Bay Street Meets Machine Learning: Predicting Stock Risk Premium

Najah Attig*
Dalhousie University
najah.attig@dal.ca

Chahine Attig
ENSAI
chahine.attig@eleve.ensai.fr

May 2025

Abstract: We evaluate the performance of machine learning (ML) methods in forecasting Canadian equity risk premiums and show that moderately flexible models—such as gradient-boosted trees (XGBoost) and mid-depth neural networks—consistently outperform deeper architectures and traditional linear benchmarks across portfolio strategies and horizons. Monthly long–short portfolios constructed using these models generate economically meaningful excess returns and notable improvements in Sharpe ratios relative to traditional benchmarks. Annual backtests further confirm the robustness of these gains across investment horizons. Our results highlight that return predictability is concentrated in small-cap and value stocks. Furthermore, predictive power is substantially enhanced in stocks characterized by market frictions, including high information asymmetry and low trading volume, where regularized and nonlinear models achieve superior risk-adjusted returns. We also show that elevated net anonymous buying and greater broker-level dispersion of anonymous order flow augment forecast accuracy, yielding 60 to 80 basis points higher monthly returns and Sharpe ratios that approach or exceed 0.9 for flexible learners. Overall, select ML techniques materially enhance return forecasting and portfolio construction in the Canadian equity market.

Keywords: Stock return prediction, Machine learning, Canadian stock market, Asset Pricing, Anonymous Trading
JEL Classification C55, C58, G11, G12, G14, G17

---

* Contact author. Najah Attig is a Full Professor of Finance and Chair of the Department of Finance at Dalhousie University, Canada. Chahine Attig is a data science engineering student at ENSAI (École Nationale de la Statistique et de l'Analyse de l'Information), France. We are grateful to Paul Brockman, and Jihene Fayala Attig. The authors are especially indebted to Tamara McAllister from TMX Group (DataLinx) for her assistance with the anonymous trading data. We also acknowledge the generous financial support provided by the Social Sciences and Humanities Research Council of Canada (SSHRC). All remaining errors are our own.

1. Introduction

This study investigates two timely and underexplored questions in the context of Canadian capital markets, with direct implications for investment management. First, it provides the first empirical evidence on the performance of machine learning (ML)[1] methods in forecasting Canadian stock risk premiums. Second, it offers novel evidence on whether anonymous trading imbalances carry predictive power for the cross-section of stock returns. Together, these contributions add the empirical asset pricing literature and offer valuable insights for Canadian capital markets. We expect the findings of this study to inform the evolving toolkit of asset managers and researchers seeking to enhance return forecasting amid increasing data complexity and shifting market dynamics. This is because return predictability remains one of the most consequential themes in modern finance, given its direct implications for asset pricing, portfolio allocation, and trading decisions (Abhyankar et al., 2012). Importantly, despite increasing market complexity—including heightened volatility and evolving information flows—the academic pursuit of return predictability has only intensified, reflecting its enduring importance in navigating today's dynamic financial landscape (Jensen et al., 2023; Gu et al., 2020).

The growing complexity of financial markets—characterized by heightened volatility and evolving information dynamics—has deepened academic interest in return predictability, underscoring its enduring importance for both theory and practice (Jensen et al., 2023; Gu et al., 2020). This sustained focus is well justified, as even marginal improvements in forecast accuracy can generate substantial economic value (Campbell and Thompson, 2008). At the same time, this expanding literature has fueled debate over a foundational question: can "empirical models accurately forecast the equity premium any better than the historical

[1] ML broadly refers to high-dimensional predictive models that incorporate regularization to prevent overfitting, along with efficient algorithms for exploring numerous model specifications (Gu et al. 2020).

mean?" (Spiegel, 2008, p. 1453). While extensive research has identified a wide range of return predictors (e.g., Campbell and Thompson, 2008; Xu and Liu, 2024), skepticism about spurious predictability and data mining—particularly when predictors are correlated (Stambaugh, 1999)—has motivated a parallel stream of work focused on econometric corrections (e.g., Lewellen, 2004; Campbell and Yogo, 2006).

A salient critique from Goyal and Welch (2003) is that strong in-sample correlations often mask systematic out-of-sample underperformance, with few predictors consistently outperforming the historical average. They further argue (Goyal and Welch, 2004) that this failure is symptomatic of the widespread reliance on simple linear regression models. More recent studies (Hou et al., 2020; Drobetz and Otto, 2021) echo this concern and suggest that model misspecification may be a key driver of poor out-of-sample performance. Importantly, models that relax the linearity assumption of the pricing kernel[2]—as imposed in traditional frameworks such as the CAPM and APT—often yield improved predictive performance. Seminal studies (e.g., Bansal et al., 1993; Dittmar, 2002; Asgharian and Karlsson, 2008) demonstrate that accounting for nonlinearities and complex interactions between variables can lead to more robust and economically meaningful forecasts.

Given the inherent noise in stock returns and the potential multicollinearity among predictors, increasing the number of predictors can cause simple linear models to overfit noise rather than extract meaningful signals, thereby undermining predictive stability (Drobetz and Otto, 2021; Gu et al., 2020). Against this backdrop, machine learning (ML) methods offer considerable promise for enhancing return predictability. Their flexibility enables them to handle high-dimensional predictor spaces and capture complex nonlinear interactions (Gu

[2] The Euler equation, $\mathrm{E}\left[\left(\left(1+\mathrm{R}_{\mathrm{i},\mathrm{t}+1}\right)\cdot \mathrm{m}_{\mathrm{t}+1}\middle|\Omega_{\mathrm{t}}\right)\right]=1$, characterizes the first-order condition of an investor's intertemporal consumption–investment decision, where $\mathrm{m}_{\mathrm{t}+1}$ denotes the stochastic discount factor (SDF), also referred to as the intertemporal marginal rate of substitution (see Cochrane, 2005).

et al., 2020). Despite this potential, applications of ML in empirical asset pricing remain relatively limited. Gu et al. (2020) provide one of the first comprehensive comparisons of ML algorithms for forecasting asset risk premiums, demonstrating that ML-based forecasts can yield substantial economic gains for investors. Drobetz and Otto (2021) similarly show that accounting for nonlinearities and interaction effects improves predictive accuracy and portfolio performance in European equity markets, with long-only ML strategies producing higher returns and Sharpe ratios. Extending this line of inquiry, Leippold et al. (2022) apply multiple ML techniques to develop a comprehensive set of return predictors for the Chinese stock market.[3]

In our empirical analysis, we build on a burgeoning body of research (e.g., Gu et al., 2020; Drobetz and Otto, 2021; Leippold et al., 2022) by comparing a suite of ML techniques to a linear benchmark model. Our baseline specification incorporates 37 firm-level and 15 macroeconomic predictors, including liquidity, investment, volatility, and valuation signals. We draw on a comprehensive panel of 129,047 firm-month observations covering 1,308 Canadian firms over the period January 1990 to December 2023. We evaluate each model's performance using both statistical and economic criteria, including portfolio-level backtests at monthly and annual frequencies.

Our results show that moderately flexible nonlinear models—particularly mid-depth neural networks and gradient-boosted trees (XGBoost)—consistently outperform deeper architectures and regularized linear models in predicting stock returns. For instance, monthly long–short portfolios constructed with XGBoost deliver average excess returns exceeding 1% per month, with Sharpe ratios above 0.75. In comparison, OLS benchmarks generate only 0.44% average monthly returns with Sharpe ratios around 0.30. Annual backtests reinforce

[3] Drobetz et al. (2024) show that ML-based estimators of time-varying market betas for U.S. stocks produce the lowest forecast and hedging errors.

these findings: leading models explain up to 23% of return variation and yield average annual returns over 20%, with Sharpe ratios approaching 0.9.

We show that return predictability is concentrated in small-cap and value stocks. Equal-weighted long–short portfolios consistently outperform their value-weighted counterparts by 10 to 20 basis points monthly, suggesting that smaller firms contribute disproportionately to forecastable alpha. Within style categories, value stocks exhibit particularly strong predictive signals—monthly returns exceed 3% with Sharpe ratios above 1.1—whereas growth stocks show weaker performance, with simpler models (e.g., LASSO, Elastic Net) outperforming complex ones due to the noisier nature of the input signals.

We further find that market microstructure frictions enhance predictive performance. Stocks with high information asymmetry (proxied by bid–ask spreads) and low trading volume exhibit higher forecastability, with flexible models extracting more reliable signals under these conditions.

Importantly, a novel insight from our analysis is the relevance of anonymous trading in return predictability. Stock of elevated net anonymous buying are associated with monthly return gains of 60 to 80 basis points, with Sharpe ratios nearly doubling for penalized regression models. Moreover, greater broker-level dispersion in anonymous order flow—capturing heterogeneity in private information—amplifies the performance of flexible, interaction-based learners such as Random Forest, XGBoost, and mid-depth neural networks. These models achieve Sharpe ratios above 0.9 in such environments, while simpler and overly deep models show diminished performance.

From an investment management perspective, these findings underscore that the effectiveness of predictive models is highly contingent on firm characteristics. Small-cap returns are largely driven by liquidity and fast-moving fundamentals, consistent with the

view that microstructure inefficiencies dominate in these segments. In contrast, large-cap returns are more influenced by systematic risk exposures and capital structure variables. Similarly, value stocks are well suited to flexible ML models that can capture nonlinear patterns, whereas growth stocks benefit from regularized approaches that mitigate overfitting in noisy signal environments. These results suggest that investors should tailor their forecasting and allocation strategies accordingly. For example, stable and implementable portfolios may benefit from shallow neural networks or penalized linear models when applied to illiquid stocks, while complex models require rigorous risk controls when deployed in more liquid, large-cap universes.

The remainder of the paper is organized as follows. Section 2 provides the literature background relevant to our research context. Section 3 describes the data and outlines the methodological approach used for return prediction. Section 4 presents the empirical results, including the assessment of out-of-sample predictability, the identification of key predictors, and model selection based on unconditional and conditional predictive ability tests. Section 5 investigates whether predictive performance translates into portfolio gains. Section 6 concludes.

2. Literature Background

Forecasting the equity risk premium has a long history, dating back to Graham and Dodd (1934), who argued that high valuation ratios signal undervalued markets and should predict higher future returns. Most empirical asset pricing models—both time-series and cross-sectional[4]—typically assume linear relationships between financial variables and

[4] Stock return prediction can be viewed through the lens of time-series (TS) models, which forecast aggregate returns using macroeconomic variables and technical indicators (e.g., Cochrane 2011; Rapach and Zhou 2013), and cross-sectional (CS) models, which explain return variation across individual stocks based on firm-level characteristics (e.g., Fama and French 1993; Jegadeesh and Titman 1993), typically estimated via Fama–MacBeth regressions (Drobetz and Otto 2021).

subsequent stock returns (see Drobetz and Otto, 2021, for a discussion). A classic example is the Capital Asset Pricing Model (CAPM) developed by Sharpe (1964), Lintner (1965), and Mossin (1966), which posits that expected returns are driven solely by exposure to market risk, implying a linear dependence on a single factor. However, as Campbell and Thompson (2008) note, academic finance began to rigorously examine this premise only in the 1980s, following empirical studies by Rozeff (1984), Fama and French (1988), and Campbell and Shiller (1988) showing that valuation ratios such as the dividend-price and earnings-price ratios possess substantial predictive power, particularly over long horizons. Parallel research identified other return predictors, including interest rates (Fama and Schwert, 1977; Campbell, 1987), corporate issuance activity (Baker and Wurgler, 2000), consumption-wealth ratios (Lettau and Ludvigson, 2001), and relative valuations (Polk et al., 2003).

Yet, the robustness of return predictability has long been debated. Early concerns centered on the persistence of predictors, biases in estimated coefficients (Stambaugh, 1999), and the risk of data mining (Ferson et al., 2003). Goyal and Welch (2004) notably show that many predictors fail out-of-sample and rarely outperform the historical average. This sparked an ongoing debate about whether equity premium predictability reflects a genuine market feature or results from in-sample overfitting. Part of the issue may stem from the common assumption of a linear relationship between the equity premium and its predictors. However, evidence suggests that this relationship is often non-linear and time-varying due to structural breaks and model instability (Rapach, Strauss, and Zhou 2010; Pettenuzzo and Timmermann 2011).[5] Subsequent studies have addressed these concerns through improved econometric techniques (e.g., Campbell and Thompson 2008; Rapach et al. 2010; Henkel et al. 2011), the

[5] As the number of predictors increases, simple linear models often overfit noise rather than capture meaningful signals, making them unreliable for forecasting future stock returns given the inherent noise in financial data (Drobetz and Otto, 2021). Moreover, parameter instability and structural breaks further exacerbate estimation uncertainty in standard return forecasting models (Rapach, Strauss, and Zhou, 2010; Pettenuzzo and Timmermann, 2011).

use of technical indicators (Neely et al. 2014), and the adoption of modern approaches such as ridgeless regression (Kelly et al. 2024).

While these efforts are valuable, as even modest predictive power can enhance investment decisions (Campbell and Thompson, 2008), still, consensus remains elusive. Recent evidence (e.g., Dichtl et al. 2021) warns that many advanced models fail to consistently outperform the historical average in out-of-sample tests, especially when accounting for data-snooping biases. Most recently, Goyal et al. (2024) reassess variables from 26 post-2008 studies and find that over one-third lose in-sample significance, half perform poorly out-of-sample, and only a small subset exhibit robust predictive power in both settings.

Interstginyly, the recent advancements in ML have reinvigorated research on equity premium predictability, as ML models are capable of capturing complex, non-linear relationships without relying on rigid assumptions. This flexibility makes ML particularly useful in asset pricing (Gu et al. 2020; Akbari et al. 2021, Leippold et al. 2022). Gu et al. (2020), for instance, highlight three key benefits of ML in return premium prediction: (1) enhanced flexibility to model intricate relationships, (2) the ability to address high dimensionality and multicollinearity through variable selection and dimensionality reduction, and (3) methodological diversity that allows for modeling nonlinearities and interactions, reducing overfitting and false discovery through penalization and model selection techniques.[6]

Despite this paradigm shift, research on using ML methods to predict equity risk premiums remains limited. Gu et al. (2020) conducted a comparative analysis of ML techniques for measuring equity risk premiums and identified significant economic gains.

[6] See Drobetz and Otto (2021) for a discussion of the literature on applied machine learning techniques in asset pricing and return prediction. For a discussion of the growing importance of machine learning and textual information processing in finance, see Heston and Sinha (2017).

Similarly, Leippold et al. (2022) applied these methods to the Chinese stock market, finding high predictability for large stocks and state-owned enterprises over longer horizons. Drobetz and Otto (2021) demonstrated that accounting for nonlinearities and interactions, effectively captured by ML, enhances predictive performance over linear models in the European stock market, with ML-based trading strategies leading to notable improvements in return and risk metrics.

Building on this new line of research, we evaluate the performance of ML methods in forecasting Canadian equity risk premiums. Our focus on Canada provides a compelling and underexplored research setting for several reasons. First, restricting the analysis to a single country offers a more homogeneous environment in terms of financial development, legal institutions, corporate governance, and industrial composition—all of which influence the effectiveness of return predictors (Assoe et al., 2024). Second, while the asset pricing literature remains heavily U.S.-centric, incorporating non-U.S. evidence is critical for mitigating the pervasive home bias in academic research (Karolyi, 2016). With a significant share of global equity market capitalization (SIFMA, 2023), Canada represents an ideal setting for extending findings beyond the U.S. context. This is particularly relevant given the segmentation between Canadian and U.S. equity markets, which differ meaningfully in valuation levels and cost of capital (King and Segal, 2008).

Despite perceived economic integration and overlapping institutional features (Irvine, 2000; La Porta et al., 2006; Bargeron et al., 2010), important differences remain—especially in regulatory regimes and corporate governance structures (Attig et al., 2006; Nicholls, 2006; Kryzanowski and Zhang, 2013). For instance, Canadian firms often exhibit higher ownership concentration and operate within a largely voluntary governance framework, typically viewed as weaker than that of the U.S. (Baker et al., 2011). These institutional distinctions shape the relevance and performance of return predictors and help explain valuation gaps between

Canadian-listed firms and those cross-listed across both markets (Athanassakos and Ackert, 2021; King and Segal, 2008). Exploring these nuances is particularly important amid growing concerns about the replicability and robustness of asset pricing models (Harvey et al., 2016; Hou et al., 2022; Chen and Zimmermann, 2022; Jensen et al., 2023).

3. Data

3.1 Sample Construction

Our analysis begins with the universe of Canadian stocks from the Canadian Financial Markets Research Center (CFMRC-TSX) monthly database. We focus on TSX-listed ordinary common shares, excluding REITs, income trusts, and exchangeable shares. For firms with multiple share classes, we aggregate market capitalizations and assign the stock to the class with the highest market capitalization. Penny stocks are excluded. We then match TSX-listed firms with COMPUSTAT and eliminate observations with missing or negative total assets or sales, as well as firms classified as financial institutions (SIC codes 6000–6999), utilities (4900–4999), or non-operating entities (9000–9999). To ensure consistency, we further restrict the sample to common stocks identified by COMPUSTAT as each firm's primary security. Finally, we merge this dataset with Jensen et al. (2023)[7] to obtain a final sample comprising 129,047 firm-month observations, covering 1,308 firms over the period January 1990 to December 2023.

We define the stock risk premium (SRP) as the difference between a stock's total return and the risk-free rate, proxied by the one-month return on three-month Government of Canada Treasury bills (from CFMRC). Building on Jensen et al. (2023), we start with a

[7] Available at https://jkpfactors.com/stock-char and downloadable via WRDS: https://wrds-www.wharton.upenn.edu/pages/get-data/contributed-data-forms/global-factor-data/

comprehensive set of 85 stock-level predictive characteristics[8] widely used in the asset pricing literature and demonstrated to have strong explanatory power for the cross-section of expected returns (e.g., Chordia et al., 2017; Harvey et al., 2016; Lewellen, 2015; McLean and Pontiff, 2016; Leippold et al., 2022). The variables and their definitions are primarily based on Jensen et al. (2023). Notably, accounting characteristics are assumed to become available four months after the fiscal year-end, and we use the most recent accounting data—annual or quarterly—to construct predictors. To mitigate seasonality effects, quarterly income and cash flow items are aggregated over the trailing four quarters (see Jensen et al., 2023 for further details).[9] In a supplementary analysis (Section 3.4), we investigate whether anonymous trading imbalances contribute to explaining the cross-section of Canadian stock returns.

Complementing our set of firm-level variables, we construct 28 capital market and macroeconomic predictors based on prior studies (e.g., Welch, 2008; Champagne et al., 2018; Gu et al., 2020; Drobetz and Otto, 2021; Leippold et al., 2022; Jensen et al., 2023). Most macroeconomic variables are obtained from the CANSIM database.[10]

3.2 Data preparation:

Following Campbell and Thompson (2008) and Gu et al. (2020), we impose a **four-month effective lag** on all accounting variables: **three months** to accommodate standard reporting delays and **one additional month** due to the release lag already embedded in our lag1_ series. We define the stock risk premium (SRP) as the **one-month-ahead excess**

[8] We began with a set of 153 firm-level predictors. To mitigate multicollinearity concerns, we retained only one variable from each pair (of the same category) exhibiting a correlation coefficient greater than 0.50.

[9] Jensen et al. (2023) construct characteristics separately from annual and quarterly Compustat data and retain the most recent value available from either dataset.

[10] CANSIM (short for Canadian Socio-Economic Information Management System) is the main socioeconomic database maintained by Statistics Canada

return, ensuring that the predictive design remains free from look-ahead bias. To mitigate the influence of extreme values and potential data errors, we winsorize all monthly firm characteristics, including excess returns, at the 1st and 99th percentiles (Jensen et al., 2023; Drobetz and Otto, 2021).[11]

We then standardize macroeconomic indicators (e.g., inflation, exchange rates, term spreads, money growth) using pre-2009 means and standard deviations. Firm-level features are cross-sectionally rank-normalized each month to the [–1, +1] interval. This hybrid approach preserves the time-series dynamics of broad economic trends while ensuring comparability and robustness to heavy tails across firms. To capture state-dependent effects, we construct pairwise interaction terms between each standardized macro indicator and each normalized firm-level variable.

To reduce dimensionality, we drop predictors with near-zero variance or a high number of missing observations. We then run 100 bootstrap LASSO regressions (using cv.glmnet with $\alpha = 1$) on pre-2009 data, retaining only those predictors that receive nonzero coefficients in at least 70% of resamples. Finally, we identify all variable pairs with Pearson correlations above 0.70 and drop one member of each pair. Macroeconomic indicators and calendar dummies are always retained. This yields a final design matrix of 381 predictors for use in our machine learning and linear model analyses.

[11] As highlighted by Drobetz et al. (2019) and Drobetz and Otto (2021), two key issues arise when using firm-level characteristics to predict returns. First, many predictors exhibit strong temporal persistence—either as slowly changing level variables (e.g., firm size) or as aggregated flow variables (e.g., book equity)—implying that return predictability may extend beyond short horizons (Campbell & Cochrane, 1999; Cochrane, 2008). Second, overlaps among predictors, particularly issuance and profitability measures, induce high correlations. Consistent with Drobetz and Otto (2021), multicollinearity is not a primary concern here since our focus is on overall model predictive performance rather than individual variable effects. Moreover, machine learning methods address multicollinearity via regularization and variable selection, enhancing forecast accuracy.

Table 1 provides the list of predictor variables along with their definitions and summary statistics. The table also reports time-series averages of the monthly cross-sectional means and standard deviations for excess returns, all 37 firm-level characteristics, and the 15 macroeconomic variables, along with the overall sample.

Table 1 goes here.

For the macro variables, we control for a comprehensive set of macroeconomic and market-level variables to capture systematic influences on asset returns. These include: inflation (INFLATION), measured as the monthly change in the Consumer Price Index (CPI); money supply growth (M2_GR), defined as the monthly change in the M2 monetary aggregate; unemployment rate (UNEMPLOYMENT); growth in capacity utilization (CAP_UTI_GR), capturing the percentage of production capacity in use; growth in manufacturers' sales (MANU_SALES_GR); and commodity price fluctuations, proxied by the monthly change in the Fisher Commodity Index (FISHER_PRICE_GR), which tracks global price movements in 26 key Canadian commodities.

We also include the composite leading indicator (CLI_GR), an OECD-based index designed to anticipate turning points in economic activity; short-term interest rates (RET30_TBILL), measured as the one-month return on three-month Government of Canada Treasury bills; term spread (TERM_SPREAD), calculated as the yield differential between long- and short-term government bonds; and the exchange rate (X_RATE), defined as the number of Canadian dollars per U.S. dollar.

To account for financial market conditions, we control for equity market volatility (SPTSXVOL30), computed as the 30-day annualized standard deviation of daily log returns

on the S&P/TSX 300 Index; excess market returns (SPTSX_RP), measured as the return on the S&P/TSX Composite Index in excess of the risk-free rate; equity issuance activity (TSX_NET_ISSUE), following Welch and Goyal (2008), defined as the 12-month sum of firm-level net issuance scaled by aggregate market capitalization; and trading activity (TSX_VOLUME_GR), measured as the monthly growth in total dollar trading volume on the Toronto Stock Exchange.

Additionally, we control for economic policy uncertainty (EPU), which reflects uncertainty surrounding government actions that influence the economic environment. Although inherently unobservable, EPU is commonly proxied by the news-based index developed by Baker et al. (2016), which measures the frequency of newspaper articles containing terms related to the economy, policy, and uncertainty. This index captures uncertainty regarding who will make policy decisions, what actions will be taken, when they will occur, and what their economic effects might be. EPU is included due to its documented effects on the equity risk premium (Brogaard and Detzel, 2015), firm investment and efficiency (Gulen and Ion, 2016; Drobetz et al., 2018), bank liquidity hoarding (Berger et al., 2020), M&A activity (Bonaime et al., 2018), financial reporting quality (El Ghoul et al., 2021), and dividend policy (Attig et al., 2021). Following prior research (e.g., Brogaard and Detzel, 2015; Gulen and Ion, 2016; Bonaime et al., 2018), we use the monthly aggregate EPU index.[12]

Finally, we incorporate a January dummy (JAN), set to 1 in January and 0 otherwise, which accounts for the well-documented January Effect, and a December dummy (DEC), set

[12] Baker et al. (2016) construct the EPU index by performing automated text searches of major newspapers to count monthly articles containing terms related to the economy (E), policy (P), and uncertainty (U). These counts are scaled by the total number of articles in each newspaper and month, standardized to unit standard deviation per newspaper, averaged across newspapers by country-month, and normalized to a mean of 100 over the sample period. The resulting index, available at www.policyuncertainty.com, offers a consistent cross-country measure of economic policy uncertainty.

to 1 in December and 0 otherwise, to control for possible year-end effects related to tax-loss selling or portfolio rebalancing.[13]

4. Research Design and Results

4.1 Machine learning methods

For our empirical analysis, we adapt the methodological framework of Gu et al. (2020), Drobetz and Otto (2021) and Leippold et al. (2022) to the context of the Canadian equity market. Namely, to forecast one-month-ahead excess returns $R_{i,t+1}$, we model the conditional expectation $E\left[\left(R_{i,t+1}\middle|z_{i,t}\right)\right]$ as a function $\hat{g}(z_{i,t})$ of P = 381 firm-month predictors $z_{i,t}$. We train all models using a rolling, expanding-window framework. For each calendar year y=2012,…,2023y = 2012, \dots, 2023y=2012,…,2023, we estimate the models on (i) a training sample spanning January 1990 to December 2008 (108 months), followed by (ii) a fixed validation window from January 2009 to December 2011 (36 months). We then produce out-of-sample forecasts for the twelve test months of year y, store the model parameters, and roll the test window forward by one year.[14]

One of the two primary objectives of this study is to examine whether incorporating interaction effects and nonlinearities enhances the predictability of Canadian stock returns. To this end, we assess the performance of seven modeling approaches—emphasizing predictive accuracy rather than structural interpretation—from both statistical and economic perspectives. Specifically, we evaluate whether flexible ML methods outperform standard linear benchmarks in forecasting excess stock returns using a high-dimensional set of firm-

[13] We do not control for industry fixed effects (based on the Fama-French 48 classification) in the final model, as their influence is largely absorbed by the extensive set of firm-specific predictors included. Moreover, these fixed effects were filtered out during the LASSO variable selection process, suggesting limited incremental explanatory power in this context.

[14] This procedure yields twelve non-overlapping test folds.

level and macroeconomic predictors in the Canadian equity market. The models considered are as follows:

1. **Ordinary Least Squares (OLS)**: This serves as our benchmark model and minimizes the standard squared error loss across all firm-month observations. We namely fit a standard linear model of the form:

$$\hat{g}_{OLS}(z) = \hat{\beta}_0 + z^T\hat{\beta}$$

This model includes the complete set of 381 predictors, the variables described in Table 1 and their interactions.

2. **Least Absolute Shrinkage and Selection Operator (LASSO)**:[15] is a penalized regression method that simultaneously performs variable selection and regularization, making it particularly effective in high-dimensional settings. By augmenting the traditional ordinary least squares (OLS) loss function with an ℓ_1-norm penalty on the regression coefficients, LASSO encourages sparsity—shrinking some coefficients exactly to zero—thereby improving model interpretability and reducing overfitting. This feature is particularly valuable in financial applications, where predictors tend to be numerous and potentially highly correlated. Formally, LASSO solves the following optimization problem:

$$\hat{\beta} = \arg min_{\beta}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(r_{i,t+1} - \beta_0 - z_{i,t}^T\beta\right)^2 + \lambda\sum_{j=1}^{P}|\beta_j|\right\}$$

where:

[15] Introduced by Tibshirani (1996).

- $r_{i,t+1}$ is the one-month-ahead excess return for firm I at t+1,
- $z_{i,t}$ is P-dimesntional vector of predictor variables,
- $\lambda > 0$ is a regularization parameter controlling the sparsity of the solution

The $\lambda \sum_{j=1}^{P} |\beta_j|$ includes sparsity by penalizing the absolute magnitude of the coefficients. We select λ via ten-fold cross-validation to balalnce the trade-off between model complexity and out-of-sample forecast accuracy.

LASSO provides a strong baseline for evaluating more flexible machine learning models. However, its reliance on linearity limits its ability to capture complex nonlinear interactions, which motivates the exploration of richer methods such as ensemble trees and neural networks (see Gu, Kelly, and Xiu, 2020).

3. **Elastic Net (ENet)**:[16], combines the strengths of LASSO and Ridge regression by incorporating both ℓ_1 and ℓ_2 penalties. This hybrid approach balances variable selection and coefficient shrinkage, making it especially useful when predictors are highly correlated—a common feature in financial datasets. Unlike LASSO, which may select only one variable among a group of correlated predictors, ENet can retain grouped variables, improving model stability and robustness. The ENet objective function is defined as:

$$\hat{\beta} = \arg min_{\beta} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(r_{i,t+1} - \beta_0 - z_{i,t}^{T} \beta \right)^2 + \lambda \left[\alpha \sum_{j=1}^{P} |\beta_j| + (1-\alpha) \sum_{j=1}^{P} \beta_j^2 \right] \right\}$$

where:

- $r_{i,t+1}$ is the one-month-ahead excess return for firm I at t+1,
- $z_{i,t}$ is P-dimesntional vector of predictor variables,

[16] Introduced by Zou and Hastie (2005).

- $\lambda > 0$ is a regularization parameter controlling the sparsity of the solution
- $\alpha \in [0, 1]$ governs the mix between the LASSO (ℓ_1) and Ridge (ℓ_2) penalties

We fix α=0.5 to balance sparsity and shrinkage and use ten-fold cross-validation to select the optimal λ, following standard practice in empirical finance (e.g., Gu et al., 2020).

ENet serves as an effective compromise between variable selection and multicollinearity control, offering enhanced predictive performance when the number of predictors is large and correlated. In our framework, ENet helps mitigate overfitting while preserving information embedded in grouped features.

4. **Partial Least Squares (PLS)**:[17] Extracts ten latent components from the predictor matrix before performing linear regression. PLS, is a dimension-reduction technique designed to handle high-dimensional, multicollinear predictor sets. Unlike principal components regression (PCR), which forms components solely based on the predictor variance, PLS constructs components that maximize the covariance between the predictors and the response variable, enhancing predictive relevance. In PLS, the high-dimensional predictor matrix $\mathbf{Z}_{i,t} \in \mathbb{R}^{N \times P}$ is decomposed into a set of orthogonal latent components that capture the directions in the predictor space most associated with the target variable $r_{i,t+1}$. These components are then used in a standard linear regression framework. Following prior work (e.g., Gu et al., 2020), we extract ten latent components from the firm-month predictor matrix and regress one-month-ahead excess returns on these components

$$r_{i,t+1} = \beta_0 + \sum_{j=1}^{P} \gamma_k \cdot PLS_{k,i,t} + e_{i,t+1}$$

[17] Introduced by Wold (1975).

where $PLS_{k,i,t}$ denotes the kkk-th latent factor extracted from the predictor matrix $\mathbf{Z}_{i,t}$.

PLS is particularly suited for financial applications where the number of predictors is large and where explanatory variables may be noisy or highly collinear. By projecting data into a lower-dimensional, supervised space, PLS can improve forecast stability without sacrificing signal relevance.

5. **Random Forests (RF)**:[18] Random Forests, are an ensemble learning method that improves predictive accuracy by averaging the forecasts of multiple regression trees built on different bootstrap samples of the training data. Each tree captures complex nonlinear relationships and interactions by recursively partitioning the predictor space to minimize in-sample squared error. To reduce correlation among trees and improve generalization, RF randomly select a subset of predictors—commonly referred to as mtry—to consider at each split. In our implementation, we grow 500 trees and tune mtry$\in\{$P,P/2,P/3$\}$ based on validation root mean squared error (RMSE). Each tree is grown to full depth without pruning, and the final prediction is obtained by averaging across all trees.

Random Forests are particularly well-suited for financial applications involving high-dimensional data, as they are robust to multicollinearity, accommodate both continuous and categorical predictors, and inherently model nonlinearities and higher-order interactions.

6. **Gradient-Boosted Trees (XGBoost)**: are an ensemble method that builds predictive models in a sequential manner by iteratively fitting shallow regression trees to the residuals of previous predictions. Each successive tree is trained to correct the errors of the current model, and the contributions of new trees are scaled by a learning rate $\eta$to prevent overfitting. We implement the XGBoost algorithm (Chen and

[18]Introduced by Breiman (2001).

Guestrin, 2016), tuning the learning rate $\eta \in \{0.01, 0.1\}$ and maximum tree depth $d \in \{4, 6\}$. To avoid overfitting, we apply early stopping: training halts if validation loss fails to improve for ten consecutive rounds. The final prediction is the weighted sum of the outputs from all fitted trees.

XGBoost is well-suited for high-dimensional financial data, offering strong predictive accuracy through its ability to model nonlinearities and complex interactions while controlling model complexity via regularization.

7. **Feed-forward Neural Networks (NNs)**: offer a flexible modeling approach capable of capturing complex, nonlinear interactions between predictors and returns. Inspired by biological neural systems, these models consist of layers of interconnected nodes (neurons), where each neuron applies an activation function to a weighted sum of inputs from the previous layer. We estimate four NN architectures: NN1 (single hidden layer with 64 units), NN2 (64–32), NN3 (128–64), and NN4 (256–128), each with ReLU activation functions in the hidden layers. To prevent overfitting, we implement several regularization strategies:

    - **Dropout**: applied at rates of 10% and 20% to randomly deactivate neurons during training.
    - **Weight decay**: an ℓ_2 penalty with a coefficient of 10^{-4} is used to shrink weights.
    - **Early stopping**: training halts if validation loss does not improve for 15 consecutive epochs.
    - **Learning rate**: chosen from $\{10^3, 5 \times 10^{-3}\}$, controlling step size in gradient descent.
    - **Batch size**: fixed at 256 to improve training efficiency and stabilize updates.

All networks are trained using stochastic gradient descent (SGD) and evaluated via out-of-sample mean squared forecast error (MSFE). These architectures offer a lower bound on the forecasting potential of neural models in our context and serve as a comparison benchmark against both linear and tree-based machine learning methods.

It is important to note that our approach prioritizes predictive accuracy over structural interpretation, aiming to evaluate whether flexible machine learning methods can enhance return forecasting relative to traditional linear benchmarks when applied to a high-dimensional set of predictors in the Canadian equity market.

4.2 Statistical inference:

This section describes the methods used to estimate model parameters, assess predictive accuracy, and compare forecast performance across benchmarks. In-sample coefficient estimates from OLS and Elastic-Net regressions are computed with two-way clustered standard errors by firm and month, following Petersen (2009). For average monthly portfolio returns and annualized Sharpe ratios, we use Newey–West heteroskedasticity and autocorrelation-consistent (HAC) standard errors with 11 lags to account for serial correlation from overlapping return windows.

For the out-of-sample goodness-of-fit metric R^2_{OOS}, we report point estimates only, given its bounded support and non-normal sampling distribution, which renders standard inference unreliable. To formally evaluate forecast performance, we employ Diebold–Mariano (1995) tests based on squared prediction error differentials, again using HAC standard errors with 11 lags to correct for serial dependence.

For instance, in Panel A of Table 2 we present selected in-sample coefficients from OLS and Elastic-Net models.[19] Consistent with established asset pricing anomalies, lagged

[19] Full estimates for all 381 predictors are available upon request from the authors.

log market equity (Log_me) is negatively associated with future returns, while return momentum (RET_12_1) shows a positive and statistically significant effect. These results are robust across specifications.

Table 2 goes here.

Panel B reports Diebold–Mariano statistics comparing each model's forecast errors to the OLS benchmark. Negative and significant HAC-adjusted statistics indicate that several models—notably LASSO, Elastic-Net, and Random Forest—generate superior forecasts. In line with this, the R^2_{OOS} values confirm the enhanced explanatory power of machine learning models. Together, these results underscore the potential of flexible, high-dimensional methods to improve return predictability in complex financial environments.

4.3 Results

Before presenting the empirical results, it is important to define the primary out-of-sample performance metric used throughout the paper: a pseudo out-of-sample R^2, constructed from the root mean squared error (RMSE) and the unconditional variance of the target variable. Let $\{y_i, \hat{y}_i\}_{i=1}^N$ denote the true and predicted returns in the hold-out sample (pooled across all cross-validation folds). The RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

We scale this error by the variance of the target variable computed over the full sample:

$$Var(y_i) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \text{ , where } y_i = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

The pseudo out-of-sample $R^2$is then defined as:

$$R^2_{OOS} = 1 - \frac{RMSE^2}{Var(y)}$$

which is algebraically equivalent to the forecast-MSPE ratio popularized in Goyal and Welch (2008) and Campbell and Thompson (2008).

We compute the variance using the full sample rather than fold-specific test sets. This approach ensures that R^2_{OOS} remains a single, directly comparable metric across all models and specifications. For conditional or subsample analyses, we re-estimate the denominator using the unconditional variance of the target variable within the corresponding subsample, computed over the original (non-test) observations. This adjustment serves two purposes: it maintains the meaningful interpretation of the statistic as an MSPE-to-population-variance ratio, and it reflects the distinct volatility regimes that may characterize different sample partitions. For completeness and transparency, we also report the raw RMSE alongside the pseudo-R^2 (R^2_{OOS}).

4.3.1. Out-of-sample predictability

4.3.1.1. Full sample analysis

Table 3 reports the out-of-sample forecasting performance across ten predictive models, evaluated using the root mean squared error (RMSE) and the pseudo R-squared metric R^2_{OOS}. These results underscore the inherent difficulty of predicting monthly Canadian stock returns—predictive signals are modest, yet performance differentials among models are economically meaningful.

The strongest results emerge from the two shrinkage-based linear models: Elastic-Net and LASSO, which attain the highest R^2_{OOS} values (6.20% and 6.18%, respectively), along

with the lowest RMSEs. Their ability to retain relevant, potentially correlated predictors while shrinking noise appears particularly well-suited to the high-dimensional and noisy nature of return forecasting. Deep neural networks also perform competitively—specifically, the three-layer architectures (NN4 and NN3) explain approximately 6.0% of the out-of-sample variance—closely trailing the regularized linear models. However, this marginal improvement suggests that increasing model complexity and nonlinearity yields limited additional benefit once regularization is applied.

In contrast, simpler nonlinear models perform less effectively. The two-layer neural network (NN2) and random forest achieve R^2_{OOS} values near 5.3%, while the shallow NN1 yields only 5.1%. Gradient-boosted trees (XGB) perform worst, with an R^2_{OOS} of 3.7%—less than half that of the top shrinkage methods—indicating that boosting and tree depth do not automatically translate into better performance in this context.

Traditional dimensionality-reduction techniques also underperform. Partial least squares (PLS) achieves 4.4%, while OLS—without any form of regularization—produces the lowest among linear methods at 4.3%. These findings confirm that regularization is not only beneficial but essential in high-noise, high-collinearity environments.

Taken together, the results demonstrate that well-tuned linear shrinkage methods outperform both traditional and flexible nonlinear approaches, highlighting the value of regularization in extracting stable predictive signals. Figure 1 visually reinforces these conclusions, showing a clear performance gap between Elastic-Net, LASSO, and the rest. In the sections that follow, we investigate whether these patterns hold across firm characteristics such as size and valuation, or under different market conditions.

Table 3 and Figure 1 go here.

4.3.1.2. Cross-Sectional Heterogeneity: Firm Size and Investment Style

To assess whether the performance of return prediction models varies systematically across observable firm characteristics, Table 4 reports out-of-sample R^2_{OOS} values by firm size and investment style. Namely, Panel A partitions stocks by market capitalization into Small-cap (bottom 30%) and Large-cap (top 70%) groups. Predictive accuracy is consistently stronger for small firms. Elastic-Net and LASSO deliver the highest performance in this segment, with R^2_{OOS} values of 12.78% and 12.77%, respectively. Other models—including moderately deep neural networks (NN3, NN4), Random Forest, and XGBoost—also post robust performance, ranging from 11.5% to 12.5%. In contrast, model performance deteriorates sharply in the large-cap segment. The best-performing models (Elastic-Net, LASSO) explain less than 5% of return variation, and nonlinear models such as XGBoost and NN1 fail to deliver meaningful predictive gains. The performance gap between small and large firms ranges from 6.8 to 8.6 percentage points across models.

To a large extent, these patterns align with the view that small-cap stocks, being more opaque and subject to greater frictions and limits to arbitrage, present greater opportunities for forecasting models to extract exploitable signals. Large-cap stocks, by contrast, tend to trade in more efficient, information-rich environments where return predictability is more limited.

Table 4 goes here.

In Panel B, we group firms by investment style, classifying stocks as Value (bottom 30% of price-to-book ratio) or Growth (top 30%). The predictive power of all models is highly

concentrated in the value segment. Neural nets (NN4, NN3), Elastic-Net, and LASSO all achieve R^2_{OOS} values exceeding 24%, with the deepest net (NN4) posting the highest explanatory power at 25.8%. Even simpler models such as OLS and PLS achieve over 23% predictive fit. In stark contrast, all models produce negative out-of-sample R^2_{OOS} in the growth segment—indicating that none outperform the historical mean forecast for these firms. The performance gap between value and growth stocks exceeds 34 percentage points across models. This finding corroborates prior evidence that return predictability is concentrated in value stocks, which are often under-analyzed, mispriced, or more sensitive to fundamental signals. Growth stocks, on the other hand, behave more like a random walk and remain largely unpredictable, even for sophisticated models.[20]

4.3.1.3 Predictability at Annual Horizon

In Table 5, we extend the analysis to a one-year return horizon. As reported in Table X, only non-linear models achieve positive out-of-sample R^2_{OOS} values. Random Forest (RF) leads with an R^2_{OOS} of 6.2%, followed by the three-layer neural network (NN3) at 5.6%, and the four-layer NN4 at 4.8%. Shallower networks (NN2 and NN1) deliver lower, yet still positive, R^2_{OOS} values of 4.3% and 3.2%, respectively. In contrast, gradient-boosted trees (XGB) underperform slightly (−0.2%), and all linear models—including LASSO (−2.96%), Elastic Net (−3.01%), OLS (−4.33%), and PLS (−6.39%)—fail to generate positive forecast accuracy. Despite the larger raw RMSEs due to the higher volatility of annual returns, the relative R^2_{OOS} values remain in line with shorter-horizon results. This reflects the rescaling

[20] In untabulated results, we assess the temporal robustness of return forecasting models by splitting the out-of-sample prediction period into two macroeconomically distinct subperiods: the 2010s (2012–2020) and the early 2020s (2021–2024). This division captures both the post-global financial crisis recovery and low-volatility environment of the 2010s, as well as the heightened uncertainty and market disruptions associated with the COVID-19 pandemic and subsequent policy interventions. Importantly, model parameters are held constant throughout the entire sample to isolate the effect of changing macroeconomic regimes on predictive performance. While detailed results are omitted for brevity, they are available upon request.

by the increased variance of the one-year return target, which compresses gains in predictability.

Table 5 goes here.

Panel B of Table 5 examines cross-sectional heterogeneity by firm size. Small-cap stocks continue to exhibit greater predictability than large caps, though the gap is modest. For small firms, RF reaches 9.2% R^2_{OOS} , NN3 7.5%, NN4 6.8%, and NN2 6.1%. Among large firms, the same models yield between 4.3% and 5.5%. While linear models collapse on large caps (e.g., LASSO and Elastic Net both around −5.6%), they maintain mildly positive performance for small caps (≈2.5%). Thus, while small-cap returns remain more predictable, non-linear models still outperform within the large-cap segment.

Panel C highlights style effects based on valuation. The disparity between value and growth stocks is pronounced. In the value portfolio, all non-linear models post strong performance—NN3 and NN4 reach 21.8% and 21.2% R^2_{OOS} , respectively, followed by RF at 20.6% and NN2 at 19.1%. Linear methods also perform reasonably well, all exceeding 5% R^2_{OOS} . Conversely, in the growth segment, no model achieves positive predictability; all R^2_{OOS} values are negative, ranging from −7.1% for RF to −12.2% for XGB. This yields substantial style gaps of 18–31 percentage points.

In sum, the evidence in Table 5 indictes that non-linear models dominate at the annual horizon. Predictability is strongest among small-cap and value stocks. However, no model delivers consistent, horizon-invariant performance across firm size and investment style segments.

4.4 Which Predictors Matter?

To better understand the sources of return predictability, we next investigate the relative importance of the predictors included in our models. Given the breadth of the feature set, we differentiate between two broad categories: macroeconomic indicators and firm-specific characteristics. This analysis enables us to identify which variables contribute most significantly to forecasting performance, and whether the predictive signal is primarily driven by aggregate economic conditions or idiosyncratic firm-level information.

In this analysis, we focus on the eight models most relevant for interpretability—six machine learning methods (NN1–NN4, Random Forest, and XGBoost) and two penalized linear models (LASSO and Elastic Net). We exclude classical linear benchmarks, including PLS and OLS variants, from the heatmap visualization, as our discussion centers on the machine learning approaches that have demonstrated the strongest predictive performance. Among linear models, only LASSO and Elastic Net are retained to serve as regularized baselines. Results are visualized in the heatmap "Macroeconomic Variable Importance" (Figure 2).

Figure 2 goes here.

We note that for the two shrinkage regressions the pattern remains highly concentrated. Elastic-Net and LASSO both give their maximum normalised weight (1.00) to OECD leading-indicator growth and assign the next-highest scores to the CAD / USD exchange rate (≈ 0.75) and M2 growth (≈ 0.70). Inflation, labour-market slack, volatility and every equity-supply or trading variable receive values below 0.35, indicating that the

penalised linear models continue to collapse the macro set to one dominant cyclical gauge plus a liquidity-currency pair.

The neural networks distribute importance more broadly but still reveal a clear hierarchy. In all four nets the single strongest series is 30-day S&P/TSX realised volatility, which attains the maximum score in every architecture. The shallow configurations (NN1, NN2) complement volatility with sizeable weight on economic-policy uncertainty, the exchange rate and short-rate returns; NN2 in particular draws heavily on the term spread and on high-frequency equity-market signals such as TSX trading-volume growth and net share issuance. The deeper nets (NN3, NN4) keep volatility at the top while adding moderate emphasis to relative-performance momentum, inflation and capacity-utilisation growth. No single variable other than volatility dominates these deeper networks; instead they rely on a mixture of cyclical, rate-sensitive and market-microstructure measures.

The tree ensembles pivot to yet another subset. Random Forest and XGBoost both score the unemployment rate at 1.00, making labour-market slack their primary macro input. The exchange rate follows at roughly 0.90 in Random Forest and 0.70 in XGBoost, while the OECD index still contributes (0.57 and 0.88, respectively). Short-rate returns and capacity-utilisation growth offer smaller support, and most other series—including the money supply, term spread and equity-market additions—are effectively ignored (normalised scores near zero). This narrow, high-impact selection is characteristic of trees' tendency to pick a handful of decisive non-linear thresholds.

These findings underscore that there is no single "best" macro predictor. Linear shrinkage models gravitate toward the leading indicator and a currency-liquidity factor, neural nets elevate equity-market volatility while blending in several complementary signals, and tree-based learners focus on labour-market conditions and exchange-rate moves. Each

algorithm's inductive bias therefore channels a different macro signal into return forecasts, reinforcing the value of ensemble approaches that blend these complementary perspectives.

We next examine which firm-level characteristics most influence annual return forecasts.[21] Starting from the full, cleaned design matrix, we compute variable importance at the fold level for each model: absolute coefficients for Elastic Net and LASSO, permutation-based loss increases for neural networks, gain-based splits for XGBoost, and mean-decrease-in-MSE for Random Forest. These importance scores are normalized to the [0,1] range within each fold and then averaged across the rolling forecast windows. Table 6 presents the top 15 predictors for each model, which together account for at least 54.3% (Random Forest) of total model-level importance.

Table 6 goes here.

Across all eight models, liquidity and trading-friction variables dominate the predictive structure of twelve-month TSX excess returns. In the penalized regressions (Elastic Net and LASSO), six-month idiosyncratic volatility (ivol_capm_60m) and the 21-day zero-trade ratio (bidaskhl_21d) are the top-ranked features, followed closely by net working capital scaled by sales (Sale_nwc) and one-year cumulative return (RET_12_1). These four variables consistently appear in the top five across all neural networks (NN1–NN4) and tree-based models (XGBoost and Random Forest), alongside turnover over 126 trading days (turnover_126d) and the O-score distress measure. In total, 12 of the 15 most influential predictors recur in every architecture, underscoring the stability of liquidity and past-return effects across diverse estimation paradigms.

[21] We do not report the heatmap for firm-level predictors due to the large number of variables involved.

Model-specific distinctions nonetheless emerge. The shrinkage regressions produce nearly identical rankings—unsurprising given their shared regularization framework. The shallow neural network (NN1) shifts greater importance toward working-capital ratios and volatility-adjusted issuance metrics (e.g., EQNPO_GR1A). Deeper networks (NN3, NN4) give increased weight to accrual-based signals (Eqnetis_gr1a), leverage (debt_bev), and structural distress (Z_score). The tree-based models surface unique non-linear drivers, including firm age, analyst forecast dispersion (earnings variability), and long-horizon return skewness (RET_36_1). These differences highlight the capacity of trees to extract signal from discrete and interaction-heavy variables often overlooked in linear models.

In summary, while liquidity, volatility, and investment-related variables constitute a common predictive core, each algorithm's inductive bias subtly reshapes the importance hierarchy—revealing how different architectures extract and prioritize firm-level information for return forecasting.

4.5 Robustness Test: Alternative Model Evaluation

In this section, we go beyond conventional out-of-sample performance metrics—such as R^2_{OOS} and RMSE—to assess the statistical robustness of our forecasting models. These measures, while informative, can obscure important performance differentials, especially under varying macroeconomic conditions. We therefore conduct formal tests of superior predictive ability (SPA) to evaluate whether observed gains in forecast accuracy are statistically meaningful.

Specifically, we implement two frameworks: (i) the Unconditional SPA (USPA) test of Hansen (2005), which evaluates average loss differentials over the full sample; and (ii) the Conditional SPA (CSPA) test of Li et al. (2020), which examines whether model performance holds under different macroeconomic regimes. Both tests are applied using squared-error loss.

Table 7 reports the number of pairwise model rejections under the USPA and CSPA frameworks. The USPA row summarizes how often each model is significantly outperformed by others across 45 one-versus-one comparisons. The following rows break down CSPA rejections across 13 macroeconomic conditions, including inflation, exchange rates, economic policy uncertainty (EPU), and interest rate changes. Each cell reflects how frequently a given model is rejected under a specific regime, while the final column aggregates rejections across all regimes.

Table 7 goes here.

The USPA results indicate that NN2 and NN3 are the most robust models, each incurring only 4 rejections, compared to 5 for Elastic Net and LASSO, 6 for OLS, PLS, and NN4, 6 for Random Forest, and 7 for XGBoost. Even the single-layer NN1 performs relatively well, with just 3 rejections.

The performance gap widens in the conditional setting. Under CSPA, NN3 and NN2 are rejected only 28 and 31 times, respectively, across 260 regime-specific tests—substantially lower than the 114–123 rejections accumulated by Random Forest, XGBoost, and all linear models. These results demonstrate that neural networks, particularly NN3 and NN2, maintain forecast superiority even when macroeconomic conditions vary, while linear and tree-based models are far more sensitive to such shifts.

These robustness checks offer two key insights. First, models with similar average R^2_{OOS} may differ significantly in their statistical reliability. Second, model performance is regime-dependent—OLS, PLS, Elastic Net, and Random Forest often lose their predictive advantage under inflationary, uncertain, or capacity-constrained environments. Taken

together, the evidence positions NN3 as the most statistically robust model, followed closely by NN2. These architectures not only lead in average accuracy but also deliver stable and resilient forecasts across a broad range of macroeconomic conditions.

4.6 Dissecting the Leading Model's Predictive Signals

To uncover the economic drivers underlying the performance of our top-performing annual-horizon model—Random Forest—we analyze its variable importance across firm size segments. We freeze the model trained on the full sample and apply it separately to test observations from the bottom 30% (small-cap) and top 70% (large-cap) of the market capitalization distribution. Within each segment, we permute one predictor at a time, re-estimate returns using the frozen model, and record the change in root-mean-squared error (ΔRMSE). These values are scaled to the [0,1] interval within each bucket, and we compute the difference (Small – Large) to identify variables more important for small versus large firms.

Panel A of Figure 3 displays the top 15 predictors with the largest absolute differences. Panel B presents the same analysis at the theme level, grouping predictors into ten broad economic categories (excluding interaction terms). At the variable level, small-cap predictability is primarily driven by short-horizon profitability and price momentum. Permuting the three-month change in common shares outstanding (CHCSHO), asset turnover growth, twelve-month momentum (both absolute and relative to the S&P/TSX), standardized ROE, and free cash flow to enterprise value materially increases RMSE for small firms (by 40–60 basis points) but has negligible effects on large caps.

Figure 3 goes here.

In contrast, large-cap forecasts rely more on balance sheet structure and macroeconomic indicators. Variables such as headline inflation, composite leading indicator (CLI) growth, EBITDA-to-sales, debt-to-book equity, asset tangibility, and SG&A growth meaningfully affect large-cap predictions, raising RMSE by 25–35 basis points, with minimal small-cap impact. Measures of trading activity and external financing (e.g., turnover, TSX net issuance, and sales-to-working-capital ratio) also exhibit a large-cap tilt.

Theme-level results reinforce this segmentation. Small-cap forecasts are most sensitive to liquidity-related earnings, volatility, and momentum signals, which together explain approximately 60% of the total ΔRMSE in that segment. Large-cap forecasts are more influenced by valuation, leverage, classical risk factors, and firm characteristics (e.g., size, age), each contributing 20–30 basis points to forecast error. Growth and macro themes display limited size-based differentiation, suggesting uniform macro exposure at the annual horizon.

Overall, these findings indicate that return predictability stems from different economic channels across firm size segments. For small caps, predictability reflects frictions in microstructure and delayed incorporation of fast-moving fundamentals. For large caps, it arises from balance sheet fundamentals, systematic risk exposure, and valuation. These patterns align with heterogeneous-attention theories: greater analyst coverage and liquidity facilitate rapid information incorporation for large firms, whereas small firms remain prone to inefficiencies tied to trading frictions and incremental profitability updates.

4.7 Portfolio analysis

So far, our evaluation of predictive performance has been purely statistical, based on out-of-sample $\boldsymbol{R^2}$ and formal hypothesis testing. We now assess the economic value of return forecasts by examining their implementability in portfolio strategies that reflect practical

constraints in the Canadian market, such as short-selling limitations. Specifically, we use each model's one-month-ahead predictions to rank stocks and compute realized returns based on actual one-month-ahead total returns (RET_1_0). This exercise translates predictive accuracy into economic terms, comparing the performance of value-weighted and equal-weighted long–short portfolios, as well as a long-only top-decile strategy, over the period January 2012 to December 2023.

4.71. Portfolio sorts

Monthly-rebalanced portfolios: Table 8 reports the results of a total return backtest using one-month-ahead forecasts over the period January 2012 to December 2023. At each month-end, stocks are sorted into deciles based on predicted total returns. We evaluate three portfolio strategies: (i) a long–short strategy that buys the top decile and shorts the bottom decile, constructed using both value and equal weights; and (ii) a long-only strategy that invests in the top decile only. For each strategy, we report the average monthly return (Avg, %), monthly volatility (Std, %), annualized Sharpe ratio (scaled by $\sqrt{12}$), skewness, kurtosis, maximum drawdown (Max DD, %), and worst single-month return (MinMonthly, %).

Table 8 goes here.

All ten forecasting models generate positive average monthly long–short spreads under value-weighting, though performance levels and statistical significance vary. XGBoost delivers the highest mean spread, at 1.05% per month ($\sigma \approx 4.88\%$, Sharpe ≈ 0.75, NW $t = 2.58$), followed by NN1 at 0.92% (Sharpe ≈ 0.69, $t = 2.67$) and Random Forest at 0.81% (Sharpe ≈ 0.64, $t = 2.20$). Regularized linear models also perform competitively: LASSO and

Elastic Net each yield spreads of approximately 0.74% (Sharpe ≈ 0.57, $t \approx 1.70$). Mid-depth neural networks (e.g., NN3 at 0.85%) and the deepest model (NN4 at 0.31%) occupy the middle and lower range, respectively. The OLS benchmark generates the lowest spread (0.44%, Sharpe ≈ 0.30, $t = 1.03$), consistent with its relatively weak predictive performance.

Switching to equal-weighting increases all spreads by approximately 10 to 20 basis points. XGBoost again leads with a 1.19% average monthly spread (Sharpe ≈ 0.91, NW $t = 3.15$), followed by NN1 at 0.98% (Sharpe ≈ 0.77, $t = 2.67$) and Random Forest at 0.94% (Sharpe ≈ 0.74, $t = 2.55$). LASSO and Elastic Net each deliver spreads of approximately 0.82% (Sharpe ≈ 0.58), while OLS posts a lower 0.55% (Sharpe ≈ 0.40, $t \approx 1.37$). NN3 performs comparably at 0.84%, whereas NN4 trails at 0.34%.

The long-only leg amplifies gross returns substantially. XGBoost again achieves the highest return, averaging 1.61% per month ($\sigma \approx 7.04\%$, Sharpe ≈ 0.79, $t = 2.73$). Elastic Net and LASSO follow closely at 1.53% (Sharpe ≈ 0.74, $t \approx 2.56$). NN1 and NN3 post 1.49% (Sharpe ≈ 0.72, $t = 2.50$) and 1.36% (Sharpe ≈ 0.62, $t \approx 2.13$), respectively. Random Forest yields 1.46% (Sharpe ≈ 0.75, $t = 2.58$), while OLS and PLS generate 1.19% and 1.16%, respectively (Sharpe ≈ 0.57, $t \approx 1.91$). The deepest networks—NN2 (1.23%) and NN4 (0.97%)—finish slightly behind.

In summary, flexible nonlinear models (XGBoost, NN1, NN3) consistently outperform linear benchmarks across both long–short and long-only strategies. The relatively weaker performance of NN4 suggests diminishing returns to network depth in this setting. See Figure 4 for a visualization of the annual portfolio strategy performance across models and weighting schemes.

Figure 4 goes here.

Annual Portfolio Backtests: To evaluate the effectiveness of the return forecasts at a longer horizon, we conduct an annual portfolio backtest based on one-year-ahead predictions. Each January from 2012 to 2023, we rank all stocks according to each model's forecast and implement three strategies using the resulting cross-sectional rankings: (i) a long–short portfolio that buys the top decile and shorts the bottom decile using both value and equal weighting; and (ii) a long-only strategy that invests in the top decile. We hold each position for 12 months and then compute the realized total return over the horizon t+1 to t12. We report the average annual return (Avg, %), annualized standard deviation (Std, %), Sharpe ratio (unscaled), and Newey–West t-statistic with an 11-month lag. Results are presented in Table 9.

Table 9 goes here.

Across all three strategies, the most flexible models generally deliver the highest risk-adjusted returns. In the value-weighted long–short portfolios, the three-layer neural network (NN3) leads with an average annual return of **20.3%** ($\sigma \approx 21.1\%$, Sharpe ≈ 0.89, $\mathbf{t} \approx 5.02$), followed by NN4 at 18.0% and the shrinkage methods—Elastic Net and LASSO—at approximately 18.7%–18.9%. Random Forest lags behind with a 14.0% return (Sharpe ≈ 0.50, $\mathbf{t} \approx 4.01$). Switching to equal-weighting amplifies both returns and Sharpe ratios across the board. NN3 again tops the performance table at 22.1% ($\sigma \approx 20.9\%$, Sharpe ≈ 0.99, $\mathbf{t} \approx 5.46$), closely matched by LASSO and Elastic Net. Interestingly, even the basic OLS model achieves a competitive return of 19.8% (Sharpe ≈ 0.73, $\mathbf{t} \approx 4.97$), highlighting that equal-weighting

increases the contribution of smaller firms and sharpens alpha signals from extreme forecast ranks.

In long-only portfolios, absolute returns rise further, albeit with increased volatility. Elastic Net and LASSO again perform strongly, generating 26.9%–27.0% annual returns (Sharpe ≈ 0.59, **t** ≈ 4.88–4.91), with NN3 just behind at 26.0% (Sharpe ≈ 0.68, **t** ≈ 4.11). Several models—including OLS, PLS, and XGBoost—cluster near 24%–24.6% (Sharpe ≈ 0.56–0.58), while Random Forest trails slightly at 23.1%.

Notably, XGBoost and NN3 deliver the strongest Newey–West t-statistics under equal weighting—6.31 and 5.46, respectively—signaling robust statistical performance even in relatively noisy one-year horizons. Maximum drawdowns are largest in the long-only strategies, particularly for LASSO (≈ 37.7%) and NN4 (≈ 36.9%), but remain moderate in long–short portfolios, where the short leg provides meaningful downside protection.

To recap, two core insights emerge from this analysis: first, **m**oderately deep neural networks (NN2–NN4) and shrinkage regressions (LASSO, ENet) consistently rank among the best performers, suggesting they strike an effective trade-off between model complexity and overfitting. Second, **e**qual-weighted long–short portfolios outperform value-weighted **ones** on both return and risk-adjusted metrics, implying that small-cap stocks drive a disproportionate share of forecastable alpha. Even OLS delivers economically meaningful performance when applied to the extremes of the predicted return distribution—demonstrating that forecast rank, not precision, is the primary driver of cross-sectional portfolio value.

4.7.2 Performance by Investment Style

In this section we investgate whether model performance varies systematically across investment styles. Specifically, we partition the universe each month into value stocks (top

30% by book-to-market) and growth stocks (bottom 30%), and implement long-only top-decile portfolios within each group. Results are reported in Table 10.

Table 10 goes here.

In the growth-stock subsample (Panel A), XGBoost stands out as the top performer, delivering an average excess return of 1.22 % per month ($\sigma \approx 7.44$ %, Sharpe ≈ 0.57, NW t ≈ 1.96). Regularized linear models follow close behind: LASSO posts 0.99 % pm (Sharpe $\approx$ 0.46, NW t ≈ 1.59) and Elastic Net 0.97 % pm (Sharpe ≈ 0.45, NW t ≈ 1.56). Among the neural networks, the three-layer network (NN3) generates 0.86 % pm (Sharpe ≈ 0.43, NW t ≈ 1.49), while the shallow one-hidden-layer model (NN1) delivers a more modest 0.81 % pm (Sharpe ≈ 0.37, NW t ≈ 1.28). Traditional linear benchmarks are comparable: OLS earns 0.98 % pm (Sharpe ≈ 0.46, NW t ≈ 1.59) and PLS 0.86 % pm (Sharpe ≈ 0.39, NW t ≈ 1.36). Random Forest lags at 0.75 % pm (Sharpe ≈ 0.34, NW t ≈ 1.17). Increasing network depth does not help: NN2 slips to 0.60 % pm (Sharpe ≈ 0.30, NW t ≈ 1.03) and the four-layer NN4 is essentially flat at 0.10 % pm (Sharpe ≈ 0.05, NW t ≈ 0.17). This evidence appears to favour boosted trees and simpler, regularised linear models over capacity-heavy neural architectures.

In the value-stock subsample (Panel B), forecast accuracy improves markedly. NN1 again leads, posting an impressive 2.99% per month (volatility $\approx 9.06\%$, Sharpe ≈ 1.14, NW t ≈ 3.94), followed by Random Forest (2.65%, Sharpe ≈ 0.99, t ≈ 3.41) and PLS (2.47%, Sharpe ≈ 0.91, t ≈ 3.14). OLS and XGBoost also deliver robust returns near 2.35%-2.37% (Sharpe ≈ 0.88 and 0.79, respectively). LASSO and Elastic Net yield similar performance ($\approx$2.30%, Sharpe ≈ 0.85). Mid-depth networks (NN2, NN3) produce slightly lower returns ($\approx$1.97%, Sharpe $\approx$ 0.70-0.73), while NN4 trails at 1.69% (Sharpe ≈ 0.61). In sum, value stocks

exhibit richer cross-sectional predictability, and both nonlinear and regularized linear models are effective in extracting that signal. However, excessive model complexity, as seen with NN4, tends to degrade performance.

4.7.3. Additional Cross-Sectional Analyses

We further test the robustness of our findings by conditioning on two dimensions of market frictions: information asymmetry and trading volume. Table 11 reports results for both dimensions. To evaluate performance under different information environments (Panel A), we sort stocks each month into high- and low-asymmetry groups using the cross-sectional median of the 21-day bid–ask spread proxy (BIDASKHL_21D).[22] Within each group, we form top-decile, value-weighted long-only portfolios. Model performance improves substantially in the high-asymmetry group. For instance, Random Forest's mean return nearly triples—from 0.66% to 1.89%—boosting its Sharpe ratio to 0.85 and NW t-stat to 2.93. XGBoost improves from 1.18% to 1.93%, yielding the highest Sharpe ratio at 0.90 ($t \approx 3.11$). Elastic Net and LASSO also show sharp gains: mean returns rise from 0.82% to over 2.10%, Sharpe ratios nearly double to 0.90, and t-stats breach 3.0. Shallow networks like NN1 benefit modestly (Sharpe $\approx$ 0.67), while deeper networks such as NN4 see performance deteriorate, consistent with earlier results. These findings underscore the advantage of nonparametric and regularized models in information-scarce environments.

Table 11 goes here.

---

[22] Following prior literature (e.g., Heflin and Shaw, 2000; Amihud, 2002; Attig, Fong, Lang, and Gadhoum, 2006), we proxy information asymmetry using the relative bid–ask spread.

In Panel B, we examine whether predictive efficacy varies by market liquidity. [23] Stocks are sorted each month by lagged 21-day trading volume into high- and low-volume groups, and top-decile, value-weighted long-only portfolios are implemented in each segment. A distinct pattern emerges: low-volume stocks yield superior risk-adjusted returns across nearly all models. Elastic Net's mean return falls slightly in the low-volume segment (2.19% to 1.57%), but lower volatility boosts the Sharpe ratio from 0.57 to 0.72. LASSO and PLS display similar dynamics. Neural networks respond more strongly to the lower volatility regime. NN1's Sharpe nearly triples to 0.74, while NN2's performance moves from a loss to a 1.49% gain (Sharpe $\approx$ 0.75). Even NN4 sees its Sharpe increase, albeit modestly.

Tree-based models show mixed reactions. Random Forest performs best in low-volume stocks (1.87% return, Sharpe $\approx$ 0.74), while XGBoost's highest nominal return is in the high-volume group (2.39%), yet its Sharpe remains higher in the low-volume tier (0.84 vs. 0.69). OLS mirrors the behavior of regularized models: slightly lower returns in thinly traded stocks are offset by reduced volatility, resulting in improved risk-adjusted performance.

In sum, evidence in Table 11 indicates that predictive signals manifest differently across liquidity regimes. In thinly traded stocks, models with built-in regularization or simpler architectures perform best. In contrast, high-volume stocks offer larger nominal returns but greater risk; only highly flexible models like XGBoost manage to harness that return effectively. Panels A and B of Figure 5 visualize the results of Table 11.

Figure 5 goes here.

[23] We consider the role of trading volume, given its longstanding association with price discovery and investor disagreement (Harris and Raviv, 1993). Wang (1994) further links volume behavior to heterogeneity in investor beliefs.

4.8 The Role of Anonymous Trading in Return Predictability

In a novel extension of our analysis, we examine whether the structure of anonymous trading activity enhances the cross-sectional predictability of stock returns. Specifically, we examine two key facets of anonymous trading behavior: (i) the net volume of anonymous buying (NET_ANO_BUY_VOLUM), and (ii) the dispersion of broker-level anonymous order flow (VOLATILITY_BROKER_VOLUM).

These dimensions are motivated by theoretical and empirical research linking trading activity, information asymmetry, and price discovery. Anonymous trades, by design, conceal trader identity and are frequently used by informed institutional investors to avoid revealing private signals. We thus posit that anonymous buying is more likely to reflect informed trading than anonymous selling, consistent with the empirical finding that buy orders tend to be more informative than sell orders (Madhavan, 1995; Keim and Madhavan, 1995). This asymmetry may arise because purchases typically signal selective optimism or stronger preferences for individual securities, while sales may reflect portfolio rebalancing or liquidity needs rather than informational advantage.

The dispersion in anonymous trading across brokers captures the heterogeneity of private information and belief disagreement among informed traders. Building on Anderson, Ghysels, and Juergens (2005), we hypothesize that differing broker-level anonymous flows may reflect access to distinct information sets or divergent interpretations of public and private signals. This heterogeneity can be understood as a proxy for the distribution of beliefs about future returns across market participants. Chordia et al. (2001) show that the volatility of trading activity may proxy for investor clientele heterogeneity, which can in turn affect

asset pricing dynamics. To formally quantify this dispersion, we construct a cross-sectional volatility measure of anonymous trading imbalance at the broker level:

$$Sigma_{it}^{ATI} = \left[\sum_{j=1}^{N} \frac{\left(ATI_{ijt} - \overline{ATI}_{it} \right)^2}{N} \right]^{1/2}$$

where ATI_{ijt} is the anonymous trading imbalance for broker j, stock i, and month t, and $\overline{ATI}_{it}$ is the average imbalance across all NNN brokers actively trading stock i during t. A higher value of $Sigma_{it}^{ATI}$ indicates greater dispersion and thus greater heterogeneity of private information or expectations. Informed trading that is more fragmented across brokers may reflect greater uncertainty or disagreement, which could dampen the aggregate predictive signal.

To assess the return implications of these two trading conditions, we perform monthly portfolio sorts. For each month, we split the stock universe at the median value of the relevant proxy—either net anonymous buying or broker dispersion—and re-rank stocks within each subset based on the model's predicted return. We then construct value-weighted, top-decile long-only portfolios, holding positions for 12 months.

Tables 12 present the average monthly return, annualized volatility, and Newey–West t-statistics for the models in each sub-sample. The results are striking. When net anonymous buying is elevated (Panel A), model performance improves substantially. Average monthly returns increase by 60 to 80 basis points across most forecasting models, while Sharpe ratios also rise despite modest increases in volatility. For instance, Elastic Net improves from 1.48% to 2.13% per month, with its Sharpe ratio nearly doubling to 0.88. Similar gains are observed for LASSO, XGBoost, and shallow neural networks (NN1–NN3). Even models with previously weaker performance—such as Random Forest and deeper architectures like NN4—register improvements, albeit to a lesser extent.

Table 12 goes here.

These findings suggest that concentrated hidden buying leaves an identifiable footprint in price dynamics that both linear and non-linear predictive models can exploit. Importantly, they also point to the salience of microstructural frictions and trading anonymity as conditioning variables in return predictability. Future research may further disentangle whether the observed patterns are driven by informed trading, strategic order placement, or institutional preferences under opacity.

In Panel B, we examine whether broker-level dispersion in anonymous order flow—our proxy for the heterogeneity of informed trading—modulates return predictability. The results in Table 16 show that models with flexible learning architectures benefit disproportionately from higher dispersion. When broker disagreement is high, Random Forest's average return rises from 1.40% to 2.09%, and its Sharpe ratio climbs to 0.97. XGBoost exhibits a similarly pronounced gain, with its Sharpe ratio exceeding one. NN2 and NN3 also realize substantial improvements, posting return increases of 70–90 basis points and materially stronger risk-adjusted performance.

By contrast, penalized regressions and PLS regressions experience moderate improvements, while OLS and NN4 exhibit only marginal changes. These patterns suggest that greater heterogeneity in broker-level anonymous activity creates a richer and more complex information environment, which interaction-based learners—such as tree-based models and mid-depth neural networks—are better equipped to exploit. In contrast, linear models and deeper networks may underperform in these settings, either due to their limited flexibility or overfitting sensitivity, respectively.

Taken together, the findings underscore that the efficacy of anonymous trading signals is highly state-dependent. Returns are strongest when anonymous buy-side activity is elevated but not pervasive, and when broker-level order flows are highly dispersed—conditions indicative of both informational intensity and belief heterogeneity. Regularized linear models and moderate-complexity neural networks can translate such structure into stable profits, while higher-capacity models—such as ensemble learners—offer enhanced upside only in dispersion-rich regimes, albeit with higher volatility. These insights suggest that tilting predictive strategies toward stocks with concentrated but non-uniform anonymity and broker heterogeneity can significantly improve long-only portfolio performance while mitigating estimation risk. Figure 6 visualizes the findings reported in Table 12.

Figure 6 goes here.

---

5. Conclusion

This study provides the first comprehensive evaluation of machine learning methods for forecasting the cross-section of Canadian equity returns. Using a rich set of firm-level and macroeconomic predictors, we benchmark a diverse suite of models—including penalized linear methods, tree-based algorithms, and neural networks—against traditional linear forecasting approaches. Our analysis demonstrates that moderately flexible nonlinear models, particularly shallow neural networks and gradient-boosted trees, offer consistently superior performance across both statistical and economic dimensions. Return predictability is strongest in small-cap and value stocks, where information frictions and delayed price adjustment create conditions favorable to machine learning. By contrast, simpler regularized

methods such as Elastic Net and LASSO perform more reliably in growth stocks and large caps, where signal-to-noise ratios are lower and data-generating processes are more linear.

A novel contribution of our study is the identification of market microstructure variables—particularly net anonymous buying and broker-level dispersion in order flow—as critical sources of forecastable alpha. These features, indicative of informed institutional trading and information heterogeneity, substantially improve model accuracy and portfolio performance, especially for flexible, interaction-based learners. The effectiveness of such models, however, is sensitive to model depth and firm-level liquidity, reinforcing the importance of matching model architecture to market conditions.

From a practical perspective, our findings underscore the implementability of ML-based return forecasts. Even under long-only constraints, top-decile portfolios constructed using ML signals yield materially higher Sharpe ratios and excess returns than traditional approaches. These results suggest that ML is not a substitute for economic theory but a complementary tool—particularly valuable in high-dimensional, high-noise settings where conventional models underperform. While our analysis centers on the Canadian equity market, the broader implications extend to other contexts, especially emerging markets where structural inefficiencies and microstructure frictions are more pronounced. As financial data becomes increasingly complex and heterogeneous, machine learning is poised to play a central role in both academic asset pricing research and applied investment strategy. More broadly, our findings support the growing call for ML integration in addressing novel questions in finance—ranging from climate risk pricing and ESG-related mispricing to the influence of retail sentiment and algorithmic trading on return dynamics. Future research should aim not only to enhance predictive accuracy, but also to interpret ML-generated signals and link them to underlying economic mechanisms. Doing so will help bridge the gap between data-driven

inference and theory-consistent modeling, a critical step toward the broader adoption of ML techniques in financial economics.

References

Abhyankar A, Basu D, Stremme A. 2012. The Optimal Use of Return Predictability: An Empirical Study. Journal of Financial and Quantitative Analysis 47, 973-1001.

Aitken, M. J., Henk B., and D. Mak. 2001. The Use of Undisclosed Limit Orders on the Australian Stock Exchange." Journal of Banking & Finance 25, 1589–1603.

Akbari, A., Ng, L. and B. Solnik. 2021. "Drivers of economic and financial integration: A machine learning approach." Journal of Empirical Finance 61:82-102.

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

Asgharian, H., and S. Karlsson. 2008. Evaluating a Non-Linear Asset Pricing Model on International Data. International Review of Financial Analysis 17 (3): 604–621.

Asness, C.S. , Porter, R.B. , Stevens, R.L. , 2000. Predicting Stock Returns Using Industry-Relative Firm Characteristics. Available at SSRN 213872 .

Assoe, K., Attig, N., & Sy, O. (2024). The battle of factors. Global Finance Journal, 62, Article 101004. https://doi.org/10.1016/j.gfj.2024.101004

Athanassakos, G., Ackert, L.F., 2021. The value premium is NOT dead in Canada. Journal of International Finance and Economics 21, 5–19.

Attig, N., El Ghoul, S., Guedhami, O. and X. Zheng. 2021. Dividends and economic policy uncertainty: International evidence. Journal of Corporate Finance 66, 101785, ISSN 0929-1199, https://doi.org/10.1016/j.jcorpfin.2020.101785.

Attig, N., Fong, W-M., Gadhoum, Y., Lang, L.H.P., 2006. Effects of large shareholding on information asymmetry and stock liquidity. Journal of Banking and Finance 30, 2875–2892.

Baker M., and Wurgler J.. 2000. The equity share in new issues and aggregate stock returns. Journal of Finance 55:2219–57.

Baker M., and Wurgler J.. 2007. Investor sentiment in the stock market. Journal of Economic Perspectives 21:129–52.

Baker, H.K., Dutta, S., Saadi, S., 2011. Corporate finance practices in Canada: Where do we stand? Multinational Finance Journal 15, 157–192.

Baker, S.R. , Bloom, N. and S.J. Davis. 2016. Measuring economic policy uncertainty Quarterly Journal of Economics 131, 1593-1636

Bali, T.G. , Cakici, N. , Whitelaw, R.F. , 2011. Maxing out: stocks as lotter- ies and the cross-section of expected returns. Journl of Financial Economics 99, 427–446 .

Bansal, R., and S. Viswanathan. 1993. No Arbitrage and Arbitrage Pric- ing: A New Approach. The Journal of Finance 48 (4): 1231–1262. Bollerslev, T., B. Hood, J. Huss, and L.H. Pedersen. 2018. Risk Eve- rywhere: Modeling and Managing Volatility. Review of Financial Studies 31 (7): 2729–2773.

Bansal, R., D.A. Hsieh, and S. Viswanathan. 1993. A New Approach to International Arbitrage Pricing. The Journal of Finance 48 (5): 1719–1747.

Barbee, W., Mukherji, S. , Raines, G. , 1996. Do sales-price and debt-equity explain stock returns better than book-market and firm size? Financial Analysts Jorunal 52, 56–60 .

Barclay, M., Hendershott, T., and McCormick, T. (2003) Competition among trading venues: Information and trading on electronic communications networks, Journal of Finance 58, 2639-2667.

Bargeron, L., Lehn, K., Zutter, C., 2010. Sarbanes-Oxley and corporate risk-taking. Journal of Accounting and Economics 49, 34–52.

Berger, A.N. , Guedhami, O. , Kim, H.H. and X. Li. 2020. Economic policy uncertainty and bank liquidity hoarding. Journal of Financial Intermdiation 49, p. 100893

Bessembinder, H., Panayides, M., and Venkataraman, K. (2009) Hidden liquidity: An analysis of order exposure strategies in electronic stock markets, Journal of Financial Economics 94(3), 361-383.

Bin, L. , Chen, J. , Puclik, M. , Su, Y. , 2017. Predicting extreme returns in Chi- nese stock market: an application of contextual fundamental analysis. J. Account. Finance 17 (3), 10 .

Black, F. (1986). Noise, The Journal of Finance 41, 529-543.

Bloomfield, R. and O'Hara, M. (2000) Can transparent markets survive? Journal of Financial Economics 55, 425-459.

Bloomfield, R., O'Hara, M. and Saar, G. (2009) How noise trading affect markets: An experimental analysis, Review of Financial Studies 22, 2275-2302.

Bloomfield, R., O'Hara, M. and Saar, G. 2015. Hidden liquidity: Some new light on dark trading, Journal of Finance 70, 2227-2274.

Boehmer, E., O'Hara, M., & Saar, G. (2015). Hidden liquidity: Some new light on dark trading. Journal of Finance, 70(5), 2223–2274.

Bonaime, A.A. , Gulen, H. and M. Ion. 2018. Does policy uncertainty affect mergers and acquisitions? Journal of Financail Economics 129, 531-558

Boulatov, A. and George, T.J. (2013) Hidden and displayed liquidity in securities markets with informed liquidity providers, Review of Financial Studies 26, 2095-2137.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Brennan, M.J., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics 49, 345-373.

Brogaard, J. and A. Detzel. 2015. The asset-pricing implications of government economic policy uncertainty. Management Science 61, 3-18

Campbell J. Y. 1987. Stock returns and the term structure. Journal of Financial Economics 18:373–99.

Campbell, J. Y. and M. Yogo. 2005. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27-60.

Campbell, J. Y., and S. B. Thompson. 2008. Predicting the equity premium out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Campbell, J.Y., and J.H. Cochrane. 1999. By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behav- ior. Journal of Political Economy 107 (2): 205–251.

Champagne, C., Chrétien, S. and F. Coggins. 2018. Equity Premium Predictability: Combination Forecasts versus Multivariate Regression Predictions. Retrived from https://www.cfatoronto.ca/docs/default-source/default-document-library/equity-premium-predictability.pdf?sfvrsn=ff77281e_0

Chen, A.Y., Zimmermann, T., 2022. Open source cross-sectional asset pricing. Critical Finance Review 11, 207–264.

Chen, L. , Pelger, M. , Zhu, J. , 2019. Deep Learning in Asset Pricing. Available at SSRN 3350138

Chen, N., Roll, R., & Ross, S. A. (1986). Economic Forces and the Stock Market. The Journal of Business, 59(3), 383–403

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chen, T. , Gao, Z. , He, J. , Jiang, W. , Xiong, W. , 2019. Daily price limits and destructive market behavior. J. Econom. 208 (1), 249–264 .

Chordia, T., & Swaminathan, B. (2000). Trading volume and cross-autocorrelations in stock returns. The Journal of Finance, 55(2), 913–935.

Chordia, T., A. Goyal, and A. Saretto. 2017. p-Hacking: Evidence from Two Million Trading Strategies. Swiss Finance Institute Research Paper No. 17–37: 1–53.

Chordia, T., Roll, R., and Subrahmanyam, A. (2001) Market liquidity and trading activity, The Journal of Finance 56, 501-530.

Chordia, Tarun, Amit Goyal, and Alessio Saretto, 2020, Anomalies and false rejections, Re of Financial Studies 33, 2134–2179.

Cochrane J. H. 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. Journal of Finance 46:209–37.

Cochrane, J.H. 2005. Asset Pricing. Revised. Princeton (NJ), US: Princeton University Press.

Cochrane, J.H. 2008. The Dog That Did Not Bark: A Defense of Return Predictability. The Review of Financial Studies 21 (4): 1533–1575. Cochrane, J.H. 2011. Presidential Address: Discount Rates. The Journal of Finance 66 (4): 1047–1108.

Cochrane, John H., 2011, Presidential address: Discount rates, Journal of Finance 66, 1047–1108.

Comerton-Forde, C., Malinova, K., and Park, A. (2018) Regulating dark trading: Order flow segmentation and market quality, Journal of Financial Economics 130, 347-366.

Comerton-Forde, C., Putniņš, T.J., and Tang, K.M. (2011) Why do traders choose to trade anonymously? Journal of Financial and Quantitative Analysis 46, 1025-1049.

De Winne, R. and D'Hondt, C. (2007) Hide-and-seek in the market: Placing and detecting hidden orders, Review of Finance 11(4), 663-692.

Deb, S.S. , Kalev, P.S. , Marisetty, V.B. , 2010. Are price limits really bad for equity markets? Journal of Banking & Finance 34, 2462–2471 .

Dichtl, H., Drobetz, W., Neuhierl, A. and V-S. Wendt. 2021. Data snooping in equity premium prediction. International Journal of Forecasting 37 (1):72-94.

Diebold, F., and R. Mariano. 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13 (3): 253–263.

Dittmar, R.F. 2002. Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence From the Cross Section of Equity Returns. The Journal of Finance 57 (1): 369–403.

Drobetz W, Hollstein F, Otto T, and M. Prokopczuk. 2024. Estimating Stock Market Betas via Machine Learning. Journal of Financial and Quantitative Analysis. Published on:1-37.

Drobetz, W. , El Ghoul, S., Guedhami, O. and M. Janzen. 2018. Policy uncertainty, investment, and the cost of capital. Jorunal of Financial Stability 39, 28-45

Drobetz, W., and T. Otto. 2021. Empirical asset pricing via machine learning: evidence from the European stock market. Journal of Asset Management 22, 507–538

Drobetz, W., R. Haller, C. Jasperneite, and T. Otto. 2019. Predictability and the Cross Section of Expected Returns: Evidence from the European Stock Market. Journal of Asset Management 20 (7): 508–533.

El Ghoul, S., Guedhami, O., Kim, Y. and H-J. Yoon. 2021. Policy Uncertainty and Accounting Quality. The Accounting Review 96, 233–260.

Fama E. F., and French K. R.. 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25:23–49.

Fama E. F., and Schwert G. W.. 1977. Asset returns and inflation. Journal of Financial Economics 5:115–46.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. Journal of Financial Economics, 22(1), 3–25.

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

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

Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55-84.

Fama, E.F., French, K.R., 2006. The value premium and the CAPM. Journal of Finance 61, 2163-2185.

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

Foucault, T., Moinas, S., and Theissen, E. (2007) Does anonymity matter in electronic limit order markets, Review of Financial Studies 20, 1707-1747.

Frino, A., Johnstone, D., and Zheng, H. (2010) Anonymity, stealth trading, and the information content of broker identity. The Financial Review 45, 501-522.

Goyal A., and Welch I.. 2003. Predicting the equity premium with dividend ratios. Management Science 49:639–54.

Goyal A., and Welch I.. 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21:1455–508.

Goyal, A., Welch, I. and A. Zafirov. 2024. A comprehensive look at the empirical performance of equity premium prediction II. Review of Financial Studies, Forthcoming, DOI: http://dx.doi.org/10.2139/ssrn.3929119.

Goyal, A., Welch, I., and A. Zafirov. 2024. A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction. The Review of Financial Studies 37, 3490–3557.

Graham, B., & Dodd, D. L. (1934). Security Analysis. New York: McGraw-Hill.

Gu, S., Kelly, B., and D. Xiu. 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33, 2223–2273.

Gulen, H. and M. Ion. 2016. Policy uncertainty and corporate investment. Review of Financial Studies 29, 523-564

Guo H. 2006. On the out-of-sample predictability of stock market returns. Journal of Business 79, 645–70.

Han, B., Tang, Y., and Yang, L. (2016) Public information and uninformed trading: Implications for market liquidity and price efficiency, Journal of Economic Theory 163, 604-643.

Hansen, P.R. , 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23, 365–380 .

Harris, L. (1996). Does a large minimum price variation encourage order exposure? Working Paper, University of Southern California.

Harris, M. and A. Raviv. 1993. Differences of Opinion Make a Horse Race, The Review of Financial Studies 6, 473–506

Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. Review of Financial Studies, 6(3), 473–506.

Harvey C. R., Liu Y., and Zhu H.. 2016. … and the cross-section of expected returns. Review of Financial Studies 29:5–68.

Heflin F, and KW Shaw. 2000. Blockholder Ownership and Market Liquidity. Journal of Financial and Quantitative Analysis 35, 621-633

Henkel, Sam James, J Spencer Martin, and Federico Nardari. 2011. "Time-varying short-horizon predictability." Journal of Financial Economics 99 (3):560-580.

Heston L. S. & N. R. Sinha. 2017. News vs. Sentiment: Predicting Stock Returns from News Stories, Financial Analysts Journal, 73:3, 67-83.

Hou, K., Mo, H., Xue, C., Zhang, L., 2021. An augmented q-factor model with ex-pected growth. Review of Finance 25 (1), 1-41.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. Re-view of Financial Studies 28 (3), 650-705.

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

Huber, P.J. , 2004. Robust Statistics, vol. 523. John Wiley & Sons .

Irvine, P.J.A., 2000. Do analysts generate trade for their firms? Evidence from the Toronto stock exchange. Journal of Accounting Economics 30, 209–226.

Jegadeesh, N., and S. Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance 48 (1): 65–91.

Jensen T. H., Kelly, B. and L. H. Pedersen. 2023. Is There a Replication Crisis in Finance? Jorunal of Finance 78, 2465-2518

Jiang, F. , Kim, K.A. , 2020. Corporate governance in China: a survey. Rev. Finance 24 (4), 733–772 .

Karolyi, A.G., 2016. Home bias: An academic puzzle. Review of Finance 20, 2049–2078.

Kelly, Bryan, Semyon Malamud, and Kangying Zhou. 2024. The virtue of complexity in return prediction. Journal of Finance 79, 459-503.

Kim, O., & Verrecchia, R. E. (2001). The relation among disclosure, returns, and trading volume. The Accounting Review, 76(4), 633–654.

King, R.M., Segal, D., 2008. Market segmentation and equity valuation: Comparing Canada and the United States. Journal of International Financial Markets, Institutions and Money 18, 245–258.

Kryzanowski, L., Zhang, Y., 2013. Financial restatements and Sarbanes–Oxley: Impact on Canadian firm governance and management turnover. Journal of Corporate Finance 21, 87–105.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2006. What works in securities laws? Journal of Finance 61, 1–32.

Lee, C. M. C., & Swaminathan, B. (1998). Price momentum and trading volume. Journal of Finance, 53(5), 2041–2069.

Lee, C. M. C., & Swaminathan, B. (2000). Price momentum and trading volume. The Journal of Finance, 55(5), 2017–2069. https://doi.org/10.1111/0022-1082.00280

Lee, W. , Wang, L. , 2017. Do political connections affect stock price crash risk? Firm-level evidence from China. Rev. Quant. Finance Account. 48 (3), 643–676 .

Leippold, M.; Wang, Q.; Zhou, W. 2022. Machine learning in the Chinese stock market, Journal of Financial Economics 145, 64-82, ISSN 0304-405X.

Lettau M., and Ludvigson S.. 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56:815–49.

Lewellen, J. 2015. The Cross-Section of Expected Stock Returns. Criti- cal Finance Review 4 (1): 1–44.

Lewellen, J., 1999. The time-series relations among expected return, risk, and book-to-market. Journal of Financial Economics 54, 5-43.

Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289-314.

Li, G., Li, F.W., Wei, K.C.J., 2020. Security analysts and capital market anomalies. Journal of Financial Economics 137, 204-230.

Linnainmaa, J.T. and Saar, G. (2012) Lack of anonymity and the inference from order flow, Review of Financial Studies 25, 1414-1456.

Liu, D. , Gu, H. , Xing, T. , 2016. The meltdown of the Chinese equity market in the summer of 2015. Int. Rev. Econ. Finance 45, 504–517 .

Liu, J. , Stambaugh, R.F. , Yuan, Y. , 2019. Size and value in China. J. Financ. Econ. 134 (1), 48–69 .

Liu, Y. , Zhou, G. , Zhu, Y. , 2020. Trend Factor in China. Available at SSRN 3402038 .

Lo, A. W., & Wang, J. (2000). Trading volume: Definitions, data analysis, and implications of portfolio theory. The Review of Financial Studies, 13(2), 257–300.

Madhavan, A. (1996) Security prices and market transparency, Journal of Financial Intermediation 5, 255-283.

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

Meling, T. G. (2020). Anonymous Trading in Equities. Journal of Finance 76, 707-754.

Neely C. J., Rapach D. E., Tu J., and Zhou G.. 2014. Forecasting the equity risk premium: The role of technical indicators. Management Science 60:1772–91.

Nicholls, C., 2006. The characteristics of Canada's capital markets and the illustrative case of Canada's legislative regulatory response to Sarbanes–Oxley. Research study commissioned by the Task Force to Modernize Securities Legislation in Canada, 129–203.

Pettenuzzo, D., and A. Timmermann. 2011. Predictability of stock returns and asset allocation under structural breaks. Journal of Econometrics 164 (1):60-78.

Pham, T. and Westerholm, J. (2020) A survey of research into broker identity and limit order book transparency, Australasian Accounting Business and Finance Journal 9 (forthcoming)

Polk, C., Thompson, S., & Vuolteenaho, T. (2003). Cross-sectional forecasts of the equity premium. Journal of Financial Economics 81, 101-141,

Pontiff, J. , 2006. Costly arbitrage and the myth of idiosyncratic risk. J. Ac- count. Econ. 42 (1-2), 35–52 .

Rapach D. E., Ringgenberg M. C., and Zhou G.. 2016. Short interest and aggregate stock returns. Journal of Financial Economics 121:46–65.

Rapach D. E., Strauss J. K., and Zhou G.. 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23:821–62.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the United States? *Journal of Finance*, 68(4), 1633–1662.

Rozeff, M. S. (1984). Dividend yields are equity risk premiums. Journal of Portfolio Management, 11(1), 68–75.

Shleifer, A. , Vishny, R.W. , 1997. The limits of arbitrage. Journal of Finance 52 (1), 35–55 .

SIFMA. 2023. Top 10 Takeaways from SIFMA's 2023 Capital Markets Fact Book. The Securities Industry and Financial Markets Association (July 24, 2023), retrieved on May 05, 2024 from https://www.sifma.org/resources/news/top-10-takeaways-from-sifmas-2023-capital-markets-fact-book/

Spiegel, Matthew. 2008. Forecasting the equity premium: Where we stand today. Review of Financial Studies 21 (4):1453-1454.

Stambaugh, R.S., 1999. Predictive regressions. Journal of Financial Economics 54, 315-421.

Theissen, E. (2002). Price discovery in floor and screen trading systems. Journal of Empirical Finance, 9(4), 455–474.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288.

Wang, J. 1994. A Model of Competitive Stock Trading Volume. Journal of Political Economy 102, 127–168.

Welch I. 2014. Referee recommendations. Review of Financial Studies 27:2773–804.

Welch I. 2016. The (time-varying) importance of disaster risk. Financial Analysts Journal 72:14–30.

Welch, I. 2019, Reproducing, extending, updating, replicating, reexamining, and reconciling, Critical Finance Re 8, 301–304.

Wold, H. (1975). Soft modeling by latent variables: The non-linear iterative partial least squares (NIPALS) approach. In *Perspectives in Probability and Statistics*.

Wurgler, J. , Zhuravskaya, E. , 2002. Does arbitrage flatten demand curves for stocks? J. Bus. 75 (4), 583–608 .

Xu, X., & Liu, W-H. 2024. Forecasting the equity premium: can machine learning beat the historical average? Quantitative Finance, 24(10), 1445–1461.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320.

Table 1. Variable Descriptions & Descriptive Statistics

This table provides detailed descriptions of all firm-level, capital market, and macroeconomic variables used in the analysis, along with their respective data sources.

Variable	Description	Source	Mean	Stdev	Skewness
	Firm Level Variables				
AP_TURNOVER	Account Payables Turnover	Jensen et al. (2023)	5.7829	5.4531	2.3241
AT_GR1	Sales Growth 1yr	As above	0.2226	0.6773	3.8665
BE_ME	Book Equity scaled by Market Equity	As above	0.8381	0.8519	3.0092
BETA_60M	60-month rolling market beta	As above	0.9357	0.6877	0.5244
BIDASKHL_21D	The high-low bid-ask spread	As above	0.0103	0.0081	2.1867
CASH_BEV	Cash and Short-Term Investments scaled by BEV	As above	0.397	1.302	6.3396
CASH_GR1A	Gross Profit Change 1yr	As above	-0.0091	0.1397	-0.131
CHCSHO_12M	Change in Shares - 12 Month	As above	0.0932	0.3063	4.3949
CHCSHO_3M	Change in Shares - 3 Month	As above	0.0193	0.0794	4.9619
COSKEW_21D	Decimal Coskewness (coskew_21d)	As above	-0.0275	0.329	-0.0881
DEBT_BEV	Total Debt scaled by BEV	As above	0.2986	0.277	1.4287
DIV_GR1A	Dividend Payout Ratio Change 1yr	As above	-0.0059	0.0319	-0.9584
DIV_ME	Total Dividends scaled by ME	As above	0.0233	0.0452	3.401
EBITDA_SALE	Operating Profit Margin before Depreciation	As above	-2.3364	7.5068	-2.6843
EMP_GR1	employment groth	As above	-0.2351	0.4749	0.0125
EQNETIS_GR1A	Equity Net Issuance Change 1yr	As above	-0.0456	0.2298	-0.8852
EQNPO_1M	Net Equity Payout - 1 Month	As above	-0.0051	0.0334	-5.0036
EQNPO_GR1A	Equity Net Payout Change 1yr (eqnpo_gr1a)	As above	-0.0643	0.2428	-1.1059
FCF_ME	Free Cash Flow scaled by ME	As above	-0.0715	0.3219	-2.0217
INT_DEBT	Interest scaled by Total Debt	As above	0.0938	0.2401	7.1379
INTAN_GR1A	growth in intangible	As above	-0.0109	0.0963	-0.4959
ISKEW_CAPM_21D	Idiosyncratic skewness from the CAPM	As above	0.1666	0.9337	-0.2342
IVOL_CAPM_60M	Idiosyncratic volatility from the CAPM (60 months)	As above	0.1169	0.0744	1.539
LOG_ME	Log market value of equity	As above	5.6899	1.7655	0.3156
NWC_AT	Working Capital scaled by Assets	As above	0.1597	0.2157	0.5244
O_SCORE	Ohlson O-Score	As above	-3.5006	3.3521	0.7711
RESFF3_12_1	Residual Momentum - 12 Month	As above	-0.1144	0.3744	-0.5673
RET_12_1	Price momentum t-12 to t-1 (ret_12_1)	As above	0.1674	0.6798	2.0748
RET_3_1	Price momentum t-3 to t-1 (ret_3_1)	As above	0.022	0.2045	0.6896

RET_36_1	Momentum 1-36 Months (ret_36_1)	As above	0.6502	2.1196	3.7874
ROE_BE_STD	ROE Volatility	As above	0.1204	0.2898	5.6292
SALE_NWC	Sales scaled by Working Capital	As above	23.3995	88.6359	6.7261
SGA_GR1	Cost of Goods Sold Growth 1yr	As above	0.0277	0.5784	2.3793
SRP_TSX	Stock risk premium, the difference between a stock's total return and the risk-free rate, proxied by the one-month return on three-month Government of Canada Treasury bills	CFMRC and authors' calculations.	0.0095	0.1284	0.692
TANGIBILITY	Tangibility	Jensen et al. (2023)	0.6202	0.2711	-0.2212
TURNOVER_126D	Share turnover	As above	0.0022	0.0021	1.7558
Z_SCORE	Altman Z-Score	As above	6.0153	14.7842	5.6455
	Macro Variables				
CAP_UTI_GR	Growth in capacity utilization, which captures the percentage of production capacity in use	CANSIM	-0.0008	0.0148	-1.3861
CLI_GR	The growth in the composite leading indicator, an OECD index designed to anticipate turning points in business cycles by tracking fluctuations in economic activity relative to its long-run trend	www.oecd.org	-0.0001	0.0024	-0.0387
EPU (LOG_EPU)	A news-based index developed by Baker et al. (2016) that reflects uncertainty surrounding government actions that affect the economic environment.	www.policyuncertainty.com	4.8343	0.6186	0.1314
FISHER_PRICE_GR	Change in the Fisher Commodity Index, which reflects price movements in 26 key Canadian commodities sold globally	CANSIM	0.0026	0.0505	-0.5204
INFLATION	The inflation rate, measured as the monthly change in the Consumer Price Index	CANSIM	0.0017	0.0036	-0.0681
M2_GR	The money supply growth, defined as the monthly change in the M2 money supply index	CANSIM	0.0048	0.0039	0.1754
MANU_SALES_GR	The growth in manufacturers' sales	CANSIM	-0.0019	0.0485	-0.6332
RET30_TBILL	The one-month return on three-month Government of Canada Treasury bills	CFMRC	0.0025	0.0021	1.3663
SPTSX_RP	The excess return on the S&P/TSX Composite Index over a risk-free asset, measured as the difference between the index's total return and the yield on short-term Government of Canada Treasury bills.	CFMRC	0.0036	0.0424	-0.9772
SPTSXVOL30	The 30-day volatility of the S&P/TSX 300 index (SPTSXVOL30), computed as the annualized standard deviation of daily log price changes (Bloomberg, item RK002); TSX_volatility30day_Bloomberg	Bloomberg	13.62	7.8884	2.3945
TERM_SPREAD	The difference between long-term and short-term government bond yields.	CFMRC	0.0026	0.0244	-0.0134
TSX_NET_ISSUE	Builds on Welch and Goyal 2008 and captures the extent of aggregate corporate equity financing activity in the market. We compute firm-level	CFMRC	18.5519	16.5727	0.8397

	net issuance as the change in market capitalization not explained by stock returns. The market-level measure (TSX_EQUITY_ISSUE) is defined as the ratio of the 12-month moving sum of firm-level net equity issuances to the aggregate market capitalization in the current month.				
TSX_VOLUME_GR	The monthly growth rate in the total dollar trading volume on the Toronto Stock Exchange (TSX), capturing fluctuations in market trading activity.		0.0229	0.177	0.3802
UNEMPLOYMENT	The unemployment rate	CANSIM	7.5436	1.4571	0.876
X_RATE	The Canada–U.S. exchange rate, measured as the number of Canadian dollars per U.S. dollar		1.2572	0.1711	0.002
	Calendar Anomalies				
DECEMBER_EFFECT	A December dummy, set to 1 in December and 0 otherwise.	CFMRC			
JANUARY_EFFECT	A January dummy, set to 1 in January and 0 otherwise, which accounts for the well-documented January Effect	CFMRC			

Table 2: In-Sample Coefficient Estimates with Two-Way Clustered Standard Errors

Panel A of this table reports selected in-sample coefficients for OLS and Elastic-Net regressions with two-way clustered SEs. Panel B reports two-sided Diebold–Mariano tests of squared-error loss differentials between each model and the OLS benchmark, with Newey–West (11-lag) standard errors to account for serial correlation. Negative HAC-adjusted t-statistics indicate that the challenger's forecast errors are significantly lower than OLS's (i.e. it "beats" the benchmark), while a positive t-stat would imply the opposite.

Panel A: In-Sample Coefficient Estimates with Two-Way Clustered Standard Errors

	Predictor	Estimate	Std. Error	t-stat	p-value
OLS					
	(Intercept)	0.0275496	0.00439455	6.269046	3.654E-10
	LOG_ME	-0.0041938	0.00078598	-5.335744	9.545E-08
	BE_ME	0.0008192	0.00127544	0.6422875	0.5206886
	RET_12_1	0.0050575	0.00185447	2.7271952	0.0063891
Elastic-Net					
	(Intercept)	0.0275496	0.00439455	6.269046	3.654E-10
	LOG_ME	-0.0041938	0.00078598	-5.335744	9.545E-08
	BE_ME	0.0008192	0.00127544	0.6422875	0.5206886
	RET_12_1	0.0050575	0.00185447	2.7271952	0.0063891

Panel B:

	LASSO	ENET	PLS	RF	XGB	NN1	NN2	NN3	NN4
DM_t_HAC	-21.1144	-21.5397	1.311228	-14.1739	-6.64468	-12.106	-12.89	-13.305	-12.592
p_val_HAC	5.90E-99	6.60E-103	0.189781	1.33E-45	3.04E-11	9.81E-34	5.14E-38	2.16E-40	2.34E-36

Table 3. Out-of-Sample Forecast Performance across Predictive Models

This table reports in column 1 the root mean squared error (RMSE) and out-of-sample R^2_{OOS} in column 2 for a set of machine learning and linear models.

Model	RMSE	R^2_{OOS}
enet	0.12134	0.061999
lasso	0.12135	0.06184
NN4	0.121214	0.060363
NN3	0.121533	0.059951
NN2	0.121889	0.052957
rf	0.122031	0.052517
NN1	0.122079	0.050731
pls	0.12251	0.043934
ols	0.122553	0.043295
xgb	0.122975	0.036963
OLS3	0.194	0.164

Figure 1. Comparative Out-of-Sample Performance of Predictive Models

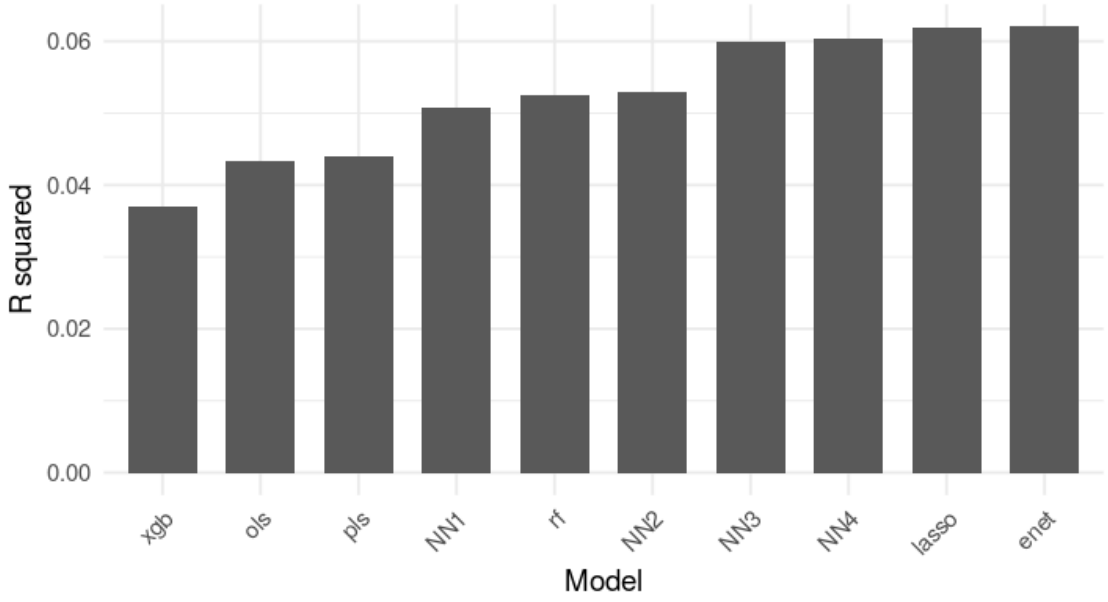


Table 4. Cross-Sectional Variation in Out-of-Sample Predictive Performance (R^2_{OOS})

This table evaluates the out-of-sample predictive performance (R^2_{OOS}) of various return forecasting models across firm size (small vs. large Panel A) and investment style (growth vs. value stocks, Panel B).

Panel A: Size Effect										
	enet	lasso	NN3	NN4	NN2	rf	xgb	NN1	pls	ols
Small	0.127782	0.127677	0.125202	0.120013	0.119163	0.115684	0.115096	0.114191	0.11416	0.11358
Large	0.047999	0.047823	0.043038	0.052344	0.037575	0.037539	0.014481	0.036972	0.02837	0.02772
gap_pp	7.978377	7.985429	8.216407	6.766906	8.158831	7.814406	10.06153	7.721933	8.57851	8.58643
Panel B: Investment Style										
Model	enet	lasso	NN3	NN4	NN2	rf	xgb	NN1	pls	ols
Value	0.257661	0.249942	0.249807	0.248833	0.246253	0.239226	0.233242	0.232855	0.231058	0.224672
Growth	-0.10303	-0.09292	-0.09306	-0.09837	-0.10999	-0.10158	-0.10938	-0.11022	-0.10189	-0.11511
gap_pp	36.06913	34.28623	34.28713	34.7199	35.62396	34.08023	34.26256	34.30732	33.29439	33.97873

Table 5. Predictive Performance at the Annual Horizon

This table reports out-of-sample root mean squared error (RMSE) and R^2_{oos} for one-year-ahead return forecasts across ten models, including neural networks (NN1–NN4), tree-based learners (RF, XGB), and linear benchmarks (OLS, PLS, LASSO, Elastic Net). Panel A reports results for the full sample. Panel B presents performance conditional on firm size (small vs. large market capitalization). Panel C evaluates models by investment style (value vs. growth).

Panel A: Full Sample										
	enet	lasso	NN3	NN4	NN2	rf	xgb	NN1	pls	ols
RMSE	0.439613	0.44075	0.442449	0.444421	0.445967	0.454591	0.46221	0.46232	0.46989	0.46999
R2	0.062004	0.056166	0.048169	0.042936	0.031854	-0.00204	-0.0296	-0.0301	-0.0634	-0.0639
Panel B: Size Effect										
	enet	lasso	NN3	NN4	NN2	rf	xgb	NN1	pls	ols
Small	0.091897	0.075314	0.06754	0.061246	0.059302	0.053178	0.025241	0.024778	0.000251	-0.00011
Large	0.051847	0.054833	0.048516	0.0434	0.028864	-0.02544	-0.05603	-0.05646	-0.09464	-0.09521
gap_pp	4.005031	2.048129	1.902406	1.784554	3.043871	7.861722	8.127026	8.123557	9.489304	9.509247
Panel C: Investment Style										
Model	enet	lasso	NN3	NN4	NN2	rf	xgb	NN1	pls	ols
Value	0.218078	0.211501	0.206039	0.190845	0.177026	0.153965	0.09108	0.09021	0.05235	0.05105
Growth	-0.09464	-0.10082	-0.07071	-0.09191	-0.09142	-0.12239	-0.1143	-0.1142	-0.1364	-0.1355
gap_pp	31.27184	31.23224	27.6754	28.27509	26.8446	27.63519	20.5391	20.4425	18.8738	18.6551

Figure 2. Heatmap: Macroeconomic Variable Importance

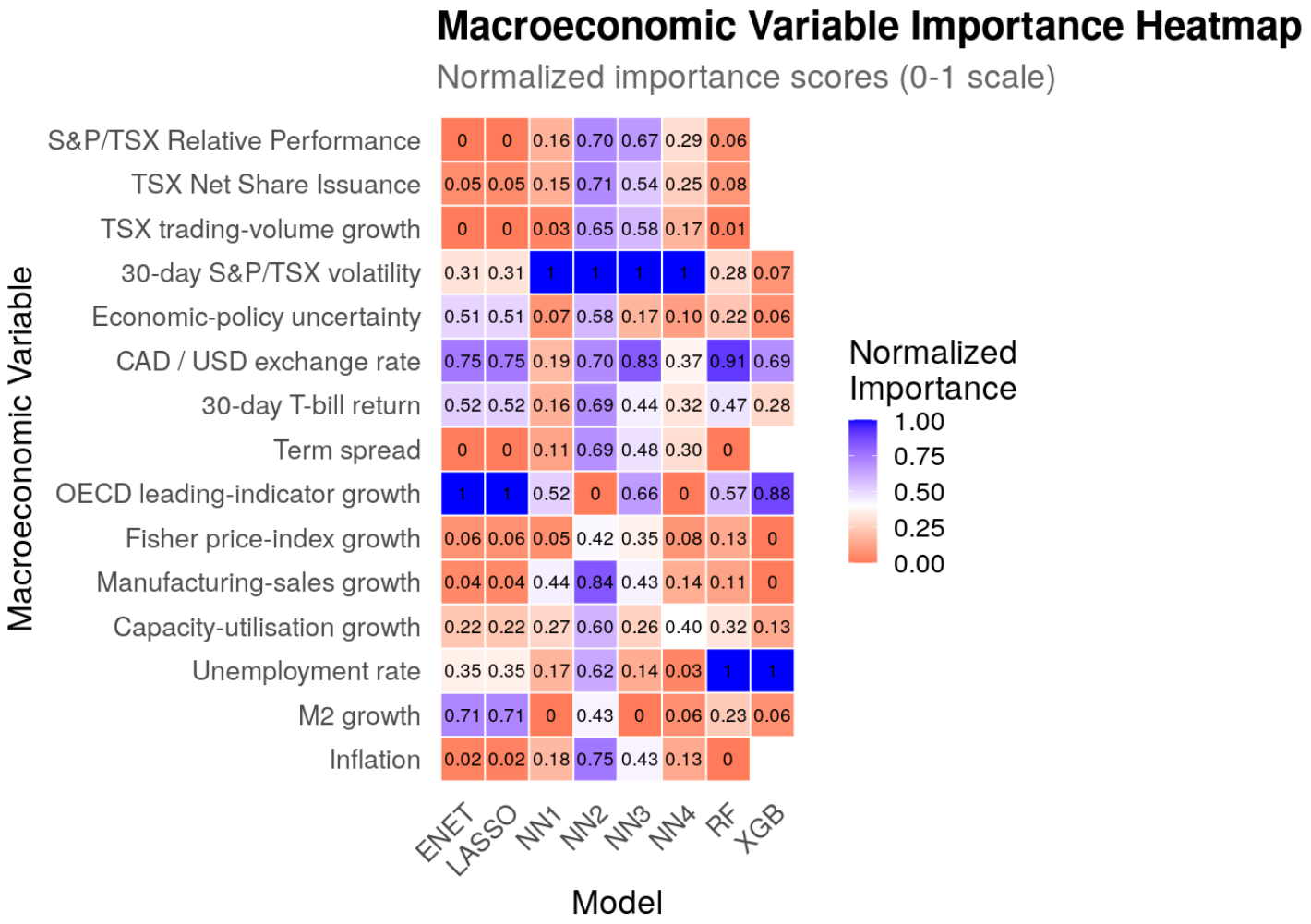


Table 6. Top 15 Firm-Level Predictors of One-Year Excess Returns across Models

This table reports the 15 most influential firm characteristics for each forecasting model, based on normalized importance scores averaged across rolling windows.

rank	ENET	LASSO	NN1	NN2	NN3	NN4	XGB	rf
1	ivol_capm_60m	ivol_capm_60m	Sale_nwc	CHCSHO_3M	turnover_126d	Sale_nwc	bidaskhl_21d	bidaskhl_21d
2	bidaskhl_21d	bidaskhl_21d	turnover_126d	O_score	Sale_nwc	turnover_126d	ivol_capm_60m	ivol_capm_60m
3	Sale_nwc	Sale_nwc	EQNPO_GR1A	turnover_126d	CHCSHO_3M	CHCSHO_3M	RET_36_1	RET_36_1
4	RET_12_1	RET_12_1	Nwc_at	Sale_nwc	RET_3_1	RET_3_1	Intan_gr1a	ebitda_sale
5	Intan_gr1a	Intan_gr1a	CHCSHO_3M	Nwc_at	O_score	RET_36_1	Roe_be_std	RET_12_1
6	Ap_turnover	Ap_turnover	div_gr1a	EQNPO_GR1A	EQNPO_GR1A	O_score	Fcf_me	Roe_be_std
7	Debt_bev	Debt_bev	RET_3_1	ebitda_sale	Nwc_at	Nwc_at	Div_me	Fcf_me
8	turnover_126d	turnover_126d	O_score	Int_debt	Z_score	div_gr1a	AT_GR1	beta_60m
9	Z_score	Z_score	AT_GR1	Cash_gr1a	resff3_12_1	EQNPO_GR1A	Sga_gr1	O_score
10	EQNPO_GR1A	EQNPO_GR1A	Eqnetis_gr1a	Sga_gr1	ebitda_sale	Z_score	tangibility	Intan_gr1a
11	tangibility	tangibility	Roe_be_std	Z_score	AT_GR1	Emp_gr1	RET_12_1	Nwc_at
12	RET_3_1	RET_3_1	Int_debt	RET_3_1	RET_36_1	beta_60m	Emp_gr1	AT_GR1
13	Div_me	Div_me	resff3_12_1	Emp_gr1	Eqnetis_gr1a	AT_GR1	Z_score	turnover_126d
14	RET_36_1	RET_36_1	ebitda_sale	TSX_NET_ISSUE	div_gr1a	Int_debt	beta_60m	Z_score
15	div_gr1a	div_gr1a	Z_score	SPTSX_RP	SPTSX_RP	SPTSX_RP	Sale_nwc	Emp_gr1

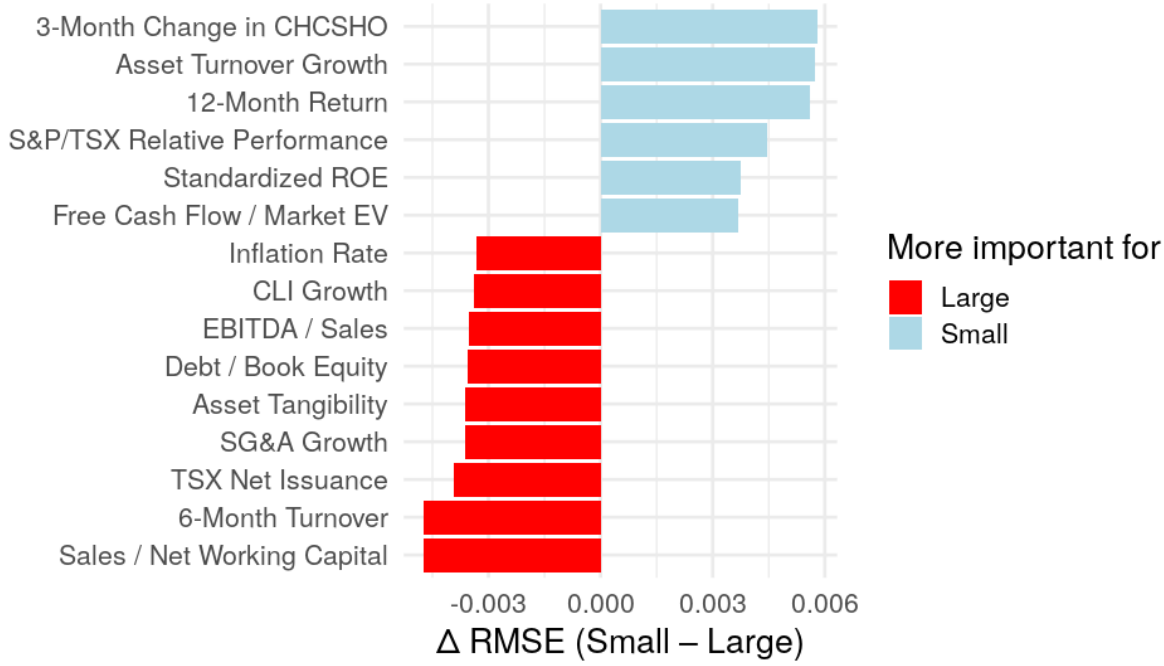
Table 7: Unconditional and Conditional Superior Predictive Ability (SPA) Test Results

This table summarizes model comparisons based on the unconditional SPA (USPA) test of Hansen (2005) and the conditional SPA (CSPA) test of Li et al. (2020), all evaluated using squared-error loss over the full sample. The first row summarizes USPA test results, indicating how often each model is significantly outperformed by an alternative across 55 one-versus-one comparisons at the 5% level. The next 13 rows report CSPA results, where predictive losses are evaluated within macroeconomic regimes. Each cell shows how many times the corresponding model is rejected in favor of a competing model (out of 10) under a given regime. The final row aggregates total CSPA rejections across all 13 regimes (maximum possible: 260). Lower values indicate greater statistical robustness.

METRIC	NN1	NN2	NN3	NN4	ENET	LASSO	OLS	OLS3	PLS	RF	XGB
uspa_rejections	3	4	4	6	5	5	6	6	7	2	8
regime_log_EPU	7	2	3	4	10	10	8	8	8	0	18
regime_CAP_UTI_GR	2	1	2	4	10	10	9	9	9	5	18
regime_CLI_GR	4	2	1	4	9	9	8	8	9	1	18
regime_FISHER_PRICE_GR	4	2	0	4	7	7	7	7	5	1	16
regime_INFLATION	4	4	4	9	11	11	12	12	12	3	20
regime_M2_GR	3	2	2	5	8	8	9	9	7	3	15
regime_MANU_SALES_GR	6	3	5	9	11	11	12	12	9	6	21
regime_RET30_TBILL	6	3	2	5	9	9	9	9	9	5	18
regime_SPTSXVOL30	4	1	1	3	9	9	8	8	9	1	19
regime_TERM_SPREAD	4	4	4	9	10	10	12	12	11	2	18
regime_TSX_VOLUME_GR	5	1	2	5	9	9	9	9	10	4	15
regime_UNEMPLOYMENT	5	3	2	5	8	8	9	9	10	6	19
regime_X_RATE	3	3	0	4	9	9	8	11	6	2	16
total_cspa_rejections	57	31	28	70	120	120	120	123	114	39	231

Figure 3: Size-Based Decomposition of Predictive Signals in the XGBoost Model

Panel A: Small – Large permutation impact (top 15 variables)



Panel B: Small – Large permutation impact by economic theme

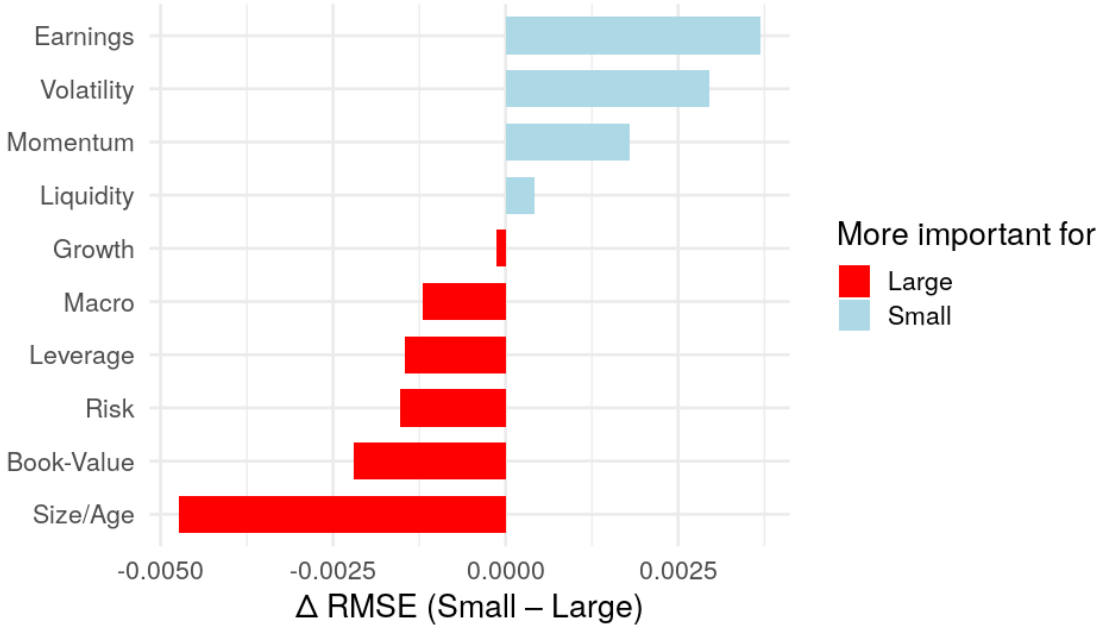


Table 8. Portfolio Performance Based on Predicted Monthly Returns (Jan 2012–Dec 2023)

This table reports backtested performance of monthly rebalanced portfolios sorted into deciles based on one-month-ahead predicted total returns. For each of ten forecasting models, we report results for: (i) a value-weighted long–short portfolio (top–bottom decile), (ii) an equal-weighted long–short portfolio, and (iii) a long-only portfolio of the top decile. Performance metrics include average monthly return (%), monthly volatility (%), annualized Sharpe ratio, skewness, Newey-West t-statistic, maximum drawdown (%), and worst monthly return. The sample covers January 2012 to December 2023.

Model		**ols**			**lasso**	
Strategy	Long-Short	Long-Short (EW)	Long-Only	Long-Short	Long-Short (EW)	Long-Only
Avg	0.435599	0.549422	1.18975	0.739377	0.8226	1.530304
Std	5.064644	4.794009	7.280027	5.187392	4.955132	7.15459
Sharpe	0.29794	0.397007	0.566126	0.49375	0.575074	0.740941
Skew	0.337616	0.351651	0.416236	0.433592	0.531367	0.644688
NW_tstat	1.028528	1.370521	1.954346	1.704494	1.985236	2.55783

Model		**enet**			**rf**	
Strategy	Long-Short	Long-Short (EW)	Long-Only	Long-Short	Long-Short (EW)	Long-Only
Avg	0.741863	0.827843	1.531891	0.81449	0.943474	1.456139
Std	5.198507	4.969048	7.157872	4.417631	4.424014	6.762365
Sharpe	0.494351	0.577119	0.741369	0.638685	0.738761	0.745925
Skew	0.465333	0.563263	0.648735	0.884759	1.140327	0.252205
NW_tstat	1.706569	1.992295	2.559309	2.204829	2.550306	2.575035

Model		**xgb**			**pls**	
Strategy	Long-Short	Long-Short (EW)	Long-Only	Long-Short	Long-Short (EW)	Long-Only
Avg	1.053095	1.189423	1.606937	0.46843	0.612037	1.158568
Std	4.878962	4.52226	7.042709	5.010269	4.776396	7.252679
Sharpe	0.747706	0.911111	0.790405	0.323873	0.443882	0.553368
Skew	0.159896	0.19318	0.435141	0.378309	0.401366	0.458976
NW_tstat	2.581183	3.145283	2.728587	1.118053	1.532343	1.910301

Model		NN1			NN2	
Strategy	Long-Short	Long-Short (EW)	Long-Only	Long-Short	Long-Short (EW)	Long-Only
Avg	0.92425	0.98282	1.494611	0.280399	0.217911	1.225918
Std	4.629584	4.394408	7.160244	4.53958	4.397355	6.762025

Sharpe	0.691573	0.774755	0.723088	0.213969	0.171664	0.628023
Skew	0.46672	0.183805	0.622726	0.098789	-0.28918	0.691765
NW_tstat	2.387405	2.67456	2.496199	0.738652	0.592607	2.168021

Model	**NN3**			**NN4**		
Strategy	Long-Short	Long-Short (EW)	Long-Only	Long-Short	Long-Short (EW)	Long-Only
Avg	0.845313	0.835909	1.363991	0.313619	0.339524	0.968938
Std	4.971174	4.740617	7.678023	4.719472	4.763638	7.219239
Sharpe	0.589046	0.610822	0.615393	0.230197	0.246901	0.464938
Skew	0.447628	0.32326	0.854872	0.478444	0.317664	0.509699
NW_tstat	2.033468	2.108643	2.124422	0.794671	0.852336	1.605031

Figure 4. Sharpe Ratios of Long–Short Portfolios across Forecasting Models

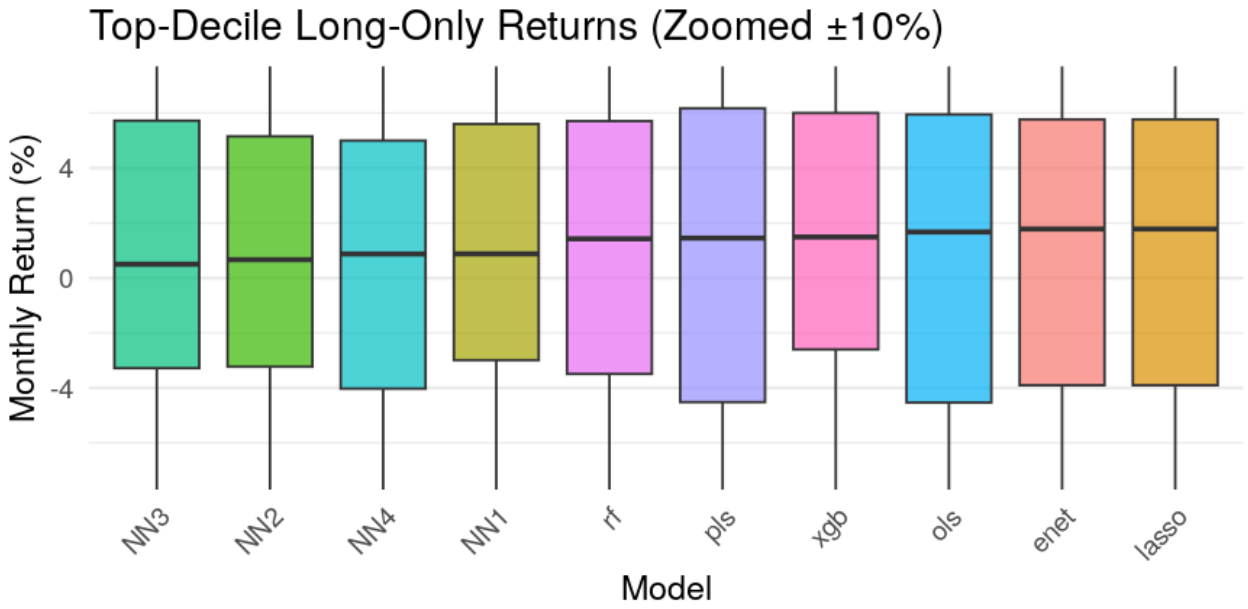


Table 9. Annual Portfolio Performance Based on One-Year-Ahead Return Forecasts

This table reports the results of an annual backtest evaluating the predictive power of each model's one-year-ahead return forecasts. Portfolios are rebalanced every January from 2012 to 2023 based on predicted returns and held for 12 months. For each model and strategy—value-weighted long–short, equal-weighted long–short, and long-only top decile—we report the average annual return (Avg, %), annualized volatility (Std, %), Sharpe ratio, and Newey–West t-statistic (using 11-month lags). The results benchmark model efficacy across horizon-relevant investment implementations.

Model	enet			lasso		
Strategy	Long-Short (VW)	Long-Short (EW)	Long-Only	Long-Short (VW)	Long-Short (EW)	Long-Only
Avg	18.72149	22.13005	26.85594	18.92345	22.14987	26.98488
Std	27.78043	26.59831	37.32658	28.12048	26.57046	37.72341
Sharpe	0.568246	0.734038	0.590676	0.566697	0.735819	0.585781
Skew	25.82541	17.84663	20.77017	25.25983	17.05399	20.77017
NW_tstat	4.604931	5.309338	4.912317	4.587324	5.359944	4.876368

Model	**nn1**			**nn2**		
Strategy	Long-Short (VW)	Long-Short (EW)	Long-Only	Long-Short (VW)	Long-Short (EW)	Long-Only
Avg	13.20897	15.7019	21.5695	17.0512	18.35683	22.71778
Std	21.6907	20.61262	35.16769	24.80303	23.37541	35.81436
Sharpe	0.524469	0.684048	0.483626	0.590763	0.69544	0.508119
Skew	19.97142	14.95767	27.97314	23.30898	18.91023	28.45517
NW_tstat	4.496086	5.455502	4.786551	5.442424	6.278412	4.486378

Model	**nn3**			**nn4**		
Strategy	Long-Short (VW)	Long-Short (EW)	Long-Only	Long-Short (VW)	Long-Short (EW)	Long-Only
Avg	20.32535	22.11019	26.04654	18.00878	19.91413	24.94937
Std	21.07325	20.87246	32.9453	29.81789	28.20777	41.14497
Sharpe	0.891371	0.986988	0.683703	0.489901	0.600712	0.458748
Skew	4.908218	4.133957	14.86534	27.57105	21.88901	36.86851
NW_tstat	5.024181	5.457084	4.114493	2.839982	3.267567	3.102701

Model	**ols**			**pls**		
Strategy	Long-Short (VW)	Long-Short (EW)	Long-Only	Long-Short (VW)	Long-Short (EW)	Long-Only

Avg	16.582	19.7709	24.63925	16.28321	19.44866	24.31278
Std	25.0708	24.03844	35.92022	24.58591	23.54643	35.0547
Sharpe	0.566031	0.734417	0.562564	0.568849	0.739158	0.572959
Skew	22.19451	15.66012	21.03396	20.32234	14.82506	19.86695
NW_tstat	4.369733	4.966779	4.442999	4.451163	5.169279	4.607469

Model	**rf**			**xgb**		
Strategy	Long-Short (VW)	Long-Short (EW)	Long-Only	Long-Short (VW)	Long-Short (EW)	Long-Only
Avg	13.9633	17.70423	23.09565	15.87037	17.12981	24.26346
Std	24.07486	21.87753	34.05067	22.80627	22.21803	34.65966
Sharpe	0.495657	0.734301	0.5703	0.611969	0.689671	0.583421
Skew	10.97327	3.598918	12.51025	14.72188	14.92981	25.24476
NW_tstat	3.124535	4.008527	4.033743	6.310758	6.293569	5.69015

Table 10. Annual Portfolio Performance by Investment Style

This table reports the performance of annual long-only portfolios constructed separately for growth (Panel A) and value (Panel B) stocks, based on each model's one-year-ahead return forecasts. At the start of each year from 2012 to 2023, stocks are ranked by forecasted return and sorted into style terciles based on book-to-market ratios. We focus on the top decile within the growth (bottom 30%) and value (top 30%) universes. Each portfolio is held for 12 months. Reported metrics include average annual return (Avg, %), annualized standard deviation (Std, %), Sharpe ratio, skewness, and the Newey–West t-statistic (11-month lag). All returns are value-weighted.

Panel A: Growth Stocks

Model	Strategy	Avg	Std	Sharpe	Skew	NW_tstat
enet	Long-Only	0.970073	7.413553	0.453282	0.66875	1.564793
lasso	Long-Only	0.986977	7.403155	0.461828	0.66385	1.594296
NN1	Long-Only	0.810595	7.567602	0.371053	0.103873	1.280928
NN2	Long-Only	0.598256	6.965887	0.29751	0.408386	1.027046
NN3	Long-Only	0.856025	6.860097	0.432262	0.440278	1.492226
NN4	Long-Only	0.098722	6.777942	0.050455	0.163699	0.174178
ols	Long-Only	0.976871	7.361076	0.459713	0.333759	1.586991
pls	Long-Only	0.857633	7.555786	0.393199	0.360549	1.357378
rf	Long-Only	0.750367	7.684258	0.338269	0.022199	1.167752
xgb	Long-Only	1.220287	7.440906	0.568103	-0.27105	1.961

Panel B: Value stocks

Model	Strategy	Avg	Std	Sharpe	Skew	NW_tstat
enet	Long-Only	2.297327	9.313985	0.854433	0.42563	2.949621
lasso	Long-Only	2.29662	9.311966	0.854355	0.426985	2.949352
NN1	Long-Only	2.985229	9.064037	1.140897	0.653598	3.938535
NN2	Long-Only	1.966438	9.388487	0.725563	0.667091	2.504745
NN3	Long-Only	1.969547	9.770599	0.69829	0.604966	2.410593
NN4	Long-Only	1.68676	9.564984	0.610885	0.680892	2.10886
ols	Long-Only	2.372261	9.341232	0.879729	0.415592	3.036947
pls	Long-Only	2.468023	9.408048	0.908741	0.447534	3.137101
rf	Long-Only	2.646394	9.269531	0.988979	0.32768	3.414094
xgb	Long-Only	2.345859	9.364822	0.867747	0.382923	2.995582

Table 11. Annual Portfolio Performance by information aysmmtry quality

This table reports the performance of annual long-only top-decile portfolios formed each January from 2012 to 2023 using one-year-ahead return forecasts from various models. We evaluate conditional predictability by sorting the stock universe into two subsamples based on: Information asymmetry (Panel A), proxied by the 21-day bid–ask spread (BIDASKHL_21D). Stocks are classified each month into High and Low Asymmetry groups based on the cross-sectional median. Trading volume (Panel B), proxied by the 21-day lagged trading volume. Stocks are similarly split into High and Low Volume groups. Within each group, we form value-weighted long-only portfolios of the top forecast decile and report average monthly return (Avg, %), annualized volatility (Std, %), Sharpe ratio, skewness, and the Newey–West t-statistic (11-month lag). The results highlight how market frictions and liquidity conditions modulate model effectiveness.

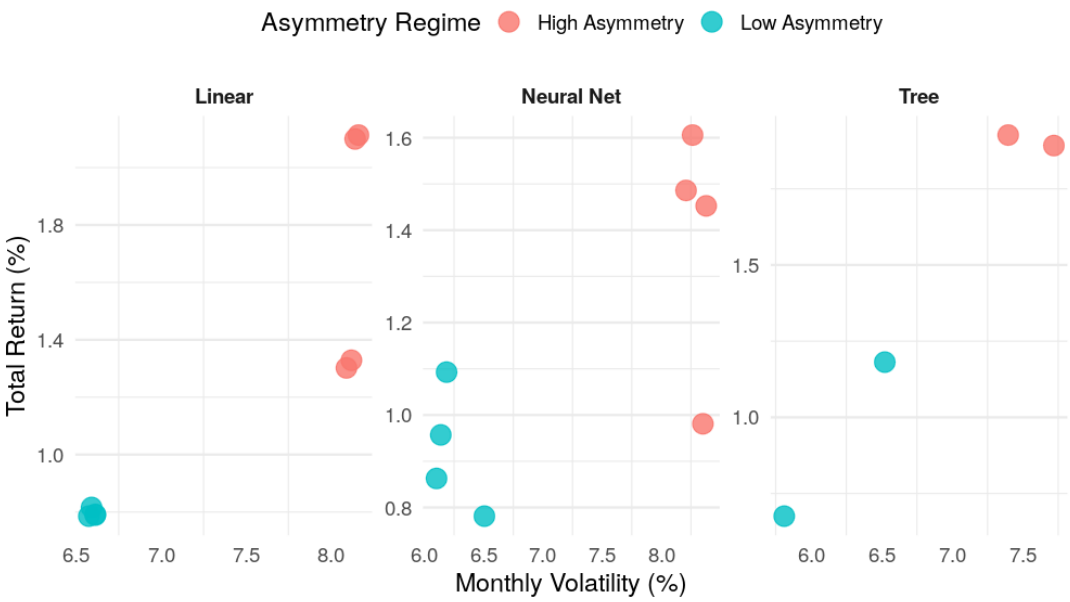
Panel A: Information Asymmrty

Model	Strategy	Avg	Std	Sharpe	Skew	NW_tstat	Group
enet	Long-Only	0.815757	6.58982	0.428823	0.173923	1.480355	Low Asymmetry
enet	Long-Only	2.113352	8.163209	0.896812	0.486136	3.095919	High Asymmetry
lasso	Long-Only	0.791649	6.612389	0.414729	0.173321	1.431703	Low Asymmetry
lasso	Long-Only	2.099719	8.143631	0.893169	0.493747	3.083343	High Asymmetry
NN1	Long-Only	1.092829	6.186741	0.6119	0.17542	2.112365	Low Asymmetry
NN1	Long-Only	1.605729	8.263441	0.673135	0.631076	2.323754	High Asymmetry
NN2	Long-Only	0.862787	6.101706	0.489828	0.883847	1.690952	Low Asymmetry
NN2	Long-Only	1.485923	8.208349	0.627092	0.640711	2.164807	High Asymmetry
NN3	Long-Only	0.781206	6.505993	0.415951	-0.39022	1.435921	Low Asymmetry
NN3	Long-Only	1.452501	8.378382	0.600547	0.622252	2.073171	High Asymmetry
NN4	Long-Only	0.956989	6.137755	0.540117	0.107445	1.864558	Low Asymmetry
NN4	Long-Only	0.981008	8.350729	0.406948	0.49657	1.404841	High Asymmetry
ols	Long-Only	0.789063	6.611921	0.413404	0.196755	1.427128	Low Asymmetry
ols	Long-Only	1.301924	8.09228	0.557321	0.625493	1.923949	High Asymmetry
pls	Long-Only	0.785495	6.573599	0.413934	0.149642	1.428957	Low Asymmetry
pls	Long-Only	1.328633	8.121734	0.566692	0.522632	1.956298	High Asymmetry
rf	Long-Only	0.675737	5.810652	0.40285	-0.56894	1.390693	Low Asymmetry
rf	Long-Only	1.891752	7.719472	0.848921	0.349364	2.930593	High Asymmetry
xgb	Long-Only	1.181142	6.522524	0.627303	-0.46783	2.157951	Low Asymmetry
xgb	Long-Only	1.926235	7.395874	0.902215	0.33797	3.114573	High Asymmetry

Panel B: Trading Volume							
Model	**Strategy**	**Avg**	**Std**	**Sharpe**	**Skew**	**NW_tstat**	**volume_group**
enet	Long-Only	2.187959	13.21769	0.573422	0.988521	1.90907	High Volume
enet	Long-Only	1.574268	7.53666	0.723587	0.61136	2.497921	Low Volume
lasso	Long-Only	2.202908	13.21251	0.577566	0.986931	1.922867	High Volume
lasso	Long-Only	1.564671	7.540182	0.718839	0.61395	2.481533	Low Volume
NN1	Long-Only	0.86284	11.2814	0.264946	1.061254	0.882075	High Volume
NN1	Long-Only	1.532862	7.149638	0.742693	0.729751	2.56388	Low Volume
NN2	Long-Only	-0.07465	12.78333	-0.02023	1.357902	-0.06735	High Volume
NN2	Long-Only	1.491467	6.882727	0.750661	0.67858	2.591385	Low Volume
NN3	Long-Only	1.055365	13.07236	0.279666	0.910584	0.93108	High Volume
NN3	Long-Only	1.207392	7.740494	0.540344	0.944563	1.865342	Low Volume
NN4	Long-Only	1.302944	12.85822	0.351023	1.168405	1.168646	High Volume
NN4	Long-Only	1.190717	7.503728	0.549695	0.59782	1.897624	Low Volume
ols	Long-Only	1.939541	12.46605	0.538965	0.933155	1.794355	High Volume
ols	Long-Only	1.266292	7.496179	0.585174	0.380762	2.0201	Low Volume
pls	Long-Only	1.805671	12.47605	0.501363	0.954844	1.669167	High Volume
pls	Long-Only	1.387541	7.553093	0.636373	0.415264	2.196847	Low Volume
rf	Long-Only	1.413671	12.98841	0.377036	0.849691	1.255251	High Volume
rf	Long-Only	1.489649	7.032008	0.73383	0.330173	2.533281	Low Volume
xgb	Long-Only	2.398334	12.03343	0.690416	1.22845	2.298575	High Volume
xgb	Long-Only	1.74607	7.227946	0.83683	0.499251	2.888854	Low Volume

Figure 5. Annual Portfolio Performance by information aysmmtry quality

Panel A. Bid Ask Spread



Panel B. Trading Volume

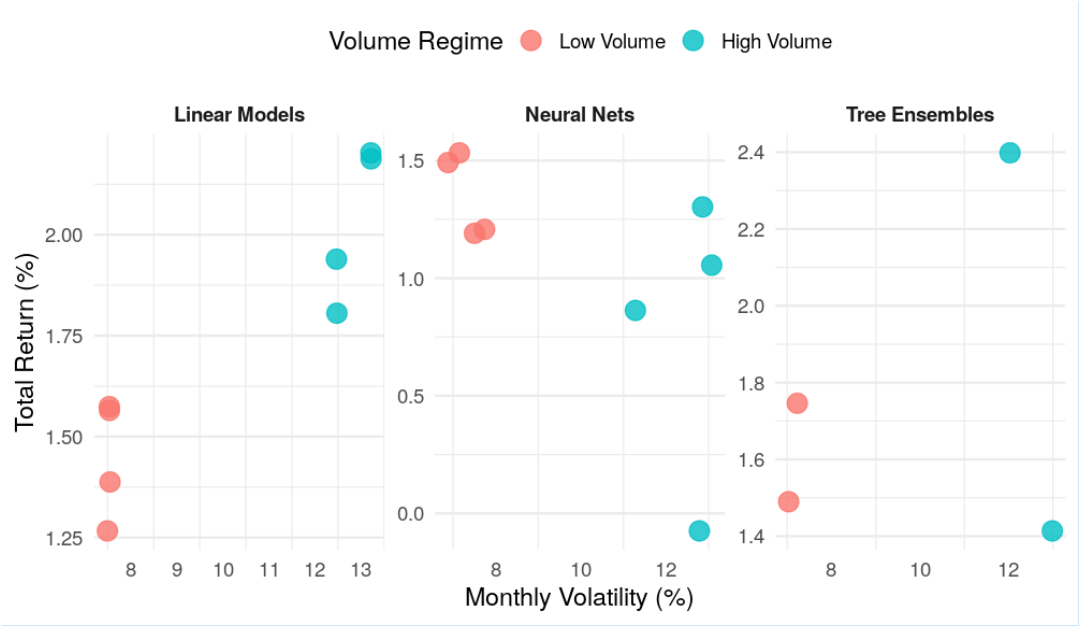


Table 12. Return Predictability Conditional on Anonymous Trading Activity

This table reports the performance of long-only top-decile portfolios sorted on predicted returns, conditional on two dimensions of anonymous trading: Panel A splits the sample based on the level of *net anonymous buying* (NET_ANO_BUY_VOLUM), and Panel B based on the dispersion in broker-level anonymous order flow (VOLATILITY_BROKER_VOLUM). For each month, stocks are ranked above and below the median of the relevant proxy, and portfolios are constructed from the top decile within each subset. Reported statistics include average monthly return (Avg), annualized volatility (Std), Sharpe ratio, return skewness, and Newey–West adjusted t-statistics. The results show that elevated anonymous buying and greater broker-level dispersion significantly enhance return predictability, particularly for non-linear and interaction-rich models.

Panel A: Net Anonymous Buying

Model	Strategy	Avg	Std	Sharpe	Skew	NW_tstat	model	group
enet	Long-Only	2.133751	8.40025	0.879918	0.406331	2.840015	enet	High net ANO buy volume
enet	Long-Only	1.474761	7.603962	0.67185	0.831419	2.31932	enet	Low net ANO buy volume
lasso	Long-Only	2.109819	8.402037	0.869864	0.416092	2.807565	lasso	High net ANO buy volume
lasso	Long-Only	1.467827	7.595477	0.669438	0.826256	2.310993	lasso	Low net ANO buy volume
NN1	Long-Only	1.845217	7.892401	0.809895	0.604504	2.614011	NN1	High net ANO buy volume
NN1	Long-Only	1.467306	7.673394	0.662405	0.73969	2.286715	NN1	Low net ANO buy volume
NN2	Long-Only	1.776104	8.451384	0.728	0.876097	2.349685	NN2	High net ANO buy volume
NN2	Long-Only	1.06037	7.105806	0.516933	0.670663	1.784525	NN2	Low net ANO buy volume
NN3	Long-Only	1.942621	8.753044	0.768811	0.82283	2.481407	NN3	High net ANO buy volume
NN3	Long-Only	1.303928	7.966342	0.567003	0.862469	1.957372	NN3	Low net ANO buy volume
NN4	Long-Only	1.423307	8.475588	0.581727	0.505376	1.877578	NN4	High net ANO buy volume
NN4	Long-Only	0.965735	7.610519	0.439576	0.71321	1.517479	NN4	Low net ANO buy volume
ols	Long-Only	1.649505	8.637099	0.661571	0.321886	2.135281	ols	High net ANO buy volume
ols	Long-Only	1.203475	7.746783	0.538154	0.768416	1.857781	ols	Low net ANO buy volume
pls	Long-Only	1.583398	8.560724	0.640723	0.434756	2.067991	pls	High net ANO buy volume
pls	Long-Only	1.05974	7.765875	0.472715	0.817696	1.631878	pls	Low net ANO buy volume
rf	Long-Only	1.914885	7.59455	0.873436	0.735433	2.819095	rf	High net ANO buy volume
rf	Long-Only	1.801671	7.558078	0.825762	0.475767	2.850643	rf	Low net ANO buy volume
xgb	Long-Only	2.021443	8.189639	0.855042	0.57448	2.759726	xgb	High net ANO buy volume
xgb	Long-Only	1.706651	7.421448	0.796612	0.418186	2.750014	xgb	Low net ANO buy volume

Panel B: The Dispersion in Broker-Level Anonymous Order Flow

Model	Strategy	Avg	Std	Sharpe	Skew	NW_tstat	model	group

enet	Long-Only	1.749672	7.598932	0.797617	0.513023	2.753485	enet	Low Heterogeneity
enet	Long-Only	1.429367	8.260424	0.599421	0.770145	1.934686	enet	High Heterogeneity
lasso	Long-Only	1.756061	7.590009	0.801471	0.513441	2.766789	lasso	Low Heterogeneity
lasso	Long-Only	1.439218	8.265999	0.603145	0.766164	1.946706	lasso	High Heterogeneity
NN1	Long-Only	1.605133	7.369122	0.754546	0.458277	2.604799	NN1	Low Heterogeneity
NN1	Long-Only	1.660819	8.124862	0.708104	0.677582	2.28547	NN1	High Heterogeneity
NN2	Long-Only	1.210014	7.229609	0.579784	0.712727	2.001494	NN2	Low Heterogeneity
NN2	Long-Only	1.475537	8.272993	0.617843	0.944776	1.994144	NN2	High Heterogeneity
NN3	Long-Only	1.476274	7.767243	0.658401	0.7725	2.272893	NN3	Low Heterogeneity
NN3	Long-Only	1.866488	8.896007	0.72681	0.857149	2.345844	NN3	High Heterogeneity
NN4	Long-Only	1.244981	7.445965	0.579205	0.473816	1.999496	NN4	Low Heterogeneity
NN4	Long-Only	1.366564	8.604659	0.550157	0.638208	1.775683	NN4	High Heterogeneity
ols	Long-Only	1.499162	7.76266	0.669004	0.272335	2.309494	ols	Low Heterogeneity
ols	Long-Only	1.598408	8.765681	0.631674	0.702253	2.038784	ols	High Heterogeneity
pls	Long-Only	1.362165	7.62103	0.619166	0.327716	2.137445	pls	Low Heterogeneity
pls	Long-Only	1.511662	8.63166	0.606668	0.783327	1.958075	pls	High Heterogeneity
rf	Long-Only	1.396808	7.287161	0.664002	0.093993	2.292225	rf	Low Heterogeneity
rf	Long-Only	2.091438	7.504632	0.965398	0.991879	3.115909	rf	High Heterogeneity
xgb	Long-Only	1.323103	7.718013	0.593853	0.206296	2.050061	xgb	Low Heterogeneity
xgb	Long-Only	2.391055	7.942981	1.04279	0.932135	3.365698	xgb	High Heterogeneity

Figure 6. Impact of Anonymous Trading Intensity and Dispersion on Portfolio Returns