.84	0.91	1.51	1.69***
		(-0.25)	(0.48)	(1.13)	(1.17)	(1.14)	(2.10)	(2.58)	(2.92)	(2.77)	(3.41)	(3.45)
FF3	Alpha	-1.09	-0.50	-0.19	-0.14	-0.17	0.11	0.23	0.28	0.31	0.80	1.88***
		(-3.90)	(-2.35)	(-1.29)	(-0.84)	(-1.40)	(1.03)	(2.11)	(2.19)	(2.13)	(2.98)	(5.12)
	MKT	1.58	1.37	1.05	1.02	0.94	0.88	0.87	0.87	0.96	1.17	-0.42***
		(16.66)	(22.35)	(27.41)	(24.40)	(24.30)	(35.85)	(25.12)	(19.08)	(23.91)	(17.60)	(-3.06)
	SMB	0.69	0.41	0.34	0.00	-0.13	-0.07	-0.16	-0.02	-0.01	0.24	-0.45
		(4.01)	(4.94)	(6.26)	(0.04)	(-2.49)	(-2.68)	(-4.57)	(-0.51)	(-0.12)	(1.28)	(-1.39)
	HML	-0.85	-0.56	-0.28	-0.22	0.02	-0.03	0.08	0.13	0.05	-0.19	0.66***
		(-6.24)	(-4.57)	(-5.11)	(-2.36)	(0.50)	(-0.65)	(1.34)	(1.69)	(0.66)	(-1.61)	(3.62)
C4	Alpha	-1.02	-0.44	-0.19	-0.11	-0.18	0.09	0.22	0.26	0.28	0.90	1.91***
		(-3.63)	(-2.15)	(-1.24)	(-0.69)	(-1.45)	(0.86)	(1.90)	(2.04)	(1.92)	(2.97)	(4.65)
	MKT	1.52	1.32	1.05	1.00	0.95	0.90	0.87	0.88	0.99	1.08	-0.44**
		(15.75)	(22.66)	(23.14)	(23.17)	(23.97)	(28.94)	(22.99)	(19.09)	(23.94)	(10.95)	(-2.59)
	SMB	0.72	0.43	0.34	0.01	-0.13	-0.08	-0.16	-0.03	-0.02	0.29	-0.43
		(3.99)	(5.66)	(6.03)	(0.28)	(-2.48)	(-3.18)	(-4.68)	(-0.62)	(-0.36)	(1.57)	(-1.31)
	HML	-0.90	-0.60	-0.28	-0.24	0.03	-0.01	0.09	0.14	0.07	-0.26	0.64***
		(-7.16)	(-5.08)	(-5.36)	(-2.64)	(0.68)	(-0.33)	(1.37)	(1.86)	(0.92)	(-2.04)	(3.41)
	MOM	-0.13	-0.10	-0.01	-0.05	0.02	0.03	0.02	0.03	0.06	-0.19	-0.06
		(-2.11)	(-1.60)	(-0.14)	(-1.17)	(0.72)	(1.36)	(0.50)	(0.70)	(1.57)	(-1.65)	(-0.37)
FF5	Alpha	-0.49	-0.14	-0.05	-0.15	-0.27	0.00	0.02	0.18	0.10	0.93	1.43***
		(-1.83)	(-0.65)	(-0.32)	(-1.04)	(-2.13)	(0.02)	(0.16)	(1.67)	(0.74)	(3.17)	(3.61)
	MKT	1.27	1.18	0.98	1.03	0.99	0.94	0.98	0.92	1.07	1.10	-0.18
		(14.86)	(23.03)	(21.46)	(26.70)	(24.34)	(33.67)	(37.25)	(21.89)	(28.82)	(12.61)	(-1.17)
	SMB	0.36	0.29	0.25	0.03	-0.08	-0.02	-0.08	0.06	0.05	0.30	-0.06
		(3.08)	(3.51)	(4.33)	(0.37)	(-1.56)	(-0.41)	(-2.29)	(0.96)	(0.94)	(1.63)	(-0.26)
	HML	-0.45	-0.27	-0.24	-0.23	-0.06	-0.11	-0.11	0.06	-0.18	-0.03	0.42*
		(-3.57)	(-2.14)	(-2.77)	(-3.10)	(-1.03)	(-2.38)	(-2.19)	(0.85)	(-2.63)	(-0.21)	(1.95)
	RMW	-1.01	-0.47	-0.27	0.06	0.16	0.18	0.28	0.22	0.25	0.00	1.01***
		(-8.91)	(-4.86)	(-2.89)	(0.70)	(2.57)	(2.39)	(4.53)	(2.71)	(3.60)	(0.01)	(5.10)
	CMA	-0.42	-0.50	-0.04	-0.04	0.11	0.08	0.30	-0.02	0.34	-0.51	-0.09
		(-1.85)	(-3.52)	(-0.47)	(-0.38)	(1.73)	(1.15)	(3.90)	(-0.15)	(3.39)	(-1.91)	(-0.24)
Q4	Alpha	-0.40	0.09	0.15	-0.04	-0.21	0.02	0.05	0.24	0.20	1.18	1.59***
		(-1.08)	(0.35)	(0.91)	(-0.20)	(-1.51)	(0.18)	(0.41)	(1.68)	(1.21)	(3.24)	(2.97)
	MKT	1.21	1.13	0.93	0.99	0.98	0.94	0.95	0.92	1.02	1.03	-0.18
		(8.47)	(18.09)	(17.06)	(18.93)	(20.25)	(34.50)	(29.91)	(25.19)	(23.26)	(9.59)	(-0.79)
	SMB	0.50	0.36	0.29	0.01	-0.13	-0.02	-0.12	0.02	0.01	0.13	-0.37
		(2.35)	(4.20)	(4.53)	(0.16)	(-2.59)	(-0.39)	(-3.33)	(0.26)	(0.07)	(0.72)	(-1.05)
	I2A	-1.21	-1.04	-0.51	-0.33	0.12	0.01	0.24	0.09	0.22	-0.41	0.79***
		(-6.11)	(-5.84)	(-6.05)	(-2.31)	(1.91)	(0.19)	(3.61)	(0.79)	(2.30)	(-2.41)	(3.10)
	ROE	-0.74	-0.35	-0.21	-0.02	0.04	0.17	0.18	0.15	0.10	-0.30	0.44
		(-4.30)	(-3.74)	(-2.40)	(-0.28)	(0.75)	(3.33)	(3.05)	(2.02)	(1.48)	(-1.78)	(1.57)

4.2.1. ML portfolio versus ML-mimicking portfolio K1

We constructed a time-varying factor that utilizes information from the ML portfolio to explain-away the α_{MLA}. Each month, we identified characteristics that were significantly different between the PW and PL portfolios. Using the Stambaugh and Yuan (2017) mispricing factor approach, we combined all significant characteristics to form an ML-mimicking portfolio, K1. In the characteristic domain, K, the ML portfolio exhibits time-varying characteristic exposures. However, unlike static factor models, the K1 factor can track the ML portfolio over time.

Using the Stambaugh and Yuan (2017) approach also allows us to address two potential concerns. First, FF5 and Q4 are similar benchmarks. Other than market and size factors, the models also contain an investment factor (CMA_t vs. $I2A_t$) and a profitability factor (RMW_t vs. ROE_t). Hence, if the ML portfolio outperforms FF5, it is likely to beat Q4 as well. This somewhat dilutes our claim of a pervasively significant α_{MLA} against entrenched factor models. Second, for the portfolio analysis, we could include α_{MLA} estimates against Stambaugh and Yuan (2017) mispricing factors M4, which have been shown to explain more anomalies than FF5 or Q4 factors. This may enhance the robustness of a significant α_{MLA}, but it does not offer additional insights on the likely sources of α_{MLA}. As with other factor models, the M4 model is static in K, albeit covering a wider set of characteristics than FF5 or Q4. This is because the two mispricing factors were constructed from 11 well-documented anomalies. As such, it is unsurprising for M4 to subsume the explanatory power of FF5 and Q4. Rather than another robustness check using M4 factors, we used the M4 methodology to construct K1 from significant characteristics in the ML portfolio.

Each month over the testing period, and for each characteristic $k \in K$, we conducted an unpaired t-test on the difference in characteristic mean between the PW and PL portfolios. Those with t-stat > 1.96 were shortlisted, after which they were sorted on the magnitude of the difference in normalized characteristic mean dk_t between the PW and PL portfolios. The means were normalized by the cross-sectional variation in characteristic values across PW and PL stocks. Normalization was required prior to ranking as the levels differed across characteristics. Doing so, we obtained a monthly ranking of dominant characteristics in the ML portfolio during the testing period. We denote $w_{kt} = \frac{dk_t}{\sum_{k=1}|dk_t|}$ as the weight of the characteristic k. Each month, we combined the list

of dominant characteristics into a single feature $K1_{it} = \sum_{k=1}(w_{kt}k_{it})$, where k_{it} was the normalized characteristic for each firm i. We sorted the entire firm sample on $K1_{it}$ to form a long-short K1 portfolio with return $r_{k1,t}$. The Stambaugh and Yuan (2017) mispricing factors were formed on firms' simple-average ranking across 11 anomalies. Using $K1_{it}$ is equivalent to sorting firms on weighted-average ranking. If the ML portfolio loads heavily on a characteristic k, it would rank highly in the shortlist; hence, it is assigned a heavier weight w_{kt} in $K1_{it}$.

We ran the regression, $r_{ml,t} = \alpha_1 + \beta_1 r_{k1,t} + \varepsilon_{ml,t}$, to evaluate the ML portfolio against K1. By design, K1 was endowed with inside information on the ML portfolio, such that it exhibits strong explanatory power over the testing period. Furthermore, we used weighted-average characteristic rankings to sort firms, which should be more informative than the equally-weighted rankings of Stambaugh and Yuan (2017). We found a significant $\beta_1 = 0.43$ on K1, with an adjusted R^2 of 0.424. The estimated monthly α_1 remained significantly positive, albeit smaller in magnitude (1.71%) compared with the average α_{MLA} against entrenched factors. We showed that simply combining the ML portfolio's significant monthly characteristics is insufficient to beat it. This suggests that the ML portfolio's implied weights in dominant characteristics (i.e., its timing on characteristic exposure over time) are informative. This is an important clue that motivates a detailed analysis of the ML portfolio's time-varying dominant characteristics for the likely source of α_{MLA}.

4.2.2. ML portfolio versus futuristic portfolio K2

The second benchmark K2 was derived from the factor zoo. Each month during the testing period, we performed an unpaired t-test on the spread return, r_{kt}, for each characteristic $k \in K$. Those with t-stats > 1.96 were shortlisted and ranked on the magnitude of r_{kt}. This generated a monthly rank list of characteristic portfolios with significant spread returns over the testing period. We denote $w'_{kt} = \frac{r_{kt}}{\sum_{k=1}|r_{kt}|}$ as the weight in characteristic k. As with $K1_{it}$, each month, we combine ranked characteristics into a single feature, $K2_{it} = \sum_{k=1}(w'_{kt+1}k_{it})$. We sorted firms on $K2_{it}$ to form a long-short K2 portfolio with return $r_{k2,t}$. Note that $K2_{it}$ is a function of w'_{kt+1}, such that the K2 portfolio assigns heavier weights on characteristic portfolios with larger next-month spread returns.

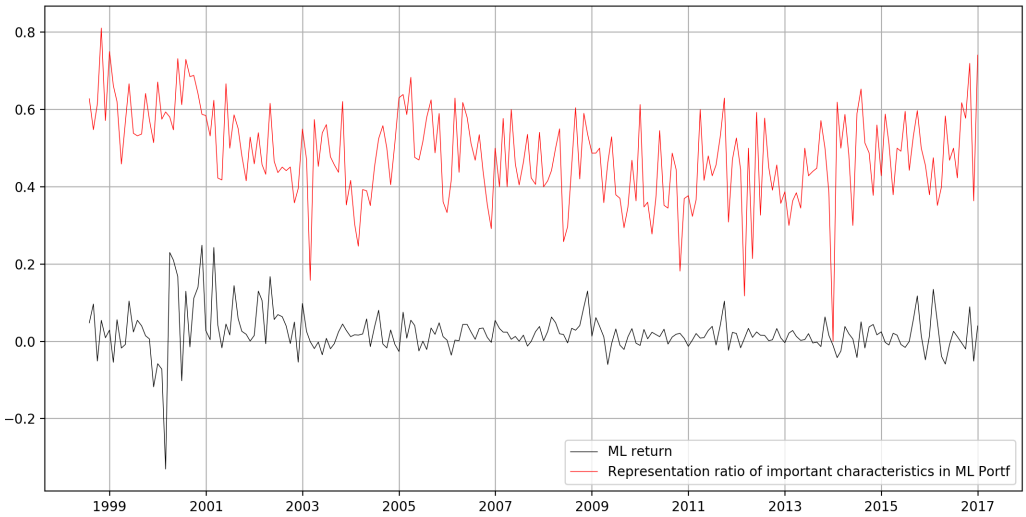


Figure 4: ML portfolio monthly hit rate on characteristic portfolios with significant spread returns. The bottom graph plots the ML portfolio monthly returns over the testing period. The top graph shows the proportion of monthly significant characteristics in the ML portfolio (K1) that also exhibit significant next-month spread returns (K2).

We ran the regression, $r_{ml,t} = \alpha_2 + \beta_2 r_{k2,t} + \varepsilon_{ml,t}$, to evaluate the ML portfolio against K2, which is endowed with perfect foresight on next-month spread returns for all characteristic portfolios. In month t, K2 is formed on $K2_{it}$, whose weights w'_{kt+1} are calculated using next-month characteristic portfolio return, r_{kt+1}. As such, K2 systematically loads on characteristic portfolios with significant next-month spread return. The regression has a substantially lower R^2 of 0.022. The loading on $r_{k2,t}$ is also smaller at $\beta_2 = 0.14$, albeit significant. Now, we have an insignificant $\alpha_2 = 0.4$. Given that it is practically impossible to form K2, the latter benchmark beats the ML portfolio by cheating. We can confirm that by changing to $K2_{it} = \sum_{k=1}(w'_{kt} k_{it})$, α_2 becomes significantly positive.

Figure 4 plots two graphs. The bottom graph is the ML portfolio monthly return over the testing period. The top graph shows the proportion of monthly significant characteristics in the ML portfolio (K1) that also exhibit significant next-month spread return (K2). Put simply, the top graph plots the ML portfolio monthly 'hit rate' on K2 characteristics that exhibit significant next-month spread return. Over the testing period, the ML portfolio hit rate averaged around 50%. The two graphs show some visible comovements in a number of peaks and troughs. This was expected as the hit rate refers to characteristic portfolios that exhibit significant next-month spread returns.

4.3. Dominant characteristics in the ML portfolio

Our analysis thus far demonstrates a pervasively significant α_{MLA} against (i.e., FF3, C4, FF5, and Q4), as well as an ML-mimicking portfolio K1, which uses Stambaugh and Yuan (2017) to combine significant characteristics in the ML portfolio. It takes a cheating K2 portfolio that peeks into the next-month spread return of all factor-zoo characteristics to render α_{MLA} insignificant. Granted, we did not consider the transaction cost associated with a monthly rebalancing ML portfolio over 18 years. However, the magnitude of α_{MLA}, which ranges from 1.43 to 2.48% per month, is too large to be explained away by transaction costs.

In this section, we dissect the ML portfolio to identify the nature and pattern in dominant characteristics that trained ML models uncovered during the 18-year testing period.

4.3.1. Likely patterns in dominant characteristics

We discuss potential sources of α_{MLA} and how this could manifest as patterns in dominant characteristics in the ML portfolio. First, if training-sample anomalies in θ_{1998} survive during the testing period, it would be an obvious source of α_{MLA}. The ensemble forecast from ML models would onload θ_{1998} characteristics, which could then manifest as dominant characteristics in the ML portfolio. However, this scenario is unlikely, given that Mclean and Pontiff (2016) documented a postpublication decline in anomalies. Furthermore, we have shown that most training-sample anomalies in θ_{1998} were no longer significant in the testing sample.

Second, according to Harvey et al. (2015), the proliferation of anomalies began around 2003. This suggests that many of our K characteristics were published during the 1998–2016 testing period. Hence, a potential source of α_{MLA} could stem from the ML portfolio loading on prepublication testing-sample anomalies. Hence, as their spread returns diminish postpublication, the ML portfolio shifts onto other prepublication anomalies. If this is the likely source of α_{MLA}, then the ML portfolio's dominant characteristics would cover a large subset K. We argue that this is also unlikely, as all characteristics in K were published by 2016. Hence, FF5 and Q4 should suffice in explaining returns for most characteristics, regardless of when they were published during the 1998–2016 testing period. Later, we show that the dominant characteristics in the ML portfolio represent only a small subset of K.

Lastly, if our ML portfolio exhibits time-varying characteristic exposure, it could generate significant α_{MLA} against any factor model that, by definition, is static in the K domain. We show that this is the most likely source of significant α_{MLA} everywhere. During the 18-year testing period, the ML portfolio's top three dominant characteristics revolved around just 10 variables, although the ML models were trained on K.

4.3.2. Actual patterns in dominant characteristics

In Table 5, we report the proportion of the testing period in which each characteristic appears as a top three dominant characteristic and the average proportion of each characteristic appearing in the rankings. Panels A and B correspond to θ_{1998}, which was identified by FF5 and Q4 factors. We sorted characteristics on how often they appeared as the most dominant (Rank 1) in the ML portfolio. The table also cites the original or key paper that published the characteristic. We could not identify a published paper that focused on sstk, the sale of common and preferred stocks. Bradshaw et al. (2006) considered a total external financing characteristic that computes net share issuance as sstk-prstkc. Pontiff and Woodgate (2008) examined how share issuance explains cross-sectional stock returns. We also could not find a published paper focusing on operating income before depreciation and tax (oibdp). Simutin (2010) examined how firms' excess cash holdings can explain future stock returns, of which oibdp was used to calculate excess cash holdings. We postulate that oibdp is highly correlated with the Novy-Marx (2013) grossprofit.

Table 5: Top-10 characteristics in the ML portfolio.

This table lists details of the 10 most dominant characteristics in the ML portfolio. We report the proportion of the testing period in which a given characteristic appears in the top three ranks. Characteristics are sorted based on how frequently they appear as the most dominant (Rank 1) characteristic in the ML portfolio. We also list the original paper or key paper that highlights each characteristic. Panels A and B correspond to the ML portfolios that are based on θ_{1998} identified using FF5 and Q4 factors.

(a) Training-sample anomalies identified by FF5.

Name	Characteristic	Rank 1	Rank 2	Rank 3	Average
ivol	**Ang et al. (2006 JF)**	27.9%	21.6%	7.2%	18.9%
	Idiosyncratic volatility				
cashflow	**Da and Warachka (2009 JFE)**	22.1%	13.1%	13.1%	16.1%
	Cashflow				
gxfin	**Bradshaw et al. (2006 JAE)**	20.7%	13.5%	9.0%	14.4%
	Growth in external financing				
min	**Bali et al. (2011 JFE)**	8.1%	9.9%	12.6%	10.2%
	Min daily return in month				
max	**Bali et al. (2011 JFE)**	7.7%	11.3%	10.4%	9.8%
	Max daily return in month				
sstk	**Pontiff and Woodgate (2008 JF)**	6.3%	5.0%	5.0%	5.4%
	Sale of common or preferred stock				
grossprofit	**Novy-Marx (2013 JFE)**	3.2%	6.3%	5.4%	5.0%
	Gross profit				
earnings	**Chordia and Shivakumar (2006 JFE)**	1.8%	5.4%	10.8%	6.0%
	Earnings				
oibdp	**Simutin (2010 FM)**	0.9%	6.8%	11.7%	6.5%
	Operating income before depreciation and tax				
gequity	**Bradshaw et al. (2006 JAE)**	0.5%	3.6%	4.5%	2.9%
	Growth in equity financing				
	Total:	**99.1%**	**96.4%**	**89.6%**	**95.0%**

(b) Training-sample anomalies identified by Q4.

Name	Characteristic	Rank 1	Rank 2	Rank 3	Average
ivol	**Ang et al. (2006 JF)**	40.5%	13.5%	6.3%	20.1%
	Idiosyncratic volatility				
cashflow	**Da and Warachka (2009 JFE)**	21.2%	15.8%	11.7%	16.2%
	Cashflow				
gxfin	**Bradshaw et al. (2006 JAE)**	13.1%	12.2%	11.7%	12.30%
	Growth in external financing				
min	**Bali et al. (2011 JFE)**	7.7%	12.6%	10.4%	10.2%
	Min daily return in month				
max	**Bali et al. (2011 JFE)**	1.8%	11.3%	10.8%	8.0%
	Max daily return in month				
sstk	**Pontiff and Woodgate (2008 JF)**	5.9%	3.6%	5.0%	4.8%
	Sale of common or preferred stock				
grossprofit	**Novy-Marx (2013 JFE)**	1.8%	3.2%	1.8%	2.3%
	Gross profit				
earnings	**Chordia and Shivakumar (2006 JFE)**	1.80%	4.5%	11.7%	6.0%
	Earnings				
oibdp	**Simutin (2010 FM)**	2.3%	8.1%	13.5%	8.0%
	Operating income before depreciation and tax				
gequity	**Bradshaw et al. (2006 JAE)**	0.5%	6.8%	5.4%	4.2%
	Growth in equity financing				
	Total:	**96.4%**	**91.4%**	**88.3%**	**92.0%**

Table 5a shows that the top 10 characteristics cover 99.1% of the testing period as the ML portfolio's most dominant characteristic. For 2 months during the 18-year testing period, vroa2 and roa appeared once each as the top ranked characteristic in the ML portfolio (Rank 1). The same characteristics occupied 96.4 and 89.6% of the testing period as the second (Rank 2) and third (Rank 3) most dominant characteristic. Overall, these 10 characteristics covered 95% of the testing period as the ML portfolio's top three dominant characteristics. Apart from sstk, these characteristics were published in Journal of Finance, Journal of Financial Economics, or Journal of Accounting and Economics. The Ang et al. (2006) idiosyncratic volatility (ivol) appeared most frequently in Rank 1 at 28%, followed by the Da and Warachka (2009) cashflow risk measure at 22% and the Bradshaw et al. (2006) growth in external financing (gxfin) at 20.7%. These three characteristics also dominated at Rank 2, with ivol at 21.6%, gxfin at 13.5%, and cashflow at 13.1%. However, the pair of extreme return characteristics (i.e., max and min) of Bali et al. (2011) were close at 11.3 and 9.9%, respectively. Here, firms were monthly sorted on their maximum or minimum daily return. The relative importance of characteristics was more evenly spread in Rank 3, with the top six characteristics appearing between 13.1% (cashflow) and 9% (gxfin) of the testing period. On average over the top three ranks, ivol was the most dominant (18.9%), followed by cashflow (16.1%), gxfin (14.4%), and min and max return (10.2 and 9.77%, respectively). Two other noteworthy characteristics were the Chordia and Shivakumar (2006) earnings (6%) and operating income before depreciation, oibdp (6.47%), which was used by Simutin (2010) to compute a firm's excess cash holding.

Table 5b is similarly described. Although θ_{1998} were separately identified using FF5 and Q4 factors, there were four common training-sample anomalies related to ivol, max, amihud-d, and price-d. Consequently, the two ML portfolios were similar in various aspects. The same 10 characteristics occupied 96.4, 91.4, and 88.3% of the testing period as the top 3 ranks in the ML portfolio, averaging 92%. Rank 1 was mainly occupied by ivol (40.5%), cashflow (21.2%), and gxfin (13.1%). Although these three characteristics remained prominent in Rank 2 at 13.5%, 15.8%, and 12.2%, respectively, they were more or less on par with min (12.6%) and max (11.3%). Rank 3 was more evenly covered, with nine characteristics covering between 5 to 13.5% of the testing period. Averaging across the top three ranks, ivol remained the most dominant (20.1%), followed by cashflow

(16.2%), gxfin (12.3%), and min return (10.2%). Two other noteworthy characteristics were max return and oibdp, both at 8%.

That the ML portfolio's time-varying exposures revolved around 10 characteristics during the 18 years, suggesting that the α_{MLA} could be associated with a fundamental economic mechanism, which is not adequately explained by any of the entrenched factor models. Table 5 shows that these 10 consist of three trading characteristics (i.e., ivol, max, and min), four internal funding characteristics (i.e., cashflow, oibdp, earnings, and grossprofit), and three external funding characteristics (i.e., gxfin, sstk, and gequity). These funding characteristics can be viewed as variant measures of a firm's financial constraint. In the literature, ivol, min, and max reflect a stock's tendency to exhibit extreme returns, which makes arbitrage trading in mispriced stocks costly (i.e., they could be viewed as proxies of investor arbitrage constraint).

To further explore this issue, we plotted heatmaps to visualize the time-varying dominance of the 10 characteristics in the ML portfolio over the testing period. The heatmaps in the top half of Figures 5 and 6 both correspond to the ML portfolio formed with θ_{1998} identified using FF5 factors. The heatmap for the ML portfolio formed with Q4 anomalies in θ_{1998} have a similar pattern, which we show in Appendix A6. In the next section, we elaborate the purpose of the bottom graphs. We plotted the heatmaps by assigning different colors to the 10 characteristics, with max and min distinguished by two shades of green. The larger the circle, the higher the characteristic's rank in the ML portfolio. To plot the out-of-sample ranks in a single diagram, we converted the characteristics' monthly ranks into average quarterly rankings and used the results to generate the heatmaps.

We found several noteworthy patterns. First, all 10 heatmaps exhibited a pattern resembling the GARCH effect; some maps were more evident than others. Second, certain characteristics shared correlated rankings. Specifically, the rankings among arbitrage constraint (AC) characteristics (i.e., ivol, max, and min) visibly rose and fell around the same time. The correlated rankings among ivol, max, and min were consistent with the main findings of Bali et al. (2011), where the max-effect reversed the Ang et al. (2006) ivol puzzle. Similarly, financial constraint (FC) characteristics exhibited correlated rankings, especially among cashflow, gxfin, earnings, and oibdp. Third, the declining importance of the AC characteristics in the ML portfolio coincided with the rising importance of

FC characteristics (i.e., cashflow, gxfin, oibdp, and, to a lesser extent, grossprofit and sstk). In summary, the heatmap suggests that during the 18-year testing period, the ML portfolio's characteristic exposure alternated between investor AC characteristics and firm FC characteristics.

Earlier, we showed that the α_{MLA} remained significant against a K1 portfolio, which is formed using monthly dominant characteristics from the ML portfolio. The patterns in characteristic exposure from the heatmaps confirm our conjecture that a likely source of α_{MLA} stems from timely shifts in the ML portfolio's exposure between AC and FC characteristics. The existing literature applies risk, mispricing, information, and/or behavioral channels to explain how a given characteristic could explain cross-sectional stock return. For example, stocks with high ivol or extreme returns (i.e., max and min) impose a cost on arbitragers. As such, variations in the arbitrage cost affect the degree of mispricing across stocks, thereby explaining cross-sectional stock returns. Livdan et al. (2009) applied a risk argument for FCs affecting stock returns. Financially constrained firms face reduced investment choices, and binding debt collateral impedes their ability to manage exogenous earning shocks through dividend smoothing. As such, variations in FC across firms generate cross-sectional stock returns over time. However, it is unclear whether any of the economic channels could readily explain the alternating importance of AC and FC characteristics in explaining stock returns beyond the entrenched factor models.

Is there an economic mechanism that can accommodate the importance of investor AC, firm FC, as well as their alternating significance in explaining stock return over a long period? In the next section, we provide a possible answer with supporting empirical evidence. The purpose is to convince readers that our main finding is not random. The rise and fall of characteristics in the factor zoo relates to a fundamental explanation of cross-sectional stock returns.

4.4. Credit cycle and the rise and fall of factor-zoo characteristics

Funding liquidity is an important market friction that affects asset markets. Longstaff and Wang (2012) extended the canonical asset pricing of Cox et al. (1985) to allow heterogeneous agents to achieve optimal risk-sharing between credit markets and other assets. In their model, the size of the credit sector varied over economic cycles in response to

risk-sharing, which affects asset prices. Brunnermeier and Pedersen (2009) showed that, under certain conditions, traders' funding liquidity and assets' market liquidity are mutually reinforcing. Studies on investor funding liquidity include borrowing constraints (Black, 1972), asset margin constraints (Gârleanu & Pedersen, 2011), and financial intermediary capital constraints (He & Krishnamurthy, 2013).

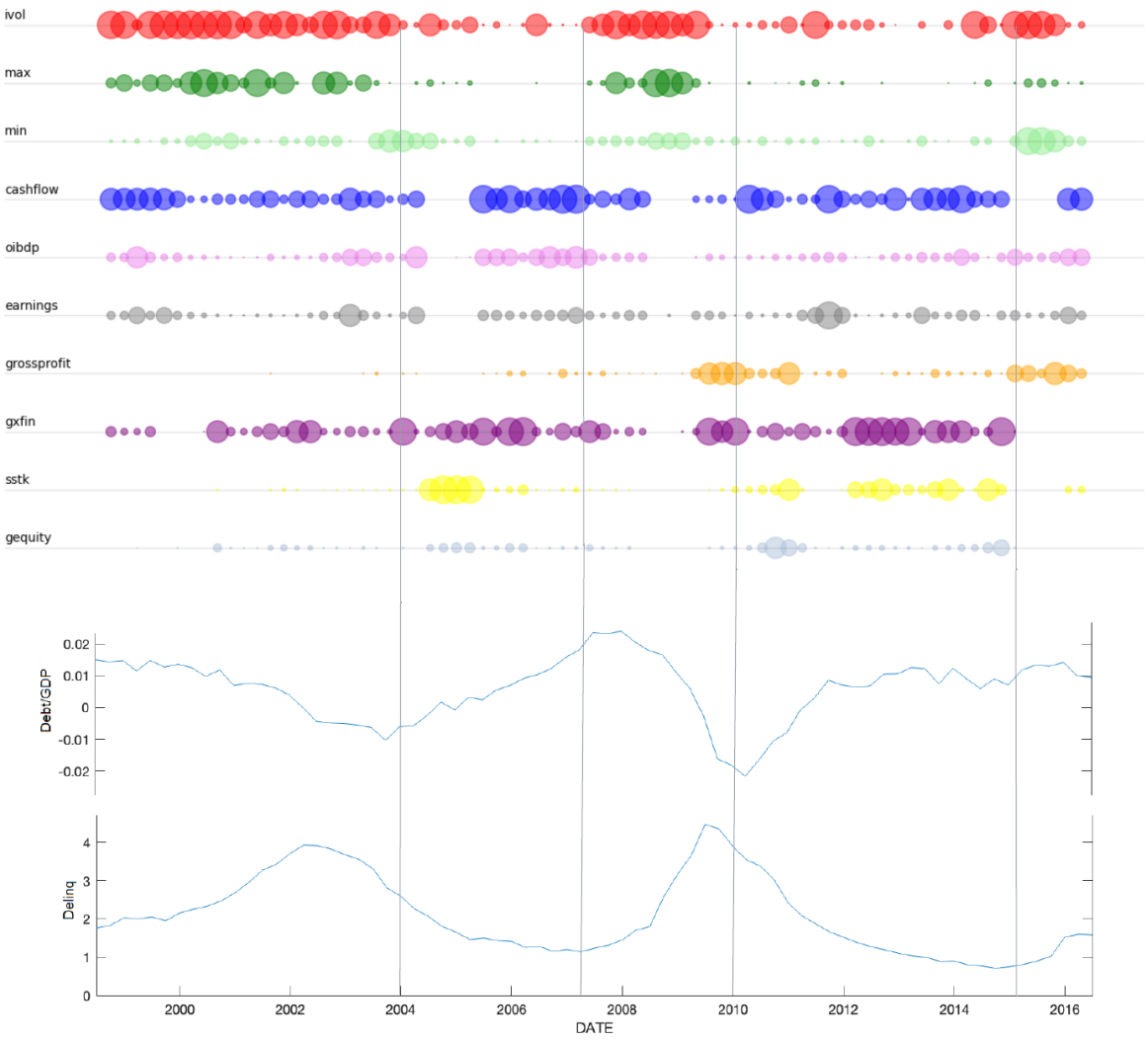


Figure 5: Aligning heatmaps with Debt/GDP and delinquency rate.
The heatmaps illustrate how the ranking of dominant characteristics changes over time. Here, θ_{1998} is identified using FF5. The most dominant characteristics are idiosyncratic volatility (ivol), cash flow (cashflow), external financing (gxfin), minimum daily return per month (min), maximum daily return per month (max), sale of common and preferred stock (sstk), earnings, gross profit (grossprofit), operating income before depreciation (oibdp), and equity financing (gequity). The heatmap is aligned with variables commonly associated with the US credit cycle, represented by corporate debt-to-GDP ratio (Debt/GDP) and delinquency rate (Delinq).

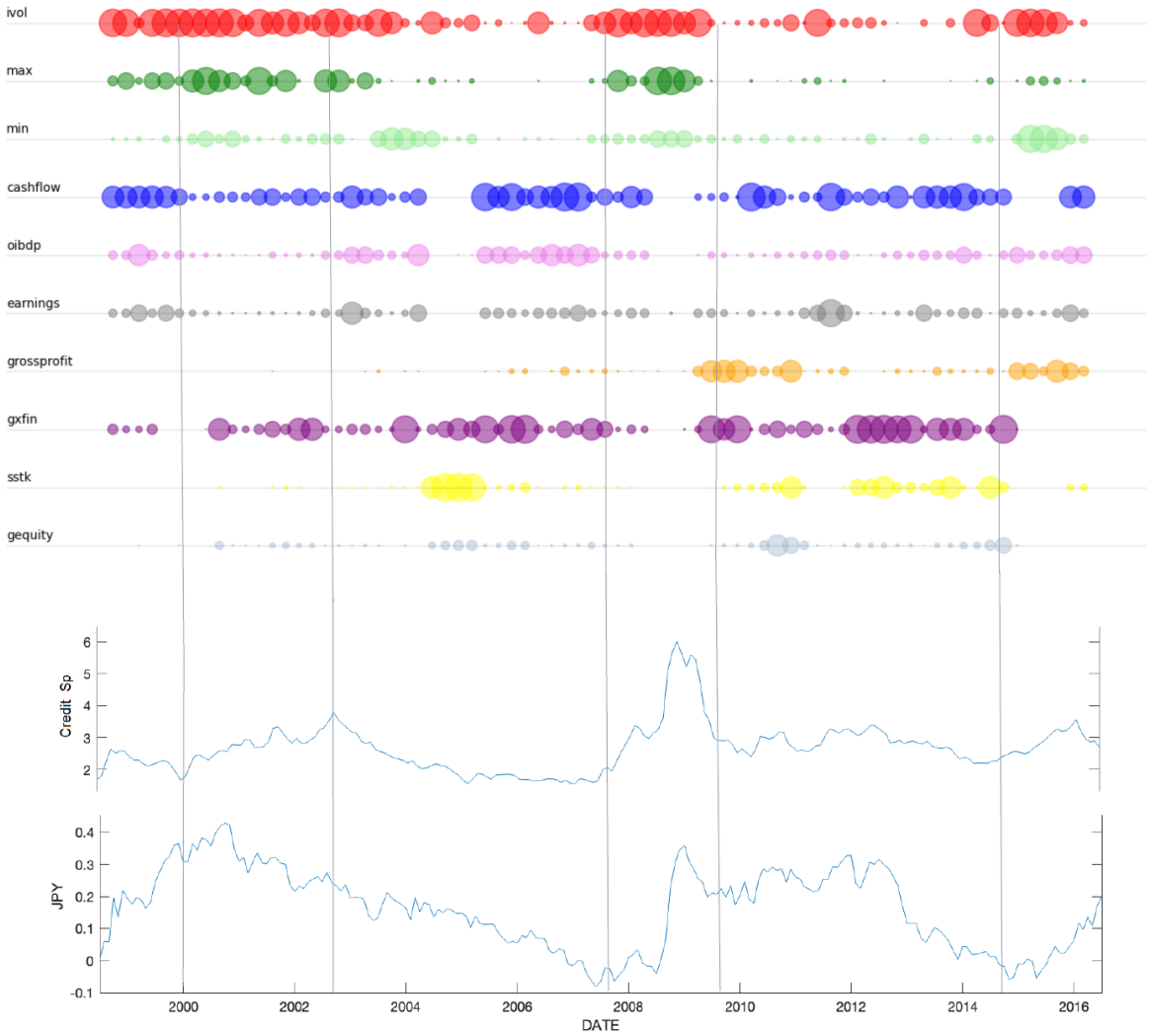


Figure 6: Aligning heatmaps with credit spread and value of JPY.
The heatmaps illustrate how the ranking of dominant characteristics changes over time. Here, θ_{1998} is identified using FF5. The most dominant characteristics are idiosyncratic volatility (ivol), cash flow (cashflow), external financing (gxfin), minimum daily return within each month (min), maximum daily return within each month (max), sale of common and preferred stock (sstk), earnings (earnings), gross profit (grossprofit), operating income before depreciation (oibdp), and equity financing (gequity). The heatmap is aligned with variables that are commonly associated with the US credit cycle, represented by the corporate debt/GDP ratio, and JPY as a proxy for safe-haven markets.

liquidity includes Lamont et al. (2001), Whited and Wu (2006), and Livdan et al. (2009).

We argue that an economy-level credit cycle reflects the funding liquidity of both investors and firms. Credit-cycle fluctuations over time generate the alternating importance of investor arbitrage and firm FCs over time. Consider two economic states:

- **As the US economy moves toward a credit-cycle peak,**

  1. With improving funding liquidity, arbitragers (e.g., hedge funds and invest-

ment banks) take advantage of capital access to trade mispriced stocks. This includes those that impose large arbitrage costs (e.g., high ivol or extreme returns (Max/Min)), to which capital-constrained arbitragers are usually sensitive under normal or illiquid funding conditions. As such, arbitrage constraints become less important in explaining cross-sectional stock returns. Investors may also take advantage of available low-cost credit and move funds into US markets from safe havens such as Japanese Yen (JPY) or gold.

2. As credit becomes easier to access, firms engage in capital-raising regardless of size, credit-rating, or expected profitability. This has two effects. First, subprime firms avoid or delay their default, causing a build-up of systematic distress risk. Second, cross-sectional variations in FC characteristics are expected to reduce over time. Hence, if a firm exhibits FC during a credit boom, it is particularly informative on expected distress risk. Both increase the relevance of FC characteristics in explaining stock return.

- **As the US economy moves toward a credit-cycle trough,**

1. As funding conditions deteriorate, capital-constrained arbitragers become sensitive to stocks that are costly to arbitrage (i.e., high ivol and extreme returns (Max/Min)). These stocks are also likely to require greater margins from investors, further increasing assets' sensitivity to funding illiquidity. Accordingly, AC characteristics become important in explaining cross-sectional stock return. With diminishing leverage opportunities, investment capital could also exit US markets into JPY or gold.
2. As funding liquidity dries up, firms that possibly over-borrowed during the credit boom start to experience delinquency in debt-servicing obligations, or they exhibit observable symptoms of financial distress. This directly leads to rising ivol and extreme returns (Max/Min). More importantly, as the symptoms of financial distress manifest, FC characteristics become less important indicators of distress risk.

4.4.1. Alignment graphs of heatmap and credit-cycle measures

To verify our conceptual argument, we aligned the heatmap with variables that are commonly associated with the US credit cycle. Following Altman (2020), we examined corporate debt/GDP ratio, delinquency rate, and credit spread. We chose JPY and gold as proxies for safe-haven asset markets. The top reserve currency was USD; however, we needed to identify a well-accepted reserve currency that is potentially affected by investment capital flow in and out of US markets but is not directly related to the US credit cycle. Gold is predominantly traded in USD, hence it is possible for gold returns to be directly related to the US credit cycle. In Appendix A6, we outline various proxy variables relating to the US credit cycle, sourced from the Federal Reserve Economic Database of the Federal Reserve Bank of St. Louis website[4].

In Figure 5, we plotted Debt/GDP as the one-year moving average quarterly change in corporate debt/GDP ratio. Altman (2020) associated an upward trending ratio with an expansionary credit cycle, which indicates improving funding liquidity, during which higher-risk firms can further onload substantial debt with relative ease. Figure 5 shows that periods of rising Debt/GDP ratio are associated with the importance of FC characteristics in the ML portfolio. We also plotted Delinq in Figure 5. A rise in corporate delinquency indicates deteriorating funding condition in terms of higher levels of expected default or greater difficulty to refinance existing debt. Altman (2020) associated this with a contractionary credit cycle, which occurs when capital-constrained arbitragers become sensitive to stocks with high arbitrage costs. Figure 5 shows that periods of rising delinquency correspond to the increasing (decreasing) importance of arbitrage (financial) constraint characteristics in the ML portfolio.

In Figure 6, we plotted Credit Sp directly. A downward trending BAA10YM indicates improving funding liquidity, during which riskier borrowers can onload considerable debt with relative ease. Altman (2020) associated this with an expansionary credit cycle. Figure 6 shows that periods of declining (rising) Credit Sp correspond to the importance of financial (arbitrage) constraint characteristics in the ML portfolio. We also plotted the JPY return against a major currency basket in Figure 6. An upward trending JPY could indicate deteriorating funding conditions in the US due to investors withdrawing

[4]https://fred.stlouisfed.org/

capital from US markets into a reserve currency. When arbitragers become more capital-constrained, AC characteristics become important in explaining stock returns. Figure 6 shows that periods of a rising JPY correspond to the rising importance of AC characteristics. Conversely, periods of a falling JPY may be associated with investors returning to US markets as funding conditions improve. Figure 6 also shows that periods of a falling JPY correspond to the rising importance of FC characteristics in the ML portfolio.

4.4.2. Factor models' explanatory power over time

To complement Figures 5 and 6, we conducted two sets of analysis to contrast the importance of AC and FC characteristics between credit-cycle peaks and troughs. Using the three indicators of the US credit cycle, we partitioned the testing period into the following subsamples.

- Trough 1 (T1): July 1998 to July 2003 = 61 months
- Peak 1 (P1): August 2003 to October 2008 = 63 months
- Trough 2 (T2): November 2008 to February 2010 = 16 months
- Peak 2 (P2): March 2010 to February 2015 = 60 months
- Trough 3 (T3): March 2015 to June 2016 = 16 months

First, we examined factor models' explanatory power on AC and FC portfolio returns over time. When the economy is in credit contraction (i.e., T1, T2, and T3), AC characteristics become important in explaining cross-sectional stock return. If this is the case, we expect factor models to exhibit greater explanatory power on AC characteristic portfolio (ACP) returns relative to FC characteristic portfolios (FCP). Conversely, during credit expansion periods, factor models should exhibit greater explanatory power on FCP returns compared with ACP returns. Notably, our purpose is not to identify which factor model(s) is (are) better at explaining ACP or FCP returns and the respective subsample periods. Instead, our aim is simply to ascertain whether ACP (FCP) returns are more important than FCP (ACP) returns during periods of credit contraction (expansion).

We constructed ACP as an equally-weighted portfolio in ivol, max, and min portfolios. To follow, FCP is an equally-weighted portfolio in cashflow, gxfin, sale of common stock

(sstk), and gross profit portfolios. As both T2 and T3 had only 16 observations, we focused our analysis on the T1, P1, and P2 subsamples.

Table 6: Average adjusted-R^2 from AC and FC portfolio return regressions against factor models.

The table reports the average adjusted-R^2 from regressing arbitrage constraint portfolio (ACP), financial constraint portfolio (FCP) and the ML portfolio (MLP) return on different factor models for subsamples July 1998 to July 2003 (T1), August 2003 to October 2008 (P1), and March 2010 to February 2015 (P2).

Average adj-R^2	T1	P1	P2
ACP	0.76	0.24	0.21
FCP	0.45	0.28	0.24
MLP	0.35	0.31	0.15

The table reports the average adjusted-R^2 from regressing ACP, FCP, and ML portfolio (MLP) return on different factor models, for subsamples T1, P1, and P2. During T1, factor models produce an average R^2 0.76 on ACP return. In contrast, the average R^2 for FCP was substantially lower at 0.45. However, when we moved the estimation window from T1 to P1, the average R^2 from factor models was now higher for the FCP return at 0.28, compared with 0.24 for ACP returns. Similarly, for the P2 sample period, factor models produced a higher average R^2 of 0.24 for FCP return compared with 0.21 for ACP return.

4.4.3. Conditional volatility of AC and FC portfolios

Second, if investors become sensitive to AC (FC) characteristics during credit contraction (expansion), then following the Daniel and Titman (1997) argument, it is possible for stocks with similar AC (FC) characteristics to covary more strongly during the trough (peak) subsample periods. If so, we would expect the conditional volatility of ACP $\sigma_{ac,t}$ and FCP $\sigma_{fc,t}$ to affect the ML portfolio's conditional volatility $\sigma_{mlp,t}$ differently, as the estimation window moves between credit expansion and contraction subsamples. Specifically, if the covariance structure among high AC stocks is expected to increase during T1, it is possible that $\sigma_{ac,t}$ becomes dominant in $\sigma_{mlp,t}$. To follow, when the economy moves into credit expansion (P1 or P2), we expect the conditional covariance structure among AC stocks to weaken, causing $\sigma_{ac,t}$ to decline. At the same time, the covariance structure among high FC stocks is expected to strengthen, increasing $\sigma_{fc,t}$ and its impact on $\sigma_{mlp,t}$.

If the above mechanism is evident throughout the testing period, it is possible for $\sigma_{ac,t}$ and $\sigma_{fc,t}$ to exhibit GARCH effects that resemble the heatmap clustering of AC and FC characteristic rankings in the ML portfolio. More importantly, the impact of $\sigma_{ac,t}$ and $\sigma_{fc,t}$ on $\sigma_{mlp,t}$ could vary over time as the US economy transits between credit expansion and contraction states. To test this, we modeled $\sigma_{ac,t}, \sigma_{fc,t}$ and $\sigma_{mlp,t}$ as GARCH(1,1) processes and performed causality tests based on different sample periods.

For the full testing period, both $\sigma_{ac,t}$ and $\sigma_{fc,t}$ were significant in $\sigma_{mlp,t}$. In the credit contraction subsample T1, $\sigma_{ac,t}$ significantly Granger-caused $\sigma_{mlp,t}$ with a p-value of 0.068. However, $\sigma_{fc,t}$ was insignificant, with a p-value of 0.485. Conversely, for both credit expansion subsamples P1 and P2, causality tests confirm that $\sigma_{ac,t}$ was no longer significant in $\sigma_{mlp,t}$, with p-values of 0.237 and 0.468 respectively. To follow, $\sigma_{fc,t}$ became significant in $\sigma_{mlp,t}$ at the 10% level, with a p-value of 0.059 for P1 and 0.061 for P2.

Our two findings complement each other. During credit contraction T1, AC characteristics were more important than FC characteristics. This is shown by a relatively higher adjusted-R^2 range from factor models in explaining increased return covariance among AC stocks, as well as $\sigma_{ac,t}$ having a significant causal effect on $\sigma_{mlp,t}$. FC characteristics became relatively more important during credit expansions P1 and P2. Here, factor models exhibited a higher R^2 range for FCP over ACP, and $\sigma_{fc,t}$ had a significant causal effect on $\sigma_{mlp,t}$, but $\sigma_{ac,t}$ did not.

4.5. Implications of the main findings

We discussed two implications that are of interest to academics and practitioners.

4.5.1. Evaluation of portfolios formed using ML methods

Given the proliferated usage of ML methods by the investment community in recent years, our main finding that an ML portfolio exhibits evident time-varying characteristic exposure has a timely and important implication for the evaluation of ML portfolios. As our results show, if an ML portfolio loads *alternatively* on two distinct sets of arbitrage- and financial constraint-related characteristics over time, it is unlikely that any characteristic-sparse factor model can span the full range of dominant characteristics to which the ML portfolio is exposed at different segments of the investment horizon.

Consistent with Kozak et al. (2020), we noted that any factor model is static in

its number and type of characteristics, by definition. Over time, new cross-sectional predictors emerge as the stock-market experiences different economic conditions or market states. Accordingly, the literature has accumulated an array of factor models that have been modified, expanded, or augmented to accommodate new anomalies over time. If an ML portfolio loads on two distinct sets of weakly correlated characteristic portfolios, a given factor model could explain the portfolio's characteristic exposure, but only for part of the evaluation period.

One possible solution is to construct a dynamic factor that can potentially "stalk" the ML portfolio in the K domain (i.e., the K1 factor). After identifying an ML portfolio's dominant characteristics, we aggregated them into a benchmark portfolio, allowing for different weighting schemes (e.g., equally-weighted, mean-variance optimized (Stambaugh & Yuan, 2017)). However, this approach requires detailed stock holdings in the ML portfolio being evaluated, which may not be practically feasible.

Another approach is to exclude all entrenched factors from the factor zoo and see if the resultant ML portfolio can still generate α_{MLA} everywhere. Although our ML portfolio does not onload any of the entrenched factors, it is uncertain whether excluding the latter from the training sample would lead to a different ML portfolio. More importantly, this approach has limited practical relevance, as it requires dictating how ML models are to be trained by different fund managers. Lastly, one could use existing funds that utilize ML methods to construct an ML portfolio index, and use the result to peer-evaluate a given ML portfolio.

4.5.2. Credit cycle related dominant characteristics and implications for investment styles and tactical asset allocations

Our finding of a long-run cyclical importance between AC and FC characteristics, which covary with several economic variables that proxy the US credit cycle, has a potentially relevant applications to investment styles and tactical asset allocation decisions.

1. The alternating importance between AC and FC characteristics offers a reconciliation in the age-old debate between fundamental analysis and technical analysis. The AC characteristics (i.e., IVOL and Max/Min) are trading-related variables, whereas FC characteristics (i.e., cashflow, growth in external financing, and earnings) are financial statement variables that indicate a firm's fundamental value.

The alternating pattern between AC and FC characteristics in the ML portfolio suggests that trading and fundamental variables should be viewed as complements, not substitutes. Both matter in generating significant α_{MLA}, but not all the time and not at the same time.

2. Our finding that the time-varying importance of AC and FC characteristics is associated with credit contractions and expansion stages of the US economy, and it has implications for fund managers' investments and asset allocation decisions. Experienced fund managers often base their investment decisions on the current economic state. Although they may not be able to ex ante pinpoint credit-cycle peaks and troughs, they have a good sense of whether the economy is currently in a credit expansion or contraction stage. Our results suggest that during a credit contraction, portfolio strategies based on trading variables are likely to be more informative than valuation-type strategies. Conversely, when the economy undergoes credit expansion, valuation-type portfolio strategies are likely to be more successful in generating abnormal returns against factor models.
3. Lastly, our results provide insights into the conditional covariance structure of asset classes. Not only does this paper document a visibly evident comovement among credit spread, JPY returns against a global currency basket, and the gold–silver ratio (results not reported), we also showed that their fluctuation, co-moves with the dominance of AC and FC characteristics in the ML portfolio (equity market).
    - This suggests that with tactical asset allocations, greater diversification benefits can potentially come from excluding stocks with strong AC and FC characteristics from the equity portfolio. Based on our findings, stocks with AC or FC characteristics are likely to exhibit stronger covariance with debt, currency, and precious metal markets, which imply lower diversification benefits.
    - For studies on return and volatility spillover across asset classes, our results suggest that the cross-market trading linkages between equity and other markets may not occur at the market index level. Rather, it may occur at the characteristic portfolio level. Hence, rather than equity index returns, using AC and FC characteristic portfolio returns to estimate the conditional covariance matrix may reveal more insights into the nature of comovements among equity, debt, currency, and precious metal markets.

5. Conclusion

In this paper, we undertook a comprehensive out-of-sample analysis to examine the rise and fall of characteristics in explaining stock returns without imposing any assumptions on the underlying factor structure. The analysis involved a large database that spanned 36 years and 2,500 stocks per year on average for 212 characteristic portfolios formed based on the level and change in 106 trading and firm variables. Estimations on the large three-dimensional database were feasibly achieved using ML methods.

For the 1998–2016 testing period, the ML portfolio generated significant α_{MLA} everywhere. Both PW and PL portfolios exhibited highly significant α across estimation specifications, and the magnitude and t-stat of α monotonically increased from the PL to the PW portfolio. To ascertain the source of α_{MLA}, we dissected the ML portfolio to uncover patterns in its monthly dominant characteristics. Although ML models are trained on the factor zoo, the ML portfolio alternates its exposure between two small characteristic subsets that proxy for investor AC and firm FC.

Because the full K was published by the end of 2016, it is unlikely that these characteristics could produce significant α against FF5 and Q4 factors. We attributed significant α_{MLA} everywhere to implied ML portfolio weights that shift between arbitrage and FC characteristics over time. The rise and fall of characteristics in the factor zoo are very interesting, which we relate to an economic explanation of cross-sectional stock returns over a long sample period, beyond factor models. Our conceptual argument and empirical evidence suggest that the alternating dominance of arbitrage and FC characteristics in the ML portfolio is associated with contraction and expansion phases of the US credit cycle.

We developed a method to find dominant characteristics upon which ML portfolios load. As a limitation of our study, the dominant characteristics and rotation patterns may be contingent on the particular ML model. A formal theoretical framework to show how credit cycle affects different characteristics' ability to explain cross-sectional stock returns and their dependence on employed ML models, as well as rigorous empirical testing, is still needed. Even if we establish a micro-foundation for how credit cycles affect cross-sectional stock returns, we still must empirically harness its explanatory power. This may be pursued as a conditioning state variable (e.g., investor sentiment) or an economy-wide funding liquidity factor that is tradable, orthogonal to entrenched factors, and exhibits

cyclical covariance structures with arbitrage and FC characteristics. Nevertheless, our research continues.

References

[1] Altman, E., 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. Journal of Finance 23, 589-609.

[2] Altman, E., 2020. Covid-19 and the credit cycle. Journal of Credit Risk 16, 67-94.

[3] Alshater, M., Kampouris, I., Marashdeh, H., Atayah, O., Banna, H., 2022. Early warning system to predict energy prices: The role of artificial intelligence and machine learning. Annals of Operations Research, 1-37.

[4] Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31-56.

[5] Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259-299.

[6] Avramov, D., Cheng, S., Metzker, L., 2023. Machine learning versus economic restrictions: Evidence from stock return predictability. Management Science 69 (5), 2547-3155.

[7] Bali, T., Cakici, N., Whitelaw, R., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. Journal of Financial Economics 99, 427-446.

[8] Barillas, F., Shanken, J., 2018. Comparing asset pricing models. Journal of Finance 73, 715-754.

[9] Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444-455.

[10] Bradshaw, M., Richardson, S., Sloan, R., 2006. The relation between corporate financing activities, analysts' forecasts and stock returns. Journal of Accounting and Economics 42, 53-85.

[11] Brunnermeier, M., Pedersen, L., 2009. Market liquidity and funding liquidity. Review of Financial Studies 22, 2201-2238.

[12] Bryzgalova, S., Huang, J., Julliard, C., 2023. Bayesian solutions for the factor zoo: We just ran two quadrillion models. Journal of Finance 78, 487-557.

[13] Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57-82.

[14] Chan, K., Chen, N., 1991. Structural and return characteristics of small and large firms. Journal of Finance 46, 1467-1484.

[15] Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Computers and Electrical Engineering 40, 16-28.

[16] Chen, N., Roll, R., Ross, S., 1986. Economic forces and the stock market. Journal of Business 59, 383-403.

[17] Chen, L., Pelger, M., Zhu, J., 2024. Deep learning in asset pricing. Management Science 70, 714-750.

[18] Chordia, T., Shivakumar, L., 2006. Earnings and price momentum. Journal of Financial Economics 80, 627-656.

[19] Cochrane, J., 2011. Presidential address: Discount rates. Journal of Finance 66, 1047-1108.

[20] Conrad, J., Dittmar, R., Ghysels, E., 2013. Ex ante skewness and expected stock returns. Journal of Finance 68, 85-124.

[21] Cong, L., Tang, K., Wang, J., Zhang, Y., 2021. Alpha Portfolio: Direct construction through deep reinforcement learning and interpretable, AI, Working Paper.

[22] Cox, J., Ingersoll, J., Ross, S., 1985. An intertemporal general equilibrium model of asset prices. Econometrica 53, 363-384.

[23] Covas, F., Haan, W., 2011. The cyclical behavior of debt and equity finance. American Economic Review 101, 877-899.

[24] Da, Z., Warachka, M., 2009. Cashflow risk, systematic earnings revisions, and the cross-section of stock returns. Journal of Financial Economics 94, 448-468.

[25] Fairfield, P., Whisenant, J., Yohn, T., 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. The Accounting Review 78, 353-371.

[26] Fama, E., French, K., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427-465.

[27] Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

[28] Fama, E., French, K., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55-84.

[29] Fama, E., French, K., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1-22.

[30] Fama, E., French, K., 2018. Choosing factors. Journal of Financial Economics 128, 234-252.

[31] Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607-636.

[32] Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. Journal of Finance 75, 1327-1370.

[33] Feng, G., He, J., Polson, N., 2018. Deep learning for predicting asset returns, 1804.09314, arXiv.org.

[34] Foster, F., Smith, T., Whaley, R., 1997. Assessing goodness-of-fit of asset pricing models: The distribution of the maximal R 2. Journal of Finance 52, 591-607.

[35] Freyberger, J., Neuhierl, A., Weber, M., 2017. Dissecting characteristics nonparametrically. Technical report, University of Wisconsin-Madison.

[36] Friedman, J., 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29, 1189-1232.

[37] Fu, F., 2009. Idiosyncratic risk and the cross-section of expected stock returns. Journal of Financial Economics 91, 24-37.

[38] Gârleanu, N. and Pedersen, L., 2011. Margin-based asset pricing and deviations from the law of one price. Review of Financial Studies 24, 1980-2022.

[39] Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63, 3-42.

[40] Giglio, S., Kelly, B., Xiu, D., 2022. Factor models, machine learning, and asset pricing. Annual Review of Financial Economics 14, 337-368.

[41] Green, J., Hand, J., Zhang, X., 2017. The characteristics that provide independent information about average U.S. monthly stock returns. Review of Financial Studies 30, 4389-4436.

[42] Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223-2273.

[43] Hahn, J., Lee, H., 2009. Financial constraints, debt capacity, and the cross-section of stock returns. Journal of Finance 64, 891-921.

[44] Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance, 55, 1263-1295.

[45] Harvey, C., Liu, Y., Zhu, H., 2015. ldots and the cross-section of expected returns. Review of Financial Studies 29, 5-68.

[46] He, Z., Krishnamurthy, A., 2013. Intermediary asset pricing. American Economic Review 103, 732-770.

[47] Ho, T., Hull, J., 1994. Decision combination in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 1-12.

[48] Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. Review of Financial Studies 28, 650-705.

[49] Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. Review of Financial Studies 33, 2019-2133.

[50] Hou, K., Mo, H., Xue, C., Zhang, L., 2021. An augmented q-factor model with expected growth. Review of Finance 25, 1-41.

[51] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T., 2017. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems 30, 3149-3157.

[52] Kelly, B., Pruitt, S., Su, Y., 2017. Some characteristics are risk exposures, and the rest are irrelevant. Technical report, University of Chicago.

[53] Kittler, J., Hatef, M., Duin, R., Matas, J., 1998. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1-8.

[54] Kozak, S., Nagel, S., Santos, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271-292.

[55] Kraus, A., Litzenberger, R., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31, 1085-1100.

[56] Lamont, O., Polk, C., Saaá-Requejo, J., 2001. Financial constraints and stock returns. Review of Financial Studies 14, 529-554.

[57] Leippold, M., Wang, Q., Zhou, W., 2022. Machine learning in the Chinese stock market. Journal of Financial Economics 145, 64-82.

[58] Li, L., Jamieson, K., 2018. Hyperband: A novel bandit-based approach to hyperparameter optimization. Journal of Machine Learning Research 18, 1-52.

[59] Light, N., Maslov, D., Rytchkov, O., 2017. Aggregation of information about the cross-section of stock returns: A latent variable approach. Review of Financial Studies 30, 1339-1381.

[60] Liu, W., 2006. A liquidity-augmented capital asset pricing model. Journal of Financial Economics 82, 631-671.

[61] Livdan, D., Sapriza, H., Zhang, L., 2009. Financially constrained stock returns. Journal of Finance 64, 1827-1862.

[62] Longstaff, F., Wang, J., 2012. Asset pricing and the credit market. Review of Financial Studies 25, 3169-3215.

[63] McLean, R., Pontiff, J., 2016. Does academic research destroy stock return predictability? Journal of Finance 71, 5-32.

[64] Novy-Marx, R., 2013. The other side of value: The gross profitability premium. Journal of Financial Economics 108, 1-28.

[65] Pástor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642-685.

[66] Polyzos, S., Samitas, A., Kampouris, I., 2021. Economic stimulus through bank regulation: Government responses to the COVID-19 crisis. Journal of International Financial Markets, Institutions and Money 75, 101444.

[67] Pontiff, J., Woodgate, A., 2008. Share issuance and cross-sectional stock returns. Journal of Finance 63, 921-945.

[68] Samitas, A., Kampouris, E., Kenourgios, D., 2020. Machine learning as an early warning system to predict financial crisis. International Review of Financial Analysis 71, 101507.

[69] Simutin, M., 2010. Excess cash and stock returns. Financial Management 39, 1197-1222.

[70] Stambaugh, R., Yuan, Y., 2017. Mispricing factors. Review of Financial Studies 30, 1270-1315.

[71] Tobin, J., Brainard, W., 1976. Asset markets and the cost of capital. Cowles Foundation Discussion Papers 427, Yale University.

[72] Whited, T., Wu, G., 2006. Financial constraints risk. Review of Financial Studies 19, 531-559.

[73] Wolpert, D., 1992. Stacked generalization. Neural Networks 5, 241-259.