Exploring the factor zoo with a machine-learning portfolio

Halis Sak[*], Tao Huang[†] and Michael T. Chng[‡]

21th November 2022

Abstract

Over the years, top journals have published hundreds of characteristics to explain stock return, but many have lost significance. What fundamentally affects the time-varying significance of characteristics that survive? We combine machine-learning (ML) and portfolio analysis to uncover patterns in significant characteristics. From out-of-sample portfolio analysis, we back out important characteristics that ML models uncover. The ML portfolio's exposure alternates between investor arbitrage constraint and firm financial constraint characteristics, the timing of which aligns with credit contraction and expansion states. We explain and show how the credit cycle affects different characteristics' ability to explain cross-sectional stock return over time.

JEL classification: G12, G32.

Keywords: Factor models; Firm characteristics; Return predictability.

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

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

[‡]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. 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.

Exploring the factor zoo with a machine-learning portfolio

21th November 2022

Abstract

Over the years, top journals have published hundreds of characteristics to explain stock return, but many have lost significance. What fundamentally affects the time-varying significance of characteristics that survive? We combine machine-learning (ML) and portfolio analysis to uncover patterns in significant characteristics. From out-of-sample portfolio analysis, we back out important characteristics that ML models uncover. The ML portfolio's exposure alternates between investor arbitrage constraint and firm financial constraint characteristics, the timing of which aligns with credit contraction and expansion states. We explain and show how the credit cycle affects different characteristics' ability to explain cross-sectional stock return over time.

JEL classification: G12, G32.

Keywords: Factor models; Firm characteristics; 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

According to Harvey et al. (2015), when Fama and French (1992, 1993) were published, there were around 40 anomalies that cannot be explained by the CAPM. A decade later, discovered anomalies in 2003 had doubled to 84. By 2012, total anomalies jumped to 240. This means that, within ten years, twice as many new anomalies were published as there were total known anomalies in 2003. Cochrane (2011) refers to the discovered anomalies collectively as a factor zoo, in which the market portfolio is 'Factor 0'. The post-2000 anomaly explosion may be partly attributed to data-mining made easy by rich comprehensive databases that are processed with continuously improving computational efficiency. The literature offers a 'hedged portfolio' of economic channels to motivate any given characteristic[1]. Not surprisingly, subsets of anomalies exhibit non-trivial correlations[2]. Harvey et al. (2015) "argue that most claimed research findings in financial economics are likely false."

Enormous research effort is spent on mining new anomalies, but considerably less attention is paid to examine the rise and fall of characteristics in the factor zoo over time. Many characteristics are published in top accounting, economics or finance journals. So once upon a time, they were all anomalies to some factor models, for specific control variables, and over certain sample periods. Is the evolution of the factor zoo affected by fundamental changes in how characteristics relate to stock returns? Or is it due to cross-sectional variation in published articles (that contribute to the factor zoo) in their chosen factor models, control variables and sample periods? The McLean and Pontiff (2016) post-publication effect explains a once-off rise and fall in dominant characteristics. How would one interpret the reoccurring significance of certain dominant characteristics, over a long sample period?

To address the above would require a comprehensive out-of-sample factor zoo analysis over a long sample period, without imposing any assumption on the underlying factor structure. A factor zoo analysis involves a database that spans 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 can only accommodate a small variable choice

[1] A significant characteristic can be explained by risk (relative distress; firm quality), mispricing (limits to arbitrage), information (option metrics) and/or behavioral (herding/anchoring) channels. An insignificant characteristic can be blamed on data snooping or market inefficiency.

[2] For example, illiquidity [Amihud (2002) vs. Pastor and Stambaugh (2003) vs. Liu (2006)], idiosyncratic volatility [Ang et al. (2006) vs. Fu (2009)], or co-skewness [Kraus and Litzenberger (1976) vs. Harvey and Sidique (2000) vs. Conrad et al. (2013)]

set. This poses a problem when K is large, as in a factor zoo. The estimation is normally in-sample, and assumes a linear factor structure, with some 'added' non-linearity[3].

This is where machine-learning (ML) serves our paper as a valuable research tool. We train ML models on a factor zoo to generate a wide range of linear and non-linear factor structures that relate characteristics to stock returns. ML models specialize in prediction tasks, which facilitate an out-of-sample analysis[4]. With an objective function to maximize stock return forecast accuracy over a 1980-1998 sample, we train different ML models to choose from $K = 106$ firm and trading characteristics to estimate factor structures that maximize its objective function. The trained ML models, which are not readily observable, are applied out-of-sample on a 1998-2016 test period to generate monthly stock return forecasts. We use these firm-level return forecasts to form a ML portfolio in predicted winner and loser stocks, which we show beats all of the entrenched factor models. There is a noticeable performance difference between long and short side of machine learning portfolio. This is in line with the observation of Avramov et al. (2021) that long position generates a significant and economically larger payoff than the short position for their neural network-based models. We also tested performance of machine learning portfolios employing feed-forward neural networks. It is our finding (numerical results are omitted for saving the space) that gradient boosting decision trees outperform NN3 model used in Gu et al. (2020).

The ML portfolio analysis uncovers a noteworthy pattern on the rise and fall of characteristics in the factor zoo. We find that only two small subsets of (3 to 4) characteristics play an alternating dominant role in generating the ML portfolio return. One subset relates to investor-level arbitrage constraint, while the other relates to firm-level financial constraint. Our finding allows for a focused examination to uncover the economics that underlay the ML portfolio's alternating exposure to these two distinct characteristic subsets. We explain and empirically confirm that the rotational importance of arbitrage and financial constraint characteristics coincide with different states of the credit cycle. Our paper offers a fundamental insight on a longer-horizon explanation of cross-sectional stock return.

During credit expansion, arbitragers take advantage of leverage opportunities to trade mispriced stocks, including those which are costly to arbitrage e.g. high idiosyncratic volatility or extreme returns. This reduces investor sensitivity to arbitrage constraint characteristics. At the same time, firms engage in capital raising, regardless of size, credit rating, or expected profitability. This has two effects: i) 'Sub-prime' firms delay on default, causing systemic distress risk to build up in the economy; ii) If a firm exhibits financial constraint, this is particularly informative of its distress risk. Either effect increases investor sensitivity to financial constraint characteristics, which dominate arbitrage constraint characteristics in

[3]E.g Interacting and/or squared variables. In terms of the functional form, the number and types of variables to include, the return-generating process's true factor structure is unobservable. A given factor structure may fit some characteristic subsets, but not others. Lastly, a non-linear factor structure is potentially more unstable when taken out-of-sample. This explains why entrenched factor models follow a parsimonious linear structure, with 3 to 5 characteristics between them e.g. {FF3, C4, FF5, Q4, M4}.

[4]In stark contrast, over-fitted econometric models degrade rapidly when they are applied out-of-sample.

generating the ML portfolio return.

During credit contraction, capital-constrained arbitragers become sensitive to stocks that are costly to arbitrage. In contrast, some firms (which possibly over-leveraged during credit expansion) could already exhibit observable symptoms of financial distress. This reduces investor sensitivity to financial constraint characteristics, such that arbitrage constraint characteristics dominate in generating the ML portfolio return.

Our paper complements recent studies. That our ML portfolio loads on a small set of characteristics is consistent with Hou et al. (2020) that the majority of published anomalies cannot be replicated. Many studies that apply ML methods find pervasively significant ML portfolio alpha α_{ml} against factor models. But our emphasis is to trace the source of α_{ml} to the ML portfolio's exposure to two distinct subsets of characteristics, and to understand the underlying economics. Avramov et al. (2021) attribute α_{ml} to the ML portfolio loading on microcap and distress stocks. They argue that an unconstrained α_{ml} could be driven by mispricing associated with limits to arbitrage, rather than abnormal return that could be practically achieved. Leippold et al. (2021) form ML portfolios using Chinese stocks. They confirm that illiquidity characteristics dominate in a market that is crowded with retail investors. Both studies examine potential sources of α_{ml} to better understand the risks and rewards that ML portfolios offer, but did not focus on the portfolio's out-of-sample conditional characteristic exposure. Our ML portfolio exhibit characteristic exposures that are not just time-varying, they are rotating. We explain, and empirically show that the alternating patterns in our ML portfolio's exposure to arbitrage and financial constraint characteristics, align with contraction and expansion states of the U.S. credit cycle.

To shed light on the rise and fall of characteristics in the factor zoo, we examine the following: i) Does the ML portfolio generate a significant alpha α_{ml} against entrenched factor models, namely Fama and French (1993 FF3; 2015 FF5), Carhart (1997 C4), and Hou et al. (2015 Q4)? ii) What are the likely sources of a significant α_{ml}? iii) Is there a pattern in dominant characteristics that drive the ML portfolio return, out of sample over a long sample period? First, we find significant α_{ml} everywhere. None of entrenched factor models can fully explain the ML portfolio return over the test period. The annualized α_{ml} ranges between 17.16%∼29.76% across portfolio weighting schemes, training procedures and factor models. All of the decile-sorted predicted winner portfolio's estimated αs, and almost all of the predicted loser portfolio's estimated αs, are highly significant. In all cases, the magnitude and t-stat of α both increase monotonically from the predicted loser to predicted winner portfolio.

Second, what is the source of α_{ml}? We explain that it is not surprising for a ML portfolio to beat entrenched factor models, given that ML models are trained on the factor zoo. By construction, factor models are static, locking onto target zones in the factor zoo where subsets of correlated characteristic portfolios are 'caged'. A model with more factors potentially cover a wider target zone, thereby explaining more characteristics[5]. But if the ML portfolio's

[5]For example, the market model targets only β, while FF3 triangulates on β, size and book-to-market, and

factor exposures vary over time, it would dynamically 'roam cage-free' inside the factor zoo. As such, the ML portfolio does not remain in any factor-model target zone, long enough for factor loadings to materialize and render its α_{ml} insignificant.

To ascertain the above potential α_{ml} source, we construct a ML-mimicking portfolio K1 that copies the ML portfolio's significant characteristics each month, but combines them following Stambaugh and Yuan (2017). Unlike entrenched factors, the K1 target zone stalks the ML portfolio around the factor zoo. The α_{ml} remains significantly positive, albeit smaller in magnitude. Hence simply knowing its monthly significant characteristics is insufficient to beat the ML portfolio.

Third, we find that the ML portfolio's top 3 dominant characteristics revolve around just 10 features, even though ML models are trained on the factor zoo. During the 18 years test period, these 10 features occupy 99% of first rank, as well as 95% of the top 3 ranks. While they move in and out of the top 3 ranks, the 10 features largely remain in the top 10 ranks. Given that all of our factor zoo characteristics are published by 2016, FF5 and Q4 factors should suffice in explaining away each of these 10 features[6]. In conjunction with the analysis against K1, this suggests that the potential source of α_{ml} could be associated with the ML portfolio's time-varying implied weights in dominant characteristics.

To investigate this, we plot heatmaps for the 10 features to highlight their rankings in the ML portfolio over time. The plots reveal that the ML portfolio's dominant characteristics rotate between two subsets of features. In the literature, these two subsets are generally viewed as attributes of arbitrage constraint on investors{Ang et al. (2006) Ivol; Bali et al. (2011) Max and Min effects}, and financial constraint on firms{Da and Warachka (2009) Cashflow risk; Bradshaw et al. (2006) Growth in external financing; Pontiff and Woodgate (2008) Sale of common or preferred stock; Novy-Marx (2013) Gross profit}. We present a conceptual argument and empirical results to show that the rotational importance of these two sets of characteristic aligns with different states of the credit cycle.

Our paper proceeds as follows. The next section outlines the ML training procedure. Section 3 presents the ML portfolio analysis. Section 4 concludes.

---

targets all characteristics inside its triangular zone e.g. earnings-price ratio, cashflow-price ratio, dividend yield, leverage, 5-year sales rank and contrarian, but not momentum. C4 factors expands the FF3 zone to a box-type zone that covers momentum and a number of other FF3 anomalies. FF5 further expands to a pentagon-type target zone. The two mispricing factors from Stambaugh and Yuan (2017), which are constructed from 11 anomalies, cast a wide net over the factor zoo that subsumes the targeted zones of both FF5 and Q4.

[6]As a counter example, if our test period ends in 2020, and K contains characteristics that are published during 2017-2019, these could be anomalous to FF5 and/or Q4, and produce a significant α_{ml}.

2 The Machine-Learning Algorithm (MLA)

We review recent studies that examine different aspects of the factor zoo. This is followed by a detailed outline of our ML algorithm (MLA). The MLA takes, as inputs, the firm sample N (averaging around 2,500 stocks listed on NYSE/NASDAQ/AMEX) with K=106 firm and trading characteristics, and train different ML models over the 1980-1998 sample period. The monthly stock return forecast from different trained models are combined to identify predicted winner and loser stocks for the 1998-2016 out of sample test period, which forms our ML portfolio.

2.1 Review of the factor zoo literature

Recent advancements in econometrics and ML methods allow more studies to address various issues relating to the factor zoo. These studies can be classified into four strands: i) Raise the acceptance hurdle to expand the factor zoo, ii) Dimension reduction to downsize/tame the factor zoo, iii) Principal component/factor analysis to extract common/latent/'zoo' factors, and more recently iv) Portfolio applications of factor-zoo analysis.

First, Foster et al. (1997) address a potential data-mining problem in asset pricing associated with the increased availability of easily-assessed data. They propose a simple procedure to adjust the critical maximal R^2 value to account for variable snooping. Recent studies have echoed data-mining concerns in empirical asset pricing. Hou et al. (2020) report that an adjusted t-stat of 2.78 at 5% significance will reject around 80% of published anomalies. Harvey et al (2015) document the proliferation of anomalies, and introduce a new multiple testing framework that infers historical acceptance thresholds from past studies. They propose that a new factor has to exhibit a t-stat greater than 3.0. Chordia et al. (2019) implement a data-mining approach, and generates over 2 million trading strategies. Using multiple hypothesis testing to account for covariance in trading signals and returns, they control for the proportion of false rejections by proposing that a 5% significance t-stat threshold should be closer to 4.0.

Second, Freyberger et al. (2018) use an adaptive group LASSO to select characteristics that provide independent information. To address model-selection bias, Feng et al. (2017) combine a double least absolute shrinkage and selection operator (LASSO) with a two-pass regression procedure e.g. 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) use dimension reduction methods to estimate and test factor pricing models. Feng et al. (2020) apply ML models to evaluate the explanatory power of any new factor, over that which is covered by a high-dimensional set of existing factors. They find that a small set of characteristics provide statistically significant explanatory power incrementally over the hundreds of known factors in the literature.

Third, Kelly et al. (2017) use characteristics as instruments to extract principal components to analyze time-varying factor loadings. Light et al. (2017) applies partial least-square estimation to extract a finite set of common latent factors from characteristics. They report that the latent factor sorted portfolio produces a larger spread return than individual characteristic portfolios. Kozak et al. (2020) use shrinkage and selection methods to estimate a stochastic discount factor that explain returns for a large number of stocks. Freyberger et al. (2020) use similar methods to approximate a nonlinear factor structure of expected returns.

Fourth, Moritz and Zimmermann (2016) apply tree-based models to conduct portfolio sorts. Gu, Kelly and Xiu (GKX 2020) train 13 ML models on a factor zoo of 98 characteristics, and identify a set of important characteristics (variants of momentum, liquidity and volatility) that are common across ML models. GKX (2020) attribute the outperformance of boosted trees and neural networks to their ability to allow nonlinear interactions among characteristics. Avramov et al. (2020) document a significant α_{ml} from a portfolio formed using ML models that are trained on an unconstrained factor zoo. However, they find that the significance of α_{ml} is sensitive to economic restrictions. Specifically, their unconstrained ML portfolio loads heavily on micro-caps and financially distressed stocks.

Our paper complements both GKX (2020) and Avramov et al. (2020) in several aspects. The majority of findings in GKX (2020) are ML model-specific, and focuses on stock returns. While they provide some portfolio analysis, the results are also ML model-specific. It is well-documented in the ML literature that ML models are sensitive to the train sample. Furthermore, a ML model-specific portfolio analysis is limited in scope, since the main findings cannot be readily aggregated across ML models. In contrast, our ML portfolio is generated from an ensemble forecast from different ML models. This facilitates a comprehensive portfolio analysis that is not ML model-specific. Indeed, this allows us to reverse-engineer out-of-sample patterns in the rise and fall of characteristics in the factor zoo, which have been jointly uncovered by different ML models. Lastly, the dominant features reported in GKX (2020) are variant measures of volatility, illiquidity and momentum. These are all trading characteristics that correlate to a certain extent with the dominant arbitrage constraint characteristics {Ivol, Max/Min} in our ML portfolio. But in addition, we also find that their dominance alternates with a contrasting set of firm-level financial constraint characteristics.

Avramov et al. (2020) report a significant α_{ml} from an unconstrained factor zoo. They proceed to test and attribute α_{ml} to 'difficult-to-arbitrage' stocks, namely microcaps and proxies for financial distress e.g. no rating, rating downgrade. Our ML portfolio is also generated from an unconstrained factor zoo. But rather than imposing and testing a potential source of α_{ml}, we dissect our ML portfolio to examine patterns in dominant characteristics over time. Like Avramov et al. (2020), we also find that characteristics that proxy arbitrage constraints are important. But we also find that their importance alternate with characteristics that proxy for firm-level financial constraint.

2.2 Production line for the ML portfolio

We provide a flowchart in Figure ?? to illustrate the stages to construct the ML portfolio. We start with a factor zoo of $K = 106$ firm and trading characteristics. We single-sort firms on the level and change in each $k = 1, 2, ..., K$ characteristics, which yield 212 spread portfolios. In Step 1, the MLA trains different ML models on these 212 characteristic portfolios. The trained models are shortlisted into the model set M based on in-sample return forecast accuracy. To obtain an ensemble forecast, we use stacking to generate a conditional probability distribution over trained models in M[7]. An effective implementation of stacking requires a shortlist of important features, which is referred to as the feature selection problem in the ML literature. Numerous feature selection methods are available[8], but their performances are entirely data-specific.

INSERT FIGURE ??

Step 2 addresses the feature selection problem by using factor models to identify train-sample anomalies. We regress the monthly return of each characteristic portfolio separately against FF5 and Q4 factors, and rank on α to identify a small subset of characteristics $\theta_{1998} \in K$ that are anomalous to each factor model. In Step 3, the MLA implements stacking by using θ_{1998} to generate a probability distribution over the 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 a predicted winner and predicted loser decile portfolio, which becomes a monthly rebalanced long-short ML portfolio over the test period. As a comparison, GKX (2020) train different ML models (Figure ?? Step 1), and use each model's predicted returns directly to sort firms into predicted winners and losers (Figure ?? Step 4). After that, their analysis focuses on a small set of common dominant trading characteristics identified by the different trained ML models.

To complement Figure ??, Algorithm 1 summarises the key steps to construct the ML portfolio. In Stage 1, we evenly partition the 36-years full sample period into a train sample (1980-1998) and test sample (1998-2016). We use the train sample to estimate different ML and linear regression models, and to identify train-sample anomalies in θ_{1998}. The training procedure strictly ends in June 1998, after which there is no further updating of model parameters or θ_{1998}, during the test period. As a robustness check, we consider a sub-sample partition for the main analysis, based on a 1980-1992 train period, and 1993-2004 test sample period. The main findings are similar, so we focus our discussion on the full-sample partition.

In Stage 2, the MLA trains three types of models: i) Extra Trees (ET) [see Geurts et al. (2006)], ii) Gradient Boosting Decision Tree (GBDT) [see Friedman (2001)], and iii) Linear Regression. ET and GBDT are two variants of ML models that we train on

[7]See Wolpert (1992); Ho and Hull (1994) and Kittler et al. (1998) for technical details.

[8]See Chandrashekar and Sahin (2014) for a comprehensive review and discussion of feature selection methods

the entire factor zoo K, taking into consideration a range of model configurations. The next subsection contains a brief technical outline of ET and GBDT models. GKX (2020) provide 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.

For each month during the train period, the MLA evaluates each model's overall in-sample ability to predict stock return for all firms $i = 1, ..., N$. We gauge forecast accuracy based on the R-value in equation (**??**), 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 $\sum_i (r_i - \hat{r}_i)^2$ for all firms, using the same month's cross-sectional return variance $\sum_i (r_i - \mu)^2$. Each month, the trained model with the highest R-value is shortlisted into the model set M[9].

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

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

From the above shortlisting procedure, we obtain a model set M that includes all trained models that are ranked best in predicting stock return, at least once during the train sample period[10]. This allows for likely variations in different trained models' predictive ability over an 18-years period. It is unlikely that any given model can dominate all other trained models for the entire train sample. period Fluctuation in predictive ability could be associated with the functional form, combination and/or relevance of characteristics that each trained model focuses on. 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 assumption here is that θ_{1998}, which contains train-sample anomalies to either FF5 or Q4, are more important than other factor-zoo characteristics in explaining stock return over time.

Here, Stage 3 represents our technical contribution to the application of trained ML

[9]A trained model that generates poor return forecasts could produce a negative R^2. To calculate R-value, we take square-root on the absolute value of R^2, and add a negative sign to indicate that it is a low score. While R^2 and R-value give similar model rankings, the latter is more commonly used in the machine-learning literature.

[10]The number of models in M is less than the total number of trained models from ET, GBDT and linear regressions. A trained model may exhibit the highest R-value over several months or even a few years. Another scenario is where a trained model never once produce the highest R-value throughout the train sample.

Algorithm 1 Key stages to construct the machine-learning (ML) portfolio.

- **Stage 1: Train and test sample partition**

  1. Divide 36 years full sample into a train (1980-1998) and a test (1998-2016) sample.
  2. Use train sample to identify a model set M.
  3. Identify train-sample anomalies $\theta_{1998} \in K$. Please see Section **??** and Algorithm 2 for details.
  4. Trained ML models are not retrained during test period i.e. no dynamic updating.

- **Stage 2: The machine-learning algorithm (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 train sample, the MLA computes each model's R-value to evaluate among the trained models.
  3. Every month, the model with the highest R-value is shortlisted into model set M.

- **Stage 3: Generate ensemble forecast from stacking on θ_{1998}**

  1. The MLA trains another decision tree to generate a probability distribution over M, conditional on train-sample anomalies θ_{1998}.
  2. Models in M, for which one (or more) characteristic in θ_{1998} is important, receives a larger probability weight.

- **Stage 4: Generate return predictions to form the ML portfolio**

  1. For each month t in the test sample, the MLA:
     (a) Use each model in M to predict next month return r_{it+1} for each Firm i.
     (b) Using the decision tree from Stage 3, compute the ensemble forecast $\hat{r}_{it+1}$ as the probability-weighted stock return forecast from models in M.
     (c) Repeat on a monthly basis, until the end of the test period.
  2. At each Month t, decile-sort firms on $\hat{r}_{it+1}$ to form a long-short ML portfolio in predicted winner (top-decile) and predicted loser (top-decile).

models in portfolio allocation. We use factor models to address the feature selection problem in stacking, in order to combine return forecasts from trained ML models in M. We argue that it is inappropriate to directly compare the R-value of different trained models, as they are measured from different parts of the train sample. It is also sub-optimal to compare R-values based on the entire train sample. These R-values would indicate each model's average return prediction performance over the train sample, but do not capture the likely time-varying importance of different characteristics over time. To note, the probabilities over M are conditional on θ_{1998}, but the ML models themselves are trained on the factor zoo K. We can confirm that the trained ML models (ET and GBDT) 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. From Stage 3, the conditional probability distribution over M is used to compute a probability-weighted monthly stock return forecast $\hat{r}_{it+1}$. Firms are sorted on $\hat{r}_{it+1}$ to form a ML portfolio that buys top-decile predicted winners (PW), and short-sells bottom-decile predicted losers (PL). To note, our paper examines two ML portfolios. This comes from separately using FF5 and Q4 to identify θ_{1998}, which the MLA applies to combine model forecasts. More than half of the anomalous characteristics in θ_{1998} are the same between FF5 and Q4. And while the main findings are consistent between the two ML portfolios, we report both sets of results for completeness[11].

2.2.1 A brief outline of ML models

The estimation procedure for ET and GBDT are complex, as they combine predictions from many decision trees. Here, a decision tree refers to a non-parametric 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 decision tree is based on squared residuals, with 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 train sample. A pre-specified weighting-scheme is used to combine predictions from different decision trees. An extra-tree (ET) algorithm use the full train sample to estimate decision trees. But 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. Gradient boosting decision tree (GBDT) is an extension of 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 [see Ke et al. (2017)] to estimate ET and GBDT models[12].

[11] The debate as to which is a better factor model is suitably left for the big names in empirical asset pricing.

[12] This approach speeds up the training procedure of conventional GBDT using gradient-based one-side

2.3 Identify train-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 test 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 next-month stock return forecasts. To note, θ_{1998} is identified as at June 1998, and is not updated during the test period.

2.3.1 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 firm and trading characteristics, which generates 212 characteristic portfolios. Firm characteristic portfolios are re-balanced at the end of every June, while trading characteristic portfolios are re-balanced 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 not surprising that a large portion of their corresponding θ_{1998} contain the same characteristics, albeit different rankings.

1. **Fama-French (2015)**: Regress characteristic portfolio excess return on 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-weight (VW) return on the market portfolio, r_{ft} is the riskfree 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. The coefficients $\{b_k, s_k, h_k, i_k, p_k\}$ are FF5 factor loadings. Characteristics are ranked on $|\alpha_{Fk}|$.

2. **Hou et al. (2015)**: Regress characteristic portfolio excess return 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 (return on equity) firms. The coefficients $\{b_k, s_k, i_k, p_k\}$ are loadings on Q4 factors. Characteristics are ranked on $|\alpha_{Qk}|$.

sampling and exclusive feature bundling.

In the preliminary analysis, we also consider other feature selection approaches that are 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. But based on in-sample model estimation, we can confirm that these industry-norm approaches are inferior to FF5 and Q4. Furthermore, our paper does not aim to evaluate among competing feature selection methods for combining ML model forecasts, we do include these results in the paper[13].

2.3.2 Optimal number of anomalous characteristics in θ_{1998}

We can use FF5 and Q4 factors to identify train-sample anomalies in the factor zoo. But the optimal number of anomalies to include in θ_{1998} is an empirical question. We consider between 7 to 10 characteristics. There are a few justifications. First, if θ_{1998} contains too many characteristics, it could introduce more noise than information when combining the return forecasts across trained models. Second, since the characteristics in K are all published by 2016, we expect only a small number of anomalies, if any, against FF5 or Q4 factors. Third, Stambaugh and Yuan (2017) highlight that the number of characteristics that can parsimoniously explain stock returns has typically been small[14].

INSERT FIGURE ??

In Figure **??**a, we plot the ML portfolio average monthly return $\bar{r}_{ML,t}$ against the number of characteristics in θ_{1998}, for various identification approaches. The graphs confirm that ML portfolios based on FF5 or Q4 approach generates greater $\bar{r}_{ML,t}$, compared to industry-norm approaches. The FF5 approach yields the highest average monthly return, which stabilizes at 2.2% for between 6 to 10 characteristics in θ_{1998}. Next is Q4, with a peak $\bar{r}_{ML,t}$ of 2.15% corresponding to 8 characteristics in θ_{1998}. Third is C4, which stabilizes at $\bar{r}_{ML,t} = 2.1\%$. A robustness check based on sub-sample partitioning in Figure **??**b confirms that $\bar{r}_{ML,t}$ does not increase from expanding θ_{1998} beyond 10 characteristics.

INSERT FIGURE ??

We repeat the above analysis, this time using the full sample period 1980-2016 to identify θ_{1998}. Our aim is to see if $\bar{r}_{ML,t}$ increases with the number of anomalies in θ_{1998}, if the latter is identified with a look-ahead bias. Both Figures **??**a and **??**b show less dispersion in $\bar{r}_{ML,t}$ across different approaches, compared to Figure **??**. More importantly, the figures confirm

[13] The comparison among various 'stacking' approaches for portfolio application may be of interest to a finance journal that targets a wide practitioner audience. Our results are available upon request.

[14] Entrenched factor models {FF3, C4, FF5, Q4, M4} all contain between 3 to 5 variables. Even the early generation factor models of Chen, Roll and Ross (1986) or Chan and Chen (1991) have between 2 to 4 variables.

Algorithm 2 Optimal number of characteristics in $\theta_{1998} \in K$

- **Stage 1: Ranking of characteristics**
  1. Separately use FF5 and Q4 factors to rank characteristics in K.
  2. Include only the top-ranked characteristic in θ_{1998}, implement stacking in Algorithm 1 Stage 3. Sort stocks on the combined monthly forecasts from trained models in M, and form the ML portfolio.
  3. Compute the average monthly ML portfolio return for $\theta_{1998} = 1$.

- **Stage 2: Sequential inclusion of ranked characteristics into θ_{1998}**
  1. Add the 2^{nd} ranked characteristics to θ_{1998}, and form the ML portfolio as per Stage 1 above, for $\theta_{1998} = 2$.
  2. Continue until adding the k^{th} ranked characteristic no longer improves the ML portfolio average return.

- **Stage 3: Sequential inclusion of lower ranked characteristics in K**
  1. Include lower ranked characteristics one at a time into θ_{1998}, and compute the ML portfolio average monthly return.
  2. Only characteristics that improve on the portfolio return are retained in θ_{1998}.
  3. After exhausting K, take note of the final number of characteristics in θ_{1998} for each approach.

---

that $\bar{r}_{ML,t}$ does not increase with θ_{1998}, even if θ_{1998} is identified with a forward-looking bias. Across various approaches, the ML portfolio average return is either stable or declining with an increasing number of characteristics in θ_{1998}. For example, Figures **??**a shows that the ML portfolio under FF5 approach has a $\bar{r}_{ML,t}$ of 2.1%, regardless of whether θ_{1998} has 6 or 10 characteristics. For the Q4 approach, $\bar{r}_{ML,t}$ is also stable at 2.1% for between 6 to 8 anomalies in θ_{1998}, after which it begins to decline for 9 or more anomalies.

To complement Figure **??**, we report in Table **??** the $\bar{r}_{ML,t}$ that is generated under different feature selection approaches and train/test sample partition. It also lists the corresponding characteristics that are included in θ_{1998}. The main findings are consistent across full- and sub-sample partition, so we focus our discussion on the former. To note, the look-ahead bias in Figure **??** and Table **??** refers to the identification of θ_{1998}. The sole purpose is to confirm the stylized fact that a small number of characteristics suffice for an optimal identification of θ_{1998} in Algorithm 1 Stage 3. The actual MLA to train and shortlist models into M, and the identification of θ_{1998} to combine forecasts from models in M, are strictly based on the 1980-1998 train sample period.

INSERT TABLE ??

In Algorithm 2 Stage 2, each approach identifies a set of (in bold) characteristics. The remaining characteristics in θ_{1998} are determined from Stage 3. Take the FF5 approach for example, which identifies four anomalies for θ_{1998}, namely, momentum (cumret11-1), long-term debt issue (dltis), change in illiquidity (amihud-d) and change in turnover volume (turnover-d). Adding a fifth ranked characteristic did not lead to an improvement in $\bar{r}_{ML,t}$=2.21%. The MLA proceeds to Stage 3 and performs an exhaustive search in K, which adds three more variables to θ_{1998}[15]. The average portfolio return increases by 0.18% to 2.39%. The Q4 approach identifies the same four anomalous characteristics as FF5, and so the same $\bar{r}_{ML,t}$=2.21% is obtained from Stage 2. Two more characteristics[Bradshaw et al. (2006) growth in equity financing gequity; Fairfield et al. (2003) indirect accruals acc-fwy] are added in Stage 3, which increases $\bar{r}_{ML,t}$ to 2.27%[16].

Table ?? confirms that the optimal number of characteristics in θ_{1998} is small. Across ten estimations (five approaches; two sample partitioning), θ_{1998} has on average 6.5 characteristics. While the three industry-norm approaches identify different top-ranked characteristics, they produce very similar average monthly returns of 1.74%∼1.77%. Adding a futher 3∼10 characteristics raise the $\bar{r}_{ML,t}$ range to 1.85%∼2.44%. For FF5 and Q4, Stage 2 identifies the same four top-ranked characteristics, which generate a higher initial average monthly return of 2.21% compared to industry norm approaches. Further adding 2 to 3 characteristics yield only a marginal improvement in $\bar{r}_{ML,t}$ of 0.18% for FF5, and 0.06% for Q4. The findings are robust to an alternative sampling partition.

INSERT FIGURE ??

Figure ?? plots the ML portfolio cumulative return over the test 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 return, compare to industry-norm approaches. This is consistent with our finding in Table ?? of higher ML portfolio average return under the FF5 and Q4 approaches.

In sum, our analysis shows that factor models help identify a small set of important features for stacking. Further adding characteristics from an exhaustive search of K yield

[15]They are i) Change in each month's minimum daily return (min-d), ii) Change in preference shares capital (pstk-d) and iii) Change in previous month's closing price (prc-d)

[16]For spread return and correlation approaches, only the top ranked characteristic is included in θ_{1998}, which are momentum and free cashflow respectively. For correlation, a search on K adds three more characteristics to θ_{1998} to increase $\bar{r}_{ML,t}$ from 1.75% to 2.23%. But for spread return, 10 more characteristics are added to θ_{1998}, increasing $\bar{r}_{ML,t}$ from 1.74% to 2.44%. Lastly, the mean-variance approach identifies three top-ranked characteristics for θ_{1998}, which gives $\bar{r}_{ML,t}$=1.77%. Further adding 3 more characteristics [Beta; Z-score of Altman (1968); Cash dividend] yields a marginal improvement to 1.85%.

only marginal improvements in $\bar{r}_{ML,t}$. The MLA uses FF5 and Q4 factors to separately identify θ_{1998}, which leads to two similar ML portfolios. In an earlier draft, we also consider C4 to identify θ_{1998}, but find that its θ_{1998} is almost identical to that of FF5. As such, there is a trivial difference between the two ML portfolios[17].

2.3.3 Are the train-sample anomalies still anomalous in the test sample?

To prevent a look-ahead bias in the ML portfolio, we identify θ_{1998} using the train period ending in June 1998. Based on the confirmed stylized fact, we include in θ_{1998} 7 to 8 top-ranked characteristics with the largest $|\alpha|$ against FF5 or Q4 factors. Table **??** lists the train-sample anomalies in θ_{1998} against FF5 and Q4 factors. Except for momentum (cumret11-1) and change in daily average turnover volume (turnover-d) from FF5, and book-to-market ratio (beme) from Q4, all other characteristics have significantly negative αs. The identified θ_{1998} is robust to a sub-sample partitioning, albeit a slightly different ranking order. Five characteristics are anomalous to both FF5 and Q4 factors: Idiosyncratic volatility [ivol; Ang et al. (2006)], Maximum daily return in each month [max; Bali et al. (2011)], Change in illiquidity [amihud-d; Amihud (2002)], Month-end closing price [prc-1] and Previous month return [retadj-1].

INSERT TABLE **??**

The test column reports each characteristic's α and t-stat, based on the 1998-2016 test sample. It shows that many train-sample anomalies subsequently lose their significant α during the test period. This is consistent with Mclean and Pontiff (2016), who document a post-publication decline in characteristic spread return. Indeed, the majority of our factor zoo characteristics are published after 1998. Under the FF5 approach, 4 out of 7 characteristics in θ_{1998} (ivol; max; prc-1; retadj-1) lose their significant α. For the other 3, the magnitude of t-stats drop substantially from 6.04 to 1.78 for cumret11-1, -7.40 to -3.92 for amihud-d, and -5.77 to -3.88 for turnover-d. The Q4 anomalies are similarly described, with 6 out of 8 characteristics losing their significant α in the test sample. BM-ratio and Amihud illiquidity both retain a significant α, but with lower t-stats of 1.67 and -3.92 respectively. The sub-sample partition reveals a similar decline in train sample anomalies over the test period.

The last row of Table **??** reports the ML portfolio's alpha α_{ml} against FF5 and Q4 for both sampling partitions. Across the four estimations, α_{ml} ranges from 2.14% to 2.74% per month, with t-stats between 4.12$\sim$5.46. The smallest α_{ml} of 2.14% is larger than the

[17] Due to space constraints, ML portfolio results for the C4 approach are available upon request. But strictly speaking, Carhart (1997) is the only feasible approach to identify θ_{1998} using the train sample, given FF5 and Q4 do not exist in 1998. But we argue that using FF5 and Q4 to identify θ_{1998} does not induce a look-ahead bias in the ML portfolio. The MLA training is strictly based on the train sample. Although θ_{1998} is identified using 'futuristic' factor models, the MLA has no concept of either factor models or α. Lastly, we confirm that ML portfolio results based on the C4 approach is consistent with those based on FF5 and Q4.

momentum portfolio's α=1.88%. The latter is the global maximal α from 848 estimated αs to produce Table **??**[18]. The results show that although the ML portfolio construction is based on a progressively outdated θ_{1998} over the 18-years test period, the α_{ml} remains significantly positive against FF5 and Q4 factors in both sampling partitions. This suggests that θ_{1998} is not an important source of the ML portfolio alpha α_{ml}.

3 The machine-learning portfolio analysis

In this section, we report main findings from the ML portfolio analysis, which examines three issues: i) Is the α_{ml} significant against entrenched factor models? ii) What is the source of α_{ml}? iii) What are the dominant characteristics in the ML portfolio over the test period. To reiterate, we have two ML portfolios corresponding to FF5 and Q4 approaches to identify θ_{1998}. But as the main findings are similar, we focus our discussion on the FF5 approach.

3.1 Significant α_{ml} everywhere

We evaluate the ML portfolio against entrenched factor models {FF3, C4, FF5, Q4}. The interactions among four factor models, two sets of θ_{1998} and equal/value weighted portfolio return $r_{ml,t}$, translate into 16 portfolio regression results in total. Tables **??** and **??** report the estimated factor loading and t-stat for decile portfolios, based on FF5- and Q4-identified θ_{1998} respectively. Each table has two panels for results, based on equally- and value-weighted $r_{ml,t}$. For both tables, the main findings are consistent between the two panels, and so we discuss mainly Panel A results.

The regressions confirm significant α_{ml} everywhere. Across 16 ML portfolio estimations, all α_{ml} are highly significantly positive. The magnitude of annualized α_{ml} ranges from 17% to 29%, with t-stats ranging from 2.97 to 9.72. On the long side, all 16 predicted winner (PW) portfolios exhibit positive α at 95% significance. On the short side, 13 predicted loser (PL) portfolio α are significantly negative at the 95% level, with another two α significantly negative at the 90% level. Most importantly, both the magnitude and t-stat of α increase monotonically from the PL to PW portfolio. This shows clear evidence that the sorting criterion i.e. model-combined stock return forecast, has a strong pattern in anomalous return against all factor models.

INSERT TABLE **??**

[18]We single-sort on the level and change in $K = 106$ characteristics to form 212 spread portfolios. Each characteristic portfolio is regressed against FF5 and Q4 factors, which produce a total of 424 estimated α. And since we consider both full- and sub-sample partition, there are 848 estimated α in total, of which momentum's $\alpha = 1.88$ is the global maximum.

In the first row of Table ??, the average return and t-stat both increase from the PL to PW portfolio. The latter has a significant average monthly return of 1.85%. The PL portfolio average monthly return of -0.36% is insignificant. This is likely due to short-sale constraints on predicted loser stocks. In the last column, the ML portfolio has the same spread return (t-stat) of 2.21% (5.46) as is reported in Table ??. The α_{ml} is 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 range between 6.88 and 9.72. Across the four models, the magnitude and t-stat of α_{ml} both increase monotonically from the PL to PW portfolio.

Against FF3 and C4 factors, $r_{ml,t}$ loads negatively on MKT and SMB i.e. the ML portfolio takes negative bets on beta and size. Both PL and PW portfolios load positively on the market and size premium. But since the loading on the PL exceeds that of the PW, the net loading is negative. The ML portfolio loads positively on HML, which comes from the PW portfolio chasing value stocks and the PL portfolio loading up on growth stocks. Against C4, MOM is not significant in the ML portfolio. Both PL and PW portfolios load negatively on MOM. In fact, all decile portfolios load negatively on MOM.

Against Q4 and FF5 factors, the ML portfolio loadings on MKT and SMB are similar to those for FF3 and C4, albeit less pronounced. The loadings on MKT are now insignificant i.e. market-neutral. The loadings on SMB remain significantly negative, albeit a smaller magnitude. In FF5, HML is no longer significant. Instead, the ML portfolio loads positively on both profitability (RMW) and investment (CMA) factors. It also loads positively on the corresponding ROE and I2A factors in Q4. All four loadings are highly significant. Estimates from PL and PW show that the loadings of PW are insignificant on RMW, CMA and I2A factors. The PW loading on ROE (-0.18) is 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 stem from the PL portfolio, which loads 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 predicted winner portfolio is either insignificant, or it also loads negatively on profitability and investment factors.

INSERT TABLE ??

Table ?? evaluates $r_{ml,t}$ based on θ_{1998} that is identified by Q4 factors. As with Table ??, the main results are consistent across equal- and value-weighted $r_{ml,t}$, and so we focus our discussion on panel A. The average return and t-stat also increase monotonically from the PL to the PW portfolio. Similar to the FF5 approach, the PW portfolio has a significant average return of 1.86%, but the PL return of -0.28% is insignificant. The ML portfolio has a significant spread return (t-stat) of 2.14% (4.93), as well as a significantly positive α_{ml} against all the factor models, ranging from 1.71% for FF5, 1.84% for Q4, and 2.37% for both FF3 and C4. The t-stats range from 5.86 to 8.73. Across the four models, the magnitude and t-stat of α both increase monotonically from the PL to PW portfolio.

Against FF3 and C4 factors, $r_{ml,t}$ loads negatively on MKT and SMB i.e. the ML portfolio takes negative bets on both beta and size. Both PL and PW portfolios load positively on the market and size premium. But as the loading on the PL portfolio those of the PW portfolio, the net loading of the ML portfolio is negative. The ML portfolio loads positively on HML, which comes from the PW portfolio chasing value stocks, while the PL portfolio loads up on growth stocks. Against C4, MOM is not significant in the ML portfolio. Both PL and PW portfolios load negatively on MOM.

Against FF5 and Q4 factors, the ML portfolio loading on MKT and SMB are similarly described as for FF3 and C4, albeit less pronounced. The loading on MKT is now insignificant i.e. market-neutral. The loading on SMB remain significantly negative, but with a smaller magnitude. In FF5, the ML portfolio loads positively on HML, RMW (profitability) and CMA (investment) factors. It also loads positively on the corresponding ROE and I2A factor of Q4. All these loadings are highly significant. Separate results based on PL and PW show that the significance is driven mainly by the PL portfolio, which exhibit significantly negative loading on HML, RMW, CMA for FF5, and on I2A and ROE for Q4. In contrast, the PW portfolio loads 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. But it is less able to predict higher returns for firms with strong profitability and/or conservative investment policies.

3.2 Potential source(s) of α_{ml}

Given that the ML portfolio is generated using ML models that are trained on the factor zoo, it would not be surprising for the ML portfolio to outperform all entrenched factor models. Although this statement is intuitive, it not academically insightful. Here we aim to identify the source(s) of significant α_{ml} everywhere.

Factor models are static, and target specific zones in the factor zoo where characteristic portfolios with correlated returns are 'caged'. Some factor models e.g. FF5 or Q4 cover a wider target zone than others e.g. FF3 or C4, which allows them to potentially explain a larger set of characteristic portfolios. But if the dominant characteristics in the ML portfolio vary over time, it would dynamically 'roam cage-free' inside the factor zoo. As such, the ML portfolio does not remain in any factor model target zones, long enough for factor loadings to properly manifest and explain away its α_{ml}. To test this proposition, we consider two alternative benchmark 'zoo factors' [K1, K2] that are engineered to beat the ML portfolio.

3.2.1 ML portfolio versus ML-mimicking portfolio K1

Each month, we conduct a paired t-test to identify characteristics that are significantly different between the predicted winner and loser portfolio. We follow the Stambaugh and Yuan (2017) mispricing factor approach to combine all the significant characteristics to form a ML-mimicking portfolio K1. Unlike factor models, the K1 target zone tracks the ML portfolio around the factor zoo over time. Using Stambaugh and Yuan (2017) to construct K1 allows us to address two potential concerns. First, FF5 and Q4 are similar benchmarks. Other than the market and size factors, both 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 one factor model, it is likely to beat the other as well. This dilutes our finding of a pervasively significant α_{ml} against entrenched factor models. Second, we could add to the previous section α_{ml} estimates against Stambaugh and Yuan (2017) M4 factors, which have been shown to explain more anomalies than FF5 or Q4 factors. This may enhance the robustness of a significant α_{ml}, but does not offer additional insight on the likely sources of α_{ml}. Like other factor models, M4 has a fixed target zone, albeit wider than FF5 or Q4 since the two mispricing factors are constructed from 11 well-documented anomalies[19]. Rather than a further robustness check using M4 factors, we use the M4 methodology to construct K1 from significant characteristics in the ML portfolio.

Each month during the test period, and for each characteristic $k \in K$, we perform a paired t-test on the difference in characteristic mean between the predicted winner and predicted loser portfolio. Characteristics with t-stats > 1.96 are shortlisted, and then sorted on the magnitude of the difference in normalized characteristic mean dk_t between the PW and PL portfolios[20]. From this, we obtain a monthly ranking of dominant characteristics that the ML portfolio loads on during the test period. Denote $w_{kt} = \frac{dk_t}{\sum_{k=1} |dk_t|}$ as the weight of characteristic k. Each month, we combine the list of dominant characteristics into a single feature $K1_{it} = \sum_{k=1}(w_{kt} k_{it})$, where k_{it} is the normalized characteristic for each Firm i. We sort 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 are formed on firms' simple-average ranking across 11 anomalies. Using $K1_{it}$ is equivalent to sorting firms on weighted-average ranking. A characteristic that the ML portfolio loads heavily on would rank high in the shortlist, and is assigned a heavier weight w_{kt} in $K1_{it}$.

We run the regression $r_{ml,t} = \alpha_1 + \beta_1 r_{k1,t} + \varepsilon_{ml,t}$ to evaluate the ML portfolio against K1. By design, the target zone for K1 tracks the ML portfolio around the factor zoo over the entire test period. As such, the K1 factor is endowed with an upward biased explanatory power on the ML portfolio return. Furthermore, we use weighted-average characteristic rankings to sort firms, instead of equally-weighted rankings in Stambaugh and Yuan (2017). We find a significant $\beta_1 = 0.43$ on K1, with an adjusted R^2 of 0.424. The estimated monthly α_1 remains significantly positive, albeit smaller in magnitude (1.71%) compared to the average α_{ml} against entrenched factors. Hence, simply combining all of the ML portfolio's significant monthly characteristics is insufficient to beat it. This suggests that the ML portfolio implied weights in dominant characteristics are informative, and very likely time-varying. This is an important clue that motivates a detailed analysis into the ML portfolio's time-varying dominant characteristics as a potential source of α_{ml}.

[19] In that sense, it is not surprising for M4 to subsume the explanatory power of FF5 and Q4.

[20] The means are normalized by the cross-sectional variation in characteristic values across predicted winner and loser stocks. Normalization is required for ranking purposes, since the levels differ across characteristics.

3.2.2 ML portfolio versus Futuristic portfolio K2

The second benchmark K2 is derived from the factor zoo. Each month during the test period, we perform a paired t-test on the spread return r_{kt} for each characteristic $k \in K$. Those with t-stats > 1.96 are shortlisted and ranked on the magnitude of r_{kt}. This generates a monthly rank list of characteristic portfolios with significant spread returns over the test period. 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 sort firms on $K2_{it}$ to form a long-short K2 portfolio with return $r_{k2,t}$. To note, $K2_{it}$ is a function of w'_{kt+1}, such that the K2 portfolio assigns heavier weights on characteristic portfolios with larger next-month spread return.

We run 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 return 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 i.e. it is practically impossible to form K2. The regression of $r_{ml,t}$ against the K2 portfolio 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. And we finally obtain an insignificant $\alpha_2 = 0.4$. Put simply, K2 beats the ML portfolio by cheating. We confirm that by changing to $K2_{it} = \sum_{k=1}(w'_{kt} k_{it})$, α_2 becomes significantly positive.

INSERT FIGURE **??**

Figure **??** has two graphs. The bottom graph plots the ML portfolio monthly return over the test 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 test period, the ML portfolio hit-rate averages around 50%. The two graphs show some visible co-movements in a number of peaks and troughs. This is expected since the hit rate refers to characteristic portfolios that exhibit significant next-month spread returns.

3.3 Dominant characteristics in the ML portfolio

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

In this section, we dissect the ML portfolio to reverse-engineer any potential patterns in dominant characteristics that trained ML models may have uncovered during the 18 years out-of-sample test period.

3.3.1 Likely patterns in dominant characteristics

We outline a few potential sources of α_{ml}, and how this could be manifested as patterns in dominant characteristics in the ML portfolio. First, if train-sample anomalies in θ_{1998} survive during the test period, this could be a source of α_{ml}. The ensemble forecast from ML models will load on θ_{1998} characteristics, which may then manifest as dominant characteristics in the ML portfolio. However, this scenario is unlikely, given that Mclean and Pontiff (2016) document a post-publication decline in anomalies. Furthermore, we can confirm that the train-sample anomalies θ_{1998} are almost completely different from the test-sample anomalies.

Second, according to Harvey et al. (2015), the proliferation of anomalies began around 2003. This suggests that the majority of our factor zoo characteristics are published during the 1998-2016 test period. Hence, a potential source of α_{ml} could stem from the ML portfolio loading on pre-publication test-sample anomalies[21]. And as their spread returns diminish post-publication, the ML portfolio shifts onto other pre-publication anomalies. If this is the likely source of α_{ml}, then the ML portfolio's dominant characteristics would cover a large subset of the factor zoo. We argue that this is also an unlikely source. All our factor zoo characteristics are published by 2016, and so FF5 and Q4 should suffice to explain the majority of them, regardless of when they are published during the test period. Empirically, we can confirm that the dominant characteristics in the ML portfolio cover only a small subset of the factor zoo.

Third, Avramov et al. (2020) attribute their α_{ml} to difficult-to-arbitrage stocks, including micro-caps and financially distressed stocks with credit downgrades or no credit rating. We could cross reference to see if our ML portfolio also exhibit similar characteristics. If it does, this would imply that our α_{ml} is associated with mispricing, rather than risk.

Lastly, if our ML portfolio shifts among a small number of dominant characteristics during the test period, such that it maneuvers in and out of the {FF3, C4, FF5, Q4} target zones, this could also generate a pervasively significant α_{ml}. We confirm that this is the likely source of our documented α_{ml} everywhere. During the 18 years test period, the ML portfolio's top three dominant characteristics revolve around just 10 features, even though the ML models are trained on the entire factor zoo.

[21] We thank a seminar participant at ANU for this suggestion.

3.3.2 Actual patterns in dominant characteristics

In Table ??, we report the proportion of the test period that each characteristic appears as a top 3 dominant characteristic, and the average proportion that each characteristic appear in the top 3 ranks. Panels A and B correspond to θ_{1998} that is identified by FF5 and Q4 factors. We present the features based on how often they appear as the most dominant characteristic (Rank 1) in the ML portfolio. The table also cites the original or key paper that publishes the characteristic. We could not identify a published paper that focuses on sstk, the sale of common and preferred stocks. Bradshaw et al (2006) considers a total external financing characteristic that computes net share issuance as [sstk-prstkc]. Pontiff and Woodgate (2008) examine how share issuance explains cross-sectional stock return. We are also unaware of a published paper that focuses on oibdp. Simutin (2010) examines how firms' excess cash holding explain future stock return, of which oibdp is used to calculate excess cash holding.

INSERT TABLE ??

Table ??a shows that the 10 features occupy 99.1% of the test period as the ML portfolio's most dominant characteristic[22] (Rank 1). They also occupy 96.4% and 89.6% of the test period as the second (Rank 2) and third (Rank 3) most dominant characteristics. Overall, these ten features cover 95% of the test period as the ML portfolio's top 3 dominant characteristics[23]. The Ang, Hoddrick, Xing and Zhang (2006) idiosyncratic volatility (ivol) appears most frequently in Rank 1 at 28%, followed by the Da and Warachka (2009) cashflow measure at 22%, and the Bradshaw, Richardson and Sloan (2006) growth in external financing (gxfin) at 20.7%. These three characteristics also dominate at Rank 2, with ivol at 21.6%, gxfin at 13.5%, and cashflow at 13.1%. However, the pair of extreme return characteristics (max, min) of Bali et al. (2011) are closing in at 11.3% and 9.9% respectively[24]. The relative importance of characteristics is more evenly spread out in Rank 3, with the top six characteristics appearing between 13.1% (cashflow) and 9% (gxfin) of the test period. Averaging across the top 3 ranks, ivol is the most dominant characteristic (18.93%), followed by cashflow (16.1%), gxfin (14.4%) and min and max return (10.2% and 9.77%). Two other noteworthy characteristics are the Chordia and Shivakumar (2006) earnings (6%) and operating income before depreciation oibdp (6.47%), which is used in Simutin (2010) to compute a firm's excess cash holding.

Table 5b is similarly described[25], with the same ten features occupying 96.4%, 91.4% and 88.3% of the test period as the top 3 ranks in the ML portfolio, averaging 92%. Rank 1

[22] In two months out of the 18-year test period, vroa2 and roa appears once each as the top ranked characteristic in the ML portfolio.

[23] Other than sstk, the dominant features in the ML portfolio are published in Journal of Finance, Journal of Financial Economics or Journal of Accounting and Economics.

[24] Here, firms are monthly sorted on their maximum or minimum daily return.

[25] Although θ_{1998} is separately identified using FF5 and Q4 factors, there are four common train-sample anomalies: {ivol, max, amihud-d, prc-1-d}. Consequently, the two ML portfolios are similar in various aspects.

is mainly occupied by ivol (40.5%), cashflow (21.2%) and gxfin (13.1%). While these three features remain prominent in Rank 2 at 13.5%, 15.8% and 12.2% respectively, they are more or less on par with min (12.6%) and max (11.3%). Rank 3 is more evenly covered, with nine features covering between 5% to 13.5% of the test period. Averaging across the top 3 ranks, ivol remains as the most dominant characteristic (20.1%), followed by cashflow (16.2%), gxfin (12.3%) and min return (10.2%). Two other noteworthy characteristics are max return and oibdp, both at 8%.

That the ML portfolio's dominant characteristics shift among 10 features over 18 years suggest that the α_{ml} could be associated with a fundamental economic mechanism, which is not adequately explained by any of the entrenched factor models. Table **??** shows that these 10 features consist of 3 trading characteristics {ivol; max; min}, 4 internal funding characteristics {cashflow; oibdp; earnings; grossprofit} and 3 external funding characteristics {gxfin; sstk; gequity}. These funding characteristics can be viewed as variant measures of a firm's financial constraint. In the literature, {ivol, min, max} reflect a stock's tendency to exhibit extreme returns, which makes arbitrage trading in mispriced stocks costly. Hence we regard them as investor arbitrage constraint.

To explore the above, in the top halves of Figures **??** and **??**, we plot heat-map to visualize the time-varying dominance of the ten characteristics in the ML portfolio over the test period. We have two ML portfolios associated with train-sample anomalies that are identified using FF5 and Q4 factors. Both heatmaps exhibit very similar patterns, so we focus our discussion on the heatmap for the ML portfolio with FF5 train-sample anomalies in both figures[26]. We shall elaborate on the bottom halves in the next section. We assign different colors to the ten characteristics, with Max and Min distinguished by two shades of green. The larger the circle, the higher the characteristic's ranking in the ML portfolio. Due to space constraint, we generate the heat-maps based on quarterly average rankings, instead of monthly rankings.

INSERT FIGURES **??** and **??**

A few patterns are noteworthy. First, the heatmap for every characteristic exhibits a pattern that resembles GARCH effect, albeit some more evident than others. Second, the ranking of certain characteristics are correlated. Specifically, the rankings among arbitrage constraint (AC) characteristics {ivol, max, min} seem to rise and fall around the same time. The correlated rankings among ivol, max and min is consistent with the main finding in Bali et al. (2011) that the max-effect reverses the Ang et al. (2006) ivol puzzle. Similarly, financial constraint (FC) characteristics exhibit correlated rankings, especially among {cashflow, gxfin, earnings, oibdp}. Third, the declining importance of AC characteristics in the ML portfolio coincides with the rising importance of FC characteristics {cashflow, gxfin, oibdp} and, to a lesser extent, grossprofit and sstk. In sum, the heatmap shows that the ML portfolio has been

[26]The heatmap for the ML portfolio with Q4 train-sample anomalies is available upon request.

alternating its factor exposure between investor arbitrage constraint characteristics {ivol, min, max}, and firm financial constraint characteristics {cashflow, gxfin, oibdp, grossprofit, sstk} during the 18-years test period.

It is unlikely for the 10 characteristic portfolios to generate significant α_{ml} everywhere. First, all of them are published between 2006 and 2013. These characteristics predate FF5 and Q4 factors, and are unlikely to produce significant α, whether individually or jointly. Second, the benchmarking against the K1 factor shows that, holding a portfolio of dominant characteristics in the ML portfolio, cannot explain the significance of α_{ml}. We conjecture that a likely source of α_{ml} comes 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[27]. However, it is unclear if any of the economic channels could readily explain the alternating importance of AC and FC characteristics in explaining stock returns, beyond entrenched factor models.

Is there an economic explanation that can accommodate the importance of investor arbitrage constraint, firm financial constraint, as well as their alternating significance in explaining stock return, over a long period? In the next section, we provide an explanation, and present supporting empirical evidence. The purpose is to convince readers that our main finding is not by chance. There is something about the rise and fall of characteristics in the factor zoo, which relates to a fundamental explanation of cross-sectional stock return.

3.4 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) extend 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 varies over economic cycles in response to risk-sharing, which affects asset prices. Brunnermeier and Pedersen (2009) show that, under certain conditions, traders' funding liquidity and assets' market liquidity are mutually reinforcing. Studies on investor funding liquidity include Black (1972) on borrowing constraint, Garleanu and Pedersen (2011) on assets' margin constraint, and He and Krishnamurthy (2013) on financial intermediary capital constraint. Studies on firm funding liquidity include 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

[27]For example, stocks with high ivol or extreme returns (max; min), impose a cost on arbitragers. As such, variation in arbitrage cost affects the degree of mispricing across stocks, thereby explaining cross-sectional stock return. Livdan et al. (2009) apply a risk argument for financial constraint to affect stock returns. Financially constrained firms face reduced investment choice sets, and binding debt collateral impede their ability to manage exogenous earning shocks through dividend smoothing. As such, variation in financial constraint across firms generate cross-sectional stock return over time.

and firms, and fluctuations in the credit cycle generate the alternating importance of investor arbitrage constraint and firm financial constraint over time. Consider two economic states:

- **As the U.S. economy moves toward a credit-cycle peak.**
    1. With improving funding liquidity, arbitragers e.g. hedge funds and investment banks take advantage of capital access to trade in mispriced stocks. This include stocks that impose large arbitrage costs e.g. high ivol or extreme returns (Max/Min), which capital-constrained arbitragers are usually sensitive to under normal or illiquid funding conditions. As such, arbitrage constraint becomes less important in explaining cross-sectional stock return. Investors may also take advantage of available low-cost credit, and move funds into U.S 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: i) 'Sub-prime' firms avoid or delay default, causing a build-up of systematic distress risk; ii) Cross-sectional variation in financial constraint characteristic is expected to reduce over time. Hence, if a firm exhibits financial constraint during a credit boom, this is particularly informative on expected distress risk. Either effect increases the relevance of financial constraint characteristics in explaining stock return.
- **As the U.S. economy moves toward a credit-cycle trough.**
    1. As funding condition deteriorates, capital-constrained arbitragers become sensitive to stocks that are costly to arbitrage i.e. high ivol; extreme returns (Max/Min). These stocks are also likely to require greater margins from investors, further increasing such assets' sensitivity to funding illiquidity. Accordingly, arbitrage constraint characteristics become important in explaining cross-sectional stock return. With diminishing leverage opportunities, investment capital could also exit U.S. markets into the 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 exhibit observable symptoms of financial distress. This directly leads to rising ivol and extreme return (Max/Min). More importantly, as the symptoms of financial distress gradually manifest in firms, financial constraint characteristics become less important indicators of distress risk.

3.4.1 Alignment graphs of heat-map and credit-cycle measures

To verify our conceptual argument, we align the heatmap with variables that are commonly associated with the U.S credit cycle. Following Altman (2020), we examine corporate debt to GDP ratio, delinquency rate and credit spread. We nominate JPY and gold as proxies for safe haven markets. The top reserve currency is USD. But we need to identify a well-accepted reserve currency that is potentially affected by investment capital flow in and out of US capital markets, but is not directly related to the U.S credit cycle. Gold is predominantly traded in USD, hence it is possible for gold return to be directly related to the credit cycle. We choose instead to examine the Gold-Silver price ratio.

Outline below, we source all proxy variables from the Federal Reserve Economic Database (FRED) of the Federal Reserve Bank St Louis website[28].

- The corporate debt to GDP ratio [BCNSDODNS/GDPUS] is used to proxy the liquidity of the U.S. credit market, where BCNSDODNS is the debt issued and loans taken by non-financial corporations, and GDPUS is the U.S. quarterly (real) GDP.
- The delinquency rate [DRBLACBS] is used to proxy the expected default on corporate debt. DRBLACBS measures the cross-sectional average delinquency rate on commercial and industrial loans that are issued by all commercial banks.
- The credit spread [BAA10YM] is used to proxy the price of risky debt over safe debt. BAA10YM measures the cross-sectional average of Moody's seasoned Baa corporate yield spread over the 10-year Treasury bond yield, with constant maturity.
- Cumulative return on JPY: We compute JPY return against an equally-weighted portfolio of 7 currencies: {AUD, CAD, CHF, EUR, GBP, NZD, SGD}. We download from FRED monthly USD cross-rates against JPY and 7 other major currencies, substitute USD with JPY as the base currency, compute the equally-weighted monthly return on JPY against the currency basket, and plot the cumulative return over the test period.
- Gold-Silver price ratio (GSR): The GSR could mitigate the effects of USD fluctuation. Furthermore, precious metal returns could be affected by the financialization of commodities. As a relative measure, GSR is a potentially less noisy indicator of safe haven capital flow. We download from FRED the London Bullion Market gold and silver monthly price fixings. Except between June 2012 to June 2014, we can confirm that the GSR exhibits a strong comovement with both credit spread and JPY over the test period[29].

Heatmap and Debt/GDP: In Figure **??**, we plot Debt/GDP as the one-year moving average quarterly change in the corporate debt to GDP ratio. Altman (2020) associates an upward trending ratio with the expansionary phase of the credit cycle. It indicates improving funding liquidity, during which even the riskier (sub-prime) firms can take on substantial debt with relative ease. Figure **??** shows that sub-periods of a rising Debt/GDP ratio is associated with the importance of financial constraint (FC) characteristics in the ML portfolio.

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

[29]Due to space constraint, We make the GSR graph and discussions available upon request.

Heatmap and delinquency rates: We plot Delinq directly in Figure ??. A rise in corporate delinquency indicates deteriorating funding conditions, in terms of higher levels of expected default and/or greater difficulty to refinance existing debt. According to Altman (2020), this corresponds to a downward trending credit cycle. Following our discussion, this is when capital-constrained arbitragers become sensitive to stocks with high arbitrage costs. Figure ?? shows that sub-periods of rising delinquency correspond to the increasing (decreasing) importance of arbitrage (financial) constraint characteristics in the ML portfolio.

Heatmap and credit spread: We plot Credit Sp directly in Figure ??. A downward trending BAA10YM indicates improving funding liquidity, during which even the riskier borrowers can load on considerable debt at a low incremental borrowing cost. Altman (2020) associates this with the rising portion of the credit cycle. Figure ?? shows that sub-periods of declining (rising) credit spread correspond to the importance of financial (arbitrage) constraint characteristics in the ML portfolio.

Heatmap and JPY: We plot JPY return against a major currency basket in Figure ??. An upward trending JPY could indicate deteriorating funding conditions in the U.S., as traders withdraw capital from U.S. markets into a reserve currency. As arbitragers become more capital constrained, arbitrage constraint characteristics become important in explaining cross-sectional stock returns. Figure ?? shows that sub-periods of a rising JPY correspond to the rising importance of arbitrage constraint characteristics. Conversely, sub-periods of a downward trending JPY may be associated with investors returning to U.S. markets, as funding conditions improve. Figure ?? also shows that sub-periods of a falling JPY correspond to the rising importance of financial constraint characteristics in the ML portfolio.

3.4.2 Factor models' explanatory power over time

To complement Figures ?? and ??, we conduct two sets of analysis to contrast the importance of arbitrage constraint (AC) and financial constraint (FC) characteristics between credit cycle peak and trough sub-samples. Using the three indicators of the U.S. credit cycle, we partition the test period into the following sub-samples.

- 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 examine factor models' explanatory power on AC and FC portfolio returns over time. When the economy is in credit contraction {T1, T2, 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) return, relative to FC characteristic portfolios (FCP). Conversely, during credit expansion periods, factor models would exhibit greater explanatory power on FCP return, compared to ACP return. To note, our purpose is not to identify which factor model(s) is better at explaining ACP or FCP returns, and for which sub-sample period. Our aim is simply to ascertain whether ACP (FCP) return is more important than FCP (ACP) return, during periods of credit contraction (expansion).

We construct 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 have only 16 observations, we focus our analysis on the T1, P1 and P2 sub-samples.

Adj-R^2	**T1**	**P1**	**P2**
ACP	0.74-0.79	0.21-0.28	0.13-0.29
FCP	0.37-0.54	0.22-0.32	0.12-0.31
MLP	0.31-0.39	0.28-0.34	0.08-0.20

The table above reports the adjusted-R^2 range from regressing ACP, FCP and ML portfolio (MLP) return on different factor models, for sub-samples T1, P1 and P2. During T1, factor models produce a R^2 range of 0.74-0.79 on the ACP return. In contrast, the R^2 range for FCP is substantially lower at 0.37-0.54. But after we shift the regression window from T1 to P1 or P2, the R^2 range for FCP become slightly higher than those for ACP.

3.4.3 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) sub-sample periods. If so, we would expect the conditional volatility for 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 sub-samples. 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 test 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 U.S economy transits between credit expansion and contraction states. To test this, we model $\sigma_{ac,t}, \sigma_{fc,t}$ and $\sigma_{mlp,t}$ as GARCH(1,1) processes, and perform causality tests based on different sample periods.

For the full test period, we confirm that both $\sigma_{ac,t}$ and $\sigma_{fc,t}$ are significant in $\sigma_{mlp,t}$. In the credit contraction sub-sample T1, $\sigma_{ac,t}$ significantly Granger-causes $\sigma_{mlp,t}$ with a p-value of 0.068. But $\sigma_{fc,t}$ is insignificant, with a p-value of 0.485. Conversely, for both credit expansion sub-samples P1 and P2, causality tests confirm that $\sigma_{ac,t}$ is no longer significant in $\sigma_{mlp,t}$, with p-values of 0.237 and 0.468 respectively. To follow, $\sigma_{fc,t}$ becomes 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 are 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 become relatively more important during credit expansion P1 and P2. Here, factor models exhibit a higher R^2 range for FCP over ACP, and $\sigma_{fc,t}$ has a significant causal effect on $\sigma_{mlp,t}$, but $\sigma_{ac,t}$ does not.

3.5 Implications of main findings

We outline two implications that are of practical interest to both academia and industry.

3.5.1 How do we evaluate portfolios that are formed using ML methods?

If factor models are included in the train sample, they would impose fixed target zones on the factor zoo. This implies that the same factor models cannot adequately explain ML portfolio return. For example, the FF3 target zone triangulates on beta, size and book-to-market, which allows it to explain all characteristic portfolios inside its target zone. But if the ML portfolio loads on different characteristics over time, it would move in and out of the FF3 target zone. Switching to C4/FF5/Q4/M4/FF6/q^5 augments the FF3 target zone, but it remains fixed in the factor zoo. Static factors cannot explain a ML portfolio that roams 'cage-free' inside the factor zoo.

A partial solution is to construct a dynamic factor that projects a moving target zone to potentially stalk the ML portfolio around the factor zoo e.g. $K1$ factor. After identifying a ML portfolio's dominant characteristics, we aggregate them into a benchmark portfolio, allowing for different weighting schemes e.g. Equally-weighted; Mean-variance optimized; Stambaugh and Yuan (2017). However, this approach requires detailed stock holdings in the ML portfolio, which is practically infeasible.

Another approach is to exclude all entrenched factors from the factor zoo, and see if the resultant ML portfolio could still generate α_{ml} everywhere. Although our ML portfolio does not load on any of the entrenched factors, it is uncertain if excluding the latter from the train sample would lead to a different ML portfolio. More importantly, this approach also has limited practical relevance. Lastly, one could use existing funds that utilize ML methods to construct a ML portfolio index, and use that to peer-evaluate a given ML portfolio.

3.5.2 What does the credit-cycle related significance of characteristics imply about investment style or tactical asset allocation?

Our finding of a long-run cyclical importance between trading (AC) and financial statement (FC) variables in the ML portfolio, offers a reconciliation of the age-old debate between technical versus fundamental analysis. Our results suggest that trading and financial statement variables are in fact complementary; Both are relevant in explaining stock return, just not all the time, and not at the same time.

Fund managers condition their tactical asset allocation or investment style choices on different economic and/or market states. While it is not easy to pinpoint credit-cycle peaks and troughs ex-ante, fund managers tend to have a good sense of whether the current economy is in credit expansion or contraction. Our finding suggests that trading-based portfolio styles are more relevant during credit contraction, while valuation-based portfolio styles are more relevant during credit expansion.

More generally, we offer an insight into the conditional covariance structure across different asset classes. Not only do we find visually evident co-movement among the credit spread, JPY and gold-silver ratio, we associate their return variation with the alternating importance of two distinct sets of characteristics that explain equity return. Our findings suggest that for tactical asset allocation, rather than stock indices, using AC and FC characteristic portfolios to estimate the conditional covariance structure with fixed income, currency and precious metal, may yield more insights on the co-movement among asset markets.

4 Conclusion

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

For the 1998-2016 test period, the ML portfolio generates significant α_{ml} everywhere. Both predicted winner and loser portfolios exhibit highly significant α across estimation specifications, and the magnitude and t-stat of α monotonically increases from the predicted loser to predicted winner portfolio. To ascertain the source of α_{ml}, we dissect the ML portfolio to reveal patterns in its monthly dominant characteristics. Although ML models are trained on the factor zoo, the ML portfolio alternates its exposure between two small subsets of investor arbitrage constraint, and firm financial constraint characteristics. Given that all of them are published before 2016, it is unlikely for these characteristics to produce significant α against FF5 and Q4 factors.

We attribute significant α_{ml} everywhere to implied ML portfolio weights that shift between arbitrage constraint and financial constraint characteristics over time. There is something about the rise and fall of characteristics in the factor zoo, which we relate to an economic explanation of cross-sectional stock return over a long sample period, beyond factor models. Our conceptual argument and empirical evidence suggest that the alternating dominance of arbitrage and financial constraint characteristics in the ML portfolio is associated with contraction and expansion phases of the U.S credit cycle.

A formal theoretical framework to show how credit cycle affects different characteristics' ability to explain cross-sectional stock return, as well as rigorous empirical testing, is suitably pursued as a separate paper. Even if we can establish the economic mechanism in which credit cycle affects stock return, how do we empirically harness its explanatory power? As a conditioning state variable, like investor sentiment? Or as an economy-wide funding liquidity factor that is tradable, orthogonal to entrenched factors, and exhibits cyclical covariance structures with arbitrage and financial constraint characteristics? 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] Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31-56.

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

[5] Avramov, D., Cheng, S., Metzker, L., 2021. Machine learning versus economic restrictions: Evidence from stock return predictability. Management Science, Forthcoming.

[6] 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.

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

[8] 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.

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

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

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

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

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

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

[15] Cochrane, J., 2011. Presidential address: Discount rates. Journal of Finance 66, 1046-1108.

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

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

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

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

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

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

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

[23] Fama, E., French, K., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 50, 55-84.

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

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

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

[27] Foster, D., 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.

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

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

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

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

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

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

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

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

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

[37] He, Z., and A. Krishnamurthy. 2013. Intermediary asset pricing. American Economic Review 103, 732-70.

[38] Ho, T., Hull, J., 1994. Decision combination in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 1-10.

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

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

[41] 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.

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

[43] Kittler, J., Hatef, M., Duin, R., Matas, J., 1998. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1-14.

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

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

[46] Leippold, M., Wang, Q., Zhou, W., 2021. Machine learning in the Chinese stock market. Journal of Financial Economics, Forthcoming.

[47] 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.

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

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

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

[51] McLean, D., Pontiff, J., 2016. Does academic research destroy stock return predictability? Journal of Finance 71, 5-31.

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

[53] Pastor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 643-685.

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

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

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

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

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

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

Table 1: Shortlisted characteristics based on different feature-selection methods. This table presents the identified characteristics for generating a conditional probability distribution over trained ML models. Characteristics in bold are shortlisted by the relevant approach, while the rest of the list is formed by sequentially adding each of the remaining characteristics in the factor zoo until there no further improvement in the ML portfolio return.

	Train: 1980-1998, Test: 1998-2016		Train: 1980-1994, Test: 1994-2004	
Approach	**List**	**MR**	**List**	**MR**
Portf	**cash** **che** **pstkrv** beta altman_z dv	$(\mathbf{1.77}^{***})\ 1.85^{***}$	**cash** **che** **ibc** capex	$(\mathbf{2.28}^{***})\ 2.29^{***}$
RawRet	**cumret11_1** turnover_d retadj_1_d sstk amihud_d pstk_d gnoa gat pe gxfin de	$(\mathbf{1.74}^{***})\ 2.44^{***}$	**cumret11_1** **fcashflow** **beme** **tobin_q** amihud_d ln_rdc_d	$(\mathbf{2.26}^{***})\ 3.10^{***}$
FF5	**cumret11_1** **dltis** **amihud_d** **turnover_d** min_d pstk_d prc_1_d	$(\mathbf{2.21}^{***})\ 2.39^{***}$	**cumret11_1** **amihud_d** **turnover_d** **dltis** txp_d capex debtconv	$(\mathbf{2.93}^{***})\ 3.16^{***}$
Qfact	**dltis** **cumret11_1** **amihud_d** **turnover_d** gequity acc_fwy	$(\mathbf{2.21}^{***})\ 2.27^{***}$	**cumret11_1** **amihud_d** **turnover_d** **dltis** txp_d ceq fcashflow	$(\mathbf{2.93}^{***})\ 3.15^{***}$
Corr	**fcashflow** debtsecu_d tobin_q dm	$(\mathbf{1.75}^{***})\ 2.23^{***}$	**fcashflow** **sstk** **m2b** turnover_d pstkrv_d rdirtysurplus dvp	$(\mathbf{2.31}^{***})\ 3.10^{***}$
Ensopt4	amihud_d, turnover_d cash che dltis fcashflow cash dltis cash fcashflow che dltis cumret11_1 dltis cumret11_1 pstkrv	2.21^{***}	amihud_d, turnover_d, sstk cumret11_1 dltis	2.94^{***}

Table 2: FF5 and Q4 train-sample anomalies, based on full-sample and sub-sample partitioning. The table reports the estimated α (t-stats) of the top 7 or 8 anomalies to FF5 and Q4 models, for both train and test sample estimations. These anomalies are identified based on train-sample estimation.

	Train: 1980-1998, Test: 1998-2016					Train: 1980-1994, Test: 1994-2004				
Apr.	**List**	**Train**		**Test**		**List**	**Train**		**Test**	
		Alpha	**t-stat**	**Alpha**	**t-value**		**Alpha**	**t-stat**	**Alpha**	**t-value**
FF5	cumret11_1	1.88***	(6.04)	0.95*	(1.78)	cumret11_1	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)
	prc_1_d	-0.94***	(-4.67)	-0.36	(-0.98)	prc_1_d	-1.09***	(-4.55)	-0.63	(-1.25)
	turnover_d	0.86***	(5.77)	0.84***	(3.88)	retadj_1	-1.06***	(-3.66)	-0.58	(-1.08)
	retadj_1	-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	retadj_1	-1.42***	(-4.62)	-0.34	(-0.88)	retadj_1	-1.73***	(-4.89)	-0.58	(-1.08)
	prc_1_d	-1.34***	(-5.26)	-0.36	(-0.98)	prc_1_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)	ln_me_d	-0.96***	(-2.88)	0.12	(0.24)
	ML Portfolio			2.14***	(4.93)	ML Portfolio			2.26***	(4.12)

Table 3: ML portfolio loading on entrenched factors
(a) FF5 anomalies and equally-weighted portfolio return

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) FF5 anomalies and value-weighted portfolio return

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 entrenched factors
(a) Q4 anomalies and equally-weighted portfolio return

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 value-weighted portfolio return

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)

Table 5: (a) Based on train-sample anomalies identified by FF5.
The table shows the percentage of test sample period that each of the 10 features appears as one of the top 3 dominant characteristics in the ML portfolio. The reference cites the main paper that publishes the characteristic. We could not identify a published paper that focuses on sstk. Bradshaw et al (2006) considers a total external financing characteristic that computes net share issuance as [sstk-prstkc]. Pontiff and Woodgate (2008) examine how share issuance explains cross-sectional stock return. We are also not aware of a published paper on oibdp. Simutin (2010) examines how firms' excess cash holding explain future stock return, of which oibdp is used to calculate excess cash holding.

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

Table 5: (continued)
(b) Based on train-sample anomalies identified by Q4.
The table shows the percentage of test sample period that each of the 10 features appears as one of the top 3 dominant characteristics in the ML portfolio.

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

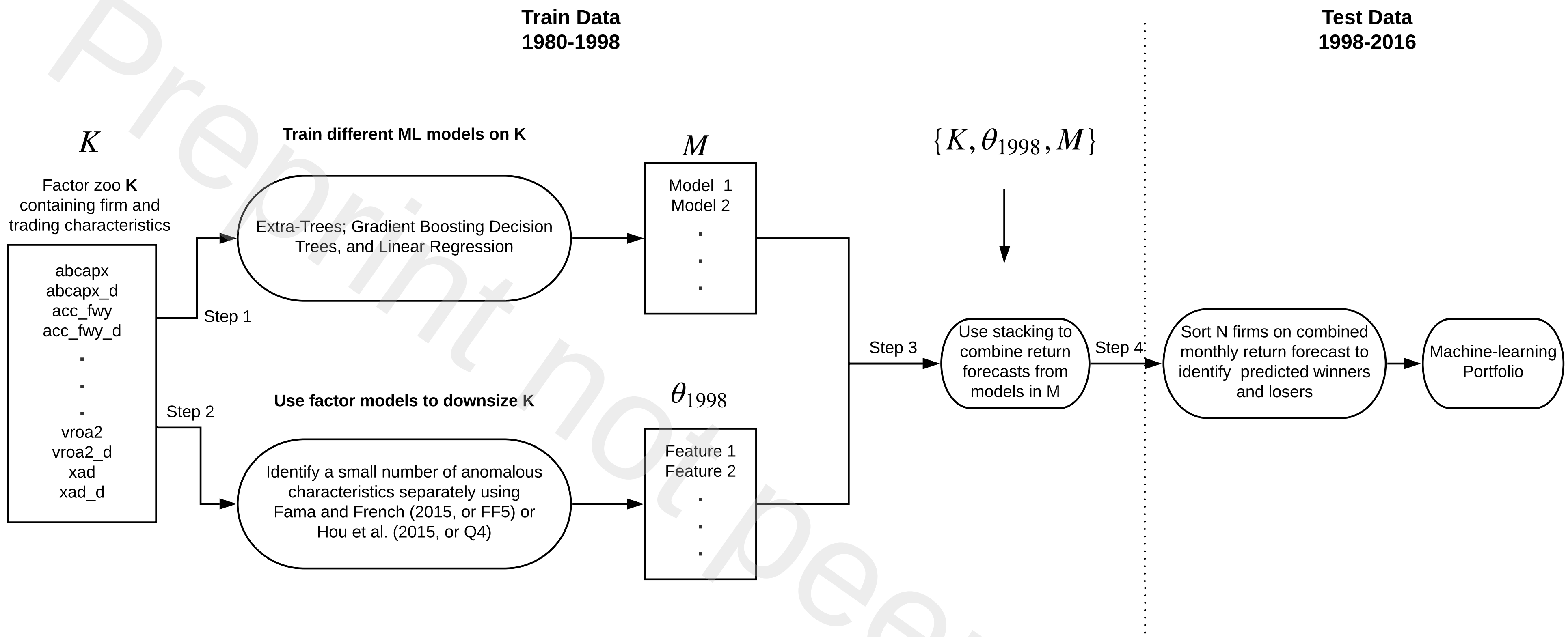
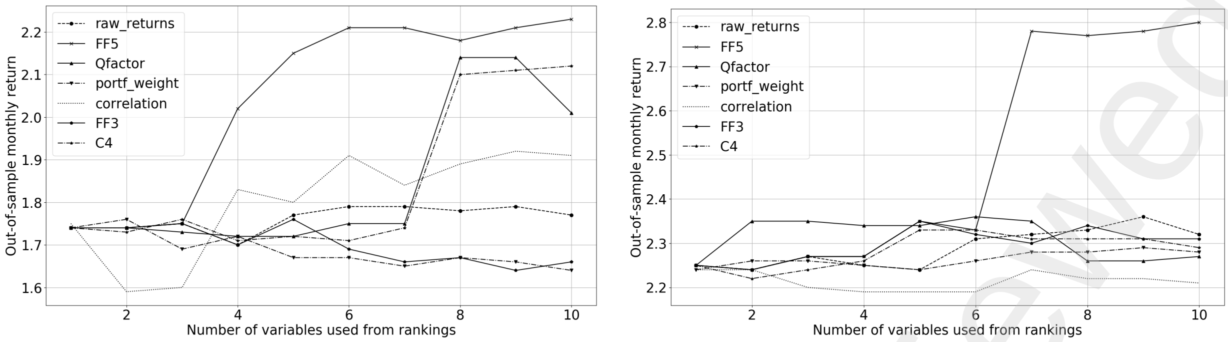
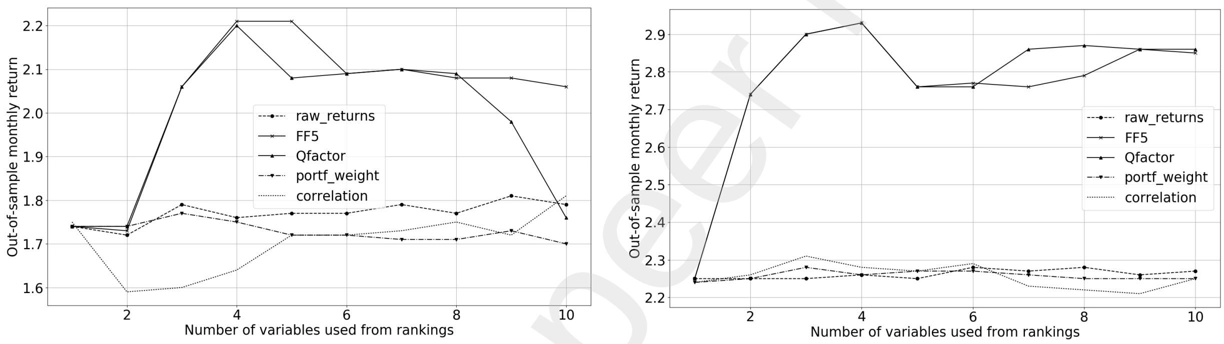


Figure 1: A blueprint for the production line of the machine-learning portfolio.



(a) **Train: 1980-1996, Test: 1997-2016** (b) **Train: 1980-1992, Test: 1993-2004**

Figure 2: Out-of-sample monthly returns for a number of variables from rankings of different ranking approaches.



(a) **Train: 1980-1996, Test: 1997-2016** (b) **Train: 1980-1992, Test: 1993-2004**

Figure 3: Out-of-sample monthly returns for a number of variables from rankings of different ranking approaches.

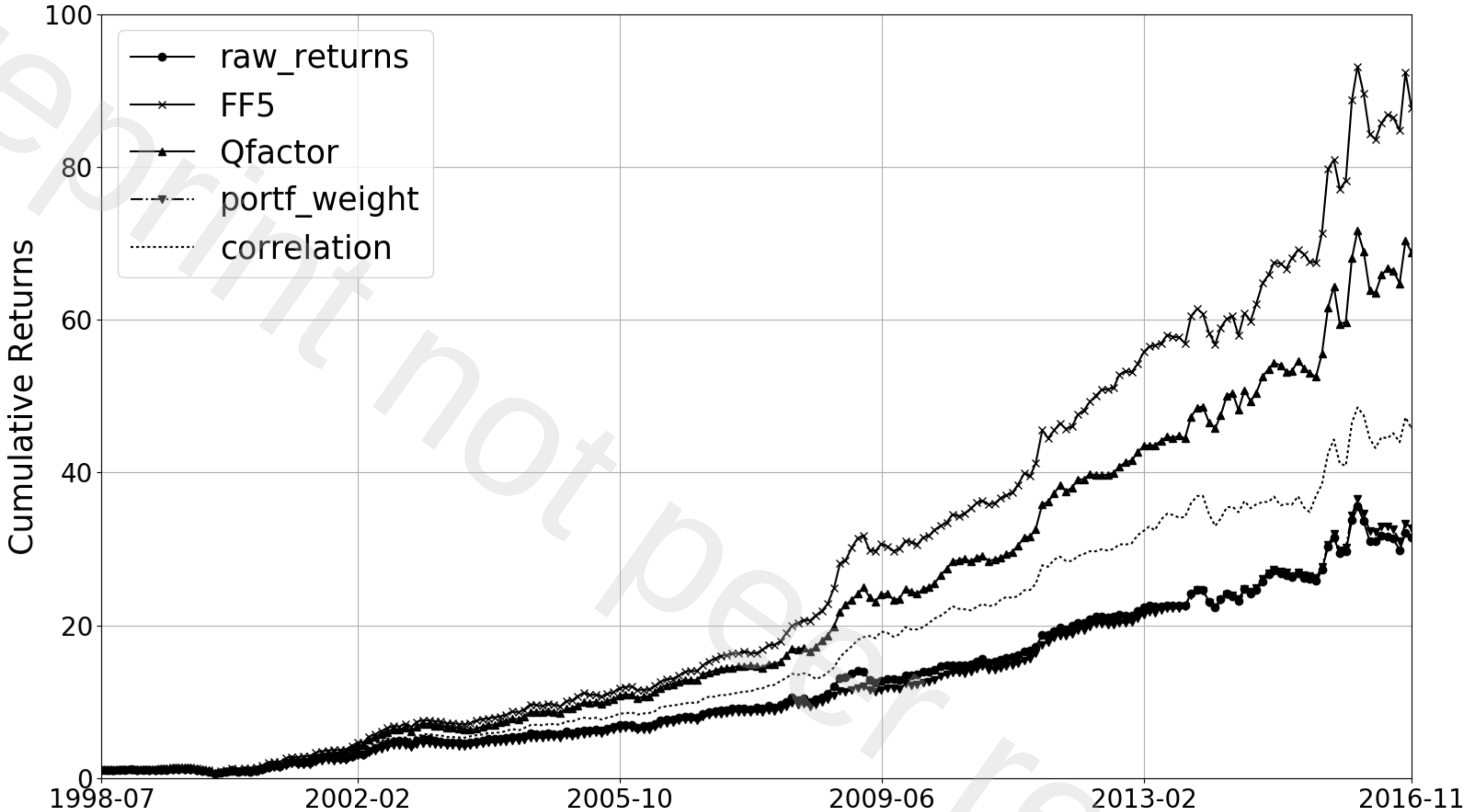


Figure 4: Out-of-sample long-short portfolio cumulated returns of different ranking approaches.

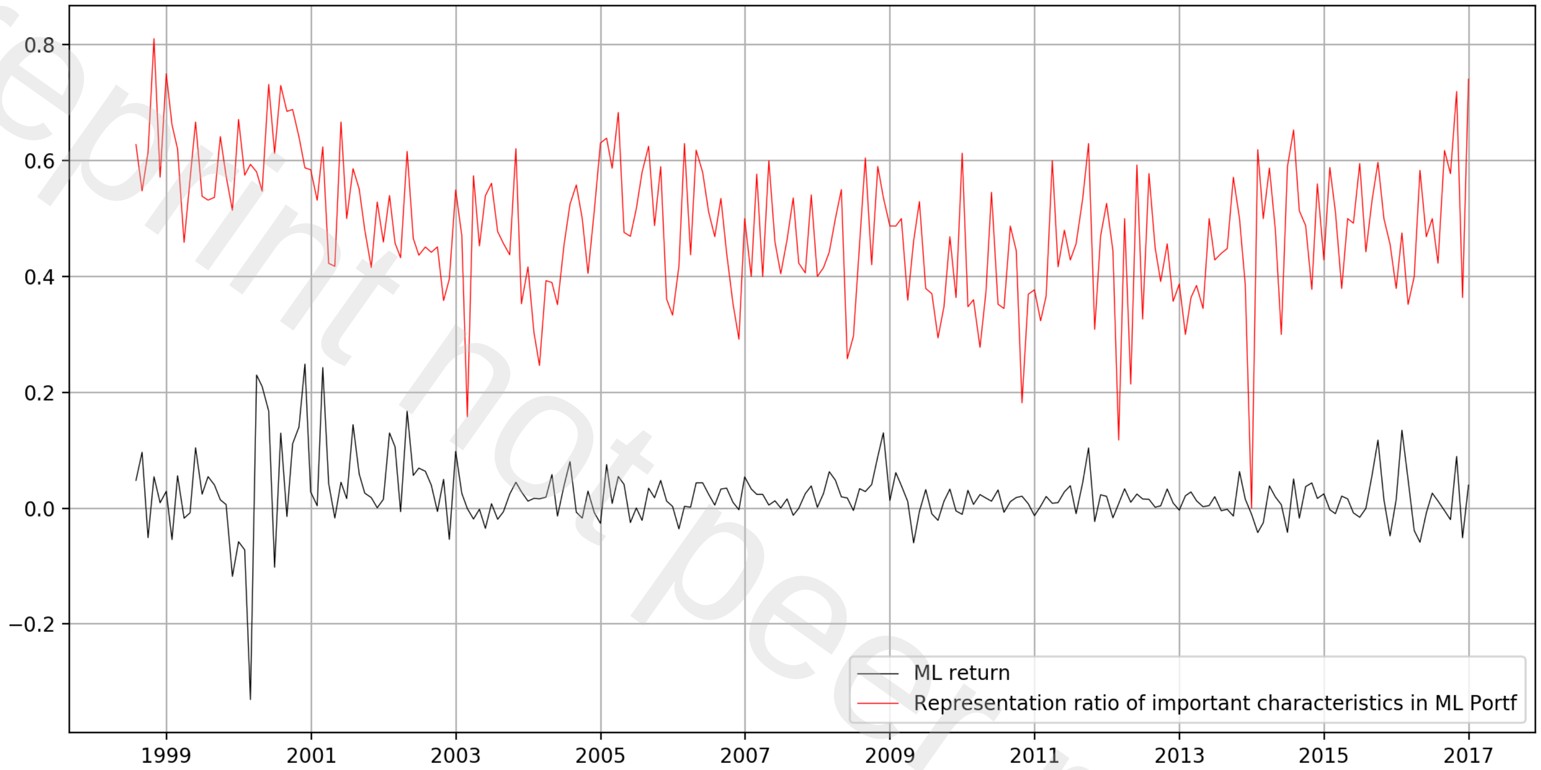


Figure 5: The ML portfolio monthly ‘hit-rate’ on characteristic portfolios with significant spread returns.

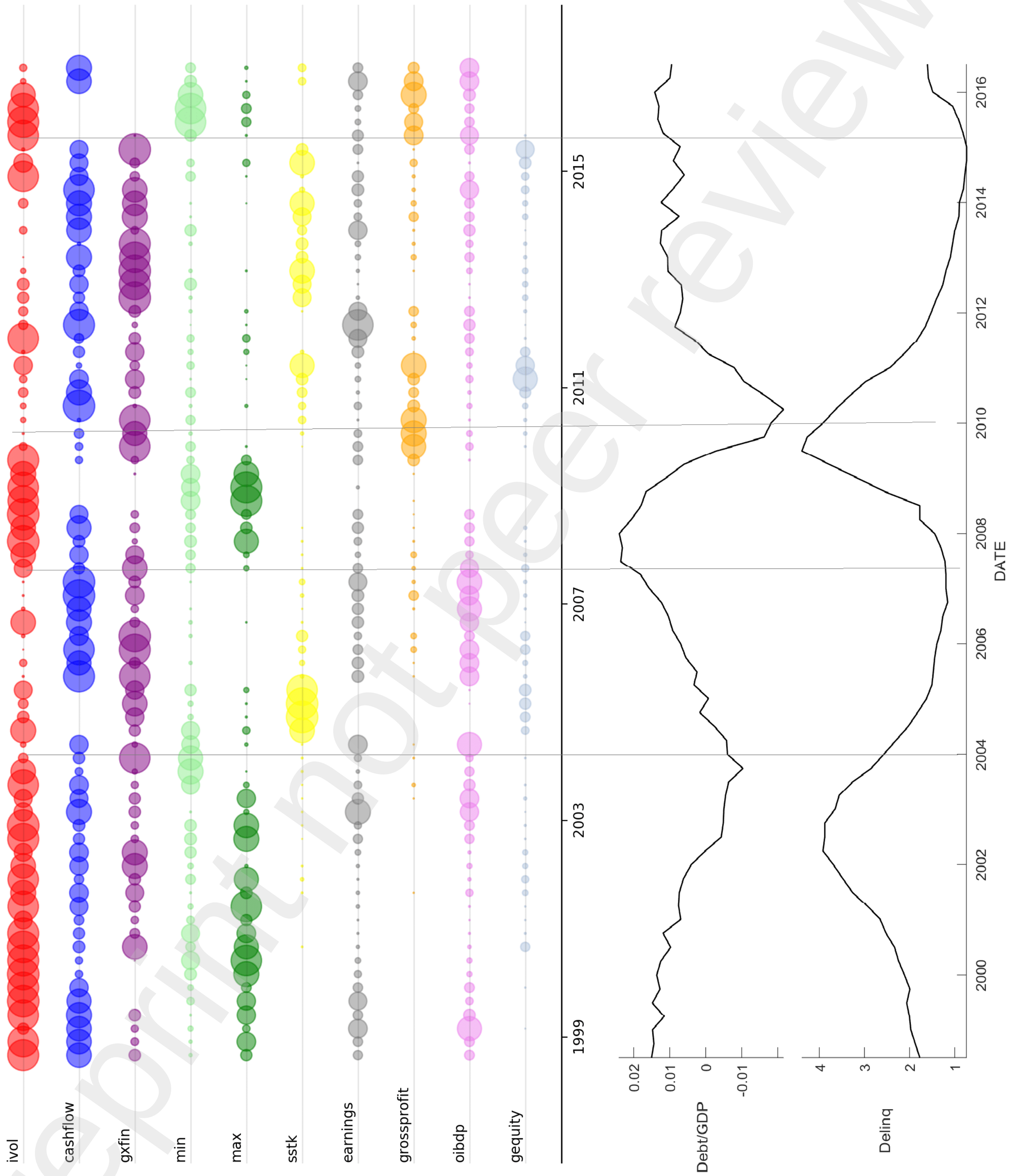


Figure 6: Aligning heatmaps with Debt/GDP and delinquency rate.

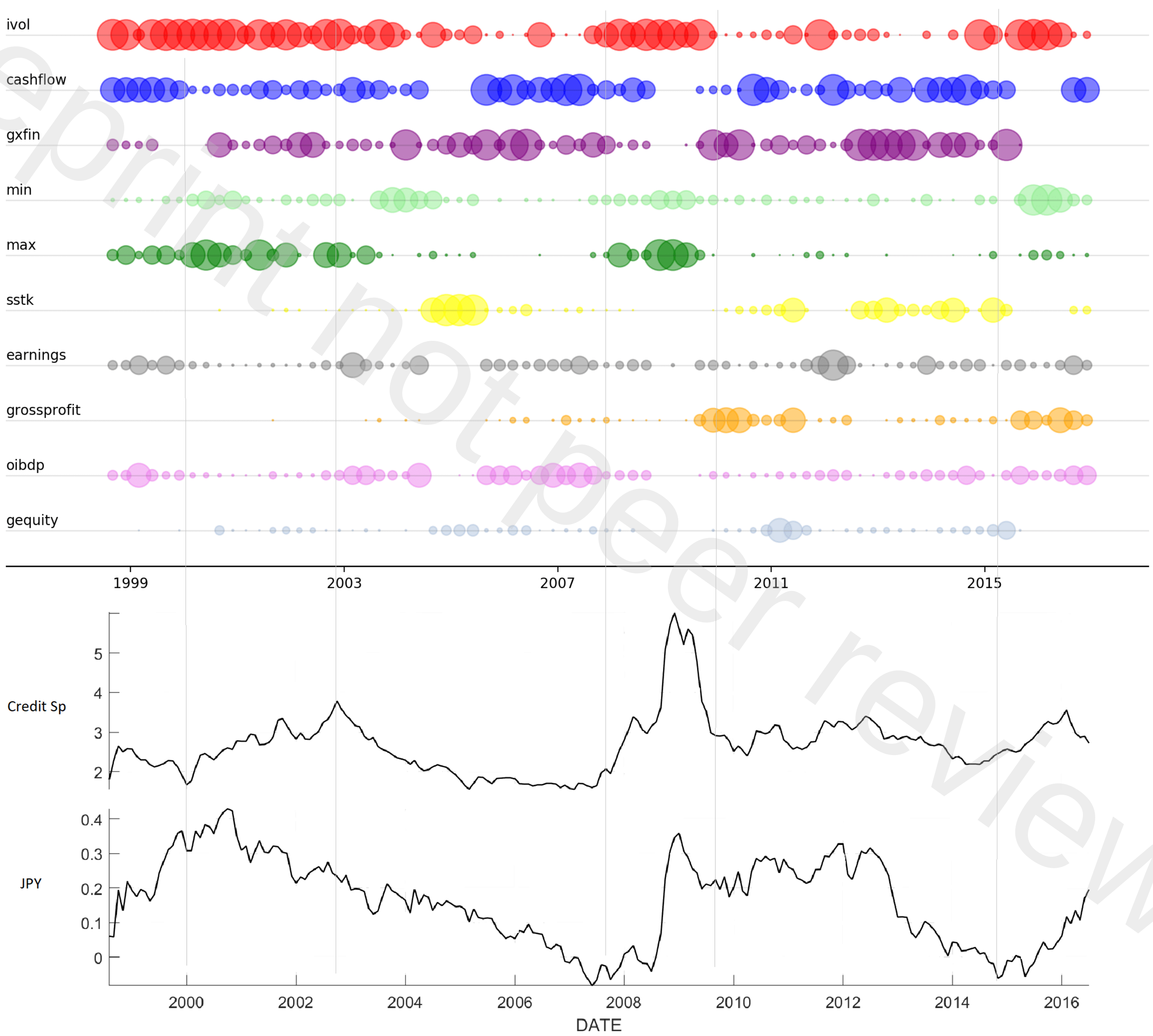


Figure 7: Aligning heatmaps with credit spread and value of JPY.

Appendix A1: Data description

The sample we use in this paper combines different data sources and spans over the period from July 1980 to December 2016. We obtain firm-specific accounting data, such as R&D expenditures, sales, and book equity from Compustat, and monthly stock re- turns, shares outstanding, and volume capitalization from Center for Research in Security Prices (CRSP). All common stocks trading on NYSE, AMEX, and NASDAQ with valid accounting and returns data are included in the sample. Firms need to be listed on Compustat for two years before including in our sample. We exclude financial firms, which have four-digit standard industrial classification (SIC) codes between 6000 and 6999 (finance, insurance, and real estate sectors). Similar to Fama and French (1992), we further discard closed-end funds, trusts, American Depository Receipts, Real Estate Investment Trusts, units of beneficial interest, and firms with negative book equity.

Following Harvey et al. (2015), we begin with 106 cross-sectional of stock return related factors. Within the 106 factors, 96 of them are accounting, 10 of them are financial. The details of the factors is provided in the list below.

Appendix A2: Variable definitions

Variables	Definition	Variables	Definition
ABCAPX	Abnormal capital investment	GLTNOA	Growth in long-term net operating assets
Acc_FWY	Indirect accruals	GNOA	Growth in net operating assets
Acc_FWY_pct	Indirect accruals percentage	gp	Gross Profit (Loss)
Accruals	Accruals	GrossProfit	Gross Profitability
Accruals_pct	Accruals percentage	GXFIN	External financing activities
aco	Current Assets – Other – Total	HHI	Industry concentration, SIC3
act	Current Assets - Total	HHI1	Industry concentration, Fama-French 48 industry
AdvEx	Advertising	ibc	Income before Extraordinary Items (Cash Flow)
Altman_Z	Altman Z-Score	intan	Intangible Assets - Total
Amihud	Iliquidity measure	invt	Inventories - Total
ao	Assets - Other	iskew	Fama-French 3 factor model residual skewness
ap	Accounts Payable - Trade	ivol	Fama-French 3 factor model residual volatility
AssetTangib	Asset tangibility	lco	Current Liabilities – Other – Total
BEME	book-to-market ratio	lct	Current Liabilities - Total
beta	CAPM beta each month	ln_at	Log value of total asset
bkvlps	Book Value Per Share	ln_me	Size
CapEx	Capital investment	ln_rdc	R&D capital
capx	Capital Expenditures	ln_xrd	Log value of Research and Development Expense
Cash	Cash holding	lo	Liabilities - Other - Total
Cashflow	Cash flow	M2B	Makret to book ratio
ceq	Common/Ordinary Equity - Total	max	Maximum daily return within each month
che	Cash and Short-Term Investments	min	Minimum daily return within each month
csho	Common Shares Outstanding	msa	Marketable Securities Adjustment
cumret11_1	Momentum, 12 month	NI	Net income (loss)
cumret6	Momentum, 6 month	NOA	Net operating asset
dcvt	Debt - Convertible	oibdp	Operating Income before Depreciation
dd3	Debt – Due in 3rd Year	PE	Price to earnings ratio
DE	Debt to equity ratio	PM	Profit margin
Debt3y	Proportion of total debt maturing within three years	ppent	Property, Plant and Equipment - Total (Net)
DebtConv	Convertible debt divided by total debt	PPEXA	Tangible asset
DebtSecu	Secured debt divided by total debt	prc_1	close price in last month end
DirtySurplus	Dirty Surplus	prstkc	Purchase of Common and Preferred Stock
dlc	Debt in Current Liabilities - Total	pstk	Preferred/Preference Stock (Capital) - Total
dlcch	Current Debt – Changes	pstkl	Preferred Stock – Liquidating Value
dltis	Long-Term Debt – Issuance	pstkrv	Preferred Stock – Redemption Value
dltr	Long-Term Debt – Reduction	RD	R&D investment
dltt	Long-Term Debt - Total	RDirtySurplus	Really dirty surplus
dm	Debt – Mortgages & Other Secured	RDS	R&D scaled by sales
dp	Depreciation and Amortization	RealEstate	Real Estate Holdings
DPM	Change of profit margin	rect	Receivables – Total
DROA	Change of ROA	recta	Retained Earnings – Cumulative Translation Adjustment
DSale	Sales growth	retadj_1	Return reversal
dv	Cash Dividends (Cash Flow)	ROA	Return on asset
dvc	Dividends Common/Ordinary	ROE	Return on equity
dvp	Dividends - Preferred/Preference	sale	Sale (net)
dvpsx_f	Dividends per Share – Ex-Date	sstk	Sale of Common and Preferred Stock
DXA	Leverage	Tobin_Q	Tobin'sQ
DY	Divident yeild	Turnover	Average daily turnover within month
Earnings	Earnings	txditc	Deferred Taxes and Investment Tax Credit
fcashflow	Free cashflow	txp	Income Taxes Payable
GAT	Asset growth	VROA1	Cash flow volatility, 3 years
GDebt	Debt financing	VROA2	Cash flow volatility, 5years
GEquity	Equity financing	xad	Advertising expense