BAYES BUSINESS SCHOOL
CITY ST GEORGE'S
UNIVERSITY OF LONDON

Degree:	BSc Business Management Cluster Bayes Business School
Project Title:	**Can machine learning models predict returns in the Vietnam equity market?**
Student Name:	**Van Duc Nguyen**
Supervisor's name:	**Siliang Wei**
Submission date:	11/05/2026

"By submitting this work, I certify that I have complied with the guidelines on plagiarism outlined in the Course Handbook in the production of this dissertation and that it is my own, unaided work"

Can machine learning models predict returns in the Vietnam equity market?

Final submission | 5,432 words

Acknowledgements

This dissertation has reminded me how important statistics, econometrics, and careful empirical thinking are in turning data into evidence. I still remember a quote introduced to me during the first year of my undergraduate degree, and this idea has stayed with me throughout the project:

"Data are widely available; what is scarce is the ability to extract wisdom from them."

Hal Varian (UC Berkeley and Chief Economist, Google)

I would like to thank the teachers who shaped my academic development at Bayes. I am grateful to Dr Harry Zhang, whose Quantitative Methods and Analytics class first helped me see how mathematics could be used to solve real-world problems. I would also like to thank Professor Ian Marsh for introducing me to asset pricing, Dr Parastoo Mousavi for introducing me to econometrics, and Dr Qinghua Li for introducing me to machine learning, which became an important inspiration for this dissertation.

I am especially grateful to my supervisor, Dr Siliang Wei, for his thoughtful feedback and guidance throughout the project. His advice helped me think more carefully and move beyond simply completing an undergraduate assignment towards conducting a more disciplined piece of academic research.

I would also like to thank my friend, Hsin Cheng Yeh, for patiently listening to my ideas, challenging my thinking, and helping me improve the project. Finally, I thank the participants at the SVUK & VIS Academic Conference: Vietnam 2045, who generously listened to my work and offered helpful feedback.

Contents

Acknowledgements **i**

Abstract **vi**

1 Introduction and Literature Review **1**

1.1 From classical asset pricing to the predictor-selection problem 1
1.2 Why machine learning enters empirical asset pricing 2
1.3 Why Vietnam is a substantive external-validity test 3
1.4 Contribution 4

2 Methodology **5**

2.1 Empirical object 5
2.2 Data inputs and predictor design 5
2.3 Rolling out-of-sample design 6
2.4 Model menu 6
2.5 Evaluation objects 8

3 Analysis and Findings **9**

3.1 Empirical setup 9
3.2 Stock-level forecast performance 10
3.3 Statistical comparison and economic significance 13
3.4 Predictor content 14
3.5 Discussion 18

4 Conclusion **20**

References **21**

A Data, Provenance, and Feature Construction **24**

A.1 Data provenance 24
A.2 Target construction and preprocessing logic 24
A.3 Bloomberg raw input fields 25
A.4 Feature construction 27
A.5 Exploratory review of the raw panel 29

B Model Implementation Backup **29**

B.1 Software and rolling mechanics 29
B.2 Estimator roles 29
B.3 Importance diagnostics 30
B.4 Comparison mechanics 30

C Backup Evidence and Empirical Detail **30**
C.1 Replication code . . . 30
C.2 Common-sample model comparison . . . 31
C.3 Full long-short table . . . 31
C.4 Long-only model portfolios and VN30-style market proxy . . . 33
C.5 Full predictor rankings . . . 35

List of Tables

1 Vietnam empirical setup 9
2 Full stock-level forecast performance on the common Vietnam sample . . 12
3 Incremental effect of the macro block relative to the matched baseline model 12
4 Diebold-Mariano comparison among leading baseline models 13
5 Condensed long-short evidence and market-reference benchmarks 14
6 Combined average predictor importance across all models 17
7 Bloomberg raw input fields at the start of the study 26
8 Feature construction and final model inputs 28
9 Full long-short decile table 32
10 Long-only model portfolios and VN30-style benchmark 34
11 Top-20 predictors by model 35

List of Figures

1 Rolling 60/24/12 train-validation-test design 10
2 Sorted stock-level forecast performance by model 11
3 Cumulative model long-short performance 15
4 Stock-level and portfolio effects of macro augmentation 15
5 Top-20 predictor importance in the leading baseline models 16
6 Average predictor importance across all models 17

Abstract

This study examines whether machine-learning methods can predict monthly stock returns in the Vietnamese equity market. Using a disciplined out-of-sample design, it compares linear benchmarks, shrinkage and dimension-reduction models, tree ensembles, and neural networks on a curated stock-characteristic panel, with a second specification that adds country-level macro factors. The evidence is positive but modest. Vietnam contains forecastable cross-sectional return variation, yet the best models form a narrow practical frontier rather than producing one decisive winner. Elastic net leads numerically on stock-level R^2_{OOS}, with random forest, PLS, OLS full, and gradient boosting close behind, while neural networks perform weakly. The macro overlay is heavily used by the models but usually does not improve stock-level fit. Economically, valuation and trading/liquidity variables carry most of the useful signal, and the strongest equal-weight long-short portfolio achieves an annualised Sharpe ratio of about 2.32, compared with about 0.37–0.38 for the VN30-style local market proxy over the same test window. Overall, the characteristic-based financial machine-learning approach of Gu et al. (2020) remains informative in Vietnam, but its strongest evidence appears in a local and moderate form.

1 Introduction and Literature Review

This study asks whether the recent machine-learning approach to empirical asset pricing remains useful when it is taken from the large U.S. setting to the Vietnamese equity market. The question is not whether a more flexible algorithm can mechanically improve in-sample fit. It is whether the broader empirical program associated with Gu et al. (2020), Kelly and Xiu (2023), and Giglio et al. (2022) remains informative when it is applied to a smaller market with thinner coverage and a less forgiving local structure of expected-return variation. Vietnam is therefore used as an external-validity test of a now-influential return-prediction framework.

The problem is naturally predictive. Researchers do not observe expected returns directly. What they observe is realized return, which mixes a conditional mean with a large amount of news and noise. Once the empirical question is phrased that way, return prediction becomes one of the main operational ways to study time-varying expected returns. The relevant issue is whether a disciplined predictive design can recover economically meaningful variation in expected returns from the information available at the end of month t (Gu et al., 2020; Kelly and Xiu, 2023).

That framing also clarifies why the numerical targets in this literature are usually modest. Expected returns are latent, realized returns are noisy, and the cross section is shaped by both firm-level information and common market states. A successful empirical design should therefore be judged not only by whether it produces a large headline forecast score, but by whether the ranking of models, signals, and economic interpretations remains coherent once the exercise is evaluated strictly out of sample.

Vietnam is a useful setting for this test because it is not a mechanical extension of the U.S. evidence. The market has a shorter modern history, thinner cross-sectional coverage, more severe liquidity frictions, and a different mix of institutional and microstructure conditions than the large developed markets that dominate the machine-learning literature. The paper is built around a narrow but substantive question: if a curated characteristic panel is evaluated out of sample in Vietnam, which model family is most useful, and does a country-level factor overlay add incremental forecasting value once the stock-level panel is already in place?

1.1 From classical asset pricing to the predictor-selection problem

The classical benchmark begins with portfolio choice. Markowitz (1952) shows that investors can be understood as choosing portfolios by trading off expected return against variance. The capital asset pricing model developed by Sharpe (1964), Lintner (1965), and Mossin (1966) translates that logic into an equilibrium statement in which expected excess return is tied to market risk. The attraction of that framework is obvious: it offers a compact economic relation between risk and return and provides a natural baseline for empirical work.

Its empirical limitations are also well known. Fama and French (2004) review the large literature documenting that the CAPM fails often enough in the data to be difficult to defend as a sufficient empirical model. The response in finance was not to abandon factor reasoning, but to expand it. The three-factor model introduced size and value (Fama and French, 1993), while the five-factor model added profitability and investment (Fama and French, 2015). This improved empirical fit, but it also changed the nature of the problem. Once the field moves beyond a small benchmark model, the empirical task becomes one of choosing among a wide set of candidate characteristics and factors that may be correlated, redundant, unstable, or all three.

That transition motivates the predictor design in this study. The issue is not simply whether more variables exist. The issue is how to organize a characteristic space in a way that remains economically interpretable and empirically disciplined. Kelly et al. (2019) are especially useful here because they argue that characteristics should not be treated as an arbitrary list of raw variables. Instead, they can be interpreted as structured summaries of variation in risk and expected return. Feng et al. (2020) reach a related conclusion from the factor side: once existing information is accounted for carefully, many proposed new factors do not survive serious post-selection testing. Together these papers argue for discipline rather than proliferation, which is why the present study uses a curated characteristic panel instead of feeding every available feature into the models.

1.2 Why machine learning enters empirical asset pricing

Machine learning enters empirical asset pricing not as a fashionable add-on, but because the return-prediction problem is already high-dimensional, weak-signal, and unlikely to be well described by a single linear specification. Gu et al. (2020) make this point directly. Their contribution is not simply to apply modern estimators to stock returns. It is to define conditional expected return estimation itself as a supervised learning problem in which nonlinearities and interactions may matter, but only if they survive strict out-of-sample evaluation.

That perspective is the conceptual backbone of the empirical design. It implies three commitments. First, the predictor set should be broad enough to cover the main characteristic families used in empirical finance, but not so unconstrained that the analysis collapses into a vendor-field sweep. Second, model comparison should be genuinely out of sample. Third, additional model flexibility should be tested empirically rather than assumed to be an improvement. These principles are restated in survey form by Kelly and Xiu (2023) and Giglio et al. (2022), both of whom emphasize that machine learning is useful in finance when it is combined with economic structure, disciplined validation, and credible model comparison.

This logic also explains why the study uses a restrained model menu rather than an exhaustive search over modern estimators. Kelly et al. (2024) show that complexity can improve return prediction when it is disciplined by shrinkage and judged out of sample. That result is important, but it is not a license to assume that deeper or less structured models will always dominate. In a smaller market such as Vietnam, the relevant question is whether the data support additional functional-form flexibility. The model menu is therefore designed to compare families that embody different views of the return-prediction surface: parsimonious linear structure, regularized linear structure, tree-based nonlinearity, and neural-network depth.

Vietnam provides a boundary-condition test for the wider literature. The practical question is not simply whether machine learning can be run in a smaller market. It is whether the same disciplined empirical program still yields a meaningful ordering of model classes and predictor blocks once coverage is thinner, liquidity is more uneven, and the local market is less likely to reward unrestrained complexity.

The machine-learning-in-finance literature is broader than the branch used here. One adjacent strand studies deep and no-arbitrage asset-pricing architectures that embed nonlinear estimation inside factor structure, as in Chen et al. (2024). Another uses alternative data, especially text, to recover return-predictive information from news and disclosures, as in Ke et al. (2019). These branches matter because they show the breadth of the field, but they are not the direct template for this paper. The current study stays within the characteristic-based cross-sectional return-prediction tradition and uses those adjacent papers only to position the project within the wider literature.

1.3 Why Vietnam is a substantive external-validity test

The literature gives good reasons not to assume one-for-one transfer from the U.S. evidence. In emerging markets more generally, Rouwenhorst (1999) shows that small stocks tend to outperform large stocks, value stocks outperform growth stocks, and turnover is strongly related to cross-sectional return differences. Yet the composition of expected-return variation is not identical across countries. Cakici et al. (2016) likewise find that value is relatively robust in emerging markets, whereas momentum is less uniformly rewarded. These papers justify the view that the general architecture of return predictability may travel, but the exact ranking of signals should not be assumed.

The Vietnam literature points in the same direction. Hoang and Phan (2019) show that liquidity is priced in the Vietnamese stock market and that adding a liquidity factor improves model fit. Huang et al. (2023) provide the most directly relevant local comparison for the present study. Their results show that a Vietnam-specific factor structure built around market, size, and earnings-price outperforms a direct local replication of the standard Fama-French benchmark. This is important for two reasons. First, it implies that value in Vietnam may be captured better by earnings-price than by book-to-market. Second, it suggests that local implementation details matter.

The local literature has developed, but the specific machine-learning question studied here remains underexplored. Existing Vietnam studies mostly compare factor structures or specific priced effects. They do not usually ask whether the same market supports tree-based nonlinearities, disciplined shrinkage, and deeper neural architectures under one unified out-of-sample design. This paper therefore sits between the local factor literature and the broader machine-learning asset-pricing literature rather than belonging fully to only one of them.

That conclusion is reinforced by broader Asian evidence. Liu et al. (2019) show that in China, size and value matter strongly, but only once the factor construction is adapted to the local market rather than copied directly from the U.S. procedure. Their key result is especially relevant here: earnings-price subsumes book-to-market in capturing the Chinese value effect. Vietnam is not China, and the Chinese evidence should not be treated as direct Vietnam evidence. But it is a strong regional comparison for the proposition that size and value structure in Asian markets may need market-specific implementation rather than mechanical Fama-French replication.

For that reason, the empirical design is a local test rather than a transplantation exercise. The stock-level characteristic panel is curated around valuation, momentum, trading and liquidity conditions, and basic risk proxies because these are the variables most plausibly linked to the local cross section. The country-level factor block is then added as a second-stage overlay rather than as the primary design. The resulting question is narrow and testable: once the stock-level panel is already in place, does common country-level factor information add stock-level predictive value?

1.4 Contribution

The paper makes three contributions. First, it evaluates whether the characteristic-based financial machine-learning approach of Gu et al. (2020) retains out-of-sample predictive value in Vietnam. Second, it compares linear, tree-based, and neural-network estimators in a market where the empirical evidence is still thin. Third, it asks whether country-level factor returns add incremental forecasting value once the stock-level characteristic panel is already present. The study is therefore best understood as a disciplined local test of a broader empirical asset-pricing framework.

The rest of the paper follows that logic. Section 2 turns the research question into a predictive design and defers technical construction detail to the appendices. Section 3 then evaluates the empirical evidence, separating stock-level forecast fit, statistical comparison, economic significance, and predictor content. Appendix A records data provenance, feature construction, and preprocessing rules; Appendix B records model formulas and methodological references; Appendix C records the full backup tables and empirical caveats.

2 Methodology

2.1 Empirical object

The empirical object is next-month stock-level excess return in Vietnam. Let $r_{i,t+1}$ denote the simple return on stock i over month $t+1$, and let r^f_{t+1} denote the one-month risk-free return derived from a daily Vietnam one-year government-bond benchmark collected from Capital IQ.[1] The prediction target is:

$$r^e_{i,t+1} = r_{i,t+1} - r^f_{t+1}. \tag{1}$$

Let $z_{i,t}$ denote the stock-level predictor vector observed at the end of month t. The study estimates:

$$\mathbb{E}_t\left[r^e_{i,t+1}\right] = g_\theta(z_{i,t}), \tag{2}$$

where the functional form $g_\theta(\cdot)$ varies by model family.

The exercise is a reduced-form conditional expected-return test in the spirit of Gu et al. (2020). Its purpose is to determine whether a disciplined predictor set produces useful out-of-sample stock-level forecasts in Vietnam, whether model family materially changes the answer, and whether a common macro factor block adds marginal predictive content beyond the stock-level panel (Kelly and Xiu, 2023).

2.2 Data inputs and predictor design

The baseline data object is a Bloomberg-derived stock-level panel exported directly from the Bloomberg Terminal, with the raw field inventory selected through the FLDS command and then transformed into a curated monthly characteristic set.[2] Sector dummies are added as controls. The final common sample retains fourteen continuous stock-level characteristics plus sector indicators.

The predictor set is kept close to the empirical asset-pricing literature so that the model comparison remains interpretable. This choice follows the logic of Kelly et al.

[1] The local risk-free series is an annualized percentage yield. Appendix A describes the conversion from daily yield observations to a monthly simple-return proxy.

[2] Appendix A reports the Bloomberg raw-input inventory, the downstream feature map, and the final inclusion or exclusion decisions. Firm age remains in the repaired panel, but it has only 13.43% non-missing coverage in the cleaned feature audit and is therefore excluded from the continuous model matrix.

(2019) and Feng et al. (2020): characteristics should be economically structured, and additional variables should be justified rather than mechanically added to the model matrix.

The experiment uses two variants of the same sample. The *baseline* variant contains stock-level characteristics only. The *macro* variant appends eight Vietnam country-level factor returns from the JKP Global Factor Data.[3] This matched-sample design makes the baseline-versus-macro comparison a marginal-information test.

2.3 Rolling out-of-sample design

The evaluation design is strictly time ordered. For each rolling split j:

$$W_j = \left(\mathcal{T}_j^{\text{train}}, \mathcal{T}_j^{\text{val}}, \mathcal{T}_j^{\text{test}}\right), \qquad |\mathcal{T}_j^{\text{train}}| = 60,\ |\mathcal{T}_j^{\text{val}}| = 24,\ |\mathcal{T}_j^{\text{test}}| = 12, \tag{3}$$

and the window advances by 12 months each time. In the Vietnam macro experiment this produces eight rolling splits and 96 scored test months.

Within each window, the estimator is fitted on the training block, tuned on the validation block, and refit on the combined training-validation sample before predictions are produced for the next test block. This prevents future information from entering model selection and forces the comparison to respect the time variation inherent in expected returns (Gu et al., 2020; Pesaran and Timmermann, 2007; Welch and Goyal, 2008).

The preparation layer rebuilds the next-month target from observed returns and standardizes the candidate predictors before estimation.

2.4 Model menu

The model menu is broad enough to test the main scientific issue, but narrow enough to remain interpretable.

The linear and robust-linear group contains ordinary least squares on the three benchmark predictors ('OLS-3'), the Huber version of that benchmark ('OLS-3 Huber'), the full ordinary least squares model ('OLS full'), and the full Huber model ('Huber full').[4] The shrinkage and dimension-reduction group contains elastic net (EN), principal component regression (PCR), and partial least squares (PLS). The nonlinear group contains

[3]The macro overlay is a compact month-level factor block spanning valuation, profitability, risk, and trading-friction signals. After those factors are merged, the baseline and macro panels are restricted to the same stock-month support so that the comparison isolates the marginal contribution of the macro block rather than a change in sample composition. Appendix A records the retained factor set and the common-sample logic.

[4]The 'OLS-3' benchmark uses size, value, and momentum, implemented as `log_mkt_cap`, `book_to_market`, and `mom1m`. It is included as a simple Fama-French-style comparison before the larger machine-learning specifications are evaluated.

random forest (RF) and gradient boosting regression trees (GBRT). The deep-learning group contains feedforward neural networks with one, two, and three hidden layers.

The core estimator forms are as follows. Linear models use:

$$\hat{r}^e_{i,t+1} = z_{i,t}^\top \hat{\theta}, \tag{4}$$

while the Huber variants replace squared loss with:

$$\hat{\theta} = \arg\min_\theta \sum_{i,t} H_\delta\left(r^e_{i,t+1} - z_{i,t}^\top \theta\right), \qquad H_\delta(u) = \begin{cases} \frac{1}{2}u^2, & |u| \le \delta, \\ \delta|u| - \frac{1}{2}\delta^2, & |u| > \delta. \end{cases} \tag{5}$$

Elastic net keeps the linear form but adds mixed shrinkage:

$$\hat{\theta} = \arg\min_\theta \left[\frac{1}{N} \sum_{n=1}^{N} (r^e_n - z_n^\top \theta)^2 + \lambda \left(\alpha \|\theta\|_1 + \frac{1-\alpha}{2} \|\theta\|_2^2 \right) \right]. \tag{6}$$

PCR and PLS summarize the predictor matrix through latent components. For PCR:

$$X = UDV^\top, \qquad T_M = XV_M, \qquad \hat{y} = \beta_0 + T_M \beta. \tag{7}$$

For PLS:

$$t_m = Xw_m, \qquad w_m = \arg\max_w \mathrm{Cov}^2(Xw, y), \qquad \hat{y} = \beta_0 + \sum_{m=1}^{M} \beta_m t_m. \tag{8}$$

The tree models estimate nonlinear functions either by averaging decorrelated trees:

$$\hat{g}_{\mathrm{RF}}(z) = \frac{1}{B} \sum_{b=1}^{B} T_b(z), \tag{9}$$

or by adding shallow trees sequentially:

$$\hat{g}_{\mathrm{GBRT}}(z) = \sum_{m=1}^{M} \nu f_m(z). \tag{10}$$

The neural networks use ReLU hidden activations with the feedforward recursion:

$$a^{(0)} = z, \qquad a^{(\ell)} = \phi(W_\ell a^{(\ell-1)} + b_\ell), \qquad \hat{r}^e = W_{L+1} a^{(L)} + b_{L+1}. \tag{11}$$

This menu compares model families that encode different views of how complicated the return-prediction surface is. If parsimonious linear models perform well, the signal is relatively stable and low dimensional. If tree methods dominate, nonlinearities and interactions matter. If neural networks add further value, the signal is more complex still. Model choice is therefore treated as an empirical question rather than a rhetorical commitment to complexity (Gu et al., 2020; Giglio et al., 2022; Kelly et al., 2024). Additional implementation notes and method references are collected in Appendix B.

2.5 Evaluation objects

The first evaluation object is stock-level forecast quality. The main summary statistic is the out-of-sample R^2 against a zero-expected-excess-return benchmark:

$$R^2_{OOS} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}^{\text{test}}} \left(r^e_{i,t+1} - \hat{r}^e_{i,t+1}\right)^2}{\sum_{(i,t)\in\mathcal{T}^{\text{test}}} \left(r^e_{i,t+1}\right)^2}. \tag{12}$$

RMSE, MAE, hit rate, and monthly Spearman information coefficients are reported alongside R^2_{OOS} because each captures a different dimension of forecasting performance.

The second evaluation object is economic usefulness. Each model produces monthly decile portfolios sorted on predicted returns, and the main spread is the top-minus-bottom portfolio:

$$r_{H-L,t} = r_{Q10,t} - r_{Q1,t}. \tag{13}$$

Stock-level point forecast quality and portfolio usefulness need not coincide. A model can improve point prediction only modestly and still meaningfully alter the cross-sectional ranking of securities.

The third evaluation object is relative model comparison. Leading models are compared using Diebold-Mariano tests under one-step squared-error loss:

$$d_n = e^2_{1,n} - e^2_{2,n}, \qquad DM = \frac{\bar{d}}{\sqrt{\widehat{\text{Var}}(\bar{d})}}. \tag{14}$$

Variable interpretation is handled with family-appropriate feature-importance diagnostics. For models where a zero-out diagnostic is available, the reported normalized importance is:

$$\Delta R^2_j = R^2_{\text{base}} - R^2_{-j}, \qquad VI_j = \frac{\Delta R^2_j}{\sum_k \Delta R^2_k}. \tag{15}$$

Tree models also retain their native importance measures, so the main text interprets feature content by broad predictor family rather than by treating all raw importance scales as directly comparable.[5]

Taken together, the design answers a focused empirical question: whether a disciplined characteristic-based return-prediction framework produces statistically and economically useful forecasts in Vietnam, and what that says about the local structure of expected-return variation.

[5] Appendix B records implementation notes and Appendix C records the backup empirical tables.

3 Analysis and Findings

This section answers the research question posed in Section 1 using the design in Section 2.[6] The aim is to evaluate whether the Vietnam evidence supports a disciplined machine-learning return-prediction framework, whether model family matters, and whether a country-level factor overlay adds marginal information beyond the stock-level characteristic panel.

3.1 Empirical setup

The empirical exercise combines valuation, size, momentum, trading and liquidity measures, basic risk proxies, and sector controls into a monthly Vietnam stock-level predictor set. The macro variant then appends a compact block of Vietnam country factors organized around valuation, profitability, risk, and trading frictions. Both variants are evaluated on the same stock-month support, so the baseline-versus-macro comparison isolates marginal information rather than changes in sample composition.

Table 1 summarizes the common Vietnam sample used throughout the manuscript, and Figure 1 shows the rolling evaluation design. The sample is large enough to permit meaningful model comparison, but it is still far smaller and thinner than the U.S. panels that dominate the modern machine-learning literature. That scale difference is important for interpretation: if complexity helps only selectively in Vietnam, the result may reflect both the local market structure and the smaller effective sample available for stable out-of-sample learning.

Table 1: Vietnam empirical setup

Item	Value
Analysis window	2008-09-30 to 2024-12-31
Months in common sample	184
Stock-month observations	99585
Stocks	695
Predictor count	23 baseline (14 stock-level + 9 sector dummies); 31 with macro
Rolling design	60 train / 24 validation / 12 test months
Number of rolling splits	8
Scored test months	96
Prediction target	Next-month stock-level excess return
Baseline versus macro comparison	Same stock-month rows; macro appends 8 country-factor series

The table reports the common sample used for both baseline and macro models.

[6] The full construction rules, feature map, backup evidence, and replication-code note are recorded in Appendix A and Appendix C.

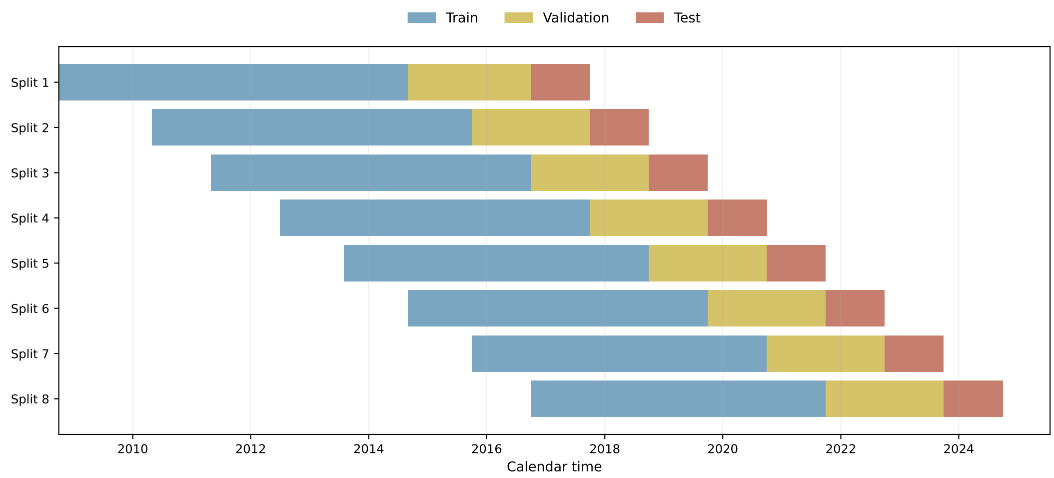

Figure 1: Rolling 60/24/12 train-validation-test design

Each horizontal bar shows one rolling window. The time ordering prevents future information from entering model selection.

3.2 Stock-level forecast performance

The full model leaderboard is reported in Table 2. Three patterns are clear. First, using the local Vietnam risk-free proxy, elastic net leads numerically on stock-level R^2_{OOS}, followed closely by random forest, PLS, OLS full, and gradient boosting. The substantive message is that the strongest models are baseline regularized-linear, dimension-reduction, and tree estimators rather than the deepest models in the menu. Second, the macro block usually worsens stock-level fit. The deterioration is especially visible for gradient boosting, random forest, elastic net, OLS full, PCR, and PLS. Third, the neural networks are weak throughout. All neural-network variants have negative R^2_{OOS}, and the macro-augmented versions are materially worse than the baseline networks.

This is a restrained but important result. Machine learning retains predictive value in Vietnam, but useful complexity is limited. Moderate flexibility helps, whereas deeper neural architectures do not translate their additional functional-form capacity into better out-of-sample forecasting performance. That pattern is consistent with the central empirical lesson in Gu et al. (2020): nonlinear structure can matter in return prediction, but only if it survives strict external evaluation. It also fits the broader argument in Giglio et al. (2022) that machine learning in asset pricing expands the toolkit without removing the basic need for regularization, interpretability, and genuine model comparison.

The magnitude of the leading R^2_{OOS} values should also be read carefully. They are positive, so the exercise contains predictive signal, but they remain small in absolute terms. That is not a weakness unique to Vietnam. Cross-sectional return prediction is a weak-signal problem even in the large U.S. datasets studied by Gu et al. (2020). In a smaller market with thinner coverage and stronger liquidity frictions, modest positive out-of-sample fit is a more realistic benchmark than large headline scores. The key evidence is the coherent ranking across model families, not a dramatic numerical improvement from any single estimator.

The table also helps separate genuine macro information from mechanical model expansion. If the country-level factor block were adding independent stock-level information, its effect should show up repeatedly across the matched baseline and macro pairs. Instead, the macro versions usually lose predictive fit, which suggests that the additional block is either too coarse for the stock-level task or largely overlapping with information already embedded in the firm-level panel.

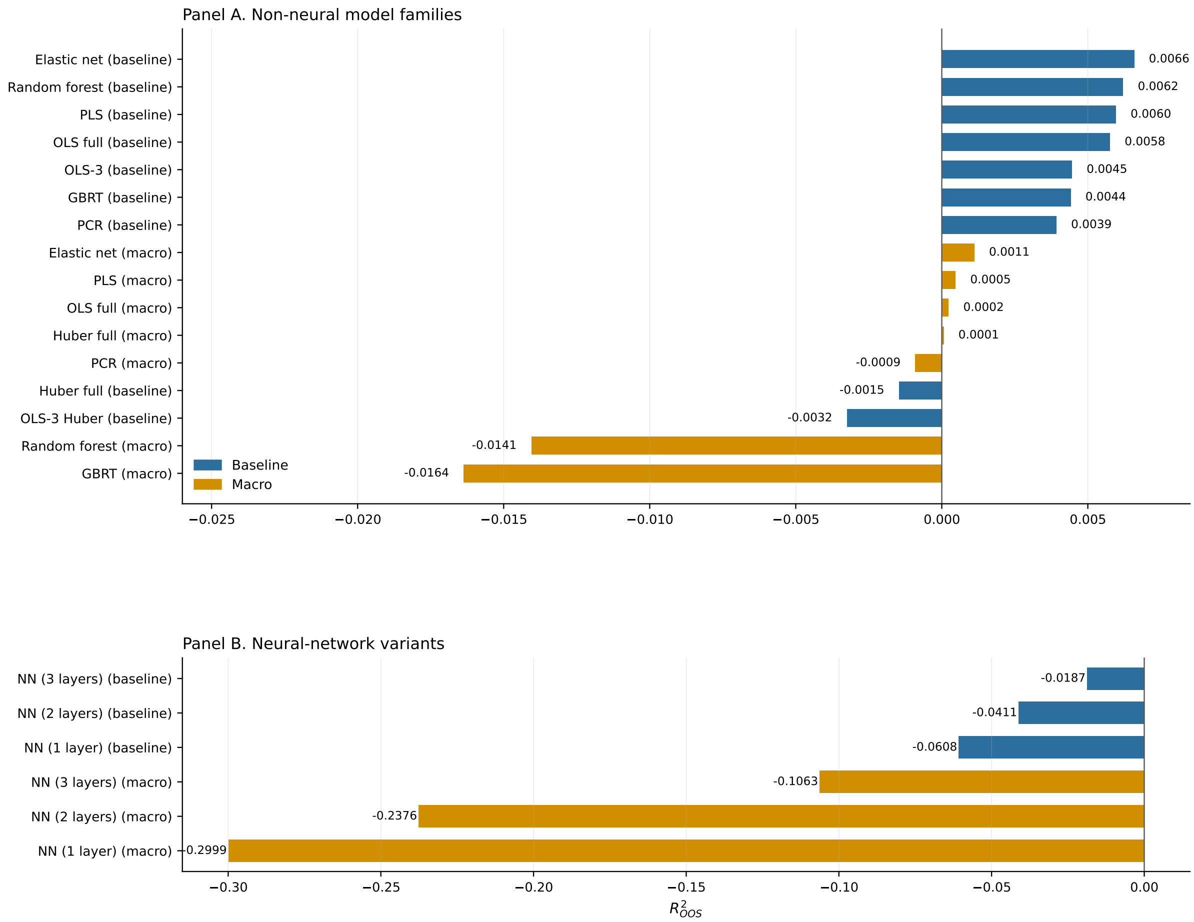

Figure 2: Sorted stock-level forecast performance by model

Bars report R^2_{OOS} from the common-sample model comparison. Panel A and Panel B use separate horizontal scales so that the neural-network losses do not compress the main non-neural comparison. Blue bars are baseline models and amber bars are macro-augmented models.

Table 3 isolates the macro question directly. The evidence is mostly negative for stock-level fit. In almost every family the macro variant raises RMSE and reduces R^2_{OOS}. The only material exception is the Huber-full family, where the differences are economically tiny. That result narrows the contribution of the macro block. Country-level factors are not irrelevant, but they do not appear to add much independent stock-level information once the firm-level panel is already carrying valuation, trading, and momentum signals.

Table 2: Full stock-level forecast performance on the common Vietnam sample

Model	RMSE	MAE	Hit rate	R^2_{OOS}	Mean IC	IC t-stat
`elastic_net_baseline`	0.1190	0.0752	0.4742	0.0066	0.0683	5.2668
`random_forest_baseline`	0.1190	0.0751	0.4390	0.0062	0.0710	6.0224
`pls_baseline`	0.1190	0.0756	0.4949	0.0060	0.0695	5.1342
`ols_full_baseline`	0.1190	0.0758	0.4965	0.0058	0.0719	5.5576
`ols3_baseline`	0.1191	0.0749	0.4548	0.0045	0.0366	2.3999
`gbrt_baseline`	0.1191	0.0746	0.5370	0.0044	0.0761	7.5001
`pcr_baseline`	0.1192	0.0755	0.4631	0.0039	0.0498	3.0570
`elastic_net_macro`	0.1193	0.0766	0.4652	0.0011	0.0689	5.5166
`pls_macro`	0.1194	0.0767	0.4629	0.0005	0.0552	3.6805
`ols_full_macro`	0.1194	0.0774	0.4863	0.0002	0.0719	5.5451
`huber_full_macro`	0.1194	0.0751	0.5417	0.0001	0.0727	7.4224
`pcr_macro`	0.1194	0.0766	0.4415	-0.0009	0.0034	0.2463
`huber_full_baseline`	0.1195	0.0743	0.5674	-0.0015	0.0728	7.5191
`ols3_huber_baseline`	0.1196	0.0736	0.6020	-0.0032	0.0230	1.3658
`random_forest_macro`	0.1202	0.0763	0.4330	-0.0141	0.0554	4.2500
`gbrt_macro`	0.1204	0.0753	0.5484	-0.0164	0.0549	5.0196
`nn_3layer_baseline`	0.1205	0.0785	0.5103	-0.0187	0.0540	7.2898
`nn_2layer_baseline`	0.1218	0.0808	0.5053	-0.0411	0.0447	6.1603
`nn_1layer_baseline`	0.1230	0.0826	0.5070	-0.0608	0.0337	4.7321
`nn_3layer_macro`	0.1256	0.0855	0.5029	-0.1063	0.0291	2.4409
`nn_2layer_macro`	0.1328	0.0941	0.5105	-0.2376	0.0219	1.5852
`nn_1layer_macro`	0.1361	0.0972	0.5006	-0.2999	0.0213	1.7100

Lower RMSE and MAE are better. Higher R^2_{OOS} and mean IC are better. The months column is omitted because all models are evaluated on the same prediction rows and the same 96 scored test months. 'OLS-3' and 'OLS-3 Huber' are shown once because these parsimonious benchmarks do not admit the macro block.

Table 3: Incremental effect of the macro block relative to the matched baseline model

Base model	Δ RMSE	Δ MAE	Δ hit rate	ΔR^2_{OOS}	Δ mean IC
`elastic_net`	0.0003	0.0014	-0.0090	-0.0055	0.0006
`gbrt`	0.0012	0.0006	0.0114	-0.0208	-0.0213
`huber_full`	-0.0001	0.0008	-0.0257	0.0015	-0.0001
`nn_1layer`	0.0132	0.0146	-0.0064	-0.2392	-0.0124
`nn_2layer`	0.0110	0.0133	0.0053	-0.1965	-0.0228
`nn_3layer`	0.0051	0.0070	-0.0074	-0.0876	-0.0249
`ols_full`	0.0003	0.0016	-0.0102	-0.0055	-0.0001
`pcr`	0.0003	0.0011	-0.0216	-0.0048	-0.0464
`pls`	0.0003	0.0011	-0.0321	-0.0055	-0.0142
`random_forest`	0.0012	0.0012	-0.0059	-0.0203	-0.0157

Negative values for ΔR^2_{OOS} indicate that the macro version underperforms the matched baseline on stock-level fit.

3.3 Statistical comparison and economic significance

Point estimates alone are not enough at the top of the leaderboard, so Table 4 reports a compact Diebold-Mariano comparison among the leading baseline contenders. The inferential result is narrow but useful. Elastic net and random forest are not statistically distinguishable from one another on common prediction rows, even though elastic net leads numerically. Several other pairwise differences are statistically visible because the common prediction sample is large, but their economic size is small. Huber-full is materially weaker than the leading group. OLS full, PLS, gradient boosting, random forest, and elastic net remain credible models, but the data do not support the stronger claim that any one of them dominates the entire practical frontier.

That matters for interpretation: the Vietnam data identify a small frontier of practical models rather than one unambiguous winner. This is the more defensible scientific conclusion, and it is preferable to overstating small ranking differences that are not statistically stable.

Table 4: Diebold-Mariano comparison among leading baseline models

Row Col	EN	GBRT	RF	OLSF	PLS	HUB
EN	0	-2.585***	-0.619	-2.249**	-1.981**	-9.768***
GBRT	2.585***	0	2.282**	1.451	1.719*	-7.080***
RF	0.619	-2.282**	0	-0.608	-0.359	-7.294***
OLSF	2.249**	-1.451	0.608	0	0.745	-7.960***
PLS	1.981**	-1.719*	0.359	-0.745	0	-8.137***
HUB	9.768***	7.080***	7.294***	7.960***	8.137***	0

Entries report the DM statistic with significance stars only: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Pairwise comparisons use the common prediction rows shared by the compared models.

The portfolio evidence partly confirms and partly complicates the forecast-ranking evidence. Table 5 shows that equal-weighted long-short performance is strongest for several baseline models, especially random forest, gradient boosting, and elastic net. Value-weighted performance is more mixed. Some macro linear variants, especially 'PLS (macro)' and 'Elastic net (macro)', improve the long-short sorting of larger stocks even though the macro block usually worsens RMSE and R^2_{OOS}.

This distinction matters because stock-level fit and implementation-oriented sorting value are related but not identical objects. A model can improve stock-level prediction only modestly and still reorder the cross section enough to affect decile spreads. The macro block therefore appears less useful for forecasting the level of stock returns than for occasionally altering the relative ordering of larger names.

This split between statistical and economic performance is one of the more informative features of the Vietnam results. The baseline frontier remains the most robust group overall, which keeps the main conclusion intact. But the value-weighted spread results also warn against reading the forecast table too mechanically. A macro variable can be weak at explaining the level of individual stock returns and still help in distinguishing larger firms on the margin. The macro block therefore contributes in a narrower and less reliable way than its feature rankings alone would suggest.

Table 5: Condensed long-short evidence and market-reference benchmarks

Panel A. Model long-short spreads

Model	EW mean	VW mean	EW Sh.	VW Sh.
RF (base)	0.0296	0.0162	2.1743	0.7932
GBRT (base)	0.0277	0.0098	2.3227	0.6775
EN (base)	0.0273	0.0199	1.7547	0.8789
PLS (base)	0.0284	0.0256	1.8258	1.0804
Huber full (base)	0.0240	0.0144	2.0332	0.8562
RF (macro)	0.0196	0.0073	1.4365	0.4320
GBRT (macro)	0.0213	0.0101	1.8035	0.5184
EN (macro)	0.0297	0.0251	1.9194	1.1659
PLS (macro)	0.0269	0.0324	1.6657	1.2873
Huber full (macro)	0.0246	0.0166	2.1248	1.0253

Panel B. Long-only market-reference benchmarks

Series	Type	Mean	Excess	Vol.	Sharpe	Cum.
VN30 EW	VN long-only	0.0089	0.0070	0.0656	0.3700	0.9013
VN30 VW	VN long-only	0.0090	0.0071	0.0639	0.3845	0.9438

Panel A reports model top-minus-bottom decile spreads. Panel B reports VN30-style local long-only market-reference benchmarks over the same 96 months. The VN30-style series are constructed from the study's Vietnam stock universe by selecting the largest 30 eligible firms and rebalancing semi-annually, because a consistent full-period constituent-level VN30 return series was not available from the project data sources. The complete 24-model long-short table and long-only model comparison are reported in Appendix C.

3.4 Predictor content

The predictor rankings make it possible to say more than which model wins.[7] They show which kinds of information the models actually use once the estimators are forced through the same out-of-sample design. Figure 5 visualizes the top 20 predictors in the leading baseline models, while Table 6 aggregates normalized feature shares across all 24 models. The full model-by-model top-20 ranking is reported in Appendix C.5.

Three features of the combined ranking are especially informative. First, valuation dominates the baseline signal. Book-to-market appears in every model and has the highest average share overall, while earnings-price is the strongest secondary valuation measure

[7] Within each model, predictor importance is normalized to sum to one. For tree models, the combined exhibits use the comparable zero-out R^2 drop when available rather than pooling incompatible raw importance scales across model families.

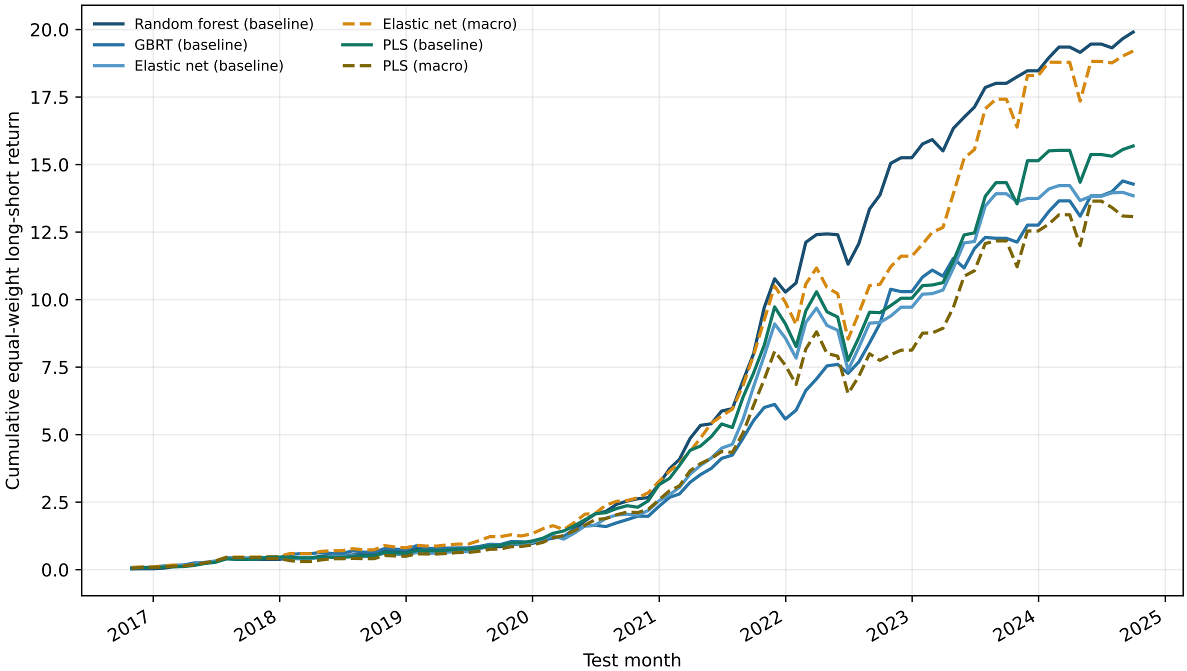

Figure 3: Cumulative model long-short performance

Solid lines are baseline model long-short spreads and dashed lines are matched macro variants. Local market-reference benchmarks are kept in Table 5 rather than plotted here because they are long-only series and are not directly comparable to model long-short spreads.

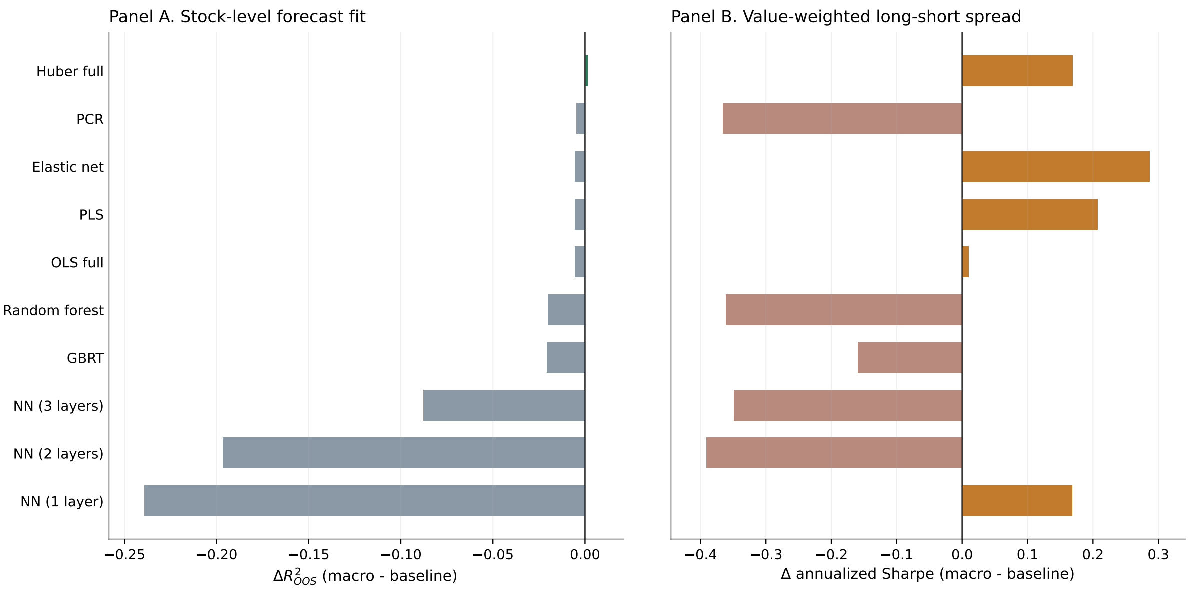

Figure 4: Stock-level and portfolio effects of macro augmentation

The left panel shows the change in stock-level R^2_{OOS} after adding the macro block. The right panel shows the change in value-weighted long-short Sharpe relative to the matched baseline model.

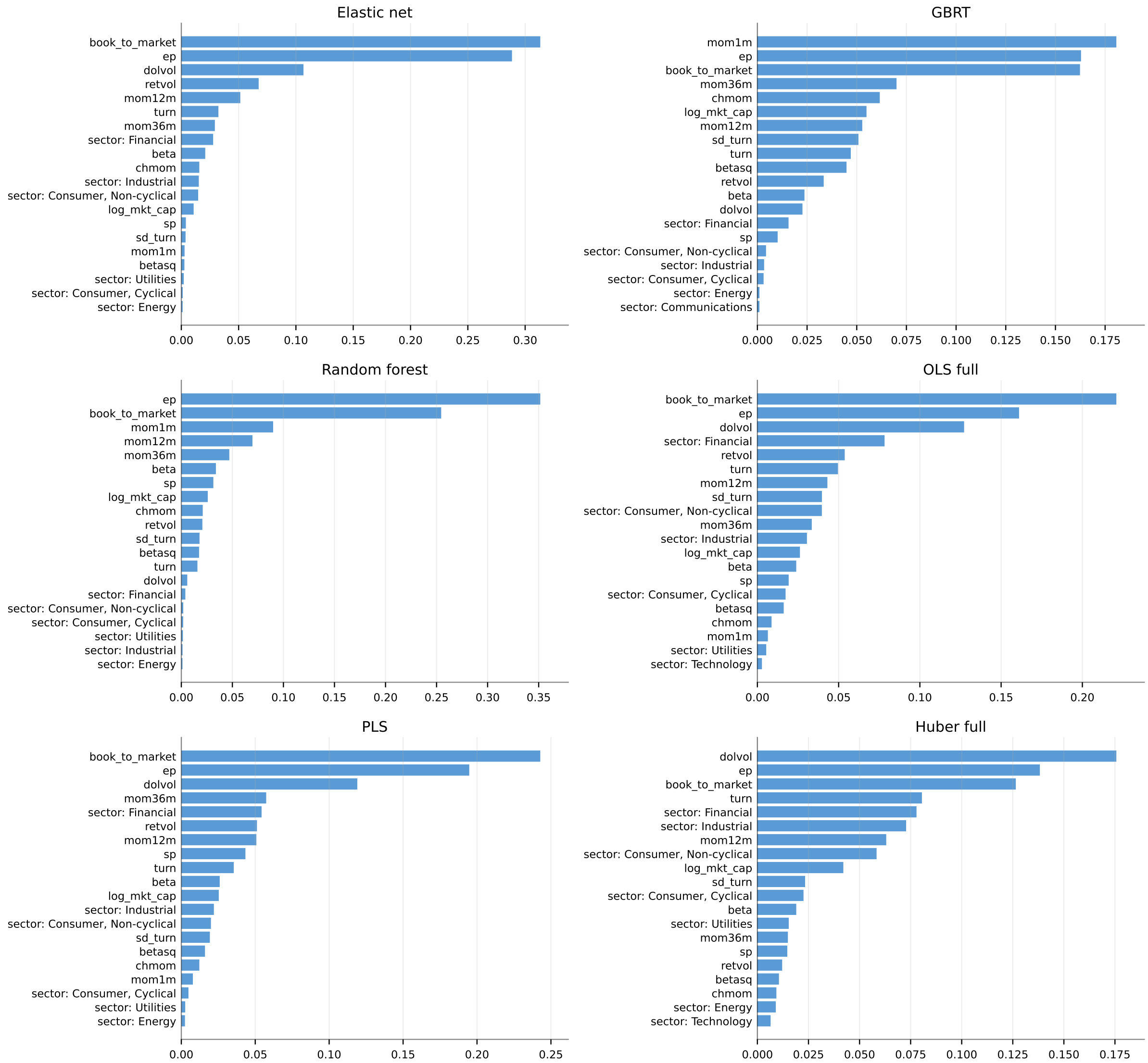

Figure 5: Top-20 predictor importance in the leading baseline models

Each panel reports the 20 most important predictors for one of the six leading baseline models. Bars show normalized within-model importance shares and are intended for within-model comparison only.

Table 6: Combined average predictor importance across all models

Feature	Family	Mean share	Median share	Models present
book_to_market	Valuation	0.2182	0.1211	24
ep	Valuation	0.1019	0.0688	20
mom12m	Momentum	0.0297	0.0269	20
mom36m	Momentum	0.0273	0.0146	20
mom1m	Momentum	0.0270	0.0071	24
dolvol	Trading/liquidity	0.0631	0.0375	20
turn	Trading/liquidity	0.0303	0.0262	20
macro_rvol_21d	Macro risk	0.0877	0.0802	10
macro_qmj_safety	Macro risk	0.0675	0.0692	10
macro_beta_dimson_21d	Macro risk	0.0580	0.0509	10
macro_be_me	Macro valuation/profitability	0.2031	0.2084	10
macro_bev_mev	Macro valuation/profitability	0.1279	0.1316	10
macro_qmj_prof	Macro valuation/profitability	0.0721	0.0666	10
macro_ocf_at_chg1	Macro valuation/profitability	0.0613	0.0588	10
macro_zero_trades_21d	Macro trading friction	0.0877	0.0733	10
betasq	Risk	0.0242	0.0110	20
retvol	Risk	0.0239	0.0127	20
beta	Risk	0.0239	0.0216	20
log_mkt_cap	Size	0.0556	0.0251	24
sector_Financial	Sector	0.0278	0.0201	20

Each row reports the mean and median normalized importance share across all 24 models and the number of models in which the feature appears in the top-20 ranking.

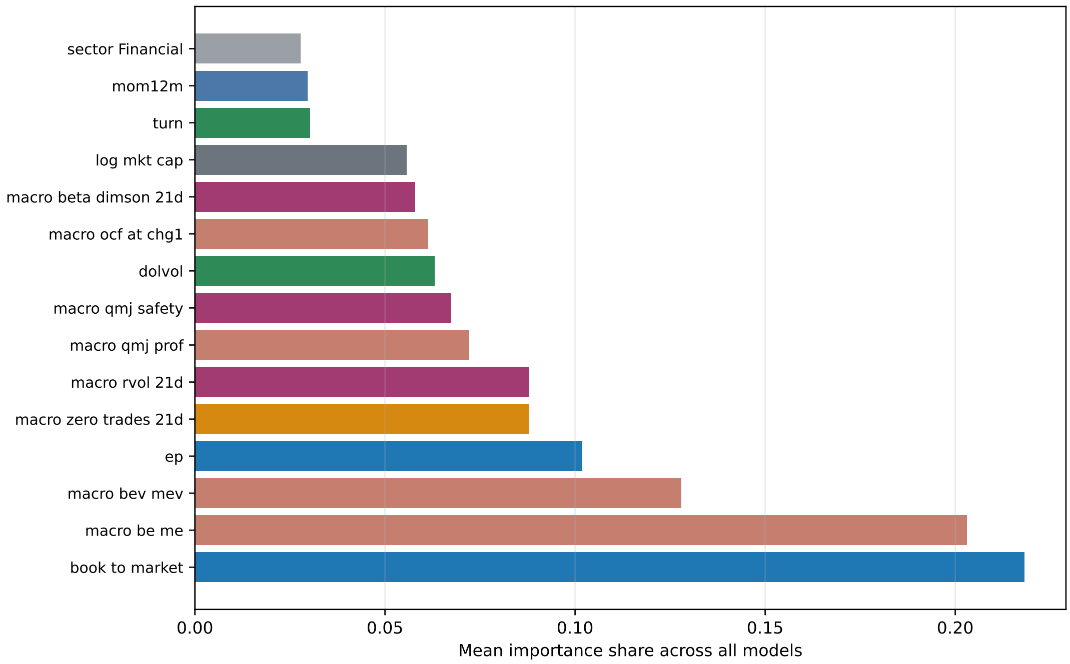

Figure 6: Average predictor importance across all models

The figure plots the top combined predictors by mean normalized importance share across all models. It is intended for within-experiment comparison only.

and is present in 20 of the 24 specifications. This is only partially aligned with the local evidence in Huang et al. (2023) and the broader emerging-market evidence in Liu et al. (2019). Those papers argue that earnings-price is the stronger value signal in Vietnam and China, whereas in the present machine-learning ranking earnings-price remains important but book-to-market still carries the largest average importance.

Second, trading and liquidity measures remain central. Dollar volume, turnover, and turnover volatility all rank near the top of the combined table. This is not a cosmetic result. It helps connect the machine-learning evidence back to a core market feature of Vietnam: the cross section is shaped not only by valuation differences, but also by trading conditions and implementation frictions. That interpretation is consistent with Hoang and Phan (2019), who show that liquidity is priced in Vietnam, and with Rouwenhorst (1999), who documents turnover-related return patterns across emerging markets. More broadly, it also fits the emerging-market evidence in Cakici et al. (2016), where value effects are more robust than momentum effects.

Third, the macro models use the country-factor block heavily even though the macro block rarely improves stock-level fit. In the combined table, `macro_be_me`, `macro_bev_mev`, `macro_rvol_21d`, `macro_zero_trades_21d`, and `macro_qmj_prof` all appear near the top. The macro variables therefore matter inside the fitted models, but they seem to summarize broad market states that overlap with information already latent in the stock-level characteristic panel. Feature importance is informative only when read alongside forecast performance: variable use alone is not evidence of incremental stock-level forecasting value.

The useful comparison with Gu et al. (2020) is therefore at the predictor-family level rather than as a one-to-one ranking of variables. In the U.S. evidence, machine learning draws on a broad characteristic set in which price ratios, recent-return measures, size, volatility, and trading-related variables all matter. The Vietnam results are narrower. Valuation and trading/liquidity carry the clearest baseline signal, momentum is present but secondary, and the country-factor block becomes important only inside the macro specifications without reliably improving stock-level fit. This difference is consistent with Vietnam being a smaller and more liquidity-constrained market rather than a direct replica of the U.S. cross section.

3.5 Discussion

Taken together, the results support a restrained external-validity claim. The Vietnam evidence agrees with Gu et al. (2020) and the related financial machine-learning literature on one central point: return prediction benefits from flexible estimators only when that flexibility is disciplined by the data and judged out of sample (Kelly and Xiu, 2023). The leading models are not the most complex ones in the menu. They are the regularized-linear, dimension-reduction, and tree specifications that can absorb moderate structure without becoming too unstable for the available sample.

At the same time, the Vietnam results are weaker than the strongest U.S. machine-learning evidence. That difference is substantively meaningful. The common Vietnam sample contains 99,585 prepared stock-month rows and 61,575 prediction rows per model, whereas the U.S. evidence in Gu et al. (2020) is built on a much deeper and longer stock universe. The local market is smaller, the effective sample is thinner, and the trading environment is less forgiving. Under those conditions, modest positive R^2_{OOS} values and a narrow frontier of competitive models are more plausible than dramatic performance gaps. The evidence supports a local version of the characteristic-based machine-learning framework rather than a claim that Vietnam replicates the quantitative scale of U.S. return-prediction results.

The predictor evidence is also consistent with both Vietnam-specific and broader emerging-market studies, but only in a partial sense. Valuation is the dominant signal, and earnings-price is clearly important, which is directionally consistent with Huang et al. (2023). However, book-to-market remains the single most important feature on average in the present design, so the local ranking does not fully match the stronger earnings-price result in that paper. Trading and liquidity variables remain important, which is consistent with Hoang and Phan (2019) in Vietnam and with the emerging-market regularities documented by Rouwenhorst (1999) and Cakici et al. (2016). Momentum is present but secondary, which is also plausible in light of the mixed momentum evidence outside the large developed markets.

The macro overlay is where the Vietnam results diverge most clearly from a simple "more information is better" reading of financial machine learning. The country-factor block is often selected by the models and ranks highly in the combined feature tables, yet it usually weakens stock-level fit. The most defensible interpretation is not that the macro variables are irrelevant. It is that broad market-state information and stock-level return prediction are different objects. The macro factors appear to summarize common states that are already partially reflected in valuation, trading, and momentum characteristics at the firm level. This is why macro augmentation can occasionally help value-weighted spreads without improving the underlying stock-level forecast problem.

These findings also sharpen how complexity should be interpreted in smaller markets. Kelly et al. (2024) show that complexity can be valuable when it is disciplined properly. In the present sample, however, additional network depth and a heavier common-factor overlay do not yield robust gains. The more useful lesson is that moderate nonlinearity is enough to matter, while excessive flexibility is not reliably rewarded.

The conclusion is deliberately specific. A disciplined characteristic panel travels to Vietnam, but in local form: the signal is modest, valuation and trading dominate the economics, regularized-linear, dimension-reduction, and tree models form the practical frontier, and macro overlays cannot replace stock-level signals.

4 Conclusion

This study examines whether the recent machine-learning approach to empirical asset pricing retains out-of-sample value in the Vietnamese equity market. The answer is qualified but positive. Vietnam contains forecastable monthly cross-sectional return structure, and a disciplined characteristic-based design can recover some of it. The signal is not large, and the leading estimators form a practical frontier rather than one definitive winner.

Using a local Vietnam risk-free-rate proxy, elastic net leads numerically on stock-level fit, with random forest, PLS, OLS full, and gradient boosting close behind. These models outperform the neural networks and generally dominate the macro-augmented variants on stock-level fit. The evidence therefore supports a restrained interpretation of financial machine learning in this setting: moderate flexibility helps, but complexity remains useful only when it is disciplined by the size and stability of the sample.

The economic content of the predictors is also coherent. Valuation dominates the baseline signal, especially through earnings-price and book-to-market. Trading and liquidity variables remain important, while momentum is present but secondary. This pattern aligns more closely with Vietnam-specific and emerging-market evidence than with the idea that the local market is a simple copy of the U.S. factor structure.

The macro overlay contributes less than the stock-level panel. Country-level factor returns are clearly used by the macro models, but they rarely improve stock-level R^2 and only selectively improve value-weighted spread performance. Common macro information is therefore a weak marginal addition once the stock-level cross section is already described by valuation, trading, and momentum information.

Future work can extend the design in three directions. First, richer risk monitoring through rolling drawdown, turnover, factor exposure, and stress/backtest diagnostics would show whether the model rankings are stable across different market states. Second, further feature engineering could test alternative accounting transformations, liquidity measures, and interactions between firm characteristics and market conditions. Third, newer methods such as quantum-enhanced feature maps may become relevant as part of the wider machine-learning toolkit, although their role in empirical asset pricing remains an open research question (Havlíček et al., 2019).

Overall, the results support a local version of the characteristic-based financial machine-learning framework. A disciplined characteristic panel travels to Vietnam, but it travels in a narrower form: the useful signal is modest, valuation and trading conditions dominate the economics, and the market rewards controlled flexibility rather than unrestricted depth.

References

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng (2016). *TensorFlow: A system for large-scale machine learning.* arXiv preprint arXiv:1605.08695. URL: `https://arxiv.org/abs/1605.08695`.

Breiman, L. (2001). "Random Forests". *Machine Learning* 45.1, pp. 5–32. DOI: `10.1023/a:1010933404324`.

Cakici, N., Y. Tang, and A. Yan (2016). "Do the size, value, and momentum factors drive stock returns in emerging markets?" *Journal of International Money and Finance* 69, pp. 179–204. DOI: `10.1016/j.jimonfin.2016.06.001`.

Chen, L., M. Pelger, and J. Zhu (2024). "Deep Learning in Asset Pricing". *Management Science* 70.2, pp. 714–750. DOI: `10.1287/mnsc.2023.4695`.

de Jong, S. (1993). "SIMPLS: An alternative approach to partial least squares regression". *Chemometrics and Intelligent Laboratory Systems* 18.3, pp. 251–263. DOI: `10.1016/0169-7439(93)85002-x`.

Diebold, F. X. and R. S. Mariano (1995). "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics* 13.3, pp. 253–263. DOI: `10.1080/07350015.1995.10524599`.

Fama, E. F. and K. R. French (1993). "Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics* 33.1, pp. 3–56. DOI: `10.1016/0304-405x(93)90023-5`.

— (2004). "The Capital Asset Pricing Model: Theory and Evidence". *Journal of Economic Perspectives* 18.3, pp. 25–46. DOI: `10.1257/0895330042162430`.

— (2015). "A five-factor asset pricing model". *Journal of Financial Economics* 116.1, pp. 1–22. DOI: `10.1016/j.jfineco.2014.10.010`.

Feng, G., S. Giglio, and D. Xiu (2020). "Taming the Factor Zoo: A Test of New Factors". *The Journal of Finance* 75.3, pp. 1327–1370. DOI: `10.1111/jofi.12883`.

Freyberger, J., A. Neuhierl, and M. Weber (2020). "Dissecting Characteristics Nonparametrically". *The Review of Financial Studies* 33.5, pp. 2326–2377. DOI: `10.1093/rfs/hhz123`.

Friedman, J. H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine". *The Annals of Statistics* 29.5, pp. 1189–1232. DOI: `10.1214/aos/1013203451`.

Giglio, S., B. Kelly, and D. Xiu (2022). "Factor Models, Machine Learning, and Asset Pricing". *Annual Review of Financial Economics* 14.1, pp. 337–368. DOI: `10.1146/annurev-financial-101521-104735`.

Gu, S., B. Kelly, and D. Xiu (2020). "Empirical Asset Pricing via Machine Learning". *The Review of Financial Studies* 33.5, pp. 2223–2273. DOI: `10.1093/rfs/hhaa009`.

Harvey, D., S. Leybourne, and P. Newbold (1997). "Testing the equality of prediction mean squared errors". *International Journal of Forecasting* 13.2, pp. 281–291. DOI: 10.1016/s0169-2070(96)00719-4.

Havlíček, V., A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta (2019). "Supervised Learning with Quantum-Enhanced Feature Spaces". *Nature* 567.7747, pp. 209–212. DOI: 10.1038/s41586-019-0980-2.

Hoang, L. T. and T. T. Phan (2019). "Is Liquidity Priced in the Vietnamese Stock Market?" *Economic Papers: A journal of applied economics and policy* 38.3, pp. 193–207. DOI: 10.1111/1759-3441.12249.

Huang, X., C. Liu, and T. Shu (2023). "Factors and anomalies in the Vietnamese stock market". *Pacific-Basin Finance Journal* 82, p. 102176. DOI: 10.1016/j.pacfin.2023.102176.

Huber, P. J. (1964). "Robust Estimation of a Location Parameter". *The Annals of Mathematical Statistics* 35.1, pp. 73–101. DOI: 10.1214/aoms/1177703732.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer US. DOI: 10.1007/978-1-0716-1418-1.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). "Is There a Replication Crisis in Finance?" *The Journal of Finance* 78.5, pp. 2465–2518. DOI: 10.1111/jofi.13249.

— (2026a). *Global Factor Data Documentation*. Accessed 2026-03-29. URL: https://jkpfactors.s3.amazonaws.com/documents/Documentation.pdf.

— (2026b). *Global Factor Data: Factor Returns*. Accessed 2026-03-29. URL: https://jkpfactors.com/factor-returns.

Ke, Z. T., B. T. Kelly, and D. Xiu (2019). *Predicting Returns With Text Data*. NBER Working Paper No. 26186. National Bureau of Economic Research. DOI: 10.3386/w26186.

Kelly, B., S. Malamud, and K. Zhou (2024). "The Virtue of Complexity in Return Prediction". *The Journal of Finance* 79.1, pp. 459–503. DOI: 10.1111/jofi.13298.

Kelly, B. and D. Xiu (2023). "Financial Machine Learning". *Foundations and Trends in Finance* 13.3–4, pp. 205–363. DOI: 10.1561/0500000064.

Kelly, B. T., S. Pruitt, and Y. Su (2019). "Characteristics are covariances: A unified model of risk and return". *Journal of Financial Economics* 134.3, pp. 501–524. DOI: 10.1016/j.jfineco.2019.05.001.

Lintner, J. (1965). "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets". *The Review of Economics and Statistics* 47.1, pp. 13–37. DOI: 10.2307/1924119.

Liu, J., R. F. Stambaugh, and Y. Yuan (2019). "Size and value in China". *Journal of Financial Economics* 134.1, pp. 48–69. DOI: 10.1016/j.jfineco.2019.03.008.

Markowitz, H. (1952). "Portfolio Selection". *The Journal of Finance* 7.1, pp. 77–91. DOI: 10.1111/j.1540-6261.1952.tb01525.x.

Mossin, J. (1966). "Equilibrium in a Capital Asset Market". *Econometrica* 34.4, pp. 768–783. DOI: 10.2307/1910098.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay (2011). "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research* 12, pp. 2825–2830. URL: https://arxiv.org/abs/1201.0490.

Pesaran, M. H. and A. Timmermann (2007). "Selection of estimation window in the presence of breaks". *Journal of Econometrics* 137.1, pp. 134–161. DOI: 10.1016/j.jeconom.2006.03.010.

Rouwenhorst, K. G. (1999). "Local Return Factors and Turnover in Emerging Stock Markets". *The Journal of Finance* 54.4, pp. 1439–1464. DOI: 10.1111/0022-1082.00151.

Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk". *The Journal of Finance* 19.3, pp. 425–442. DOI: 10.1111/j.1540-6261.1964.tb02865.x.

Welch, I. and A. Goyal (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction". *The Review of Financial Studies* 21.4, pp. 1455–1508. DOI: 10.1093/rfs/hhm014.

A Data, Provenance, and Feature Construction

A.1 Data provenance

The stock-level side of the study begins with a direct Bloomberg Terminal export. The Bloomberg FLDS command is used to identify the raw field inventory, which is then pulled at monthly frequency for each security; static sector and listing descriptors are collected separately and merged later. The raw Bloomberg layer is transformed into monthly characteristics intended to match the empirical asset-pricing literature: valuation, size, momentum, trading and liquidity conditions, simple risk proxies, and sector controls.

The main risk-free leg uses a daily Vietnam one-year government-bond benchmark collected from Capital IQ. The daily annualized percentage yield is converted to a daily simple-return proxy and then compounded to the month:

$$rf_d = \frac{\texttt{VN\_1Y\_BOND}_d}{100 \times 252}, \qquad \texttt{rf\_1m}_t = \prod_{d \in t}(1 + rf_d) - 1. \tag{16}$$

This conversion is the basis for the excess-return target used throughout the manuscript. Because the source is a proprietary Capital IQ export, the study treats it as internal evidence for target construction rather than as a public dataset to redistribute.

The macro overlay comes from the JKP Global Factor Data (Jensen et al., 2026b; Jensen et al., 2026a; Jensen et al., 2023). In the present experiment the retained Vietnam factor series cover valuation, profitability, risk, and trading frictions, and are appended to the model matrix as `macro_*` features. Panel C of Table 8 lists the exact retained mnemonics. These eight series were retained from a larger exploratory screening exercise designed to keep the overlay compact.

A.2 Target construction and preprocessing logic

The repaired monthly panel first aligns each stock-month observation with its next-month realized return using the shift rule

$$\texttt{target\_return\_1m}_{i,t} = \texttt{ret}_{i,t+1}. \tag{17}$$

The modeling layer then rebuilds the lead raw and excess targets directly from observed returns rather than relying mechanically on a stored intermediate target:

$$\texttt{ret\_lead1m\_raw}_{i,t} = r_{i,t+1}, \qquad \texttt{ret\_exc\_lead1m\_raw}_{i,t} = r_{i,t+1} - r^f_{t+1}. \tag{18}$$

This prevents stale or misaligned target fields from entering the model layer and creates a direct audit trail from observed returns to the prediction target.

The prep layer then applies the following sequence:

1. Keep the last observed row inside each ticker-month before modeling transforms;
2. Merge the monthly risk-free series;
3. Rebuild the next-month raw and excess-return targets;
4. Screen candidate continuous features at 75% non-missing coverage;
5. Impute remaining missing values with the monthly cross-sectional median;
6. Drop residual rows missing a required predictor, control, or target;
7. Winsorize selected continuous predictors and the excess-return target at the 1st and 99th percentiles by month;
8. Rank-scale selected continuous predictors to $[-1, 1]$.[8]
9. In the macro experiment, restrict both variants to the common macro month set before comparison.

This sequence matters because it keeps the baseline and macro variants on the same comparison sample. The macro block is therefore the only material difference between the two panels.

A.3 Bloomberg raw input fields

Table 7 records the Bloomberg fields used at the start of the study. The historical pull comprises 37 monthly Bloomberg fields, while three additional static Bloomberg descriptors support the sector controls and the age variable. For Vietnam, the direct Bloomberg export covers 699 securities drawn from the HOSE and HNX exchanges in Bloomberg, the two main stock exchanges in Vietnam. Each raw monthly history file contains one date column plus the 37 monthly Bloomberg fields, while the static descriptors are merged later at the panel stage. The raw extraction window is intentionally broader than the final 2008–2024 evaluation window so that lagged characteristics can be formed before the sample is restricted for forecasting.

[8] The preparation stack applies a pragmatic 75% coverage screen to continuous characteristics, imputes remaining gaps with within-month medians, winsorizes selected continuous predictors and the excess-return target at the 1st and 99th percentiles by month, and rank-scales the continuous predictors to $[-1, 1]$. The broader logic of disciplined characteristic selection and careful handling of unstable or redundant signals follows Kelly et al. (2019) and Freyberger et al. (2020). The 75% threshold itself is an implementation choice rather than a theoretical constant: it removes very thin variables, such as firm age, while preserving a workable Vietnam sample.

Table 7: Bloomberg raw input fields at the start of the study

Raw field block	Bloomberg mnemonic	Bloomberg label	Role in raw panel	Final use in study
Core price action	PX_LAST	Last price	Monthly price anchor	Target and price-based characteristics
Core price action	PX_OPEN	Open price	Opening price snapshot	Exploratory only
Core price action	PX_HIGH	High price	Intramonth high-price snapshot	Exploratory only
Core price action	PX_LOW	Low price	Intramonth low-price snapshot	Exploratory only
Core price action	DAY_TO_DAY_TOT_RETURN_GROSS_DVDS	Daily total return (gross dividends)	Total-return reference series	Exploratory only
Core volume and size	PX_VOLUME	Trading volume	Trading-activity measure	Liquidity characteristics
Core volume and size	CUR_MKT_CAP	Current market capitalisation	Firm-size anchor	Size characteristic
Core volume and size	EQY_SH_OUT	Shares outstanding	Share-base measure	Liquidity characteristics
Microstructure prices	PX_BID	Bid price	Quoted bid snapshot	Exploratory only
Microstructure prices	PX_MID	Mid price	Quoted mid snapshot	Exploratory only
Microstructure prices	PX_ASK	Ask price	Quoted ask snapshot	Exploratory only
Bid-ask spreads	AVERAGE_BID_ASK_SPREAD	Average bid-ask spread	Quoted spread diagnostic	Exploratory only
Bid-ask spreads	AVERAGE_BID_ASK_SPREAD_%	Average bid-ask spread (%)	Relative spread diagnostic	Exploratory only
Bid-ask spreads	TIME_WAVG_BID_ASK_SPREAD	Time-weighted bid-ask spread	Time-weighted spread diagnostic	Exploratory only
Bid-ask spreads	TIME_WAVG_BID_ASK_SPREAD_PCT	Time-weighted bid-ask spread (%)	Relative time-weighted spread diagnostic	Exploratory only
VWAP volume ratios	VWAP_BID_VOL_PERCENTAGE	VWAP bid volume (%)	Bid-side book-balance diagnostic	Exploratory only
VWAP volume ratios	VWAP_ASK_VOL_PERCENTAGE	VWAP ask volume (%)	Ask-side book-balance diagnostic	Exploratory only
RSI indicators	RSI_3D	RSI (3 day)	Technical momentum diagnostic	Exploratory only
RSI indicators	RSI_9D	RSI (9 day)	Technical momentum diagnostic	Exploratory only
RSI indicators	RSI_14D	RSI (14 day)	Technical momentum diagnostic	Exploratory only
RSI indicators	RSI_30D	RSI (30 day)	Technical momentum diagnostic	Exploratory only
Risk and momentum	REL_SHR_PX_MOMENTUM	Relative share price momentum	Bloomberg momentum indicator	Exploratory only
Risk and momentum	BETA_ADJ_OVERRIDABLE	Adjusted beta	Equity beta estimate	Risk characteristics
Risk and momentum	VOLATILITY_30D	30-day volatility	Short-horizon volatility diagnostic	Exploratory only
Risk and momentum	VOLATILITY_90D	90-day volatility	Longer-horizon volatility diagnostic	Exploratory only

Continued on next page

Raw field block	Bloomberg mnemonic	Bloomberg label	Role in raw panel	Final use in study
Per-share metrics	`IS_EPS`	Basic earnings per share	Per-share profitability measure	Valuation characteristics
Per-share metrics	`REVENUE_PER_SH`	Revenue per share	Per-share revenue measure	Valuation characteristics
Per-share metrics	`BOOK_VAL_PER_SH`	Book value per share	Per-share balance-sheet measure	Exploratory only
Per-share metrics	`EQY_DPS`	Dividends per share	Per-share payout measure	Exploratory only
Valuation ratios	`PX_TO_BOOK_RATIO`	Price-to-book ratio	Valuation multiple	Valuation characteristics
Valuation ratios	`AVERAGE_PRICE_TO_BOOK_RATIO`	Average price-to-book ratio	Average valuation multiple	Exploratory only
Valuation ratios	`PE_RATIO`	Price/earnings ratio	Earnings multiple	Exploratory only
Valuation ratios	`PX_TO_SALES_RATIO`	Price/sales ratio	Sales multiple	Exploratory only
Balance-sheet capital	`BS_TOT_CAP`	Total capital	Balance-sheet scale measure	Exploratory only
Balance-sheet capital	`TOTAL_EQUITY`	Total equity	Balance-sheet equity measure	Exploratory only
Balance-sheet capital	`TOT_COMMON_EQY`	Total common equity	Common equity measure	Exploratory only
Balance-sheet capital	`BS_TOT_ASSET`	Total assets	Balance-sheet asset measure	Exploratory only
Static Bloomberg descriptors	`INDUSTRY_SECTOR`	Industry sector	Static sector classification	Sector controls
Static Bloomberg descriptors	`GICS_SECTOR_NAME`	GICS sector name	Supporting sector classification	Exploratory only
Static Bloomberg descriptors	`EQY_INIT_PO_DT`	Initial public offering date	Listing-date descriptor	Age variable

The raw Bloomberg input layer is intentionally broader than the final model matrix. Several price, microstructure, technical, and balance-sheet fields were collected for exploratory screening or diagnostic review and were not retained as final predictors. Table 8 records the smaller downstream feature set that survives into the study's model matrix.

A.4 Feature construction

Table 8 records the final model inputs used in estimation. It should be read as a second layer after Table 7: the first table documents the Bloomberg extraction layer, while the second documents the transformed variables that survive into the final model matrix.

The most important inclusion and exclusion decisions are straightforward. 'OLS-3' uses `log_mkt_cap`, `book_to_market`, and `mom1m` as the parsimonious benchmark. The broader stock-level matrix adds valuation, momentum, liquidity, and risk features plus nine estimated sector controls. Firm age survives in the repaired panel but fails the 75% coverage rule and is therefore excluded from the continuous model matrix.

Table 8: Feature construction and final model inputs

Panel A. Continuous stock-level predictors

Feature	Family	Raw field(s)	Construction rule
`log_mkt_cap`	Size	`cur_mkt_cap`	Log-positive transform of current market capitalisation.
`book_to_market`	Valuation	`px_to_book_ratio`	Positive inverse: 1/`px_to_book_ratio`.
`mom1m`	Momentum	`px_last`	One-month return lag: `ret.shift(1)` after monthly return construction.
`mom12m`	Momentum	`px_last`	$(P_{t-2}/P_{t-12}) - 1$.
`chmom`	Momentum	`px_last`	Recent momentum minus earlier momentum: $(P_{t-2}/P_{t-6} - 1) - (P_{t-7}/P_{t-12} - 1)$.
`mom36m`	Momentum	`px_last`	$(P_{t-13}/P_{t-36}) - 1$.
`retvol`	Risk/volatility	`px_last`	Rolling standard deviation of monthly returns.
`turn`	Trading and liquidity	`px_volume`, `eqy_sh_out`	Share turnover: trading volume divided by shares outstanding.
`sd_turn`	Trading and liquidity	`turn`	Rolling standard deviation of turnover.
`dolvol`	Trading and liquidity	`px_last`, `px_volume`	Log of price times volume, with zero values treated as missing.
`beta`	Risk/volatility	`beta_adj_overridable`	Direct Bloomberg beta field.
`betasq`	Risk/volatility	`beta`	Squared beta.
`ep`	Valuation	`is_eps`, `px_last`	Earnings-to-price ratio.
`sp`	Valuation	`revenue_per_sh`, `px_last`	Sales-to-price ratio.

Panel B. Sector dummy variables

Dummy variable	Raw field(s)	Role in model matrix
`sector_Communications`	`industry_sector`	Indicator for Bloomberg industry sector = Communications.
`sector_Consumer,Cyclical`	`industry_sector`	Indicator for Bloomberg industry sector = Consumer, Cyclical.
`sector_Consumer,Non-cyclical`	`industry_sector`	Indicator for Bloomberg industry sector = Consumer, Non-cyclical.
`sector_Diversified`	`industry_sector`	Indicator for Bloomberg industry sector = Diversified.
`sector_Energy`	`industry_sector`	Indicator for Bloomberg industry sector = Energy.
`sector_Financial`	`industry_sector`	Indicator for Bloomberg industry sector = Financial.
`sector_Industrial`	`industry_sector`	Indicator for Bloomberg industry sector = Industrial.
`sector_Technology`	`industry_sector`	Indicator for Bloomberg industry sector = Technology.
`sector_Utilities`	`industry_sector`	Indicator for Bloomberg industry sector = Utilities.

Panel C. Country macro factors

Feature	Family	Raw field(s)	Role in model matrix
`macro_be_me`	Macro valuation/profitability	`be_me`	Monthly JKP country factor merged by month end.
`macro_beta_dimson_21d`	Macro risk	`beta_dimson_21d`	Monthly JKP country factor merged by month end.
`macro_bev_mev`	Macro valuation/profitability	`bev_mev`	Monthly JKP country factor merged by month end.
`macro_ocf_at_chg1`	Macro valuation/profitability	`ocf_at_chg1`	Monthly JKP country factor merged by month end.
`macro_qmj_prof`	Macro valuation/profitability	`qmj_prof`	Monthly JKP country factor merged by month end.
`macro_qmj_safety`	Macro risk	`qmj_safety`	Monthly JKP country factor merged by month end.
`macro_rvol_21d`	Macro risk	`rvol_21d`	Monthly JKP country factor merged by month end.
`macro_zero_trades_21d`	Macro trading friction	`zero_trades_21d`	Monthly JKP country factor merged by month end.

Notes: the final model matrix contains 14 continuous stock-level predictors, 9 estimated sector dummies, and 8 country macro factors. `sector_Basic Materials` is the omitted reference sector in the dummy set. Firm age is constructed from Bloomberg listing dates but excluded because it fails the 75% coverage screen (about 13.4% non-missing). The intercept is not listed because it is an estimated constant, not an input feature.

A.5 Exploratory review of the raw panel

Before formal modeling, the raw Vietnam panel was reviewed descriptively to understand its structure and potential weaknesses. Three points matter for the manuscript. First, sector controls are formed from the Bloomberg `INDUSTRY_SECTOR` classification, while `GICS_SECTOR_NAME` is retained only as supporting classification metadata; the exploratory review confirmed that the panel is uneven across sectors, which supports retaining sector dummies as controls. Second, company age is constructed from `EQY_INIT_PO_DT`; it remains economically interpretable, but its coverage is too thin for the final continuous model matrix. Third, the sector balance, the number of active companies over time, and the distribution of forward returns were examined as diagnostic checks before the cleaned panel was finalized. These reviews informed the data-cleaning discussion but were not used as a separate ex post model-selection device.

B Model Implementation Backup

This appendix serves as technical backup for the methodology in Section 2. The main text records the estimator and evaluation equations. The appendix keeps the implementation details needed to interpret the saved outputs.

B.1 Software and rolling mechanics

The rolling 60/24/12 split design is implemented with project-specific time-ordered window logic. Once each window is defined, the fitted estimators are drawn from standard software backends. The linear, shrinkage, latent-factor, and tree estimators are implemented with `scikit-learn`, while the feedforward neural networks are implemented with TensorFlow/Keras (Pedregosa et al., 2011; Abadi et al., 2016).

Within each split, hyperparameters are selected on the validation block, the selected specification is refit on the combined train-validation block, and predictions are then saved for the held-out test block. All model comparisons in the manuscript use common prediction rows, so the comparison is not driven by model-specific sample loss.

B.2 Estimator roles

The OLS-3 benchmark uses `log_mkt_cap`, `book_to_market`, and `mom1m`. It is included as a parsimonious size-value-momentum reference, not as a claim that three variables are sufficient for Vietnam. OLS full and Huber full then test whether the broader characteristic set improves a linear model. The Huber variants are included because monthly stock-return targets can contain extreme observations that make pure squared-loss fits unstable (Huber, 1964).

Elastic net tests whether shrinkage and sparse selection stabilize the larger linear model. PCR and PLS test whether a lower-dimensional latent representation of the characteristic set is useful, with PCR using unsupervised components and PLS choosing components with the response in view (de Jong, 1993; James et al., 2021; Kelly and Xiu, 2023). Random forest and GBRT test moderate nonlinearities and interactions through tree ensembles (Breiman, 2001; Friedman, 2001). The neural networks are deliberately modest one-, two-, and three-hidden-layer feedforward models; this scope is appropriate for testing whether the available Vietnam sample supports additional function approximation while keeping the model menu interpretable.

B.3 Importance diagnostics

For linear and neural models, the main importance diagnostic is the normalized drop in validation-window R^2 after neutralizing one standardized predictor. For tree models, native tree-importance measures are retained, and the combined exhibits use the comparable zero-out diagnostic when it is available. Because these diagnostics are not identical objects across all families, the manuscript emphasizes predictor families and broad economic content rather than over-interpreting small rank differences in individual variables.

B.4 Comparison mechanics

The Diebold-Mariano table is computed under one-step squared-error loss on common prediction rows with a finite-sample correction (Diebold and Mariano, 1995; Harvey et al., 1997). The test is used to discipline the top of the leaderboard, but it is not treated as a mechanical model-selection rule. In a weak-signal setting, statistically visible differences can still be economically small, so the manuscript combines DM evidence with forecast metrics, portfolio diagnostics, and predictor interpretation.

C Backup Evidence and Empirical Detail

C.1 Replication code

The replication code and aggregate result summaries are available at `https://github.com/vanthelearner/fyp-vietnam-ml-replication`. The repository contains the Python source code, execution notebooks, tables and figures, and aggregate model-comparison outputs used to support the empirical results.

C.2 Common-sample model comparison

The empirical comparison uses a common prediction sample across models. The full Vietnam common panel contains 99,585 stock-month observations, but after the rolling-window restrictions and common-support alignment used for model comparison, pairwise model comparisons are formed on the same 61,575 prediction rows. This matters because it means that the model ranking is not driven by different stock-month support across estimators.

The distinction between scored months and valid IC months also matters. Most models have 96 scored test months, but fewer months with defined information coefficients because Spearman IC is undefined when the relevant cross section has no variation. The main comparison therefore places more weight on RMSE, R^2_{OOS}, and the portfolio diagnostics than on small differences in mean IC.

C.3 Full long-short table

Table 9: Full long-short decile table

Model	Months	Mean excess EW	Mean excess VW	Volatility EW	Volatility VW	Sharpe EW	Sharpe VW
elastic_net_baseline	96	0.0273	0.0199	0.0540	0.0785	1.7547	0.8789
elastic_net_macro	96	0.0297	0.0251	0.0536	0.0747	1.9194	1.1659
gbrt_baseline	96	0.0277	0.0098	0.0413	0.0502	2.3227	0.6775
gbrt_macro	96	0.0213	0.0101	0.0410	0.0678	1.8035	0.5184
huber_full_baseline	96	0.0240	0.0144	0.0409	0.0581	2.0332	0.8562
huber_full_macro	96	0.0246	0.0166	0.0401	0.0561	2.1248	1.0253
nn_1layer_baseline	96	0.0124	-0.0022	0.0370	0.0577	1.1610	-0.1306
nn_1layer_macro	96	0.0095	0.0007	0.0504	0.0629	0.6536	0.0378
nn_2layer_baseline	96	0.0179	0.0087	0.0382	0.0505	1.6274	0.5993
nn_2layer_macro	96	0.0093	0.0040	0.0499	0.0671	0.6445	0.2084
nn_3layer_baseline	96	0.0178	0.0076	0.0395	0.0674	1.5572	0.3919
nn_3layer_macro	96	0.0106	0.0007	0.0448	0.0586	0.8235	0.0428
ols3_baseline	96	0.0231	0.0224	0.0634	0.0886	1.2635	0.8742
ols3_huber_baseline	96	0.0131	0.0119	0.0622	0.0783	0.7312	0.5268
ols3_huber_macro	96	0.0131	0.0119	0.0622	0.0783	0.7312	0.5268
ols3_macro	96	0.0231	0.0224	0.0634	0.0886	1.2635	0.8742
ols_full_baseline	96	0.0308	0.0216	0.0577	0.0781	1.8478	0.9576
ols_full_macro	96	0.0309	0.0223	0.0573	0.0797	1.8663	0.9676
pcr_baseline	96	0.0207	0.0177	0.0620	0.0789	1.1565	0.7782
pcr_macro	96	0.0050	0.0083	0.0568	0.0701	0.3041	0.4124
pls_baseline	96	0.0284	0.0256	0.0539	0.0820	1.8258	1.0804
pls_macro	96	0.0269	0.0324	0.0559	0.0873	1.6657	1.2873
random_forest_baseline	96	0.0296	0.0162	0.0472	0.0707	2.1743	0.7932
random_forest_macro	96	0.0196	0.0073	0.0472	0.0584	1.4365	0.4320

All rows correspond to the top-minus-bottom decile spread. Sharpe ratios are annualized from monthly excess returns.

C.4 Long-only model portfolios and VN30-style market proxy

Table 10 compares long-only top-decile model portfolios with the constructed VN30-style local market proxy over the same 96-month test window. The VN30-style benchmark is constructed from the study's Vietnam stock universe by selecting the largest 30 eligible firms and rebalancing semi-annually. This proxy is used because a consistent full-period official constituent-level VN30 return series was not available from the project data sources.

Table 10: Long-only model portfolios and VN30-style benchmark

Series	Months	EW mean	VW mean	EW vol.	VW vol.	EW Sh.	VW Sh.
OLS-3 (base)	96	0.0223	0.0233	0.0862	0.1012	0.8961	0.7970
OLS-3 Huber (base)	96	0.0147	0.0148	0.0716	0.0827	0.7096	0.6192
OLS full (base)	96	0.0258	0.0188	0.0760	0.0849	1.1751	0.7673
Huber full (base)	96	0.0219	0.0144	0.0583	0.0622	1.3046	0.8023
EN (base)	96	0.0253	0.0196	0.0744	0.0840	1.1765	0.8073
PCR (base)	96	0.0190	0.0151	0.0710	0.0774	0.9286	0.6775
PLS (base)	96	0.0255	0.0239	0.0733	0.0879	1.2064	0.9429
GBRT (base)	96	0.0231	0.0100	0.0634	0.0683	1.2616	0.5091
RF (base)	96	0.0251	0.0182	0.0731	0.0862	1.1890	0.7318
NN 1L (base)	96	0.0142	0.0019	0.0589	0.0610	0.8343	0.1068
NN 2L (base)	96	0.0188	0.0078	0.0633	0.0631	1.0265	0.4261
NN 3L (base)	96	0.0184	0.0112	0.0690	0.0713	0.9254	0.5464
OLS full (macro)	96	0.0258	0.0186	0.0759	0.0850	1.1760	0.7582
Huber full (macro)	96	0.0222	0.0153	0.0590	0.0617	1.3067	0.8581
EN (macro)	96	0.0250	0.0244	0.0731	0.0825	1.1874	1.0225
PCR (macro)	96	0.0104	0.0125	0.0702	0.0793	0.5153	0.5467
PLS (macro)	96	0.0251	0.0311	0.0725	0.0983	1.1996	1.0980
GBRT (macro)	96	0.0214	0.0142	0.0597	0.0724	1.2400	0.6782
RF (macro)	96	0.0209	0.0105	0.0641	0.0690	1.1263	0.5251
NN 1L (macro)	96	0.0106	0.0058	0.0606	0.0608	0.6039	0.3308
NN 2L (macro)	96	0.0091	0.0038	0.0701	0.0750	0.4503	0.1769
NN 3L (macro)	96	0.0126	0.0097	0.0636	0.0627	0.6872	0.5369
VN30-style benchmark	96	0.0070	0.0071	0.0656	0.0639	0.3700	0.3845

Model rows correspond to long-only top-decile portfolios sorted on predicted returns. VN30 rows are constructed market-reference benchmarks. Sharpe ratios are annualized from monthly excess returns.

C.5 Full predictor rankings

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
OLS-3 (baseline)	1	book_to_market	Valuation	0.8483
OLS-3 (baseline)	2	log_mkt_cap	Size	0.1311
OLS-3 (baseline)	3	mom1m	Momentum	0.0206
OLS-3 Huber (baseline)	1	book_to_market	Valuation	0.6857
OLS-3 Huber (baseline)	2	log_mkt_cap	Size	0.2586
OLS-3 Huber (baseline)	3	mom1m	Momentum	0.0557
OLS full (baseline)	1	book_to_market	Valuation	0.2205
OLS full (baseline)	2	ep	Valuation	0.1606
OLS full (baseline)	3	dolvol	Trading/liquidity	0.1268
OLS full (baseline)	4	sector_Financial	Sector	0.0778
OLS full (baseline)	5	retvol	Risk	0.0533
OLS full (baseline)	6	turn	Trading/liquidity	0.0492
OLS full (baseline)	7	mom12m	Momentum	0.0426
OLS full (baseline)	8	sd_turn	Trading/liquidity	0.0393
OLS full (baseline)	9	sector_Consumer,Non-cyclical	Sector	0.0392
OLS full (baseline)	10	mom36m	Momentum	0.0330
OLS full (baseline)	11	sector_Industrial	Sector	0.0300
OLS full (baseline)	12	log_mkt_cap	Size	0.0257
OLS full (baseline)	13	beta	Risk	0.0235
OLS full (baseline)	14	sp	Valuation	0.0188
OLS full (baseline)	15	sector_Consumer,Cyclical	Sector	0.0169
OLS full (baseline)	16	betasq	Risk	0.0157
OLS full (baseline)	17	chmom	Momentum	0.0083
OLS full (baseline)	18	mom1m	Momentum	0.0060
OLS full (baseline)	19	sector_Utilities	Sector	0.0050
OLS full (baseline)	20	sector_Technology	Sector	0.0024
Huber full (baseline)	1	dolvol	Trading/liquidity	0.1753
Huber full (baseline)	2	ep	Valuation	0.1379
Huber full (baseline)	3	book_to_market	Valuation	0.1261
Huber full (baseline)	4	turn	Trading/liquidity	0.0802
Huber full (baseline)	5	sector_Financial	Sector	0.0775
Huber full (baseline)	6	sector_Industrial	Sector	0.0724
Huber full (baseline)	7	mom12m	Momentum	0.0627
Huber full (baseline)	8	sector_Consumer,Non-cyclical	Sector	0.0580
Huber full (baseline)	9	log_mkt_cap	Size	0.0417
Huber full (baseline)	10	sd_turn	Trading/liquidity	0.0230
Huber full (baseline)	11	sector_Consumer,Cyclical	Sector	0.0222
Huber full (baseline)	12	beta	Risk	0.0187
Huber full (baseline)	13	sector_Utilities	Sector	0.0150
Huber full (baseline)	14	mom36m	Momentum	0.0146
Huber full (baseline)	15	sp	Valuation	0.0143
Huber full (baseline)	16	retvol	Risk	0.0118
Huber full (baseline)	17	betasq	Risk	0.0102
Huber full (baseline)	18	chmom	Momentum	0.0089
Huber full (baseline)	19	sector_Energy	Sector	0.0087
Huber full (baseline)	20	sector_Technology	Sector	0.0061

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
Elastic net (baseline)	1	book_to_market	Valuation	0.3126
Elastic net (baseline)	2	ep	Valuation	0.2879
Elastic net (baseline)	3	dolvol	Trading/liquidity	0.1060
Elastic net (baseline)	4	retvol	Risk	0.0669
Elastic net (baseline)	5	mom12m	Momentum	0.0508
Elastic net (baseline)	6	turn	Trading/liquidity	0.0317
Elastic net (baseline)	7	mom36m	Momentum	0.0286
Elastic net (baseline)	8	sector_Financial	Sector	0.0272
Elastic net (baseline)	9	beta	Risk	0.0202
Elastic net (baseline)	10	chmom	Momentum	0.0151
Elastic net (baseline)	11	sector_Industrial	Sector	0.0146
Elastic net (baseline)	12	sector_Consumer,Non-cyclical	Sector	0.0140
Elastic net (baseline)	13	log_mkt_cap	Size	0.0101
Elastic net (baseline)	14	sp	Valuation	0.0033
Elastic net (baseline)	15	sd_turn	Trading/liquidity	0.0031
Elastic net (baseline)	16	mom1m	Momentum	0.0022
Elastic net (baseline)	17	betasq	Risk	0.0021
Elastic net (baseline)	18	sector_Utilities	Sector	0.0015
Elastic net (baseline)	19	sector_Consumer,Cyclical	Sector	0.0006
Elastic net (baseline)	20	sector_Energy	Sector	0.0005
PCR (baseline)	1	book_to_market	Valuation	0.2331
PCR (baseline)	2	ep	Valuation	0.1586
PCR (baseline)	3	dolvol	Trading/liquidity	0.1337
PCR (baseline)	4	sector_Financial	Sector	0.0592
PCR (baseline)	5	retvol	Risk	0.0565
PCR (baseline)	6	mom36m	Momentum	0.0480
PCR (baseline)	7	sp	Valuation	0.0423
PCR (baseline)	8	sd_turn	Trading/liquidity	0.0420
PCR (baseline)	9	turn	Trading/liquidity	0.0406
PCR (baseline)	10	mom12m	Momentum	0.0389
PCR (baseline)	11	beta	Risk	0.0262
PCR (baseline)	12	sector_Consumer,Non-cyclical	Sector	0.0226
PCR (baseline)	13	log_mkt_cap	Size	0.0226
PCR (baseline)	14	sector_Industrial	Sector	0.0217
PCR (baseline)	15	betasq	Risk	0.0169
PCR (baseline)	16	chmom	Momentum	0.0111
PCR (baseline)	17	sector_Consumer,Cyclical	Sector	0.0110
PCR (baseline)	18	mom1m	Momentum	0.0068
PCR (baseline)	19	sector_Communications	Sector	0.0035
PCR (baseline)	20	sector_Utilities	Sector	0.0023
PLS (baseline)	1	book_to_market	Valuation	0.2423
PLS (baseline)	2	ep	Valuation	0.1943
PLS (baseline)	3	dolvol	Trading/liquidity	0.1185
PLS (baseline)	4	mom36m	Momentum	0.0569
PLS (baseline)	5	sector_Financial	Sector	0.0538
PLS (baseline)	6	retvol	Risk	0.0507
PLS (baseline)	7	mom12m	Momentum	0.0503
PLS (baseline)	8	sp	Valuation	0.0429

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
PLS (baseline)	9	turn	Trading/liquidity	0.0350
PLS (baseline)	10	beta	Risk	0.0255
PLS (baseline)	11	log_mkt_cap	Size	0.0249
PLS (baseline)	12	sector_Industrial	Sector	0.0215
PLS (baseline)	13	sector_Consumer,Non-cyclical	Sector	0.0195
PLS (baseline)	14	sd_turn	Trading/liquidity	0.0188
PLS (baseline)	15	betasq	Risk	0.0155
PLS (baseline)	16	chmom	Momentum	0.0117
PLS (baseline)	17	mom1m	Momentum	0.0073
PLS (baseline)	18	sector_Consumer,Cyclical	Sector	0.0043
PLS (baseline)	19	sector_Utilities	Sector	0.0021
PLS (baseline)	20	sector_Energy	Sector	0.0020
GBRT (baseline)	1	mom1m	Momentum	0.1803
GBRT (baseline)	2	ep	Valuation	0.1625
GBRT (baseline)	3	book_to_market	Valuation	0.1620
GBRT (baseline)	4	mom36m	Momentum	0.0697
GBRT (baseline)	5	chmom	Momentum	0.0612
GBRT (baseline)	6	log_mkt_cap	Size	0.0545
GBRT (baseline)	7	mom12m	Momentum	0.0524
GBRT (baseline)	8	sd_turn	Trading/liquidity	0.0505
GBRT (baseline)	9	turn	Trading/liquidity	0.0466
GBRT (baseline)	10	betasq	Risk	0.0445
GBRT (baseline)	11	retvol	Risk	0.0330
GBRT (baseline)	12	beta	Risk	0.0233
GBRT (baseline)	13	dolvol	Trading/liquidity	0.0223
GBRT (baseline)	14	sector_Financial	Sector	0.0153
GBRT (baseline)	15	sp	Valuation	0.0098
GBRT (baseline)	16	sector_Consumer,Non-cyclical	Sector	0.0040
GBRT (baseline)	17	sector_Industrial	Sector	0.0030
GBRT (baseline)	18	sector_Consumer,Cyclical	Sector	0.0028
GBRT (baseline)	19	sector_Energy	Sector	0.0007
GBRT (baseline)	20	sector_Communications	Sector	0.0007
Random forest (baseline)	1	ep	Valuation	0.3509
Random forest (baseline)	2	book_to_market	Valuation	0.2538
Random forest (baseline)	3	mom1m	Momentum	0.0892
Random forest (baseline)	4	mom12m	Momentum	0.0690
Random forest (baseline)	5	mom36m	Momentum	0.0463
Random forest (baseline)	6	beta	Risk	0.0332
Random forest (baseline)	7	sp	Valuation	0.0307
Random forest (baseline)	8	log_mkt_cap	Size	0.0252
Random forest (baseline)	9	chmom	Momentum	0.0203
Random forest (baseline)	10	retvol	Risk	0.0199
Random forest (baseline)	11	sd_turn	Trading/liquidity	0.0171
Random forest (baseline)	12	betasq	Risk	0.0167
Random forest (baseline)	13	turn	Trading/liquidity	0.0151
Random forest (baseline)	14	dolvol	Trading/liquidity	0.0051
Random forest (baseline)	15	sector_Financial	Sector	0.0032
Random forest (baseline)	16	sector_Consumer,Non-cyclical	Sector	0.0012

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
Random forest (baseline)	17	`sector_Consumer,Cyclical`	Sector	0.0011
Random forest (baseline)	18	`sector_Utilities`	Sector	0.0008
Random forest (baseline)	19	`sector_Industrial`	Sector	0.0005
Random forest (baseline)	20	`sector_Energy`	Sector	0.0004
NN (1 layer) (baseline)	1	`dolvol`	Trading/liquidity	0.2227
NN (1 layer) (baseline)	2	`betasq`	Risk	0.1251
NN (1 layer) (baseline)	3	`log_mkt_cap`	Size	0.1001
NN (1 layer) (baseline)	4	`beta`	Risk	0.0898
NN (1 layer) (baseline)	5	`turn`	Trading/liquidity	0.0678
NN (1 layer) (baseline)	6	`book_to_market`	Valuation	0.0550
NN (1 layer) (baseline)	7	`sector_Industrial`	Sector	0.0523
NN (1 layer) (baseline)	8	`sector_Consumer,Non-cyclical`	Sector	0.0366
NN (1 layer) (baseline)	9	`sector_Financial`	Sector	0.0356
NN (1 layer) (baseline)	10	`mom1m`	Momentum	0.0346
NN (1 layer) (baseline)	11	`mom36m`	Momentum	0.0337
NN (1 layer) (baseline)	12	`ep`	Valuation	0.0314
NN (1 layer) (baseline)	13	`mom12m`	Momentum	0.0305
NN (1 layer) (baseline)	14	`sector_Consumer,Cyclical`	Sector	0.0151
NN (1 layer) (baseline)	15	`sp`	Valuation	0.0126
NN (1 layer) (baseline)	16	`retvol`	Risk	0.0124
NN (1 layer) (baseline)	17	`sd_turn`	Trading/liquidity	0.0123
NN (1 layer) (baseline)	18	`chmom`	Momentum	0.0110
NN (1 layer) (baseline)	19	`sector_Energy`	Sector	0.0072
NN (1 layer) (baseline)	20	`sector_Communications`	Sector	0.0054
NN (2 layers) (baseline)	1	`book_to_market`	Valuation	0.1161
NN (2 layers) (baseline)	2	`sd_turn`	Trading/liquidity	0.0867
NN (2 layers) (baseline)	3	`betasq`	Risk	0.0834
NN (2 layers) (baseline)	4	`beta`	Risk	0.0730
NN (2 layers) (baseline)	5	`ep`	Valuation	0.0706
NN (2 layers) (baseline)	6	`log_mkt_cap`	Size	0.0652
NN (2 layers) (baseline)	7	`mom36m`	Momentum	0.0622
NN (2 layers) (baseline)	8	`dolvol`	Trading/liquidity	0.0594
NN (2 layers) (baseline)	9	`mom1m`	Momentum	0.0517
NN (2 layers) (baseline)	10	`turn`	Trading/liquidity	0.0467
NN (2 layers) (baseline)	11	`sector_Financial`	Sector	0.0394
NN (2 layers) (baseline)	12	`sp`	Valuation	0.0386
NN (2 layers) (baseline)	13	`retvol`	Risk	0.0365
NN (2 layers) (baseline)	14	`sector_Industrial`	Sector	0.0336
NN (2 layers) (baseline)	15	`chmom`	Momentum	0.0335
NN (2 layers) (baseline)	16	`mom12m`	Momentum	0.0313
NN (2 layers) (baseline)	17	`sector_Consumer,Cyclical`	Sector	0.0190
NN (2 layers) (baseline)	18	`sector_Consumer,Non-cyclical`	Sector	0.0186
NN (2 layers) (baseline)	19	`sector_Energy`	Sector	0.0100
NN (2 layers) (baseline)	20	`sector_Communications`	Sector	0.0087
NN (3 layers) (baseline)	1	`book_to_market`	Valuation	0.1279
NN (3 layers) (baseline)	2	`ep`	Valuation	0.0829

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
NN (3 layers) (baseline)	3	dolvol	Trading/liquidity	0.0741
NN (3 layers) (baseline)	4	log_mkt_cap	Size	0.0719
NN (3 layers) (baseline)	5	mom1m	Momentum	0.0671
NN (3 layers) (baseline)	6	mom36m	Momentum	0.0664
NN (3 layers) (baseline)	7	betasq	Risk	0.0651
NN (3 layers) (baseline)	8	turn	Trading/liquidity	0.0547
NN (3 layers) (baseline)	9	retvol	Risk	0.0531
NN (3 layers) (baseline)	10	mom12m	Momentum	0.0515
NN (3 layers) (baseline)	11	sd_turn	Trading/liquidity	0.0427
NN (3 layers) (baseline)	12	chmom	Momentum	0.0421
NN (3 layers) (baseline)	13	beta	Risk	0.0346
NN (3 layers) (baseline)	14	sp	Valuation	0.0323
NN (3 layers) (baseline)	15	sector_Financial	Sector	0.0308
NN (3 layers) (baseline)	16	sector_Consumer,Non-cyclical	Sector	0.0257
NN (3 layers) (baseline)	17	sector_Consumer,Cyclical	Sector	0.0249
NN (3 layers) (baseline)	18	sector_Industrial	Sector	0.0237
NN (3 layers) (baseline)	19	sector_Energy	Sector	0.0098
NN (3 layers) (baseline)	20	sector_Communications	Sector	0.0067
OLS-3 (macro)	1	book_to_market	Valuation	0.8483
OLS-3 (macro)	2	log_mkt_cap	Size	0.1311
OLS-3 (macro)	3	mom1m	Momentum	0.0206
OLS-3 Huber (macro)	1	book_to_market	Valuation	0.6857
OLS-3 Huber (macro)	2	log_mkt_cap	Size	0.2586
OLS-3 Huber (macro)	3	mom1m	Momentum	0.0557
OLS full (macro)	1	macro_be_me	Macro valuation/profitability	0.2732
OLS full (macro)	2	macro_bev_mev	Macro valuation/profitability	0.1287
OLS full (macro)	3	macro_rvol_21d	Macro risk	0.0746
OLS full (macro)	4	macro_qmj_prof	Macro valuation/profitability	0.0661
OLS full (macro)	5	macro_zero_trades_21d	Macro trading friction	0.0576
OLS full (macro)	6	macro_ocf_at_chg1	Macro valuation/profitability	0.0572
OLS full (macro)	7	book_to_market	Valuation	0.0571
OLS full (macro)	8	macro_qmj_safety	Macro risk	0.0547
OLS full (macro)	9	ep	Valuation	0.0413
OLS full (macro)	10	dolvol	Trading/liquidity	0.0331
OLS full (macro)	11	sector_Financial	Sector	0.0214
OLS full (macro)	12	macro_beta_dimson_21d	Macro risk	0.0178
OLS full (macro)	13	retvol	Risk	0.0137
OLS full (macro)	14	sector_Consumer,Non-cyclical	Sector	0.0134
OLS full (macro)	15	turn	Trading/liquidity	0.0127
OLS full (macro)	16	sector_Industrial	Sector	0.0116
OLS full (macro)	17	mom12m	Momentum	0.0110
OLS full (macro)	18	sd_turn	Trading/liquidity	0.0101
OLS full (macro)	19	mom36m	Momentum	0.0085

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
OLS full (macro)	20	log_mkt_cap	Size	0.0066
Huber full (macro)	1	macro_be_me	Macro valuation/profitability	0.2418
Huber full (macro)	2	macro_bev_mev	Macro valuation/profitability	0.1410
Huber full (macro)	3	macro_qmj_prof	Macro valuation/profitability	0.0776
Huber full (macro)	4	macro_qmj_safety	Macro risk	0.0718
Huber full (macro)	5	macro_zero_trades_21d	Macro trading friction	0.0684
Huber full (macro)	6	macro_ocf_at_chg1	Macro valuation/profitability	0.0604
Huber full (macro)	7	macro_rvol_21d	Macro risk	0.0519
Huber full (macro)	8	dolvol	Trading/liquidity	0.0445
Huber full (macro)	9	book_to_market	Valuation	0.0373
Huber full (macro)	10	ep	Valuation	0.0365
Huber full (macro)	11	macro_beta_dimson_21d	Macro risk	0.0261
Huber full (macro)	12	turn	Trading/liquidity	0.0207
Huber full (macro)	13	sector_Financial	Sector	0.0189
Huber full (macro)	14	mom12m	Momentum	0.0174
Huber full (macro)	15	sector_Industrial	Sector	0.0161
Huber full (macro)	16	sector_Consumer,Non-cyclical	Sector	0.0143
Huber full (macro)	17	log_mkt_cap	Size	0.0096
Huber full (macro)	18	sd_turn	Trading/liquidity	0.0063
Huber full (macro)	19	sector_Consumer,Cyclical	Sector	0.0063
Huber full (macro)	20	beta	Risk	0.0059
Elastic net (macro)	1	macro_be_me	Macro valuation/profitability	0.2760
Elastic net (macro)	2	macro_bev_mev	Macro valuation/profitability	0.1872
Elastic net (macro)	3	macro_ocf_at_chg1	Macro valuation/profitability	0.0949
Elastic net (macro)	4	macro_rvol_21d	Macro risk	0.0788
Elastic net (macro)	5	macro_qmj_prof	Macro valuation/profitability	0.0650
Elastic net (macro)	6	book_to_market	Valuation	0.0626
Elastic net (macro)	7	macro_zero_trades_21d	Macro trading friction	0.0609
Elastic net (macro)	8	ep	Valuation	0.0488
Elastic net (macro)	9	macro_qmj_safety	Macro risk	0.0223
Elastic net (macro)	10	dolvol	Trading/liquidity	0.0187
Elastic net (macro)	11	macro_beta_dimson_21d	Macro risk	0.0178
Elastic net (macro)	12	sector_Financial	Sector	0.0176
Elastic net (macro)	13	retvol	Risk	0.0099
Elastic net (macro)	14	mom36m	Momentum	0.0066
Elastic net (macro)	15	beta	Risk	0.0052
Elastic net (macro)	16	sector_Consumer,Non-cyclical	Sector	0.0045
Elastic net (macro)	17	mom12m	Momentum	0.0040

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
Elastic net (macro)	18	sector_Industrial	Sector	0.0038
Elastic net (macro)	19	log_mkt_cap	Size	0.0037
Elastic net (macro)	20	sp	Valuation	0.0019
PCR (macro)	1	macro_be_me	Macro valuation/profitability	0.3130
PCR (macro)	2	macro_bev_mev	Macro valuation/profitability	0.1423
PCR (macro)	3	macro_qmj_safety	Macro risk	0.0914
PCR (macro)	4	macro_zero_trades_21d	Macro trading friction	0.0709
PCR (macro)	5	macro_rvol_21d	Macro risk	0.0633
PCR (macro)	6	macro_ocf_at_chg1	Macro valuation/profitability	0.0567
PCR (macro)	7	macro_qmj_prof	Macro valuation/profitability	0.0540
PCR (macro)	8	book_to_market	Valuation	0.0415
PCR (macro)	9	macro_beta_dimson_21d	Macro risk	0.0333
PCR (macro)	10	ep	Valuation	0.0317
PCR (macro)	11	sector_Financial	Sector	0.0145
PCR (macro)	12	mom36m	Momentum	0.0145
PCR (macro)	13	sp	Valuation	0.0111
PCR (macro)	14	dolvol	Trading/liquidity	0.0094
PCR (macro)	15	retvol	Risk	0.0090
PCR (macro)	16	sector_Industrial	Sector	0.0072
PCR (macro)	17	log_mkt_cap	Size	0.0064
PCR (macro)	18	beta	Risk	0.0057
PCR (macro)	19	betasq	Risk	0.0040
PCR (macro)	20	sd_turn	Trading/liquidity	0.0039
PLS (macro)	1	macro_be_me	Macro valuation/profitability	0.2207
PLS (macro)	2	macro_bev_mev	Macro valuation/profitability	0.1346
PLS (macro)	3	macro_zero_trades_21d	Macro trading friction	0.1197
PLS (macro)	4	macro_qmj_prof	Macro valuation/profitability	0.0912
PLS (macro)	5	macro_rvol_21d	Macro risk	0.0816
PLS (macro)	6	macro_qmj_safety	Macro risk	0.0762
PLS (macro)	7	macro_ocf_at_chg1	Macro valuation/profitability	0.0632
PLS (macro)	8	macro_beta_dimson_21d	Macro risk	0.0594
PLS (macro)	9	ep	Valuation	0.0349
PLS (macro)	10	book_to_market	Valuation	0.0335
PLS (macro)	11	retvol	Risk	0.0131
PLS (macro)	12	mom12m	Momentum	0.0112
PLS (macro)	13	mom36m	Momentum	0.0087
PLS (macro)	14	sd_turn	Trading/liquidity	0.0079
PLS (macro)	15	sp	Valuation	0.0069
PLS (macro)	16	sector_Industrial	Sector	0.0056
PLS (macro)	17	turn	Trading/liquidity	0.0046
PLS (macro)	18	log_mkt_cap	Size	0.0042

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
PLS (macro)	19	`sector_Consumer,Non-cyclical`	Sector	0.0042
PLS (macro)	20	`sector_Financial`	Sector	0.0040
GBRT (macro)	1	`macro_be_me`	Macro valuation/profitability	0.1960
GBRT (macro)	2	`macro_bev_mev`	Macro valuation/profitability	0.1750
GBRT (macro)	3	`macro_zero_trades_21d`	Macro trading friction	0.1042
GBRT (macro)	4	`macro_qmj_prof`	Macro valuation/profitability	0.1004
GBRT (macro)	5	`ep`	Valuation	0.0938
GBRT (macro)	6	`macro_rvol_21d`	Macro risk	0.0839
GBRT (macro)	7	`macro_beta_dimson_21d`	Macro risk	0.0648
GBRT (macro)	8	`macro_qmj_safety`	Macro risk	0.0494
GBRT (macro)	9	`macro_ocf_at_chg1`	Macro valuation/profitability	0.0311
GBRT (macro)	10	`book_to_market`	Valuation	0.0265
GBRT (macro)	11	`mom12m`	Momentum	0.0209
GBRT (macro)	12	`turn`	Trading/liquidity	0.0099
GBRT (macro)	13	`sp`	Valuation	0.0090
GBRT (macro)	14	`mom36m`	Momentum	0.0076
GBRT (macro)	15	`log_mkt_cap`	Size	0.0044
GBRT (macro)	16	`chmom`	Momentum	0.0043
GBRT (macro)	17	`sd_turn`	Trading/liquidity	0.0038
GBRT (macro)	18	`retvol`	Risk	0.0035
GBRT (macro)	19	`beta`	Risk	0.0035
GBRT (macro)	20	`mom1m`	Momentum	0.0032
Random forest (macro)	1	`macro_zero_trades_21d`	Macro trading friction	0.1944
Random forest (macro)	2	`macro_be_me`	Macro valuation/profitability	0.1616
Random forest (macro)	3	`macro_rvol_21d`	Macro risk	0.1119
Random forest (macro)	4	`macro_bev_mev`	Macro valuation/profitability	0.1080
Random forest (macro)	5	`macro_beta_dimson_21d`	Macro risk	0.0989
Random forest (macro)	6	`macro_qmj_prof`	Macro valuation/profitability	0.0671
Random forest (macro)	7	`ep`	Valuation	0.0670
Random forest (macro)	8	`dolvol`	Trading/liquidity	0.0397
Random forest (macro)	9	`macro_ocf_at_chg1`	Macro valuation/profitability	0.0333
Random forest (macro)	10	`macro_qmj_safety`	Macro risk	0.0330
Random forest (macro)	11	`book_to_market`	Valuation	0.0150
Random forest (macro)	12	`sd_turn`	Trading/liquidity	0.0132
Random forest (macro)	13	`turn`	Trading/liquidity	0.0084
Random forest (macro)	14	`retvol`	Risk	0.0071
Random forest (macro)	15	`betasq`	Risk	0.0069
Random forest (macro)	16	`mom36m`	Momentum	0.0067
Random forest (macro)	17	`beta`	Risk	0.0054
Random forest (macro)	18	`mom1m`	Momentum	0.0048

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
Random forest (macro)	19	mom12m	Momentum	0.0046
Random forest (macro)	20	sp	Valuation	0.0046
NN (1 layer) (macro)	1	macro_rvol_21d	Macro risk	0.1507
NN (1 layer) (macro)	2	macro_be_me	Macro valuation/profitability	0.1255
NN (1 layer) (macro)	3	macro_bev_mev	Macro valuation/profitability	0.1088
NN (1 layer) (macro)	4	macro_zero_trades_21d	Macro trading friction	0.0758
NN (1 layer) (macro)	5	macro_qmj_safety	Macro risk	0.0666
NN (1 layer) (macro)	6	macro_qmj_prof	Macro valuation/profitability	0.0547
NN (1 layer) (macro)	7	macro_ocf_at_chg1	Macro valuation/profitability	0.0520
NN (1 layer) (macro)	8	betasq	Risk	0.0434
NN (1 layer) (macro)	9	macro_beta_dimson_21d	Macro risk	0.0424
NN (1 layer) (macro)	10	log_mkt_cap	Size	0.0357
NN (1 layer) (macro)	11	dolvol	Trading/liquidity	0.0353
NN (1 layer) (macro)	12	beta	Risk	0.0351
NN (1 layer) (macro)	13	turn	Trading/liquidity	0.0282
NN (1 layer) (macro)	14	sector_Industrial	Sector	0.0225
NN (1 layer) (macro)	15	sector_Financial	Sector	0.0202
NN (1 layer) (macro)	16	book_to_market	Valuation	0.0129
NN (1 layer) (macro)	17	sd_turn	Trading/liquidity	0.0115
NN (1 layer) (macro)	18	sector_Consumer,Cyclical	Sector	0.0111
NN (1 layer) (macro)	19	sector_Consumer,Non-cyclical	Sector	0.0098
NN (1 layer) (macro)	20	sp	Valuation	0.0084
NN (2 layers) (macro)	1	macro_beta_dimson_21d	Macro risk	0.1385
NN (2 layers) (macro)	2	macro_rvol_21d	Macro risk	0.1212
NN (2 layers) (macro)	3	macro_be_me	Macro valuation/profitability	0.0995
NN (2 layers) (macro)	4	macro_qmj_safety	Macro risk	0.0939
NN (2 layers) (macro)	5	macro_ocf_at_chg1	Macro valuation/profitability	0.0912
NN (2 layers) (macro)	6	macro_zero_trades_21d	Macro trading friction	0.0843
NN (2 layers) (macro)	7	macro_qmj_prof	Macro valuation/profitability	0.0619
NN (2 layers) (macro)	8	macro_bev_mev	Macro valuation/profitability	0.0570
NN (2 layers) (macro)	9	turn	Trading/liquidity	0.0242
NN (2 layers) (macro)	10	beta	Risk	0.0230
NN (2 layers) (macro)	11	dolvol	Trading/liquidity	0.0202
NN (2 layers) (macro)	12	sector_Financial	Sector	0.0167
NN (2 layers) (macro)	13	log_mkt_cap	Size	0.0150
NN (2 layers) (macro)	14	sd_turn	Trading/liquidity	0.0137
NN (2 layers) (macro)	15	book_to_market	Valuation	0.0123
NN (2 layers) (macro)	16	mom1m	Momentum	0.0120
NN (2 layers) (macro)	17	mom12m	Momentum	0.0114

Continued on next page

Table 11: Top-20 predictors by model

Model	Rank	Feature	Family	Importance share
NN (2 layers) (macro)	18	sector_Industrial	Sector	0.0113
NN (2 layers) (macro)	19	chmom	Momentum	0.0112
NN (2 layers) (macro)	20	ep	Valuation	0.0111
NN (3 layers) (macro)	1	macro_be_me	Macro valuation/profitability	0.1234
NN (3 layers) (macro)	2	macro_qmj_safety	Macro risk	0.1153
NN (3 layers) (macro)	3	macro_bev_mev	Macro valuation/profitability	0.0964
NN (3 layers) (macro)	4	macro_qmj_prof	Macro valuation/profitability	0.0829
NN (3 layers) (macro)	5	macro_beta_dimson_21d	Macro risk	0.0808
NN (3 layers) (macro)	6	macro_ocf_at_chg1	Macro valuation/profitability	0.0729
NN (3 layers) (macro)	7	macro_rvol_21d	Macro risk	0.0594
NN (3 layers) (macro)	8	macro_zero_trades_21d	Macro trading friction	0.0412
NN (3 layers) (macro)	9	sector_Industrial	Sector	0.0292
NN (3 layers) (macro)	10	ep	Valuation	0.0264
NN (3 layers) (macro)	11	log_mkt_cap	Size	0.0249
NN (3 layers) (macro)	12	turn	Trading/liquidity	0.0239
NN (3 layers) (macro)	13	mom12m	Momentum	0.0233
NN (3 layers) (macro)	14	sp	Valuation	0.0227
NN (3 layers) (macro)	15	book_to_market	Valuation	0.0211
NN (3 layers) (macro)	16	sd_turn	Trading/liquidity	0.0207
NN (3 layers) (macro)	17	sector_Financial	Sector	0.0201
NN (3 layers) (macro)	18	beta	Risk	0.0161
NN (3 layers) (macro)	19	mom36m	Momentum	0.0146
NN (3 layers) (macro)	20	dolvol	Trading/liquidity	0.0140