Intraday Stock Predictability Everywhere*

Fred Liu[†‡] and Lars Stentoft[§]

July 17, 2026

Abstract

With more than a billion observations, we conduct, to our knowledge, the largest study to date of intraday stock return predictability using machine learning techniques, finding consistent out-of-sample predictability across market, sector, and individual stock returns at various time horizons. For portfolios, linear models have the strongest statistical predictive power, but nonlinear models dominate them in profitability after transaction costs. Predictability is higher at shorter horizons, highest in the middle of the day, more pronounced for less liquid firms, and persists in the absence of observable news, consistent with gradual information diffusion arising from frictions such as inattention and limits to arbitrage.

JEL Classification: C45, C52, C55, C58, G0, G1, G17

Keywords: Machine Learning, Intraday Return Prediction, Virtue of Complexity, Equity Market, Big Data, Lasso, Neural Networks, Deep Learning, Fintech

*We are grateful to the Editor (Norman Schürhoff) and two anonymous referees for their insightful and detailed comments that substantially improved the paper. We thank Thierry Foucault, Albert Menkveld, Olivier Scaillet, Fabio Trojani, and participants at the 2022 Society of Financial Econometrics Summer School in Brussels and 2023 Market Microstructure Summer School in Edinburgh for their insightful and constructive comments.

[†]Department of Economics and Finance, University of Guelph, Canada, fred.liu@uoguelph.ca. Corresponding author.

[‡]Department of Economics, University of Western Ontario, Canada.

[§]Department of Economics and Department of Statistical and Actuarial Sciences, University of Western Ontario, Canada, lars.stentoft@uwo.ca.

1 Introduction

Traditionally, the asset pricing literature focuses on stock predictability at the monthly horizon or above, and a growing body of work leverages machine learning to achieve higher accuracy in these longer-horizon stock return predictions (e.g., Gu, Kelly, and Xiu (2020)).[1] Recent papers apply machine learning to intraday stock predictability, as intraday data feature large, high-dimensional panels of predictors with potentially complex functional forms, making flexible machine learning methods well-suited to this setting. Huddleston, Liu, and Stentoft (2023) and Aleti, Bollerslev, and Siggaard (2025) study market-level predictability, while Chinco, Clark-Joseph, and Ye (2019) examine individual stocks, demonstrating that intraday returns are predictable and profitable to trade on. This paper examines three questions arising from these findings. First, how pervasive is intraday stock predictability across portfolios, individual stocks, and time horizons? Second, can model complexity improve economic performance when trading on intraday stock return predictions after transaction costs? Third, what are the economic sources of this predictability?

We use machine learning techniques to study intraday stock return predictability using a high-dimensional cross-section of lagged stock returns as predictors to address these questions. First, we find systematic and pervasive evidence of intraday predictability across market, sector, and individual stock returns at multiple time horizons. Strategies based on model forecasts often produce economically significant profits after transaction costs. Second, for portfolios, linear models outperform nonlinear models in out-of-sample statistical performance; however, nonlinear models typically dominate linear models in profitability after transaction costs, which aligns with the virtue of complexity framework of Kelly, Malamud, and Zhou (2024). For individual stocks, sparse linear models outperform most nonlinear models, suggesting that the functional form for individual stocks is sparse, consistent with Chinco, Clark-Joseph, and Ye (2019). Third, we thoroughly investigate the economic sources

[1] See, among others, Ang and Bekaert (2007), Campbell and Thompson (2008), Chen, Da, and Zhao (2013), Kelly and Pruitt (2015), Neely, Rapach, Tu, and Zhou (2014), and Welch and Goyal (2008). See Koijen and Van Nieuwerburgh (2011) and Rapach and Zhou (2013) for recent surveys.

of intraday stock predictability, producing evidence that predictability arises from gradual information diffusion caused by frictions such as inattention and limits to arbitrage, which generates lead-lag effects in intraday returns. The insight gained from studying many stocks with multiple time horizons at once, a unique feature of our analysis, is what allows these intraday patterns to be discovered.

Specifically, this paper bridges the gap between studies that focus on the monthly horizon or above and those that focus on ultra-high-frequency horizons below a second, by providing a comprehensive study of the out-of-sample predictability of market and sector portfolio returns at the 1-min, 5-min, 10-min, 15-min, and 30-min horizons, as well as individual stock returns at the 1-min horizon, from January 2005 to December 2021 using machine learning techniques. We also evaluate the economic significance of trading strategies based on the machine learning predictions, both before and after transaction costs. We study the market portfolio using the SPDR S&P 500 ETF, the sector portfolios using nine S&P sector ETFs, and individual S&P 500 constituent stocks, all selected for their high liquidity and large market capitalization. For portfolios, the predictors are lagged returns of the portfolio itself and its constituents; for individual stocks, the predictors are the lagged returns of the S&P 500 and its constituents. Our dataset consists of more than a billion 1-min stock-time return observations. To our knowledge, this constitutes the largest and most comprehensive study of intraday stock predictability and its economic sources.

In the same fashion as Gu, Kelly, and Xiu (2020), we apply a total of seventeen linear and nonlinear models to predict intraday stock returns. We also consider the median ensembles of all linear models, all nonlinear models, and all models to combine the predictions of the individual models. To obtain out-of-sample predictions from our models, we proceed as follows for each stock and each time horizon in our baseline analysis. In our baseline application we estimate model parameters on a training set of ten months; then perform an extensive hyperparameter optimization, validating the model's fit on the validation set in the next month; and subsequently evaluate the prediction power on a one-month test set.

We repeat this process each month by shifting the training, validation, and test sets forward by one month from January 2005 to December 2021 for a total of 17 years or 204 months.

Our empirical results advance the understanding of intraday stock return predictability in several ways. First, we compare the predictive power of models for the market and sector portfolios using the out-of-sample R^2_{OS} metric, which benchmarks model predictions against a forecast of zero returns. Linear models generally achieve much higher R^2_{OS} values than nonlinear models, indicating that simpler models have stronger statistical predictive power for intraday portfolios. The R^2_{OS} is highest at the 1-min horizon, reaching 4.11% for the technology portfolio, and decreases as the horizon lengthens. By the 30-minute horizon, most R^2_{OS} values remain positive but have declined substantially, indicating that predictability attenuates within 30 minutes.

Second, we examine the economic significance of model predictions. Following Kelly, Malamud, and Zhou (2024), we use a strategy that buys (sells) the portfolio on positive (negative) predictions, scaled by the magnitude of the prediction. For the market and sector portfolios at the 1-min horizon, models achieve annualized Sharpe ratios of up to 3.38 for the technology portfolio and frequently outperform the buy-and-hold benchmark. Linear and nonlinear models produce similar Sharpe ratios, despite the substantially higher R^2_{OS} values of linear models, indicating that statistical predictive power does not translate directly into economic profitability.

Third, we examine the economic significance of model predictions after transaction costs. We modify the strategy following Chinco, Clark-Joseph, and Ye (2019) to trade only when the signal is strong, i.e., when the model prediction magnitude exceeds the transaction cost. For nearly every portfolio, nonlinear models dominate linear models, with median ensembles achieving after-cost Sharpe ratios of up to 2.72 for the technology portfolio. This finding may seem surprising given the strong statistical predictive power of linear models, but it aligns with the virtue of complexity framework of Kelly, Malamud, and Zhou (2024). When observed predictors are noisy proxies for the true conditioning information, nonlinear models

mitigate this model misspecification and better approximate the conditional expected return, leading to higher Sharpe ratios. Economic significance thus increases with model complexity rather than statistical predictability, as nonlinear models are flexible enough to capture complex signals that linear models miss. After-cost Sharpe ratios are highest at the 1-min horizon and decrease as the horizon lengthens. Across all portfolios, the nonlinear model median ensemble strategy at the 1-min horizon produces significant alphas after adjusting for static and dynamic risk exposures, and data mining cannot account for these results.

Turning to the cross-section of individual stocks, our fourth result is that sparse penalized linear methods, e.g. lasso and elastic net models, with R^2_{OS} values above 1%, have a much higher predictive power than other individual models, which is consistent with the results in Chinco, Clark-Joseph, and Ye (2019). Sparse models also achieve high economic significance after transaction costs, with Sharpe ratios exceeding 3.5, and the only nonlinear model to outperform them is the most complex neural network, though only slightly. These results differ from the portfolio setting, where nonlinear models frequently outperform the sparse penalized linear methods, suggesting that the functional form for individual stocks differs from that of portfolios and is likely sparse. We also find that individual stocks are less statistically predictable than portfolios but more economically profitable to trade.

Fifth, we investigate whether gradual information diffusion (Hong and Stein (1999); Hong, Torous, and Valkanov (2007)), driven by frictions such as inattention and limits to arbitrage, is an economic source of intraday stock predictability. We find a striking commonality of R^2_{OS} patterns within the trading day across portfolios. For nearly every stock portfolio, the R^2_{OS} exhibits a bell shape, where predictability is higher in the middle of the day and lower in the first and last 30 minutes. These results mirror the well-known U shape in intraday trading volume and suggest that inattention is a source of intraday predictability. We also document that firms subject to greater frictions, including smaller, illiquid firms with less analyst coverage and higher arbitrage costs, are more predictable, and that large, liquid, and actively traded firms are the most influential predictors. The results are relatively stable

across re-estimation timescales (the frequency at which models are re-estimated), ranging from yearly to daily, and yearly re-estimation achieves the strongest economic performance while requiring the least computational resources. Finally, we show that intraday stock predictability persists in the absence of macroeconomic and firm-specific news and is generally lower on news days, consistent with news raising investor attention and accelerating price discovery. A natural experiment based on a major trading glitch reinforces this finding, as the event attracted significant news coverage and investor attention, substantially reducing intraday predictability for affected stocks. Taken together, these findings support the gradual information diffusion explanation.

Our paper has implications for the growing quantitative hedge fund industry, which now manages more than a trillion dollars in investor assets and accounts for a substantial share of all stock trades (Zuckerman and Hope (2017)). These firms operate at shorter horizons than traditional funds that focus on the monthly horizon or above, but at longer horizons than high-frequency traders that operate below a second, and their proprietary nature makes them largely opaque to academic researchers (Goldstein, Spatt, and Ye (2021)). Many use machine learning techniques to trade algorithmically at intraday frequencies. While we cannot directly evaluate their proprietary strategies, our analysis provides a transparent academic benchmark for the predictability and profitability of intraday machine learning predictions, and offers indirect evidence on the potential economic sources of returns for these firms.

Our paper contributes to two active strands of literature. First, our results connect to the growing body of work on intraday return predictability using lagged returns. Chinco, Clark-Joseph, and Ye (2019) estimate a lasso model on the cross-section of NYSE lagged returns and show that 1-min returns are predictable. Chordia, Roll, and Subrahmanyam (2005) and Chordia, Roll, and Subrahmanyam (2008) find that intraday returns cannot be predicted by past prices but that order imbalances forecast short-term returns. Heston, Korajczyk, and Sadka (2010) document significant autocorrelation of intraday returns at

daily intervals. Gao, Han, Li, and Zhou (2018) show that the first half-hour market return predicts the last half-hour return. Bogousslavsky (2016) demonstrates theoretically that infrequent portfolio rebalancing by traders can generate predictability in intraday returns. These studies, however, all employ *linear* models to forecast returns, whereas our framework accommodates variable interactions and other *nonlinear* effects through more flexible functional forms. Our paper also relates to Huddleston, Liu, and Stentoft (2023) and Aleti, Bollerslev, and Siggaard (2025), who apply machine learning techniques trained on lagged returns of individual stocks and characteristic-managed portfolios, respectively, to study *market* predictability at the 5-min and 15-min horizons. We extend these analyses to the predictability of market, sector, and individual stock returns across multiple horizons. More broadly, Aleti and Bollerslev (2025) estimate an intraday stochastic discount factor using the generalized adversarial method of moments procedure from Chen, Pelger, and Zhu (2023), and Bartlett, McCrary, and O'Hara (2025) train XGBoost on odd-lot quotes to predict 1-min individual stock returns, demonstrating the predictive value of these data.

Second, our paper bridges the rapidly expanding literature applying machine learning to forecast monthly stock returns and the nascent machine learning literature on ultra-high-frequency predictability. Our paper is most closely related to Gu, Kelly, and Xiu (2020), who apply an extensive array of machine learning techniques to predict monthly risk premiums (see also Gu, Kelly, and Xiu (2021); Feng, He, Polson, and Xu (2024)). Bryzgalova, Pelger, and Zhu (2025), Chen, Pelger, and Zhu (2023), and Kozak, Nagel, and Santosh (2020) estimate the stochastic discount factor using machine learning to price stocks. At ultra-high-frequency horizons, Easley, López de Prado, O'Hara, and Zhang (2021) train a random forest on microstructure predictors to forecast measures of market quality such as bid-ask spreads and the sign of change in realized volatility; Aït-Sahalia, Fan, Xue, and Zhu (2025) train machine learning models on microstructure predictors to predict returns at millisecond horizons; and Kwan, Philip, and Shkilko (2021) apply reinforcement learning trained on the limit order book to study price discovery.

The rest of the paper is organized as follows: Section 2 introduces the data and machine learning methodology. Section 3 presents the main empirical results, including the statistical predictive power and economic significance before and after transaction costs. Section 4 examines the economic sources of predictability. Section 5 concludes.

2 Data and Methodology

Given T intraday price observations, indexed by $t = 1, ..., T$, and the natural logarithm of the t^{th} observed price denoted by p_t, the corresponding logarithmic return is given by

$$r_t \equiv p_t - p_{t-1}, \tag{1}$$

where t denotes a 1-min time interval and r_t denotes the stock return for the market portfolio, a sector portfolio, or an individual stock.[2] Of particular interest is whether the h-horizon return r_{t+h} can be predicted h periods ahead, for $h \in \{1, 5, 10, 15, 30\}$. Since t advances in 1-min increments, the h-horizon return forecasts overlap for $h > 1$.[3]

In its most general form, we describe a stock's return as the sum of expected returns plus the error term as

$$r_{t+h} = E_t[r_{t+h}] + \epsilon_{t+h}, \tag{2}$$

where

$$E_t[r_{t+h}] = g_h\left(z_t\right), \tag{3}$$

and the function $g_h(z_t)$ links the expected return at horizon h to a P-dimensional vector of

[2]For ease of presentation, we use the term "stock" to indicate either a stock portfolio or individual stock.

[3]In a previous version, we used h-min lagged returns to predict h-horizon returns. We empirically found that the current approach of using 1-min lagged returns to predict h-horizon returns, inspired by van Binsbergen, Boons, Opp, and Tamoni (2023), substantially outperforms the previous approach. We thank an anonymous reviewer for this suggestion.

predictors, denoted z_t, and varies across stocks. The predictors z_t are 1-min lagged returns and are common across all five horizons but differ across the market, sector, and individual stock settings, as detailed in Section 2.2.

2.1 Data

To assess whether stock returns can be predicted, we construct, to our knowledge, the largest sample of intraday stock returns ever considered. The primary dataset is the trade and quotes (TAQ) database, containing intraday transaction data for all stocks on the New York stock exchange (NYSE), American stock exchange (AMEX), and NASDAQ from January 1999 to December 2021. We study the market portfolio using the SPDR S&P 500 (SPY) ETF; the sector portfolios using the following nine S&P sector ETFs: Energy (XLE), Materials (XLB), Industrials (XLI), Consumer Discretionary (XLY), Consumer Staples (XLP), Health (XLV), Financials (XLF), Information Technology (XLK), and Utilities (XLU); and the individual S&P 500 constituent stocks. We obtain the list of S&P 500 constituent stocks from the Center for Research in Security Prices (CRSP), the S&P Global Industry Classification Standard (GICS) from Compustat, and update the 500 constituent stocks and sectors monthly. Our dataset consists of more than a billion stock-time 1-min return observations.

We measure transaction costs using the effective bid-ask spread, which captures the direct cost of trading. The effective bid-ask spread is a widely used transaction cost measure in the asset pricing literature (e.g., Chen and Velikov (2023), Novy-Marx and Velikov (2016)) and the intraday literature (e.g., Bartlett, McCrary, and O'Hara (2025); Jiang, Li, and Wang (2021)), where it is used to compute returns net of transaction costs.[4] See Internet Appendix Section IA1 for more details on the databases used.

[4]The effective bid-ask spread has well-known limitations. It does not capture price impact and therefore best approximates the execution costs of a small liquidity-demanding trader. For large institutional orders, the effective spread understates true trading costs. At the same time, it provides a conservative benchmark in the sense that it assumes all trades execute as marketable orders, which overstates costs for traders who use limit orders. The effective bid-ask spread is also well-suited to tests of economic significance from predictability, where the relevant question is whether the marginal arbitrageur can profitably direct capital to a strategy after covering execution costs (Novy-Marx and Velikov (2016)).

In addition to the intraday data described above, we obtain the monthly risk-free rate from Kenneth French's website. We convert the monthly risk-free rate to the 1-min frequency by dividing it with the number of 1-min intervals in the trading month to construct intraday excess returns. From the same source, we also obtain the Fama-French five factors (Fama and French (2015)) and momentum (Carhart (1997)) for risk adjustment.

To investigate the economic sources of predictability, we also incorporate 20 monthly firm characteristics from Gu, Kelly, and Xiu (2020) and Gu, Kelly, and Xiu (2021), downloaded from Dacheng Xiu's website. This data can be grouped into three categories. The first category includes liquidity variables such as firm size (mvel1), dollar volume and dollar volume volatility (dolvol, std_dolvol), turnover and turnover volatility (turn, std_turn), bid-ask spread (baspread), Amihud illiquidity (ill), price delay (pricedelay), and number of zero trading days (zerotrade). Risk measures are the second category, and include market beta (beta), beta squared (betasq), and total and idiosyncratic return volatility (retvol, idiovol). Price trend variables are the third category, including changes in momentum (chmom), industry momentum (indmom), recent maximum return (maxret), short-term reversal (mom1m), 6 and 12 month stock momentum (mom6m, mom12m), and long-term reversal (mom36m).

Finally, we obtain analyst forecast data and earnings announcement dates and times from I/B/E/S. We define macroeconomic announcement days following Savor and Wilson (2013) and Gao, Han, Li, and Zhou (2018). We identify firm-specific news days using the dataset of Boudoukh, Feldman, Kogan, and Richardson (2019), which classifies news coverage from Dow Jones Newswires through December 2015. These datasets are described in detail in Sections 4.2.1 and 4.2.2.[5]

2.2 Models Considered

For each stock portfolio, the predictors z_t are lagged returns of the portfolio itself and its constituents. For example, when predicting the market portfolio, the predictors z_t are

[5]We thank Kenneth French, the authors of Gu, Kelly, and Xiu (2020), and the authors of Boudoukh, Feldman, Kogan, and Richardson (2019) for making their datasets publicly available.

lagged returns of the S&P 500 and its constituents, resulting in approximately 501 covariates. When predicting the sector portfolios, the predictor variables z_t are lagged returns of the sector portfolio and the subset of S&P 500 constituents belonging to that sector, and the number of covariates depends on the given sector and month. For each of the individual stocks, the predictors z_t are the lagged returns of the S&P 500 and its constituents; the same set of predictors used for forecasting the market portfolio. We thus have a situation with many predictors and use machine learning techniques for estimating the functional form of $g_h(z_t)$ in the empirical model in Equation (3). Under the theoretical framework of Kelly, Malamud, and Zhou (2024), because these predictors are a subset of all relevant conditioning information, the true conditional expectation is likely nonlinear in the observed predictor space, regardless of the functional form in a richer information set. This model misspecification implies that nonlinear models may outperform linear models under certain conditions.

Following the growing literature on machine learning algorithms for predicting asset returns (e.g., Gu, Kelly, and Xiu (2020) and Bianchi, Büchner, and Tamoni (2021)), we compare a variety of linear and nonlinear machine learning models. Linear models include the autoregressive model (AR1), ordinary least squares with all covariates (OLS), penalized regression methods such as ridge (RID), lasso (LAS), elastic net (EN), and dimension reduction methods such as principal component regression (PCR). Nonlinear models include tree based methods such as random forest (RF), extremely random forest (XRF), and gradient-boosted regression trees (GBT), along with feed-forward neural networks with one to five hidden layers (NN1-NN5). We also study the median ensembles of all linear models (L-MED), all nonlinear models (NL-MED), and all models (MED) to combine the predictions of the individual models. We use the median, instead of mean, ensemble to ensure the aggregate forecasts are robust to potential outliers. Huddleston, Liu, and Stentoft (2023) provide a detailed description of these methods.

Since the nonlinear models are more complex than the linear models, this allows us to

test the virtue of complexity in intraday settings within the theoretical framework of Kelly, Malamud, and Zhou (2024). We conduct this test by comparing NL-MED against the L-MED, as ensembles reduce the estimation noise that can obscure the virtue of complexity in comparisons of individual nonlinear models. We also test the equality of Sharpe ratios between L-MED and NL-MED using the two-sided test of Leung and Wong (2008).

2.3 Estimation Procedure and Variable Importance

We estimate our models using a rolling window approach for each stock and each time horizon. In our baseline analysis, we begin by estimating model parameters on a training set of ten months $\mathcal{T}_1$ (February, 2004 - November, 2004); then perform an extensive hyperparameter optimization, validating the model's fit with the next month as the validation set $\mathcal{T}_2$ (December, 2004); and subsequently evaluate the prediction power on a one-month test set (January, 2005).[6] We repeat this procedure by shifting the training, validation, and test sets forward by one month in each iteration, for a total test set $\mathcal{T}_3$ of 17 years (January 2005 - December 2021).[7] The above is done for all stock portfolios across all five time horizons. For individual stocks, we consider only the 1-min horizon and, following Chinco, Clark-Joseph, and Ye (2019), restrict the test set to October of each year from 2005 to 2021 due to computational constraints. Internet Appendix Section IA2 details the computational framework used to estimate the models and the setup of the hyperparameter optimization.

An important element in the application of machine learning techniques is to assess variable importance. Following Gu, Kelly, and Xiu (2020), the variable importance of the j^{th} predictor each month is measured by the reduction in statistical predictability from setting all values of predictor j to zero, while holding the remaining model estimates fixed. In our setting, variables are lagged stock returns and hard to interpret as is. Thus, we propose

[6]This baseline analysis uses a monthly re-estimation timescale, with a training sample of 10 months. Section 4.1.5 considers alternatives ranging from yearly to minutely, with training samples ranging from five years to 30 minutes.

[7]Our choice to begin the test set on January of 2005 is deliberately conservative, since intraday stock market predictability has been demonstrated to be very high before this year (Huddleston, Liu, and Stentoft (2023)), and it also corresponds to the starting year used in Chinco, Clark-Joseph, and Ye (2019).

instead to conduct a monthly Fama-MacBeth regression (Fama and MacBeth (1973)) of the variable importance on each of the 20 contemporaneous firm characteristics. That is, we essentially project the hard to interpret lagged returns onto interpretable characteristics. To improve interpretability even further, we standardize both the variable importance and characteristics cross-sectionally each month to have a mean of zero and variance of one, allowing for the interpretation of the coefficient as the impact of a one standard deviation change in the characteristic on standard deviation changes in variable importance. Our machine learning interpretability technique has two major advantages. First, the sign of the coefficient provides an additional layer of understanding, helping to identify which firms possess predictive power. Second, the t-statistic of the coefficient acts as an indicator of whether the importance of the characteristic is statistically significant.

3 Main Empirical Results

This section studies the empirical results for the market and sector portfolios at the 1-min, 5-min, 10-min, 15-min, and 30-min horizons from January 2005 to December 2021. We first consider purely statistical predictability, then we assess the economic significance of any such predictability both before and after accounting for reasonable transaction costs. Finally, we assess the statistical predictability and economic significance of individual S&P 500 constituent stocks at the 1-min horizon.

3.1 Portfolio Out-of-sample Predictive Power

To evaluate the predictive power of the intraday stock return forecasts, we use the out-of-sample R^2 metric of Gu, Kelly, and Xiu (2020). Given the h-horizon stock return, r_{t+h}, and a corresponding model prediction, $\hat{r}_{t+h}$, the out-of-sample R^2 is

$$R^2_{\text{OS}} \equiv 1 - \frac{\sum_{t \in \mathcal{T}_3} \left(r_{t+h} - \hat{r}_{t+h}\right)^2}{\sum_{t \in \mathcal{T}_3} \left(r_{t+h}\right)^2}, \tag{4}$$

where $\mathcal{T}_3$ indicates that the metric is only assessed on predictions in the test set, whose data was not used for model estimation or validation. The R^2_{OS} statistic pools prediction errors over time into a grand-level assessment of each model for a given stock, and it measures the reduction in mean squared forecasting error (MSFE) for each model relative to a naive zero benchmark forecast, which assumes that the $t + h$ stock return r_{t+h} equals zero.

Table 1 presents the R^2_{OS} (in percentage) for all portfolios at the 1-min horizon. Beginning with linear models, the AR1 benchmark generally has the lowest predictive power, with a maximum R^2_{OS} of 0.09% for the finance portfolio.[8] OLS with all covariates, along with penalized linear models such as LAS, RID, and EN, achieve high R^2_{OS} values, reaching 4.11% for OLS in the technology portfolio. PCR produces lower R^2_{OS} values than OLS and penalized linear models across all sector portfolios, reaching 1.27% for consumer staples, indicating that linear dimension reduction often underperforms in this intraday setting.

Turning to nonlinear models, RF, XRF, GBT, and NN1-NN5 generally have lower R^2_{OS} values compared to OLS and penalized linear models. This finding is visually summarized in Figure 1, which shows that across all portfolios, the linear median ensemble (L-MED) has higher R^2_{OS} values than the nonlinear (NL-MED) and total (MED) median ensembles, respectively. The high R^2_{OS} values achieved by linear models may be surprising, but it is in line with previous research (e.g., Chinco, Clark-Joseph, and Ye (2019)) demonstrating that linear models such as LAS are highly successful at predicting intraday returns. The higher R^2_{OS} values of linear models over nonlinear models suggests that less complex models may have stronger statistical predictive power for intraday returns.

Next, we compare portfolio predictability across different time horizons. Figure 2 presents the R^2_{OS} values achieved by the total ensemble, MED, across portfolios at the 1-min, 5-min, 10-min, 15-min, and 30-min horizons. For all but one portfolio, the R^2_{OS} is highest at the 1-min horizon and decreases monotonically as the horizon lengthens. The exception is the market portfolio, which is most predictable at the 5-min horizon, a finding consistent with

[8]Huddleston, Liu, and Stentoft (2023) demonstrate that autoregressive models with additional lags generally perform statistically and economically worse than the AR1 for intraday returns.

the results in Huddleston, Liu, and Stentoft (2023). Nevertheless, the R^2_{OS} remains positive at the 30-min horizon across all portfolios, indicating that predictability persists beyond intervals of a few minutes, though it attenuates substantially as the horizon lengthens.

In summary, this subsection documents that most machine learning models can successfully predict intraday returns across portfolios and time horizons. However, statistical predictive power is less important than the economic significance of the model predictions. As argued by Campbell and Thompson (2008), even a small R^2_{OS} can generate economically large returns, which can be especially true given the large number of trading opportunities at the 1-min interval. In the next subsection, we examine the economic significance of the model predictions.

3.2 Portfolio Economic Significance Before Transaction Costs

To assess the economic significance of the machine learning predictions before transaction costs, we implement the timing strategy in Equation (6) of Kelly, Malamud, and Zhou (2024), which optimizes the conditional Sharpe ratio. Adopting their framework allows us to test whether their theoretical predictions extend to intraday horizons. The strategy takes a long (short) position in a stock when the model predicts a positive (negative) return, scaled by the magnitude of the prediction, i.e., the stock's conditional expected return forecast. Specifically, the strategy's return at horizon h given model prediction $\hat{r}_{t+h}$ is defined as

$$\pi\left(\hat{r}_{t+h}\right) \equiv \begin{cases} -r_{t+h}\hat{w}_t & \text{if } \hat{r}_{t+h} < 0, \\ r_{t+h}\hat{w}_t & \text{if } \hat{r}_{t+h} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

where $\hat{w}_t = \min(|\hat{r}_{t+h}|, 1)$ denotes the portfolio weight implied by the forecast $\hat{r}_{t+h}$. This timing weight scales the position up or down to exploit time variation in the stock's conditional expected return, and is capped at 100% to prevent excessive leverage. For each

stock, we subtract the intraday risk-free rate from the strategy's log returns, sum within each month, convert to arithmetic returns to form monthly excess returns, and compute annualized Sharpe ratios.

Table 2 presents the annualized Sharpe ratios of all models using the trading strategy in Equation (5) for the 1-min horizon, along with the buy-and-hold Sharpe ratio for each portfolio computed using intraday returns only and using total returns that include overnight returns. The Sharpe ratios are economically large across most portfolios and models, and frequently exceed both buy-and-hold benchmarks. The highest Sharpe ratio of 3.38 is achieved by OLS applied to the technology sector.

Figure 3 reveals that across all portfolios, the Sharpe ratios of the linear (L-MED), nonlinear (NL-MED), and total (MED) ensembles are strikingly similar, despite the earlier finding in Table 1 that linear models achieve substantially higher R^2_{OS} than nonlinear models. Comparing Sharpe ratios using the Leung and Wong (2008) test, L-MED significantly outperforms NL-MED in three portfolios but NL-MED significantly outperforms L-MED in two, yielding no clear winner before transaction costs. This disconnect between statistical and economic performance indicates that predictive power does not translate directly into profitability.

The results in this subsection demonstrate that machine learning forecasts generate economically large Sharpe ratios across all portfolios. However, they do not reveal a virtue of complexity, as nonlinear models do not consistently outperform linear models before transaction costs. One explanation is that the trading strategy in Equation (5) does not account for transaction costs, which are substantial at intraday frequencies. Linear models may capture statistically predictable but economically unprofitable patterns such as the bid-ask bounce, while lacking the complexity to identify economically meaningful predictability that exceeds the cost of trading. For machine learning forecasts to be economically meaningful, the strategy must generate significant profits net of transaction costs. The next subsection examines this question directly.

3.3 Portfolio Economic Significance After Transaction Costs

Following Chinco, Clark-Joseph, and Ye (2019), we evaluate the economic significance of returns after transaction costs, estimated as the bid-ask spread, by implementing a trading strategy that only trades when model predictions exceed the spread, scaled by the magnitude of the prediction. The return of this trading strategy is defined as

$$\gamma\left(\hat{r}_{t+h}, \text{Spread}_t\right) \equiv \begin{cases} (-r_{t+h} - \text{Spread}_t)\hat{w}_t & \text{if } |\hat{r}_{t+h}| > \text{Spread}_t \text{ and } \hat{r}_{t+h} < 0, \\ (r_{t+h} - \text{Spread}_t)\hat{w}_t & \text{if } |\hat{r}_{t+h}| > \text{Spread}_t \text{ and } \hat{r}_{t+h} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

For each stock, the strategy's log returns net of transaction costs and intraday risk-free rate are summed within each month, converted to arithmetic returns, which are then used to calculate annualized Sharpe ratios. For a model to generate profits, its predictions must clear two hurdles simultaneously: they must be correctly signed and must exceed the transaction costs in absolute value, that is, $|\hat{r}_{t+h}| > \text{Spread}_t$ with $\hat{r}_{t+h} > 0$ ($\hat{r}_{t+h} < 0$) for a long (short) position. Since predictions that are too small in magnitude yield no trade and predictions that are incorrectly signed generate large losses, this trading strategy presents a meaningful hurdle for model evaluation.

Table 3 presents the annualized Sharpe ratios of all models using the trading strategy in Equation (6) for the 1-min horizon, along with the intraday, IH, and total buy-and-hold, BH, Sharpe ratio for each portfolio. Incorporating transaction costs generally substantially reduces the Sharpe ratios relative to Table 2. For example, OLS produces Sharpe ratios of -1.85, 0.24, 0.41, and -0.09 for the market, energy, finance, and health portfolios, respectively, compared to 1.43, 1.86, 2.61, and 2.90 for the same portfolios before transaction costs. This suggests that OLS largely captures unprofitable patterns such as the bid-ask bounce rather than economically meaningful signals. Other linear models such as AR1 and PCR

also perform poorly after costs, though the penalized linear models fare better, consistent with the finding in Chinco, Clark-Joseph, and Ye (2019) that sparse models are effective for intraday return prediction.[9] In contrast, most nonlinear models retain high Sharpe ratios after transaction costs, with NL-MED reaching 2.72 in the technology sector.

Figure 4 summarizes these results visually. The Sharpe ratios for the total ensemble (MED) are generally high, reaching 1.66 for the health portfolio. Across eight out of ten portfolios, the nonlinear ensemble (NL-MED) achieves a higher Sharpe ratio than the linear ensemble (L-MED), and this outperformance is statistically significant for the finance, health, industrials, staples, technology, and utilities sectors. Figure 5 presents the annualized Sharpe ratios after transaction costs achieved by the nonlinear ensemble across portfolios for the 1-min, 5-min, 10-min, 15-min, and 30-min horizons. The 1-min horizon produces the highest Sharpe ratio in all portfolios except consumer staples, where the 5-min horizon is slightly higher. The 1-min and 5-min horizon Sharpe ratios comfortably exceed the intraday buy-and-hold benchmark across all portfolios. At the 1-min horizon, the Sharpe ratios also exceed the total buy-and-hold benchmark for all sector portfolios but not for the market portfolio, reflecting the large overnight equity premium during this period. In contrast, Sharpe ratios for the 15-min horizon are relatively low, and Sharpe ratios for the 30-min horizon are negative across all portfolios except finance.

It may seem surprising that linear models exhibit substantially higher R^2_{OS} values (Section 3.1) while nonlinear models often achieve higher Sharpe ratios after transaction costs (Section 3.3). This apparent discrepancy can be understood through the virtue of complex-

[9]While linear dimension reduction techniques like PCR underperform, nonlinear dimension reduction may still be effective. To test this, we apply autoencoder neural network models, which can be viewed as nonlinear analogues of PCR (Gu, Kelly, and Xiu (2021)), under the same empirical setup at the 1-min horizon. Internet Appendix Table IA2 reports the results, where we first apply dimension reduction using an autoencoder with tanh activation functions, then apply OLS, EN, NN1, NN2, or NN3 to the autoencoder factors. Panel A shows that autoencoders with OLS generally have negative or low R^2_{OS} values, and the other models typically underperform PCR, indicating that autoencoders do not improve statistical predictability. Panel B shows that autoencoders with OLS and EN have relatively low after-cost Sharpe ratios, but autoencoders with NN2 reach Sharpe ratios of up to 3.27 for the technology portfolio, indicating that nonlinear dimension reduction combined with NNs can improve economic significance over PCR, consistent with the findings in Gu, Kelly, and Xiu (2021) and the virtue of complexity. We do not consider autoencoder models in other settings due to computational constraints and leave this extension for future work.

ity framework of Kelly, Malamud, and Zhou (2024), which emphasizes that the R^2_{OS} is an incomplete measure of economic significance. In their framework, the R^2_{OS} measure reflects an additive trade-off between expected timing returns and the variance of the trading signal, making it sensitive to the variance of model forecasts. Nonlinear models tend to generate higher-variance forecasts as they recover more of the underlying conditional mean, which can depress the R^2_{OS} even when the resulting trading strategy delivers a high Sharpe ratio.[10] Regularized linear models, by contrast, produce smoother, lower-variance forecasts that can mechanically inflate the R^2_{OS} despite yielding lower Sharpe ratios. This effect is amplified in intraday settings, where the bid-ask bounce mechanically induces predictability that is not economically profitable. Accounting for transaction costs at intraday horizons is therefore necessary to reveal the virtue of complexity.

3.3.1 Robustness: Risk-Adjusted Returns

An important consideration is whether the trading strategies are correlated with existing risk factors. If the machine learning strategies are simply loading on known risk factors, they offer little value for generating risk-adjusted returns. Table 4 presents the alphas, betas, and adjusted R^2 values from regressing monthly excess returns of the NL-MED strategy at the 1-min horizon on a six-factor model consisting of the Fama-French five factors (Fama and French (2015)) augmented with momentum (Carhart (1997)). Similar to Chinco, Clark-Joseph, and Ye (2019), we scale the strategy's monthly returns to match the intraday buy-and-hold volatility to aid interpretation of the alphas and coefficients, noting that this scaling does not affect t-statistics or adjusted R^2 estimates.

Across portfolios, the monthly six-factor alphas range from 0.59% for the market portfolio to 4.35% for the finance sector, which are economically large given that the strategy accounts for transaction costs. The alphas are statistically significant at the 1% level for all portfolios

[10] Borrowing the illustrative example from Kelly, Malamud, and Zhou (2024), consider a single-predictor model in which the estimated coefficient is a large scalar multiple of the true value. In this case, the R^2_{OS} can be negative even though the forecasts are perfectly aligned with the true conditional mean, implying a positive expected return and Sharpe ratio for the timing strategy.

except the market and energy portfolios, which have t-statistics of 1.99 and 1.88, respectively. Most factor betas are either insignificant or negative, indicating that the strategy is largely uncorrelated with or hedges against known risk factors. Adjusted R^2 values are low across most portfolios, indicating that the strategy's returns are not well explained by the six-factor model.

3.3.2 Robustness: Time Varying Risk Exposures

The results in Table 4 use monthly factors to risk-adjust the performance of intraday strategies, which adjusts for static risk exposures. One concern is that risk exposures may vary over time (Engelberg, McLean, and Pontiff (2018)). The intraday setting, however, makes controlling for dynamic risk exposures straightforward. Patton and Verardo (2012) show that betas estimated from intraday returns capture time variation in market beta, and Pelger (2020) demonstrates that a four-factor industry model consisting of the equal-weighted market, oil, finance, and electricity portfolios explains variation in intraday returns better than the Fama-French three-factor model augmented with momentum. Pelger (2020) further shows that risk exposures change across years but are stable within each month.

Motivated by these findings, we construct intraday common factors at the 1-min interval and use them to risk-adjust returns each month, assuming that risk exposures are locally constant within each month but can vary across months. Specifically, we construct the four-factor industry model of Pelger (2020) and the Fama and French (2015) six-factor model (five factors augmented with momentum) at the 1-min horizon using portfolio returns. See Internet Appendix Section IA1 for details on intraday factor construction. Within each month, we regress 1-min strategy returns on the 1-min risk factors to obtain monthly coefficient estimates, then report Newey-West t-statistics of these estimates across months.

Table 5 presents the t-statistics for each set of intraday common factors.[11] Panel A shows that all alphas relative to the intraday six-factor model are statistically significant at the 5%

[11]We report only t-statistics and adjusted R^2 values, which are scale invariant; the coefficient estimates reflect the 1-min return scale and are difficult to interpret economically, but are available upon request.

level. Panel B shows that the four-factor alphas are significant at the 5% level for all sector portfolios, while the market portfolio has a t-statistic of 1.89. Taken together, the results in this section indicate that intraday strategies based on machine learning forecasts deliver substantial risk-adjusted returns after accounting for transaction costs and both static and dynamic risk factor exposures.

3.3.3 Robustness: Data Mining

Another natural concern is that our results reflect data mining (Engelberg, McLean, and Pontiff (2018)). We address this concern by evaluating our findings on data beyond the original sample period, following the approach in McLean and Pontiff (2016) and Jiang, Li, and Wang (2021). Specifically, we examine predictability in the most recent sample from 2017 to 2021. These five years were not available when the original version of this paper was written, nor were they included in several closely related studies of intraday return predictability for the aggregate market (Huddleston, Liu, and Stentoft (2023)) and individual stocks (Chinco, Clark-Joseph, and Ye (2019)). This period therefore provides an evaluation of intraday stock return predictability that is entirely independent of the original study.

Table 6 presents the R^2_{OS} and Sharpe ratios from the strategy in Equation (6) for the 1-min horizon across the original sample from 2005 to 2016 and the recent period from 2017 to 2021. The R^2_{OS} decreases substantially for all sector portfolios in the out-of-sample period. For example, the R^2_{OS} for the technology sector falls from 1.60 in the original sample to 0.26 out of sample. This decline in statistical predictability may partly reflect the secular narrowing of bid-ask spreads driven by technological improvements (Chordia, Roll, and Subrahmanyam (2011)), which mechanically reduces the predictable component associated with the bid-ask bounce.[12] The Sharpe ratios, however, do not exhibit the same attenuation, and in fact increase in seven out of ten portfolios, consistent with lower trading costs offsetting the

[12]Huddleston, Liu, and Stentoft (2023) show that declines in R^2_{OS} from intraday machine learning models are strongly associated with reductions in tick size and bid-ask spreads.

decline in gross predictability. Overall, these patterns are inconsistent with data mining, under which the recent Sharpe ratios should be close to zero. Instead, they suggest that intraday stock return predictability is a robust empirical phenomenon.

3.4 Empirical Results for Individual Stocks

So far, we have shown that intraday stock portfolios are consistently predictable and profitable using machine learning forecasts after accounting for transaction costs. This section evaluates the results for the pooled sample of individual S&P 500 constituent stocks at the 1-min horizon.

The left bars in Figure 6 compare machine learning methods in terms of R^2_{OS} for individual stocks. Penalized linear models that rely on sparsity outperform nonlinear models, with LAS and EN achieving R^2_{OS} values slightly above 1%, consistent with the strong predictive performance of sparse models documented in Chinco, Clark-Joseph, and Ye (2019). Non-sparse linear models such as OLS, RID, and PCR produce negative or substantially lower R^2_{OS} values. Nonlinear models generally fall between these extremes, with NN5 producing the lowest R^2_{OS} among nonlinear models due to its high complexity.

The right bars in Figure 6 compare machine learning methods in terms of annualized Sharpe ratios from the strategy in Equation (6), computed from excess returns equally weighted across the pooled sample of S&P 500 constituent stocks in each month. Among linear models, LAS and EN perform well with Sharpe ratios above 3.5, again consistent with Chinco, Clark-Joseph, and Ye (2019), while OLS and RID have negative Sharpe ratios. PCR achieves a moderate Sharpe ratio of approximately 2.5. Among nonlinear models, tree-based methods underperform PCR, while NN2 through NN4 outperform OLS, RID, and PCR but fall short of LAS and EN. NN5 achieves the highest Sharpe ratio overall, though only slightly exceeding the sparse linear models, suggesting that a modest nonlinear component exists beyond the dominant sparse signal. Driven by the strong performance of LAS and EN, the L-MED ensemble earns a higher Sharpe ratio than NL-MED, although the difference is

not statistically significant, and MED falls between the two.

We now compare the individual stock results to the portfolio results in Sections 3.1 and 3.3. While the best models achieve an R^2_{OS} of slightly above 1% for individual stocks, this is far lower than the R^2_{OS} of up to 4.11% for portfolios in Table 1. The after-cost Sharpe ratios tell the opposite story: LAS, EN, NN5, and L-MED all exceed 3.5 for individual stocks, higher than the after-cost Sharpe ratio earned by any model and portfolio combination in Table 3. Individual stocks are thus less statistically predictable than portfolios but more economically profitable to trade. This is intuitive as the cross-section of individual stocks offers many more exploitable trading signals than any single portfolio.

The model rankings also differ substantially between individual stocks and portfolios, suggesting that the functional forms governing intraday predictability differ across the two settings. OLS and RID perform well for most sector portfolios in Table 1 but produce negative or low R^2_{OS} values and Sharpe ratios for individual stocks. Most notably, nonlinear models such as NN2 through NN4 earn much higher Sharpe ratios than LAS and EN in the portfolio results (Table 3) but underperform them for individual stocks. Accordingly, NL-MED often significantly outperforms L-MED for portfolios but underperforms L-MED for individual stocks. The strong performance of LAS and EN for individual stocks, combined with the poor performance of OLS and RID, indicates that the true functional form in Equation (3) for individual stocks is likely sparse, consistent with Chinco, Clark-Joseph, and Ye (2019). In line with the theoretical framework of Kelly, Malamud, and Zhou (2024), when the true model is relatively sparse, nonlinearities contribute less to performance because sparse models are already well-specified, leaving little bias for additional complexity to reduce while incurring greater estimation variance.[13]

[13]The theoretical framework of Kelly, Malamud, and Zhou (2024) assumes that many signals each contribute a small amount of predictability, representing a functional form that is non-sparse.

4 Economic Sources of Predictability

Our main results show that lagged returns predict subsequent market, industry, and individual stock returns, a form of lead-lag effect applied to intraday markets. A large literature documents lead-lag relationships at daily, weekly, and monthly horizons. Lo and MacKinlay (1990) and DeMiguel, Nogales, and Uppal (2014) show that large-stock returns predict future small-stock returns, while Brennan, Jegadeesh, and Swaminathan (1993) and Chordia and Swaminathan (2000) demonstrate that analyst coverage and trading volume, in addition to size, shape cross-sectional lead-lag effects.

Why do lead-lag patterns not get traded away? A leading explanation is that limited information-processing capacity causes delayed price reactions to information embedded in lagged returns. Hong and Stein (1999) formalize this mechanism in a single-asset model in which information diffuses gradually across investors, prices underreact to fundamentals, and returns exhibit predictability. This slow-diffusion mechanism aligns with our evidence that predictability and profitability generally decline with the forecasting horizon (Figures 2 and 5).

Hong, Torous, and Valkanov (2007) extend this framework to a cross-asset setting and argue that frictions arising from limited market participation (Merton (1987)) generate lead-lag effects across assets. In their model, investors receive timely information about only a subset of stocks and observe information about other assets with a lag, causing news to diffuse slowly across markets. Empirically, Hong, Torous, and Valkanov (2007) show that industry portfolio returns predict future market returns at horizons of up to two months. Their theoretical framework naturally extends to predictability among sector portfolios and individual stocks, with stronger effects expected at intraday horizons where inattentive traders trade infrequently (Bogousslavsky (2016); Duffie (2010)) and information has less time to be fully processed.

Why does this cross-asset predictability not get arbitraged away in equilibrium? In Hong, Torous, and Valkanov (2007), investor inattention arising from limited information-

processing capacity, combined with limits to arbitrage (Shleifer and Vishny (1997); Pontiff (2006)), prevents investors from fully incorporating publicly available information, so prices absorb information from lagged returns of other assets only gradually. This section tests whether return predictability increases with information-processing delays due to these two frictions: limited investor attention and arbitrage costs.

A related question is what type of information drives the predictability. In intraday event studies, Chinco, Clark-Joseph, and Ye (2019) and Aleti, Bollerslev, and Siggaard (2025) show that intraday predictability of individual stocks and the market, respectively, responds strongly to firm-specific and macroeconomic news. However, while news often generates short-term predictability, it may also raise investor attention and accelerate price discovery, potentially weakening predictability outside announcement windows. We examine whether macroeconomic and firm-specific news explains intraday return predictability across the trading day, and whether predictability persists in the absence of observable news.

4.1 Frictions: Inattention and Limits to Arbitrage

In this section, we examine whether intraday return predictability reflects investor inattention and limits to arbitrage, two frictions that prevent public information from being incorporated into prices. The investor inattention hypothesis posits that limited cognitive capacity prevents investors from immediately processing all publicly available information (Barberis (2018)). As a result, information is incorporated into prices with a delay, implying stronger return predictability for stocks that receive less attention. A large empirical literature supports this view, documenting that investor inattention leads to underreaction to public information about firm fundamentals.[14]

[14]For example, Atilgan, Bali, Demirtas, and Gunaydin (2020), Bali, Peng, Shen, and Tang (2014), Bali, Beckmeyer, Moerke, and Weigert (2023), Barber and Odean (2008), Ben-Rephael, Carlin, Da, and Israelsen (2021), Cohen and Frazzini (2008), Da, Gurun, and Warachka (2014), DellaVigna and Pollet (2007), DellaVigna and Pollet (2009), Giglio and Shue (2014), Han, Hirshleifer, and Walden (2023), Hirshleifer and Teoh (2003), Hirshleifer, Hou, Teoh, and Zhang (2004), Hirshleifer, Lim, and Teoh (2009), Hirshleifer, Hsu, and Li (2013), Hou and Moskowitz (2005), Huberman and Regev (2001), Jiang, Liu, Peng, and Wang (2022), Liu, Peng, and Tang (2023), Liu (2025), Lou (2014), and Peng (2005).

Investor inattention features prominently in theories of intraday predictability. Duffie (2010) develops a model in which slow-moving capital driven by inattention generates return reversals, and Bogousslavsky (2016) extends this framework to explain regular intraday predictability patterns. Hendershott, Menkveld, Praz, and Seasholes (2022) extend the Duffie (2010) model to account for deviations from efficient prices across horizons, while Hendershott and Menkveld (2014) provide empirical evidence that price pressure partially arises from slow-moving capital and investor inattention. Limits to arbitrage (Shleifer and Vishny (1997); Pontiff (2006)) reinforce these effects by preventing investors from fully exploiting mispricings. Together, these mechanisms yield testable implications: intraday predictability should be increasing in both inattention and arbitrage costs, in both the time-series and cross-sectional dimensions.

4.1.1 Time-of-day Patterns

Beginning with the portfolio evidence, when are stocks most and least predictable within the trading day? If inattention is a source of intraday predictability, then we would expect the R^2_{OS} to be highest during the middle of the day, when traders are the least active, and lowest at the beginning and end of each day, when traders are the most active. Figure 7 presents, for each portfolio, the R^2_{OS} in each half-hour interval from 9:30am to 4:00pm, achieved by the total ensemble at the 1-min horizon. There is a striking commonality of R^2_{OS} patterns across portfolios within the day. For nearly every portfolio, the R^2_{OS} exhibits a "bell" shape, where the predictability is higher during the middle of the day, and lower in the first and last 30-min of each day. These results are in line with the well-known "U" shape exhibited by intraday trading volume, and support the hypothesis that inattention is a source of intraday predictability.

The main exceptions are the market and energy portfolios. Table 1 shows that the market portfolio has by far the lowest R^2_{OS} values, with MED achieving only 0.01%, and the energy portfolio has the second lowest (MED achieves 0.19%). Both portfolios rank

among the highest in average daily dollar trading volume. While trading volume is not the sole determinant of predictability, the weak performance of these high-volume portfolios is broadly consistent with the inattention hypothesis: portfolios attracting greater investor attention should, all else equal, exhibit less inattention-driven predictability. This aligns with the findings of Hou (2007), who shows that lead-lag effects are heterogeneous across sectors, with sectors receiving lower attention exhibiting stronger lead-lag effects.

While the lower predictability of high-attention portfolios is expected, their low signal-to-noise ratios also make the remaining predictability harder to estimate precisely. Huddleston, Liu, and Stentoft (2023) document a bell-shaped pattern in 5-min intraday market predictability, but they use expanding estimation windows that sometimes contain more than a decade of training data. Huddleston, Liu, and Stentoft (2023) show that a short rolling window of 10 months is not sufficient to accurately predict intraday market returns, and that an expanding window substantially improves R^2_{OS} values. Based on this evidence, the market and energy portfolios likely have lower signal-to-noise ratios than other portfolios due to their high attention, and therefore require longer estimation windows to reveal the bell-shaped predictability pattern.[15]

4.1.2 Influential Characteristics for Intraday Predictability

We now use our proposed method for identifying which firm characteristics are important for predicting each portfolio, using the R^2_{OS} as the natural measure of statistical importance. That is, we measure variable importance as the reduction in R^2_{OS} from setting all values of predictor j to zero, while holding the remaining model estimates fixed. We conduct monthly Fama-MacBeth regressions of the standardized variable importance on each standardized contemporaneous firm characteristic. The averaged monthly coefficients along with their Newey-West t-statistic values are reported as measures of characteristic importance.

[15]Internet Appendix Figure IA1 shows, using a monthly re-estimation timescale with a 60-month training sample, that the market and energy portfolios exhibit a pattern closer to the bell shape, though still noisy. An expanding window would likely improve results further, but is infeasible given the memory and computational constraints of working with more than a billion 1-min observations.

Figure 8 presents characteristic importances for the total ensemble model (MED) at the 1-min horizon. Liquidity variables emerge as the most important characteristics across all portfolios. Firm size (mvel1) and dollar volume (dolvol) rank as the most influential characteristics, with positive and statistically significant coefficients at the 1% level for all portfolios, indicating that lagged returns of larger firms with higher trading volume carry greater predictive power. Illiquidity (ill) ranks among the top three characteristics and is statistically significant at the 1% level for all portfolios; its negative sign indicates that lagged returns of more liquid firms are more informative. Turnover (turn), defined as shares traded as a fraction of shares outstanding, and turnover volatility (std_turn) are positive and statistically significant for the market portfolio. These results indicate that lagged returns of larger, more actively traded, and more liquid firms carry the most predictive power across all portfolios.

In summary, the most influential predictors are the lagged returns of large, liquid, and actively traded firms. These findings align with the lead-lag literature, which shows that large stocks lead small stocks (DeMiguel, Nogales, and Uppal (2014); Lo and MacKinlay (1990)) and high-volume stocks lead low-volume stocks (Chordia and Swaminathan (2000)), generating substantial short-term cross-autocorrelations in returns. These results are also consistent with the theoretical framework of Hong, Torous, and Valkanov (2007). Information quickly incorporated into large and liquid stocks, which attract more attention and face fewer limits to arbitrage, aids in forecasting the returns of stocks subject to greater inattention and arbitrage costs, as these stocks gradually absorb this information. This contrasts with the findings in Table 6 of Chinco, Clark-Joseph, and Ye (2019), who report that their variable importance measure (the characteristics of the predictors selected by LAS) is largely unrelated to size, volume, volatility, and the bid-ask spread.

4.1.3 Characteristics Affecting Stock Return Predictability

The individual stocks we examine in our study constitute the S&P500 index constituents. While these stocks represent large firms, they exhibit diverse characteristics in terms of liquidity, risk exposures, and price trends; and an obvious question is if any characteristics can explain the predictability of intraday stock returns. While Huddleston, Liu, and Stentoft (2023) have examined the role of illiquidity and volatility in intraday stock market predictability in the time-series dimension, this section provides an extensive investigation into the relationship between firm characteristics and the intraday machine learning predictability of individual stocks in the cross-sectional dimension. This question is economically important because, if the gradual information diffusion mechanism of Hong, Torous, and Valkanov (2007) is a source of intraday predictability, we would expect individual stocks with greater inattention and limits to arbitrage, such as small and illiquid firms, to be more predictable and profitable.

Table 7 presents Fama-MacBeth cross-sectional regressions of monthly firm-level outcomes (R^2_{OS} and net-of-cost returns) from the total ensemble (MED) on 20 lagged firm characteristics at the 1-min horizon. We report time-series averages of the monthly coefficients and Newey-West t-statistics. Both outcomes and characteristics are standardized cross-sectionally each month to have zero mean and unit variance, so each coefficient measures the effect of a one standard deviation change in the characteristic on the outcome in standard deviation units.

The first row presents the R^2_{OS} results. Firm size (mvel1) and dollar trading volume (dolvol) are negatively related to R^2_{OS} with t-statistics of -8.31 and -6.30, indicating that smaller firms with less trading volume are more predictable. Amihud illiquidity (ill) and price delay (pricedelay) are positively related to R^2_{OS} with t-statistics of 5.70 and 3.55, confirming that illiquid firms are more predictable. The coefficients on beta and beta squared (beta, betasq) are also positive with t-statistics of 3.48 and 3.66, suggesting that stocks with greater systematic risk are more predictable.

The second row reports results for profitability after transaction costs. The coefficients broadly align with the statistical predictability results above, though with lower t-statistics. Firm size (mvel1) and dollar trading volume levels and volatility (dolvol, std_dolvol) are negatively related to profitability with t-statistics of -3.83, -1.78, and -1.99. Amihud illiquidity (ill) and price delay (pricedelay) are positively related to profitability with t-statistics of 4.51 and 1.69.

Overall, these results align with the lead-lag literature, which shows that smaller and less liquid firms are the most predictable and profitable, and are consistent with gradual information diffusion driven by inattention and limits to arbitrage. They contrast with the results in Table 3 of Chinco, Clark-Joseph, and Ye (2019), who find that their R^2_{OS} values are largely unrelated to characteristics such as size and volume.

4.1.4 Portfolio Sorts on Firm Characteristics

The analysis in Section 4.1.3 focuses on linear cross-sectional relationships. This section investigates whether nonlinear firm-level relationships are present using portfolio sorts, providing additional evidence on the gradual information diffusion explanation. Following Hong, Lim, and Stein (2000), we proxy for the speed of information diffusion using size, analyst coverage, and trading volume. If inattention drives our results, intraday predictability should be stronger for smaller firms with lower analyst coverage and trading volume.

Furthermore, if arbitrage frictions also slow information diffusion, intraday predictability should be stronger for firms with higher arbitrage costs. To capture the multidimensional nature of arbitrage frictions, we follow Atilgan, Bali, Demirtas, and Gunaydin (2020) and construct a composite arbitrage cost from several firm-level indicators of limits to arbitrage: firm size, firm age, idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang (2006)), illiquidity (Amihud (2002)), and analyst coverage (Zhang (2006)). For each indicator, we assign stocks a decile rank, sorting in ascending order for idiosyncratic volatility and illiquidity and in descending order for size, age, and analyst coverage. The arbitrage cost for each firm is the

sum of its five decile ranks, yielding values between 5 and 50, where higher scores indicate greater limits to arbitrage.

Table 8 presents the performance of the total ensemble (MED) for individual stocks sorted into quintile portfolios by analyst coverage, size, dollar volume, and the arbitrage index, each lagged by one month. The R^2_{OS} values generally decrease across quintile portfolios for analyst coverage, size, and dollar volume, consistent with the lead-lag findings in Brennan, Jegadeesh, and Swaminathan (1993), Lo and MacKinlay (1990), and Chordia and Swaminathan (2000), respectively. The R^2_{OS} values also generally increase across the arbitrage cost quintiles, consistent with arbitrage frictions slowing information diffusion.

Table 8 also reports the quintile portfolio results for Sharpe ratios after transaction costs. For analyst coverage, Sharpe ratios generally decrease across quintiles, indicating that greater attention is associated with lower profitability. For size, the Sharpe ratios decrease monotonically from quintiles 2 through 5, but are lowest in quintile 1 containing the smallest firms. This likely reflects the fact that the smallest firms also have the highest transaction costs and volatility, which erode the Sharpe ratio despite strong gross predictability. The arbitrage index produces a similar pattern: Sharpe ratios are lower for stocks with both low and high arbitrage costs and highest for quintile 3 with moderate arbitrage costs. The lowest arbitrage cost stocks may face insufficient mispricing to exploit, while the highest arbitrage cost stocks may incur transaction costs that offset their greater predictability. For trading volume, the lowest Sharpe ratio is in quintile 5, consistent with high-volume stocks attracting more attention and exhibiting less mispricing. Quintiles 3 and 4 achieve the highest Sharpe ratios, reflecting a tradeoff between the greater predictability of low-volume stocks and the lower transaction costs of high-volume stocks.

Overall, the R^2_{OS} findings are consistent with inattention and limits to arbitrage frictions, pointing to gradual information diffusion as a key mechanism underlying our lead-lag effects. The Sharpe ratio findings are more mixed because transaction costs offset the greater predictability in stocks with the highest arbitrage costs, consistent with the prediction that

predictability persists where it is most costly to exploit.

4.1.5 Re-estimation Timescales and Limited Computational Capacity

While the attention literature focuses primarily on investors' limited processing capacity due to cognitive constraints, little work examines limitations arising from computing capacity. Many intraday trading algorithms rely on computationally intensive models, yet computing resources are finite and exhaustively searching over forecasting models using public information is costly (Hirshleifer and Teoh (2003)).[16] We study computing capacity constraints by varying re-estimation timescales and comparing performance and estimation times across models, with the implication from attention theory that greater computing requirements slow information diffusion and strengthen predictability.

To test this, we consider three additional training configurations, holding the test set fixed over the 17 years from 2005 to 2021. Under a yearly re-estimation timescale, models initially train on five years of data (1999 to 2003), validate on one year (2004), and evaluate on a one-year test set (2005); we then roll the window forward yearly. Under a daily re-estimation timescale, models initially train on 20 days of data (December 2, 2004 to December 30, 2004), validate on one day (December 31, 2004), and evaluate on a one-day test set (January 3, 2005); we then roll the window forward daily. Finally, inspired by Chinco, Clark-Joseph, and Ye (2019), we consider a minute re-estimation timescale in which models initially train on 30 minutes of data (9:32 to 10:01), validate via ten-fold cross-validation, and evaluate on a 1-min test set (10:02); we then roll the window forward minutely for 359 minutes each day. Due to computational constraints, we restrict this analysis to portfolios at the 1-min horizon, and minutely re-estimation includes only AR1, OLS, RID, LAS, and EN.

This design also tests the trade-off between estimation window length and model stal-

[16]For example, the computations for this paper required more than two years for the initial draft and an additional 2.5 years for the revision, conducted primarily on a dedicated high-performance computing cluster with 36 nodes, each containing 24 cores, totaling approximately 3,888 core-years at full capacity ($36 \times 24 \times 4.5 = 3,888$). Using a conservative on-demand core-hour rate of \$0.03 or core-year rate of \$262.80 ($0.03 \times 8,760 = 262.8$) from AWS (https://aws.amazon.com/ec2/pricing/on-demand/), the equivalent cloud computing cost is approximately \$1,021,766 USD ($262.8 \times 3,888 \approx 1,021,766$).

eness. While machine learning models benefit from larger training samples, parameters can drift as market conditions change, favoring more frequent re-estimation with shorter windows. In the extreme, Chinco, Clark-Joseph, and Ye (2019) show that LAS-based predictability is short-lived and news-driven, implying that performance should decay sharply as the re-estimation timescale lengthens. In contrast, if predictability reflects slow information diffusion arising from frictions such as inattention and arbitrage costs, performance should remain relatively stable across re-estimation timescales. Nonlinear models may even benefit from longer timescales, as larger training samples improve the estimation of complex functional forms.

Re-estimation performance. Table 9 presents the results for the different re-estimation timescales, including monthly from the baseline analysis. We report results for MED; LAS, the best-performing model in Chinco, Clark-Joseph, and Ye (2019), as a representative sparse linear model; and NN3, the best-performing model in Gu, Kelly, and Xiu (2020), as a representative nonlinear model.[17] Panel A shows that the R^2_{OS} values for MED generally increase as the re-estimation timescale shortens for all sector portfolios, but show the opposite pattern for the market portfolio. This is consistent with the finding in Huddleston, Liu, and Stentoft (2023) that the market portfolio requires a longer estimation window due to its low signal-to-noise ratio. A similar pattern emerges for NN3. For LAS, however, the monthly re-estimation timescale produces the highest R^2_{OS} for nine of ten portfolios. Minutely re-estimation for LAS produces sharply negative R^2_{OS} values for portfolios, which differs from the strong performance of minutely re-estimation for individual stocks in Chinco, Clark-Joseph, and Ye (2019).[18] A likely explanation is that this analysis uses portfolios, whereas Chinco, Clark-Joseph, and Ye (2019) study individual stocks, and Section 3.4 finds that the functional forms differ substantially between the two settings.

[17] Internet Appendix Tables IA3 through IA5 report complete results for all models under the yearly, daily, and minutely re-estimation timescales.

[18] Our analysis does not contest that minutely re-estimation performs well for individual stocks as in Chinco, Clark-Joseph, and Ye (2019), but rather demonstrates that this re-estimation timescale is unlikely to explain the intraday predictability of portfolios.

Panel B of Table 9 reports annualized Sharpe ratios after transaction costs across re-estimation timescales and models. In contrast to Panel A, yearly re-estimation produces the highest Sharpe ratios for eight of ten portfolios under MED and nine of ten portfolios under NN3. For LAS, monthly re-estimation produces the highest Sharpe ratio in six of ten portfolios. Minutely re-estimation produces negative Sharpe ratios across all portfolios for LAS, again inconsistent with the findings of Chinco, Clark-Joseph, and Ye (2019) for individual stocks. Across models, NN3 under yearly re-estimation outperforms LAS under any re-estimation timescale, highlighting the benefits of nonlinearity for portfolios.

Overall, these portfolio results support the slow information diffusion explanation. Performance is relatively stable across yearly, monthly, and daily re-estimation timescales, and economic significance is often highest under yearly re-estimation. This indicates that the signals are persistent rather than quickly becoming stale, and that the larger training samples under yearly re-estimation benefit the nonlinear models. It is unlikely that intraday stock portfolio predictability is driven by short-lived signals, as minutely re-estimation consistently performs poorly.

Estimation times. We further study limited computational capacity by reporting average estimation times for each model across re-estimation timescales. Table 10 reports training time per re-estimation to assess whether each model can feasibly be estimated in time to generate predictions, and annualized training time to quantify the total computing resources required for each timescale. Panel A reports the training time for the market portfolio in seconds per re-estimation, along with the total training time in both seconds and hours. Yearly re-estimation with a five-year training set takes the longest at 20.3 hours per re-estimation, meaning the models can be trained over a weekend once a year. The monthly and daily timescales require 4.69 and 1.39 hours per re-estimation, respectively, allowing models to be trained overnight after markets close. For the minute timescale, LAS requires only 4 seconds per re-estimation, which is feasible, but the EN re-estimation of 22 seconds

may cut significantly into the next-minute prediction window.

Panel B reports the annualized training time for the market portfolio, where all training times are standardized to a single year of predictions to compare computing costs. Because yearly re-estimation occurs only once per year, it has the lowest annualized training time. Monthly re-estimation requires 56.28 hours per year, while daily re-estimation incurs a substantially larger 350.28 hours. The minute timescale has by far the largest annualized time at 980.28 hours, incurring substantial computing costs.

Panel C presents the training time per re-estimation for the 5-min horizon with monthly re-estimation across sectors. The 5-min horizon is faster to estimate, requiring 1.27 hours per re-estimation for the market portfolio compared to 4.69 hours at the 1-min horizon in Panel A. The sector portfolios are faster to estimate than the market due to the smaller number of predictors.

Comparing estimation times across models, all panels show that OLS and penalized linear models are the fastest to estimate, whereas RF and XRF take by far the longest. PCR and GBT require moderate estimation time, generally comparable to or exceeding the training time of individual neural networks. Training times for neural networks reveal that deeper networks converge faster than shallower ones due to earlier activation of early stopping. This pattern suggests that deeper architectures learn more efficient representations of intraday return predictability, requiring fewer passes through the data than shallower networks that must approximate complex nonlinear relationships through fewer hidden layers.

The results in this section do not support limited computational capacity as a source of intraday portfolio return predictability.[19] Panel B of Table 9 shows that yearly re-estimation generally achieves the highest economic performance for MED despite requiring the least

[19]Timmermann and Vulicevic (2026) document that greater computing expenditure improves monthly cross-sectional return predictability, following stable scaling relationships. We measure total computing time across re-estimation schemes and reach a different conclusion at intraday horizons. This divergence arises because more frequent re-estimation raises computing time by refitting models on progressively shorter windows, which increases estimation noise rather than extracting additional signal. The returns to computing therefore depend on how they are deployed, and allocating computing to re-estimation frequency offers no improvement at intraday horizons.

annualized computing time (Panel B of Table 10), while the minute timescale performs worst despite consuming the most resources. Neural networks perform well despite requiring substantially less training time than RF and XRF. LAS is one of the fastest models to train and achieves the best economic performance under monthly re-estimation (Panel B of Table 9), though it consistently underperforms NN3 under yearly re-estimation. These findings suggest that the signals driving intraday portfolio predictability are sufficiently long-lived that they do not require computationally expensive, frequent model updating to exploit, and that they are complex enough to require longer training windows. This supports the slow information diffusion explanation and contrasts with the short-lived predictability explanation from Chinco, Clark-Joseph, and Ye (2019).

Is the virtue of complexity in intraday predictability robust across re-estimation timescales? Internet Appendix Figures IA2 and IA3 present the Sharpe ratios after transaction costs of the L-MED, NL-MED, and MED models under the yearly and daily re-estimation timescales, respectively. Under yearly re-estimation, the virtue of complexity result is stronger than in the baseline analysis (Figure 4), with NL-MED significantly outperforming L-MED in nine of ten portfolios, as the longer training samples allow the nonlinear models to be estimated more precisely. Under daily re-estimation, NL-MED significantly outperforms L-MED in four portfolios, reflecting the smaller training window available to nonlinear models, consistent with the finding in Kelly, Malamud, and Zhou (2024) that complex models achieve lower overall performance levels in data-scarce environments, but still improve over linear models. The virtue of complexity in intraday stock predictability thus holds across re-estimation timescales.

4.2 News

The previous section presents evidence that gradual information diffusion drives intraday stock predictability. A related question is what type of information is being diffused. Chinco, Clark-Joseph, and Ye (2019) find that intraday individual stock predictability is linked to

economically meaningful news, showing that LAS tends to identify as predictors the lagged returns of stocks with news about firm fundamentals. Similarly, Aleti, Bollerslev, and Siggaard (2025) find that intraday market predictability responds to surprise macroeconomic events. These results are based on event studies around short intraday intervals.

From these findings, one might conclude that most intraday stock predictability is driven by observable news. However, while news generates short-term predictability around announcements, it may also raise aggregate investor attention across the trading day, accelerating price discovery and potentially weakening predictability outside the announcement window. Macroeconomic and firm-specific news can both raise attention: Fisher, Martineau, and Sheng (2022) show that attention to macroeconomic news rises sharply around scheduled announcements; Ben-Rephael, Da, and Israelsen (2017) find that institutional attention increases around firm-specific news coverage, earnings announcements, and analyst recommendation changes, and that post-announcement drift following the latter two events is lower in high-attention cases, consistent with attention reducing return predictability. The net effect of news on predictability across the entire trading day therefore remains an open question. This section examines whether publicly observable macroeconomic or firm-specific news explains intraday predictability across the trading day, or whether predictability persists in the absence of observable news and even weakens when news raises attention. We also exploit a natural experiment, the Knight Capital trading glitch, to further examine the relationship between news and intraday predictability.

4.2.1 Macroeconomic News

Savor and Wilson (2013) demonstrate that stock market returns are higher on days with scheduled macroeconomic announcements about inflation, unemployment, or interest rates. In the intraday setting, macroeconomic news may also generate return predictability if new information affects certain stocks first and then diffuses slowly to related stocks or portfolios. Along these lines, Gao, Han, Li, and Zhou (2018) find that intraday market momentum is

stronger on macroeconomic announcement days. Aleti, Bollerslev, and Siggaard (2025) show that intraday market predictability responds to surprise macroeconomic events.

Following Savor and Wilson (2013), we define macroeconomic announcement days as days with scheduled releases of the Consumer Price Index (CPI) or Producer Price Index (PPI), the employment report (Employment) from the Bureau of Labor Statistics, or Federal Reserve interest rate decisions (FOMC), excluding unscheduled meetings. We augment this set with the announcements in Gao, Han, Li, and Zhou (2018), which include the Michigan Consumer Sentiment Index (MCSI), Gross Domestic Product (GDP) release dates from the Bureau of Economic Analysis, and FOMC minutes (Fed Minutes) release dates from the Federal Reserve. We also report results for all macroeconomic announcement days combined (Any Macro).

Table 11 compares the intraday stock predictability of the MED model on announcement versus non-announcement days. Panel A compares the R^2_{OS} values. The R^2_{OS} values are broadly similar between announcement and non-announcement days across portfolios and individual stocks. Panel B compares the Sharpe ratios after transaction costs. The Sharpe ratios are also similar between the two types of days. In fact, in the majority of cases, the Sharpe ratio on non-announcement days exceeds that on announcement days. These results suggest that macroeconomic news does not explain intraday stock predictability, and that the additional attention on announcement days may reduce the economic performance of intraday strategies.

4.2.2 Firm-specific Information Events

This section studies firm-specific information events. We examine how news coverage, analyst revisions, and earnings releases affect the performance of the MED model.

News coverage. Chinco, Clark-Joseph, and Ye (2019) argue that intraday return predictability in their setting is largely driven by firm-specific news, showing that LAS tends to

identify as predictors the lagged returns of stocks with news about firm fundamentals. Jiang, Li, and Wang (2021) document that markets underreact to firm-specific news, generating return drift following news arrivals. However, firm-specific news may also increase investor attention, which could dampen predictability.

Identifying firm-specific news relevant to fundamental value is challenging. Boudoukh, Feldman, Kogan, and Richardson (2019) develop a textual analysis algorithm to identify fundamental information in Dow Jones Newswires and demonstrate that relevant news events explain a large fraction of firm-level idiosyncratic volatility. We adopt their dataset to identify firm-specific news days.[20] The Boudoukh, Feldman, Kogan, and Richardson (2019) dataset classifies each firm-day observation into three categories: no news, unidentified news (days on which none of the news coverage is identified as relevant for firm valuation), and identified news (days on which at least one article is identified as relevant). They further partition news coverage across three windows: trading hours (open to close), overnight (previous close to market open), and the combination of both. Their sample ends in December 2015, which covers the sample period in Chinco, Clark-Joseph, and Ye (2019), ending in 2012.

Panels A, B, and C in Table 12 report R^2_{OS} values and Sharpe ratios after transaction costs for trading hours news, overnight news, and both, respectively. In each case, a monotonic pattern emerges: days with no news achieve the highest R^2_{OS} values and Sharpe ratios, days with unidentified news rank second, and days with identified news rank lowest. These results suggest that firm-specific news does not explain intraday stock predictability, and that attention associated with news coverage may reduce the economic performance of intraday strategies. These findings complement Chinco, Clark-Joseph, and Ye (2019), who show that the LAS predictor selection rate increases in a narrow window around news events, by demonstrating what happens over the entire news day. Together, these results suggest that while news temporarily increases predictability in a short window around its arrival, it

[20]We use their dataset over proprietary alternatives such as RavenPack for three reasons. First, their algorithm is transparent, whereas RavenPack relies on a proprietary machine learning model. Second, they demonstrate that their algorithm produces qualitatively similar results to RavenPack. Third, their dataset is freely available for replication, whereas RavenPack requires an expensive license.

reduces predictability over the full trading day by raising investor attention.

Analyst forecast revisions. The previous analysis relies on news coverage, which has two limitations. First, news coverage may not capture all information relevant to a firm. Second, news coverage does not identify whether the information is positive or negative. We therefore follow Zhang (2006) and examine an alternative measure of firm-specific information based on analyst forecast revisions in the previous month. Analysts synthesize information beyond newspaper coverage and incorporate it into earnings forecasts, so revisions reflect economically meaningful firm-level news.

We define analyst forecast revisions as the one-month change in the median analyst earnings forecast, lagged one month. Following Zhang (2006), we classify positive revisions as good news, negative revisions as bad news, and zero revisions as no news. Panel D in Table 12 reports the R^2_{OS} values and Sharpe ratios after transaction costs conditional on the previous month's analyst revision. Firms with no revision achieve the highest R^2_{OS} values and Sharpe ratios, consistent with the inattention explanation. Firms with positive revisions have only slightly lower R^2_{OS} values and Sharpe ratios than those with no revision, whereas firms with negative revisions have the lowest. This asymmetry suggests that negative analyst revisions generate greater and more persistent investor scrutiny, reducing inattention-driven predictability in the following month.

Earnings announcements. Engelberg, McLean, and Pontiff (2018) show that anomaly returns are six times higher on earnings announcement days. They find that the results are driven by biased expectations about future cash flows, which are at least partly corrected upon news arrival. Hence, we study whether earnings announcements can explain intraday predictability of individual stocks. To evaluate the impact of earnings announcements, each day we partition stocks into two subsamples: those with an earnings announcement before or during the trading day, and all remaining stocks.[21]

[21] If the earnings announcement occurs after market close, we assign it to the next trading day.

Panel E in Table 12 reports the R^2_{OS} values and Sharpe ratios after transaction costs on earnings announcement days versus non-announcement days. The results are broadly similar, though non-announcement days have slightly higher R^2_{OS} values and Sharpe ratios. These results suggest that earnings announcements do not explain intraday stock predictability, and that the increased attention on earnings days may even slightly reduce intraday predictability.

4.2.3 Evidence from a Major Trading Glitch

We further examine the relationship between news and intraday predictability by exploiting a natural experiment: the Knight Capital trading glitch of August 1, 2012. Knight Capital, at the time one of the largest market makers in U.S. equity markets, experienced a software malfunction that flooded the market with millions of unintended orders during the first half hour of trading. The episode left the firm holding long and short positions across dozens of stocks with a combined notional value of roughly $6.5 billion (see Bogousslavsky, Collin-Dufresne, and Sağlam (2021) for a comprehensive analysis). Stocks that Knight Capital inadvertently accumulated (sold) experienced sharp price increases (declines), which subsequently reversed once the nature of the glitch became widely known. This event provides a setting to study how unexpected news affects intraday predictability. News of the trading glitch first appeared around 10:00 AM and became widely known by approximately 10:30 AM. Before 10:00 AM, Knight Capital's erroneous orders generated substantial mispricing in affected stocks, creating price movements outside the range of patterns the machine learning models were trained on.

We compare intraday predictability of affected and unaffected stocks on the event day versus the remainder of August (August 2 to 31). Table 13 reports R^2_{OS} and Sharpe ratios after transaction costs for the L-MED, NL-MED, and MED models on August 1 compared to the rest of the month.[22] Panel A shows that the R^2_{OS} of affected stocks was substantially lower than that of unaffected stocks on the event day, but that affected stocks had higher R^2_{OS} values

[22]Sharpe ratios in this table are computed from 1-min returns and are not annualized, as they are not directly comparable to the annualized Sharpe ratios in Table 3, which are based on monthly excess returns.

during the rest of the month. Panel B shows that the Sharpe ratios of affected stocks were much lower on the event day, but comparable to those of unaffected stocks during the rest of the month. Affected and unaffected stocks thus exhibit similar intraday predictability under normal conditions, but the trading glitch substantially reduced predictability for affected stocks on the event day. This is consistent with the pattern documented in the preceding subsections: news raises attention and reduces predictability over the trading day. Consistent with this attention channel, Bogousslavsky, Collin-Dufresne, and Sağlam (2021) report that affected stocks experienced abnormally high volume that day, with some experiencing more share volume than the SPY ETF, an occurrence they verify is extremely rare and indicative of the heightened attention on that day.

Figure 9 examines the intraday dynamics during the event day in 30-minute intervals. During the 9:30 to 10:00 AM period, L-MED produces sharply negative R^2_{OS} and Sharpe ratio values for affected firms (Panels (a) and (c)), reflecting the extreme price dislocations caused by the glitch, while NL-MED and MED remain near zero. Between 10:00 AM and 10:30 AM, as news of the glitch began to diffuse, NL-MED and MED achieve positive Sharpe ratios for affected firms, while L-MED remains negative, suggesting that the price dynamics were sufficiently complex that only nonlinear models could capture them. For the remainder of the day, the estimates are noisy given that they are based on a single trading day, though the R^2_{OS} values for L-MED are mostly negative for affected firms but mostly positive for unaffected firms.

4.3 Discussion of Economic Source

This section investigates the economic sources of intraday stock predictability. Motivated by the finding in Sections 3.1 and 3.3 that predictability and profitability generally decrease with the time horizon, we hypothesize that intraday stock predictability arises from gradual information diffusion caused by frictions such as inattention and limits to arbitrage, generating lead-lag effects in intraday returns. We present multiple lines of evidence for this

explanation. We document a striking commonality of predictability patterns within the trading day across portfolios (Section 4.1.1), show that firms subject to greater inattention and arbitrage costs are more predictable (Sections 4.1.3 and 4.1.4), and find that large, liquid, and actively traded firms are the most influential predictors (Section 4.1.2). Taken together, these results point to a pecking order in intraday price discovery, in which information is incorporated first into the prices of large, liquid, and actively traded stocks and diffuses gradually to stocks facing greater inattention and arbitrage costs.

Section 4.1.5 also considers several re-estimation timescales ranging from yearly to minutely, with training samples ranging from five years to 30 minutes, applied to the stock portfolios. The results are relatively stable across re-estimation timescales of daily or longer, and economic significance increases with longer timescales, reaching its peak under yearly re-estimation, consistent with improved estimation from larger training samples. The virtue of complexity result strengthens as the re-estimation timescale increases. Under minutely re-estimation, however, results are consistently negative, in contrast to the findings of Chinco, Clark-Joseph, and Ye (2019) for individual stocks, indicating that the functional forms for stock portfolios and individual stocks differ substantially. Furthermore, yearly re-estimation requires the least annualized computational resources yet achieves the highest economic performance, while minutely re-estimation is the most computationally demanding yet performs worst, suggesting that computational constraints do not drive intraday predictability. Overall, these results support the view that gradual information diffusion is a systematic feature of financial markets arising from frictions.

Section 4.2 sheds new light on the question of what information drives intraday predictability. Chinco, Clark-Joseph, and Ye (2019) and Aleti, Bollerslev, and Siggaard (2025) use short event studies to show that intraday predictability of individual stocks and the market, respectively, responds strongly to firm-specific and macroeconomic news, potentially giving the impression that intraday predictability is primarily driven by public news. However, they do not capture the effect that news and the subsequent attention have on the

full trading day, which we find on average lowers intraday predictability. A natural question then arises: why is there such strong intraday predictability in the absence of news? This can be understood under the framework of Hong and Stein (1999), where information about fundamentals is initially private and gradually diffuses across the population of investors. Hong, Torous, and Valkanov (2007) extend this framework to explain lead-lag effects, where cross-autocorrelations arise from information processing constraints and limits to arbitrage. Taken together, these findings are consistent with intraday predictability being partly driven by private information that slowly diffuses into prices, in line with the well-documented fact that observable news explains only a small fraction of stock price variation (French and Roll (1986); Roll (1988)). However, more work is needed to formally investigate this channel.

5 Conclusion

A rapidly expanding literature studies the predictability of stock, bond, option, currency, mutual fund, and hedge fund returns at the monthly horizon or longer. However, research on intraday predictability is relatively scarce and its economic sources are not well understood. This paper conducts, to our knowledge, the largest study to date of intraday stock predictability using machine learning techniques, bridging the gap between studies that focus on the monthly horizon or above and those that focus on high-frequency horizons below a second. Specifically, we study the out-of-sample predictability of market and sector portfolio returns at the 1-min, 5-min, 10-min, 15-min, and 30-min horizons, as well as individual stock returns at the 1-min horizon, from January 2005 to December 2021. We also evaluate the economic significance of trading strategies based on machine learning predictions, both before and after transaction costs.

This paper advances the understanding of intraday stock return predictability in several ways. We show that stock portfolios and individual stocks are consistently predictable, and that trading strategies based on these predictions produce economically significant profits

after transaction costs. For portfolios, nonlinear models dominate linear models in profitability after transaction costs despite weaker statistical performance, consistent with the virtue of complexity framework of Kelly, Malamud, and Zhou (2024). For individual stocks, sparse linear models outperform most nonlinear models, consistent with a sparse functional form as documented in Chinco, Clark-Joseph, and Ye (2019).

We thoroughly investigate the economic sources of intraday stock predictability. We document that predictability and profitability generally decrease as the time horizon increases. Furthermore, we show a striking commonality in predictability patterns across portfolios within the trading day, where predictability is higher (lower) when traders are least (most) active, demonstrate that smaller and more illiquid firms with lower analyst coverage are more predictable and profitable, and find that large, liquid, and actively traded firms are the most influential predictors. Finally, we find that intraday predictability persists in the absence of macroeconomic and firm-specific news and is generally lower on news days, consistent with news raising investor attention and accelerating price discovery. Taken together, these results suggest that intraday predictability arises from gradual information diffusion caused by frictions such as inattention and limits to arbitrage, which generates lead-lag effects in intraday returns.

Our findings have implications for investors such as quantitative hedge funds that employ intraday algorithmic strategies. We demonstrate that nonlinear models, and in particular neural networks, achieve the strongest economic performance after transaction costs despite weaker statistical predictive power than linear models. These results provide a transparent academic benchmark for the profitability available from intraday machine learning predictions and offer new insights into how frictions slow the incorporation of information into stock prices and what types of information drive intraday predictability.

References

Aït-Sahalia, Y., J. Fan, L. Xue, and X. Zhu (2025): "How and When are High-Frequency Stock Returns Predictable?," *Management Science*, 0(0).

Aleti, S., and T. Bollerslev (2025): "News and Asset Pricing: A High-Frequency Anatomy of the SDF," *The Review of Financial Studies*, 38(3), 712–759.

Aleti, S., T. Bollerslev, and M. Siggaard (2025): "Intraday Market Return Predictability Culled from the Factor Zoo," *Management Science*, 71(9), 7731–7751.

Amihud, Y. (2002): "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of Financial Markets*, 5(1), 31–56.

Ang, A., and G. Bekaert (2007): "Stock Return Predictability: Is It There?," *The Review of Financial Studies*, 20(3), 651–707.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *The Journal of Finance*, 61(1), 259–299.

Atilgan, Y., T. G. Bali, K. O. Demirtas, and A. D. Gunaydin (2020): "Left-Tail Momentum: Underreaction to Bad News, Costly Arbitrage and Equity Returns," *Journal of Financial Economics*, 135(3), 725–753.

Bali, T. G., H. Beckmeyer, M. Moerke, and F. Weigert (2023): "Option Return Predictability with Machine Learning and Big Data," *The Review of Financial Studies*, 36(9), 3548–3602.

Bali, T. G., L. Peng, Y. Shen, and Y. Tang (2014): "Liquidity Shocks and Stock Market Reactions," *The Review of Financial Studies*, 27(5), 1434–1485.

Barber, B. M., and T. Odean (2008): "All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *The Review of Financial Studies*, 21(2), 785–818.

Barberis, N. (2018): "Psychology-Based Models of Asset Prices and Trading Volume," in *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 1, pp. 79–175. Elsevier.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009): "Realized Kernels in Practice: Trades and Quotes," *Econometrics Journal*, 12, C1–C32.

Bartlett, R. P., J. McCrary, and M. O'Hara (2025): "The Market Inside the Market: Odd-Lot Quotes," *The Review of Financial Studies*, 38(3), 661–711.

Ben-Rephael, A., B. I. Carlin, Z. Da, and R. D. Israelsen (2021): "Information Consumption and Asset Pricing," *The Journal of Finance*, 76(1), 357–394.

Ben-Rephael, A., Z. Da, and R. D. Israelsen (2017): "It Depends on Where You Search: Institutional Investor Attention and Underreaction to News," *The Review of Financial Studies*, 30(9), 3009–3047.

Bergstra, J., and Y. Bengio (2012): "Random Search for Hyper-Parameter Optimization," *The Journal of Machine Learning Research*, 13(1), 281–305.

Bianchi, D., M. Büchner, and A. Tamoni (2021): "Bond Risk Premia with Machine Learning," *The Review of Financial Studies*, 34, 1046–1089.

Bogousslavsky, V. (2016): "Infrequent Rebalancing, Return Autocorrelation, and Seasonality," *The Journal of Finance*, 71(6), 2967–3006.

Bogousslavsky, V., P. Collin-Dufresne, and M. Sağlam (2021): "Slow-Moving Capital and Execution Costs: Evidence from a Major Trading Glitch," *Journal of Financial Economics*, 139(3), 922–949.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson (2019): "Information, Trading, and Volatility: Evidence from Firm-Specific News," *The Review of Financial Studies*, 32(3), 992–1033.

Brennan, M. J., N. Jegadeesh, and B. Swaminathan (1993): "Investment Analysis and the Adjustment of Stock Prices to Common Information," *The Review of Financial Studies*, 6(4), 799–824.

Bryzgalova, S., M. Pelger, and J. Zhu (2025): "Forest Through the Trees: Building Cross-Sections of Stock Returns," *The Journal of Finance*, 80(5), 2447–2506.

Campbell, J. Y., and S. B. Thompson (2008): "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?," *The Review of Financial Studies*, 21(4), 1509–1531.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *The Journal of Finance*, 52(1), 57–82.

Chen, A. Y., and M. Velikov (2023): "Zeroing in on the Expected Returns of Anomalies," *Journal of Financial and Quantitative Analysis*, 58(3), 968–1004.

Chen, L., Z. Da, and X. Zhao (2013): "What Drives Stock Price Movements?," *The Review of Financial Studies*, 26(4), 841–876.

Chen, L., M. Pelger, and J. Zhu (2023): "Deep Learning in Asset Pricing," *Management Science*, 70(2), 714–750.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019): "Sparse Signals in the Cross-Section of Returns," *The Journal of Finance*, 74(1), 449–492.

Chordia, T., R. Roll, and A. Subrahmanyam (2005): "Evidence on the Speed of Convergence to Market Efficiency," *Journal of Financial Economics*, 76(2), 271–292.

——— (2008): "Liquidity and Market Efficiency," *Journal of Financial Economics*, 87(2), 249–268.

——— (2011): "Recent Trends in Trading Activity and Market Quality," *Journal of Financial Economics*, 101(2), 243–263.

Chordia, T., and B. Swaminathan (2000): "Trading Volume and Cross-Autocorrelations in Stock Returns," *The Journal of Finance*, 55(2), 913–935.

Cohen, L., and A. Frazzini (2008): "Economic Links and Predictable Returns," *The Journal of Finance*, 63(4), 1977–2011.

Da, Z., U. G. Gurun, and M. Warachka (2014): "Frog in the Pan: Continuous Information and Momentum," *The Review of Financial Studies*, 27(7), 2171–2218.

DellaVigna, S., and J. M. Pollet (2007): "Demographics and Industry Returns," *American Economic Review*, 97(5), 1667–1702.

——— (2009): "Investor Inattention and Friday Earnings Announcements," *The Journal of Finance*, 64(2), 709–749.

DeMiguel, V., F. J. Nogales, and R. Uppal (2014): "Stock Return Serial Dependence and Out-Of-Sample Portfolio Performance," *The Review of Financial Studies*, 27(4), 1031–1073.

Duffie, D. (2010): "Presidential Address: Asset Price Dynamics with Slow-Moving Capital," *The Journal of Finance*, 65(4), 1237–1267.

Easley, D., M. López de Prado, M. O'Hara, and Z. Zhang (2021): "Microstructure in the Machine Age," *The Review of Financial Studies*, 34(7), 3316–3363.

Engelberg, J., R. D. McLean, and J. Pontiff (2018): "Anomalies and News," *The Journal of Finance*, 73(5), 1971–2001.

Fama, E. F., and K. R. French (2015): "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116(1), 1–22.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

Feng, G., J. He, N. G. Polson, and J. Xu (2024): "Deep Learning in Characteristics-Sorted Factor Models," *Journal of Financial and Quantitative Analysis*, 59(7), 3001–3036.

Fisher, A., C. Martineau, and J. Sheng (2022): "Macroeconomic Attention and Announcement Risk Premia," *The Review of Financial Studies*, 35(11), 5057–5093.

French, K. R., and R. Roll (1986): "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17(1), 5–26.

Friedman, J., T. Hastie, and R. Tibshirani (2010): "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33(1), 1–22.

Gao, L., Y. Han, S. Z. Li, and G. Zhou (2018): "Market Intraday Momentum," *Journal of Financial Economics*, 129(2), 394–414.

Giglio, S., and K. Shue (2014): "No News is News: Do Markets Underreact to Nothing?," *The Review of Financial Studies*, 27(12), 3389–3440.

Goldstein, I., C. S. Spatt, and M. Ye (2021): "Big Data in Finance," *The Review of Financial Studies*, 34(7), 3213–3225.

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, 33(5), 2223–2273.

——— (2021): "Autoencoder Asset Pricing Models," *Journal of Econometrics*, 222(1), 429–450.

Han, B., D. Hirshleifer, and J. Walden (2023): "Visibility Bias in the Transmission of Consumption Beliefs and Undersaving," *The Journal of Finance*, 78(3), 1647–1704.

Hendershott, T., and A. J. Menkveld (2014): "Price Pressures," *Journal of Financial Economics*, 114(3), 405–423.

Hendershott, T., A. J. Menkveld, R. Praz, and M. Seasholes (2022): "Asset Price Dynamics with Limited Attention," *The Review of Financial Studies*, 35(2), 962–1008.

Heston, S. L., R. A. Korajczyk, and R. Sadka (2010): "Intraday Patterns in the Cross-Section of Stock Returns," *The Journal of Finance*, 65(4), 1369–1407.

Hirshleifer, D., K. Hou, S. H. Teoh, and Y. Zhang (2004): "Do Investors Overvalue Firms with Bloated Balance Sheets?," *Journal of Accounting and Economics*, 38, 297–331.

Hirshleifer, D., P.-H. Hsu, and D. Li (2013): "Innovative Efficiency and Stock Returns," *Journal of Financial Economics*, 107(3), 632–654.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009): "Driven to Distraction: Extraneous Events and Underreaction to Earnings News," *The Journal of Finance*, 64(5), 2289–2325.

Hirshleifer, D., and S. H. Teoh (2003): "Limited Attention, Information Disclosure, and Financial Reporting," *Journal of Accounting and Economics*, 36(1-3), 337–386.

Holden, C. W., and S. Jacobsen (2014): "Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions," *The Journal of Finance*, 69(4), 1747–1785.

Hong, H., T. Lim, and J. C. Stein (2000): "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies," *The Journal of Finance*, 55(1), 265–295.

Hong, H., and J. C. Stein (1999): "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," *The Journal of Finance*, 54(6), 2143–2184.

Hong, H., W. Torous, and R. Valkanov (2007): "Do industries lead stock markets?," *Journal of Financial Economics*, 83(2), 367–396.

Hou, K. (2007): "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns," *The Review of Financial Studies*, 20(4), 1113–1138.

Hou, K., and T. J. Moskowitz (2005): "Market Frictions, Price Delay, and the Cross-Section of Expected Returns," *The Review of Financial Studies*, 18(3), 981–1020.

Huberman, G., and T. Regev (2001): "Contagious Speculation and a Cure for Cancer: A Nonevent that made Stock Prices Soar," *The Journal of Finance*, 56(1), 387–396.

Huddleston, D., F. Liu, and L. Stentoft (2023): "Intraday Market Predictability: A Machine Learning Approach," *Journal of Financial Econometrics*, 21(2), 485–527.

Jiang, H., S. Z. Li, and H. Wang (2021): "Pervasive Underreaction: Evidence from High-Frequency Data," *Journal of Financial Economics*, 141(2), 573–599.

Jiang, L., J. Liu, L. Peng, and B. Wang (2022): "Investor Attention and Asset Pricing Anomalies," *Review of Finance*, 26(3), 563–593.

Kelly, B., and S. Pruitt (2015): "The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors," *Journal of Econometrics*, 186(2), 294–316.

Kelly, B. T., S. Malamud, and K. Zhou (2024): "The Virtue of Complexity in Return Prediction," *The Journal of Finance*, 79(1), 459–503.

Koijen, R. S., and S. Van Nieuwerburgh (2011): "Predictability of Returns and Cash Flows," *Annual Review of Financial Economics*, 3(1), 467–491.

Kozak, S., S. Nagel, and S. Santosh (2020): "Shrinking the Cross-Section," *Journal of Financial Economics*, 135(2), 271–292.

Kwan, A., R. Philip, and A. Shkilko (2021): "The Conduits of Price Discovery: A Machine Learning Approach," *Available at SSRN 3710491*.

Leung, P.-L., and W.-K. Wong (2008): "On Testing the Equality of Multiple Sharpe Ratios, with Application on the Evaluation of iShares," *The Journal of Risk*, 10(3), 15–30.

Liu, F. (2025): "Stock Distribution and Risk Premia Predictability with Quantile Machine Learning," *Available at SSRN 4491887*.

Liu, H., L. Peng, and Y. Tang (2023): "Retail Attention, Institutional Attention," *Journal of Financial and Quantitative Analysis*, 58(3), 1005–1038.

Lo, A. W., and A. C. MacKinlay (1990): "When are Contrarian Profits due to Stock Market Overreaction?," *The Review of Financial Studies*, 3(2), 175–205.

Lou, D. (2014): "Attracting Investor Attention through Advertising," *The Review of Financial Studies*, 27(6), 1797–1829.

McLean, R. D., and J. Pontiff (2016): "Does Academic Research Destroy Stock Return Predictability?," *The Journal of Finance*, 71(1), 5–32.

Merton, R. C. (1987): "A Simple Model of Capital Market Equilibrium with Incomplete Information," *The Journal of Finance*, 42(3), 483–510.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014): "Forecasting the Equity Risk Premium: The Role of Technical Indicators," *Management Science*, 60(7), 1772–1791.

Novy-Marx, R., and M. Velikov (2016): "A Taxonomy of Anomalies and Their Trading Costs," *The Review of Financial Studies*, 29(1), 104–147.

Patton, A. J., and M. Verardo (2012): "Does Beta move with News? Firm-Specific Information Flows and Learning about Profitability," *The Review of Financial Studies*, 25(9), 2789–2839.

Pelger, M. (2020): "Understanding Systematic Risk: A High-Frequency Approach," *The Journal of Finance*, 75(4), 2179–2220.

Peng, L. (2005): "Learning with Information Capacity Constraints," *Journal of Financial and Quantitative Analysis*, 40(2), 307–329.

Pontiff, J. (2006): "Costly Arbitrage and the Myth of Idiosyncratic Risk," *Journal of Accounting and Economics*, 42(1-2), 35–52.

Rapach, D., and G. Zhou (2013): "Forecasting Stock Returns," in *Handbook of Economic Forecasting*, vol. 2, pp. 328–383. Elsevier.

Roll, R. (1988): "R2," *The Journal of Finance*, 43(3), 541–566.

Savor, P., and M. Wilson (2013): "How much do Investors Care about Macroeconomic Risk? Evidence from Scheduled Economic Announcements," *Journal of Financial and Quantitative Analysis*, 48(2), 343–375.

Shleifer, A., and R. W. Vishny (1997): "The Limits of Arbitrage," *The Journal of Finance*, 52(1), 35–55.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014): "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *The Journal of Machine Learning Research*, 15(1), 1929–1958.

Timmermann, A., and L. Vulicevic (2026): "Compute, Complexity, and the Scaling Laws of Return Predictability," *Available at SSRN 6105327*.

van Binsbergen, J. H., M. Boons, C. C. Opp, and A. Tamoni (2023): "Dynamic Asset (Mis) Pricing: Build-up Versus Resolution Anomalies," *Journal of Financial Economics*, 147(2), 406–431.

Welch, I., and A. Goyal (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *The Review of Financial Studies*, 21(4), 1455–1508.

Zeiler, M. D. (2012): "Adadelta: An Adaptive Learning Rate Method," *arXiv preprint arXiv:1212.5701*.

Zhang, X. F. (2006): "Information Uncertainty and Stock Returns," *The Journal of Finance*, 61(1), 105–137.

Zuckerman, G., and B. Hope (2017): "The Quants Run Wall Street Now," *The Wall Street Journal*, 42.

Table 1: Portfolio Predictive Performance at the 1-min Horizon ($R^2_{\mathbf{OS}}$)

This table reports the R^2_{OS} (in percent) for the market and sector portfolios at the 1-min horizon from January 2005 to December 2021. The out-of-sample R^2 is calculated as

$$R^2_{\text{OS}} \equiv 1 - \frac{\sum_{t\in\mathcal{T}_3}\left(r_{t+1}-\hat{r}_{t+1}\right)^2}{\sum_{t\in\mathcal{T}_3}\left(r_{t+1}\right)^2},$$

where $\mathcal{T}_3$ indicates that the metric is only assessed on predictions in the test set, whose data was not used for model estimation or validation. The R^2_{OS} statistic pools prediction errors over time into a grand-level assessment of each model for a given stock, and it measures the reduction in mean squared forecasting error (MSFE) for each model relative to a naive forecast of the zero benchmark, which assumes that the $t+1$ stock return r_{t+1} equals zero. We compare the performance of seventeen models in total, including the fourteen individual models: autoregressive model (AR1), ordinary least squares (OLS), ridge (RID), lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), extremely random forest (XRF), gradient-boosted regression trees (GBT), and artificial neural networks with 1-5 layers (NN1-NN5); and the median ensembles of all linear models (L-MED), all nonlinear models (NL-MED), and all models (MED). We begin by estimating model parameters on a training set of ten months $\mathcal{T}_1$ (February, 2004 - November, 2004); then perform an extensive hyperparameter optimization, validating the model's fit in the next month as the validation set $\mathcal{T}_2$ (December, 2004); and subsequently evaluate the prediction error on an one-month test set (January, 2005). We repeat this procedure by shifting the training, validation, and test sets forward by one month in each iteration, for a total test set $\mathcal{T}_3$ of 17 years (January 2005 - December 2021).

Portfolio	AR1	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	L-MED	NL-MED	MED
Market	0.00	-1.45	0.01	-0.28	-0.01	0.00	-0.10	-0.08	-0.03	-11.19	0.00	0.00	0.00	0.00	0.03	0.01	0.01
Materials	0.00	2.39	1.81	2.12	1.83	0.69	0.55	0.51	0.02	1.50	0.96	0.25	0.11	0.00	1.92	0.33	0.67
Energy	-0.05	0.50	0.24	0.27	0.27	0.13	0.14	0.16	0.04	0.26	0.18	0.08	0.03	0.01	0.27	0.13	0.19
Finance	0.09	2.52	2.20	2.35	2.28	0.38	1.01	0.99	0.61	1.88	1.40	0.56	0.26	0.06	2.34	0.83	1.13
Industrials	0.00	2.21	1.64	1.97	1.75	0.70	0.30	0.35	0.03	-2E4	0.91	0.20	0.10	0.04	1.82	0.34	0.67
Technology	0.01	4.11	3.58	3.66	3.70	1.27	1.11	1.41	0.18	-3E2	1.96	0.60	0.31	0.14	3.70	0.87	1.60
Staples	-0.03	3.97	2.84	3.47	2.97	0.97	1.22	1.34	0.49	2.97	2.12	0.75	0.34	0.09	3.09	1.06	1.54
Utilities	-0.02	2.28	2.00	2.14	2.02	0.95	0.74	0.74	0.07	1.62	1.12	0.39	0.18	0.03	2.05	0.51	0.85
Health	-0.06	1.71	1.84	1.73	1.84	0.81	0.18	-0.54	0.11	-1E3	1.20	0.28	0.16	0.08	1.98	0.42	0.85
Discretionary	-0.07	3.12	2.48	2.80	2.60	0.99	1.24	1.30	0.08	1.83	1.42	0.46	0.20	0.08	2.66	0.70	1.28

Table 2: Portfolio Strategy Performance Before Transaction Costs at the 1-min Horizon (Annualized Sharpe Ratio)

This table reports the annualized Sharpe ratio of a strategy based on model forecasts for the market and sector portfolios at the 1-min horizon from January 2005 to December 2021. To assess the economic significance of the machine learning predictions, we implement a trading strategy that takes a long (short) position in a stock when the model predicts a positive (negative) return, scaled by the magnitude of the prediction, i.e., the stock's conditional expected return forecast. Specifically, the strategy's return at horizon h given model prediction $\hat{r}_{t+h}$ is defined as

$$\pi\left(\hat{r}_{t+h}\right) \equiv \begin{cases} -r_{t+h}\hat{w}_t & \text{if } \hat{r}_{t+h} < 0, \\ r_{t+h}\hat{w}_t & \text{if } \hat{r}_{t+h} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

where $\hat{w}_t = \min(|\hat{r}_{t+h}|, 1)$ denotes the portfolio weight implied by the forecast $\hat{r}_{t+h}$. This timing weight scales the position up or down to exploit time variation in the stock's conditional expected return, and is capped at 100% to prevent excessive leverage. For each stock, we subtract the intraday risk-free rate from the strategy's log returns, sum within each month, convert to arithmetic returns to form monthly excess returns, and compute annualized Sharpe ratios. As benchmarks, we also include the annualized Sharpe ratio of the buy-and-hold strategy computed using intraday returns only, IH, and using total returns that include overnight returns, BH.

Portfolio	AR1	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	L-MED	NL-MED	MED	IH	BH
Market	0.92	1.43	0.60	1.20	0.62	0.87	0.18	0.34	0.14	-0.22	1.09	0.45	0.37	0.38	0.82	1.49	1.28	0.19	0.68
Materials	0.46	3.07	2.54	2.91	2.60	2.29	1.67	1.64	0.44	2.88	2.61	1.20	2.10	0.23	2.67	1.62	2.23	-0.12	0.48
Energy	0.00	1.86	1.07	1.46	1.20	0.99	1.32	1.29	0.49	1.76	1.49	1.03	1.03	0.62	1.19	1.10	1.11	-0.22	0.29
Finance	0.98	2.61	2.28	2.62	2.17	2.26	2.86	2.93	1.89	2.23	2.58	2.11	1.95	2.54	2.28	2.58	2.73	-0.15	0.28
Industrials	0.84	2.62	2.00	2.33	2.17	1.97	1.21	1.31	0.57	0.15	1.74	1.72	1.69	1.62	2.17	1.67	1.70	0.03	0.53
Technology	0.84	3.38	2.79	3.08	2.94	2.40	2.73	2.49	1.16	0.86	3.16	2.67	2.87	2.84	2.93	3.12	3.11	0.32	0.83
Staples	0.48	2.85	2.22	2.82	2.36	2.52	1.86	1.92	1.03	2.56	2.76	2.21	2.32	2.07	2.42	1.92	2.11	0.42	0.80
Utilities	0.39	2.44	2.12	2.18	2.11	1.72	2.54	2.47	0.76	2.36	2.95	2.47	2.35	1.27	2.15	2.79	2.88	0.25	0.65
Health	0.12	2.90	2.39	2.47	2.51	1.49	1.43	1.39	0.69	0.39	2.27	1.90	1.93	2.58	2.42	2.18	2.10	0.27	0.78
Discretionary	-0.18	2.12	1.68	1.89	1.79	2.00	1.41	1.48	0.96	1.69	1.67	1.14	1.33	1.57	1.81	1.53	1.49	0.24	0.68

Table 3: Portfolio Strategy Performance After Transaction Costs at the 1-min Horizon (Annualized Sharpe Ratio)

This table reports the annualized Sharpe ratio of a strategy based on model forecasts after transaction costs for the market and sector portfolios at the 1-min horizon from January 2005 to December 2021. Following Chinco, Clark-Joseph, and Ye (2019), we evaluate the economic significance of returns after transaction costs, estimated as the bid-ask spread, by implementing a trading strategy that only trades when model predictions exceed the spread, scaled by the magnitude of the prediction. The return of this trading strategy is defined as

$$\gamma\left(\hat{r}_{t+h}, \text{Spread}_t\right) \equiv \begin{cases} (-r_{t+h} - \text{Spread}_t)\hat{w}_t & \text{if } |\hat{r}_{t+h}| > \text{Spread}_t \text{ and } \hat{r}_{t+h} < 0, \\ (r_{t+h} - \text{Spread}_t)\hat{w}_t & \text{if } |\hat{r}_{t+h}| > \text{Spread}_t \text{ and } \hat{r}_{t+h} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\hat{w}_t = \min(|\hat{r}_{t+h}|, 1)$ denotes the portfolio weight implied by the forecast $\hat{r}_{t+h}$. For each stock, the strategy's log returns net of transaction costs and intraday risk-free rate are summed within each month, converted to arithmetic returns, which are then used to calculate annualized Sharpe ratios. As benchmarks, we also include the annualized Sharpe ratio of the buy-and-hold strategy computed using intraday returns only, IH, and using total returns that include overnight returns, BH.

Portfolio	AR1	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	L-MED	NL-MED	MED	IH	BH
Market	-0.09	-1.85	0.13	-1.21	-0.34	-0.09	-0.50	-0.58	-0.46	-0.24	-0.26	-0.32	-0.24	-0.32	0.26	0.48	0.13	0.19	0.68
Materials	-0.13	1.81	1.49	1.63	1.50	0.34	1.21	0.93	-0.26	1.04	2.27	0.82	2.04	0.05	1.57	1.24	1.60	-0.12	0.48
Energy	-0.25	0.24	0.68	0.17	0.70	0.05	0.54	0.28	0.39	0.45	1.17	1.02	0.92	0.14	0.60	0.79	0.69	-0.22	0.29
Finance	0.36	0.41	0.82	0.66	0.74	1.00	2.32	1.48	1.84	0.69	1.12	1.38	2.17	2.01	0.79	2.55	1.29	-0.15	0.28
Industrials	0.25	1.32	1.14	1.24	1.21	0.59	-0.19	-0.15	0.04	-0.21	1.41	1.83	1.98	1.24	1.18	1.83	1.26	0.03	0.53
Technology	0.20	1.72	1.08	1.25	1.26	0.06	0.93	0.99	0.48	-0.05	0.89	2.39	2.61	2.13	1.31	2.72	1.36	0.32	0.83
Staples	-0.16	0.60	0.75	0.29	0.85	-0.21	1.32	1.36	1.02	0.59	2.62	2.18	2.01	2.01	0.76	1.85	1.59	0.42	0.80
Utilities	0.28	0.57	0.75	0.65	0.75	0.18	0.58	0.45	0.12	0.72	1.67	1.79	1.72	0.95	0.75	1.96	1.13	0.25	0.65
Health	-0.26	-0.09	-0.07	-0.09	-0.03	0.21	0.16	-0.11	0.50	-0.35	1.94	1.61	1.40	1.98	-0.02	1.65	1.66	0.27	0.78
Discretionary	-0.26	1.26	1.29	1.08	1.37	0.43	1.45	1.42	0.57	0.86	1.01	1.07	1.11	1.11	1.44	1.28	0.95	0.24	0.68

Table 4: Risk-Adjusted Portfolio Strategy Performance After Transaction Costs at the 1-min Horizon

This table reports the monthly alphas (in percent), betas, and adjusted R^2 values (in percent) for the monthly regression of the NL-MED strategy excess returns (after transaction costs at the 1-min horizon) onto a six-factor model that includes the Fama-French five-factors (Fama and French (2015)) augmented with the momentum (Carhart (1997)) factor. β_{mkt}, β_{smb}, β_{hml}, β_{rmw}, β_{cma}, β_{umd} denote betas with respect to the market, size, value, profitability, investment, and momentum factors, respectively. The t-statistics are calculated using Newey-West standard errors. Similar to Chinco, Clark-Joseph, and Ye (2019), we scale the strategy's monthly returns to match the intraday buy-and-hold volatility to aid interpretation of the alphas and coefficients, noting that this scaling does not affect t-statistics or adjusted R^2 estimates.

	Market	Discretionary	Energy	Finance	Health	Industrials	Materials	Staples	Technology	Utilities
α	0.59	1.64	1.45	4.35	1.68	2.21	1.82	1.73	3.25	2.29
t(α)	(1.99)	(3.53)	(1.88)	(5.36)	(3.77)	(4.76)	(4.71)	(2.75)	(6.68)	(4.75)
β_{mkt}	-0.12	-0.24	-0.13	0	0	0.04	-0.12	-0.02	-0.06	-0.02
t(β_{mkt})	(-1.5)	(-1.62)	(-1.5)	(0)	(-0.05)	(0.65)	(-1.54)	(-0.43)	(-0.89)	(-0.37)
β_{smb}	0.08	0.32	-0.07	-0.18	0.02	-0.03	0.18	-0.05	-0.07	-0.12
t(β_{smb})	(1)	(1.88)	(-0.38)	(-1.19)	(0.22)	(-0.28)	(1)	(-0.69)	(-0.6)	(-1.11)
β_{hml}	0.09	-0.02	0.38	0.31	-0.07	-0.12	-0.04	-0.06	0.12	0.1
t(β_{hml})	(0.77)	(-0.17)	(2.1)	(2.08)	(-0.56)	(-0.92)	(-0.28)	(-0.4)	(1.18)	(0.82)
β_{rmw}	0	0.34	-0.15	-0.31	-0.01	-0.1	0.14	-0.05	-0.17	-0.25
t(β_{rmw})	(-0.03)	(1.44)	(-0.53)	(-1.09)	(-0.07)	(-0.59)	(0.67)	(-0.25)	(-0.89)	(-1.22)
β_{cma}	-0.28	-0.08	-0.54	-0.32	0.26	0.11	-0.22	0.37	0.08	0.23
t(β_{cma})	(-0.77)	(-0.48)	(-1.39)	(-1.22)	(1.15)	(0.65)	(-1.35)	(1.29)	(0.38)	(1.17)
β_{umd}	0.04	-0.11	0.01	0.13	0.04	-0.3	-0.1	0.05	0.05	0.01
t(β_{umd})	(0.91)	(-1.91)	(0.33)	(1.38)	(1.27)	(-1.21)	(-1.31)	(1.09)	(0.93)	(0.3)
Adj. R^2	0.28	4.58	0.48	-0.48	-1.62	7.06	-1.15	0.93	-1.34	-0.16

Table 5: Dynamic Risk-Adjusted Portfolio Strategy Performance After Transaction Costs at the 1-min Horizon

Panel A reports t-statistics of the alphas and betas, as well as adjusted R^2 values (in percent), from regressing NL-MED strategy excess returns (after transaction costs) at the 1-min horizon onto the Fama-French five factors (Fama and French (2015)) augmented with momentum (Carhart (1997)). All variables are constructed at the 1-min interval. Within each month, we regress strategy returns on the risk factors to obtain monthly coefficient estimates, then report Newey-West t-statistics of these estimates across months. $\beta_{mkt_{vw}}$, β_{smb}, β_{hml}, β_{rmw}, β_{cma}, and β_{umd} denote betas with respect to the value-weighted market, size, value, profitability, investment, and momentum factors, respectively. Panel B reports the same analysis with respect to the Pelger (2020) factors, where $\beta_{mkt_{ew}}$, β_{oil}, $\beta_{finance}$, and $\beta_{electricity}$ denote betas with respect to the equal-weighted market, oil, finance, and electricity factors, respectively.

Panel A: Fama-French Five Factors and Momentum

	Market	Discretionary	Energy	Finance	Health	Industrials	Materials	Staples	Technology	Utilities
t(α)	(2.01)	(3.51)	(2.09)	(5.19)	(4.09)	(4.49)	(3.88)	(3.88)	(6.87)	(5.27)
t($\beta_{mkt_{vw}}$)	(3.42)	(0.7)	(-1.65)	(2.3)	(1.44)	(3.84)	(-0.42)	(0.74)	(1.81)	(1.7)
t(β_{smb})	(-1.31)	(-1.93)	(0.41)	(-0.27)	(-1.47)	(1.18)	(0.23)	(-1.75)	(-2.85)	(1.25)
t(β_{hml})	(0.91)	(0.47)	(-1.48)	(0.9)	(-0.33)	(0.15)	(1.65)	(0.08)	(-1.96)	(1.62)
t(β_{rmw})	(-0.67)	(0.79)	(1.12)	(-0.8)	(-1.3)	(0.56)	(0.32)	(-1.24)	(1.64)	(2)
t(β_{cma})	(-1.01)	(-0.92)	(-0.52)	(-1.03)	(1.56)	(0.25)	(-1.05)	(1.55)	(-1.27)	(-0.48)
t(β_{umd})	(0.99)	(0.84)	(-0.62)	(-0.42)	(-1.15)	(-0.93)	(1.04)	(1.21)	(1.75)	(-0.14)
Adj. R^2	0.85	0.31	0.62	0.3	0.57	0.6	0.36	0.31	0.51	0.41

Panel B: Pelger Factors

	Market	Discretionary	Energy	Finance	Health	Industrials	Materials	Staples	Technology	Utilities
t(α)	(1.89)	(3.45)	(2.06)	(5.19)	(4.16)	(4.45)	(3.88)	(3.91)	(6.87)	(5.26)
t($\beta_{mkt_{ew}}$)	(2.41)	(0.96)	(-0.28)	(-0.21)	(-0.39)	(1.88)	(-1.58)	(0.43)	(3.11)	(-1.14)
t(β_{oil})	(0.92)	(0.59)	(-1.07)	(2.37)	(-0.14)	(-0.36)	(1.54)	(-1.1)	(-1.3)	(0.04)
t($\beta_{finance}$)	(0.55)	(-0.31)	(-0.47)	(1.48)	(1.74)	(0.24)	(1.32)	(0.7)	(-3.06)	(1.72)
t($\beta_{electricity}$)	(-1.08)	(-0.48)	(1.16)	(-2.15)	(1.29)	(0.86)	(1.78)	(0.56)	(-1.75)	(2.19)
Adj. R^2	0.69	0.24	0.72	0.27	0.43	0.58	0.3	0.26	0.36	1.06

Table 6: Data Mining Tests at the 1-min Horizon

This table compares portfolio performance of the MED model between the original sample from 2005 to 2016 and the recent sample from 2017 to 2021. Panel A reports R^2_{OS} (in percent) and Panel B reports annualized Sharpe ratios after transaction costs.

Panel A: R^2_{OS} (in percent)										
	Market	Materials	Energy	Finance	Industrials	Technology	Staples	Utilities	Health	Discretionary
2005-2016	0.01	0.67	0.19	1.13	0.67	1.60	1.54	0.85	0.85	1.28
2017-2021	0.01	0.10	0.11	0.79	0.04	0.26	0.08	0.02	0.04	0.35
Panel B: Annualized Sharpe Ratios After Transaction Costs										
	Market	Materials	Energy	Finance	Industrials	Technology	Staples	Utilities	Health	Discretionary
2005-2016	0.48	1.24	0.79	2.55	1.83	2.72	1.85	1.96	1.65	1.28
2017-2021	0.65	2.59	1.38	2.61	1.51	2.27	2.55	0.96	1.79	1.76

Table 7: Predictive Fama-MacBeth Regressions on Firm Characteristics

This table reports Fama-MacBeth cross-sectional regressions of monthly firm-level R^2_{OS} and net returns (returns after transaction costs) from the MED model on 20 lagged firm characteristics at the 1-min horizon. We report average monthly coefficients and Newey-West t-statistics. The outcome variable and characteristics are each standardized cross-sectionally within each month to have zero mean and unit variance, so that coefficients represent the effect of a one-standard-deviation change in the characteristic on the outcome variable in standard deviation units.

		Liquidity Variables								
Outcome		mvel1	dolvol	std_dolvol	turn	std_turn	baspread	ill	pricedelay	zerotrade
R^2_{OS}	coef	-0.2	-0.21	-0.01	0	-0.01	0.03	0.14	0.03	-0.02
	t-stat	(-8.31)	(-6.3)	(-0.59)	(0.07)	(-0.38)	(0.89)	(5.7)	(3.55)	(-0.42)
Net returns	coef	-0.04	-0.03	-0.03	0.01	-0.01	0.04	0.04	0.03	-0.02
	t-stat	(-3.83)	(-1.78)	(-1.99)	(0.3)	(-0.41)	(0.81)	(4.51)	(1.69)	(-1.11)

Risk Variables				Price Trend Variables						
beta	betasq	idiovol	retvol	chmom	indmom	maxret	mom1m	mom6m	mom12m	mom36m
0.1	0.11	-0.01	0.03	-0.02	0.02	0.02	0.02	-0.01	0	0
(3.48)	(3.66)	(-0.25)	(0.99)	(-1)	(1.9)	(1.47)	(1.06)	(-0.77)	(0.08)	(-0.07)
0.04	0.04	0	0.02	-0.03	0.02	0.01	-0.02	-0.01	0.01	0.02
(1.28)	(1.24)	(0.12)	(0.44)	(-1.15)	(1.33)	(0.2)	(-1.77)	(-0.68)	(0.43)	(1.05)

Table 8: Quintile Portfolio Sorts on Firm Characteristics

This table presents the performance of the total ensemble (MED) for individual stocks sorted into quintile portfolios by analyst coverage, size, dollar volume, or the arbitrage index, each lagged by one month. For each characteristic, the first row reports the R^2_{OS} (in percent) and the second row reports the annualized Sharpe ratio (SR) after transaction costs.

		Quintile Portfolio				
Characteristic	Metric	1 (Low)	2	3	4	5 (High)
Analyst Coverage	R^2_{OS}	0.53	0.54	0.41	0.41	0.45
	SR	2.34	2.17	2.00	1.40	1.75
Size	R^2_{OS}	0.59	0.45	0.54	0.38	0.23
	SR	1.63	2.45	2.38	1.95	1.74
Dollar Volume	R^2_{OS}	0.66	0.50	0.46	0.47	0.27
	SR	1.94	1.70	2.55	2.74	1.02
Arbitrage Index	R^2_{OS}	0.27	0.38	0.48	0.57	0.52
	SR	1.71	1.65	2.45	2.00	1.70

Table 9: Different Re-estimation Timescales

This table presents 1-min horizon results across different re-estimation timescales from January 2005 to December 2021, including the baseline monthly re-estimation as well as yearly, daily, and minutely alternatives. Panel A reports the R^2_{OS} values (in percent) for each portfolio and the average across portfolios, for the MED, LAS, and NN3 models. Panel B reports the annualized Sharpe ratios after transaction costs. Due to computational constraints, minutely re-estimation is reported only for LAS.

Panel A: R^2_{OS} (in percent)

	MED			LAS				NN3		
Portfolio	Yearly	Monthly	Daily	Yearly	Monthly	Daily	Minutely	Yearly	Monthly	Daily
Market	0.02	0.01	-0.03	-0.03	0.01	-0.18	-3.91	0.01	0.00	-0.13
Materials	0.26	0.67	0.83	0.69	1.81	1.69	-3.61	0.08	0.25	0.34
Energy	0.04	0.19	0.21	0.03	0.24	0.24	-3.62	0.01	0.08	-0.03
Finance	0.88	1.13	1.28	1.21	2.20	2.18	-3.66	0.51	0.56	0.65
Industrials	0.29	0.67	0.71	0.82	1.64	1.53	-3.73	0.06	0.20	0.25
Technology	1.10	1.60	1.64	2.54	3.58	3.14	-3.52	0.43	0.60	0.74
Staples	0.70	1.54	1.67	1.79	2.84	2.69	-3.34	0.40	0.75	1.00
Utilities	0.50	0.85	1.01	1.63	2.00	1.81	-3.96	0.17	0.39	0.52
Health	0.38	0.85	0.98	1.12	1.84	1.91	-3.70	0.15	0.28	0.39
Discretionary	0.32	1.28	1.09	0.78	2.48	1.98	-3.58	0.06	0.46	0.47
Average	0.45	0.88	0.94	1.06	1.86	1.70	-3.66	0.19	0.36	0.42

Panel B: Annualized Sharpe Ratio After Transaction Costs

	MED			LAS				NN3		
Portfolio	Yearly	Monthly	Daily	Yearly	Monthly	Daily	Minutely	Yearly	Monthly	Daily
Market	0.68	0.13	-0.45	-0.93	0.13	-0.50	-2.11	0.32	-0.32	-0.45
Materials	3.29	1.60	1.13	-0.30	1.49	1.09	-1.40	2.33	0.82	0.46
Energy	1.29	0.69	0.29	-0.39	0.68	0.39	-2.33	0.99	1.02	-0.29
Finance	1.36	1.29	0.92	0.40	0.82	0.62	-1.09	2.96	1.38	2.65
Industrials	2.49	1.26	0.70	1.17	1.14	1.10	-1.26	2.22	1.83	-0.01
Technology	0.98	1.36	1.40	0.15	1.08	1.70	-1.22	3.11	2.39	0.61
Staples	2.45	1.59	0.55	1.05	0.75	0.14	-0.64	2.54	2.18	1.11
Utilities	2.09	1.13	1.06	0.53	0.75	0.67	-0.99	2.16	1.79	1.33
Health	1.34	1.66	1.07	0.28	-0.07	0.09	-1.39	2.79	1.61	0.38
Discretionary	2.25	0.95	1.35	0.59	1.29	1.01	-1.40	2.37	1.07	0.54
Average	1.82	1.17	0.80	0.26	0.81	0.63	-1.38	2.18	1.38	0.63

Table 10: Model Training Times

This table reports model training times. Individual model times are in seconds; the total column reports the sum across models in both seconds and hours. Panel A reports the time per re-estimation across different re-estimation timescales at the 1-min horizon for the market portfolio, including the baseline monthly re-estimation as well as yearly, daily, and minutely alternatives. Panel B reports the same quantities annualized to a single year of predictions. Panel C reports the time per re-estimation at the 5-min horizon across portfolios with monthly re-estimation.

Panel A: Market Training Times at 1-min Horizon Re-estimation Timescales

Re-Estimation	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	Total	Total (Hours)
Yearly	7	4	5	10	551	32973	38784	273	128	95	87	90	81	73088	20.30
Monthly	4	3	3	10	405	6609	8573	308	220	225	194	178	138	16870	4.69
Daily	3	3	3	10	104	2183	2341	177	37	37	37	37	36	5008	1.39
Minutely	4	4	4	22	NA	NA	NA	NA	NA	NA	NA	NA	NA	34	0.01

Panel B: Annualized Market Training Times at 1-min Horizon with Different Re-estimation Timescales

Re-Estimation	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	Total	Total (Hours)
Yearly	7	4	5	10	551	32973	38784	273	128	95	87	90	81	73088	20.30
Monthly	48	36	36	120	4860	79308	102876	3696	2640	2700	2328	2136	1656	202440	56.28
Daily	756	756	756	2520	26208	550116	589932	44604	9324	9324	9324	9324	9072	1262016	350.28
Minutely	392112	392112	392112	2156616	NA	NA	NA	NA	NA	NA	NA	NA	NA	3332952	980.28

Panel C: Portfolio Training Times at the 5-min Horizon with Monthly Re-estimation Timescale

Portfolio	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	Total	Total (Hours)
Market	4	2	2	10	122	1491	2207	174	130	143	118	95	64	4562	1.27
Energy	2	1	1	4	33	195	208	97	117	122	82	67	45	974	0.27
Materials	2	2	1	4	46	240	250	88	92	99	70	58	44	996	0.28
Industrials	2	2	1	5	63	319	372	114	92	107	76	66	52	1271	0.35
Discretionary	2	2	1	5	53	376	395	141	98	119	84	69	47	1392	0.39
Staples	2	2	1	4	41	297	272	74	63	63	56	54	47	976	0.27
Health	2	2	1	5	57	301	339	110	75	82	66	61	54	1155	0.32
Finance	2	2	2	6	69	538	554	140	130	162	106	85	54	1850	0.51
Technology	2	2	2	5	46	331	368	178	106	125	85	70	51	1371	0.38
Utilities	2	1	1	4	53	189	203	74	73	77	59	54	42	832	0.23

Table 11: Macroeconomic News at the 1-min Horizon

This table compares the performance of MED on macroeconomic announcement days versus non-announcement days for portfolios and individual stocks at the 1-min horizon. Macroeconomic announcement days are defined following Savor and Wilson (2013) and Gao, Han, Li, and Zhou (2018); see Section 4.2.1 for details. We also report results for all macroeconomic announcement days combined (Any Macro). Panel A reports the R^2_{OS} values (in percent), and Panel B reports the annualized Sharpe ratios after transaction costs.

Panel A: R^2_{OS} (in percent)

Announcement	News	Market	Materials	Energy	Finance	Industrials	Technology	Staples	Utilities	Health	Discretionary	Individual
CPI	Yes	0.02	0.6	0.17	0.96	0.59	1.4	1.5	0.68	0.79	0.94	0.4
CPI	No	0.01	0.67	0.19	1.14	0.67	1.61	1.55	0.86	0.86	1.3	0.47
PPI	Yes	0.01	0.55	0.2	1.06	0.55	1.56	1.35	0.81	0.82	1.12	0.59
PPI	No	0.01	0.67	0.19	1.14	0.68	1.6	1.56	0.85	0.85	1.29	0.46
Employment	Yes	0.02	0.8	0.18	1.14	0.64	1.51	1.73	0.95	0.88	1.3	0.56
Employment	No	0.01	0.66	0.19	1.13	0.67	1.6	1.54	0.85	0.85	1.28	0.46
FOMC	Yes	0	0.51	0.19	1.1	0.57	1.24	1.28	0.82	0.76	0.73	0.36
FOMC	No	0.01	0.67	0.19	1.13	0.67	1.61	1.55	0.85	0.86	1.3	0.47
GDP	Yes	0.02	0.74	0.19	1.16	0.68	1.58	1.72	0.76	0.86	1.03	0.47
GDP	No	0.01	0.66	0.19	1.13	0.67	1.6	1.54	0.86	0.85	1.3	0.47
MCSI	Yes	0.01	0.62	0.2	1.17	0.7	1.63	1.69	0.82	0.93	1.31	0.6
MCSI	No	0.01	0.67	0.19	1.13	0.67	1.6	1.54	0.85	0.85	1.28	0.46
Fed Minutes	Yes	0.02	0.54	0.05	0.87	0.62	1.37	1.7	0.72	0.89	0.86	0.39
Fed Minutes	No	0.01	0.67	0.2	1.14	0.67	1.6	1.54	0.85	0.85	1.3	0.47
Any Macro	Yes	0.01	0.63	0.17	1.09	0.62	1.48	1.52	0.8	0.84	1.05	0.48
Any Macro	No	0.01	0.68	0.2	1.15	0.69	1.64	1.55	0.87	0.86	1.37	0.46

Panel B: Sharpe Ratio After Transaction Costs

Announcement	News	Market	Materials	Energy	Finance	Industrials	Technology	Staples	Utilities	Health	Discretionary	Individual
CPI	Yes	0	0.49	0.25	0.99	0.18	1.65	0.62	0.49	0.54	0.42	2.27
CPI	No	0.12	1.59	0.71	1.23	1.08	1.32	1.72	1.11	1.54	0.91	2.17
PPI	Yes	-0.23	0.47	0.3	2	0.5	0.76	0.93	0.45	1.29	0.95	2.12
PPI	No	0.18	1.7	0.67	1.25	1.31	1.3	1.61	1.12	1.64	0.94	2.19
Employment	Yes	-0.03	0.68	0.37	1.81	1.13	0.71	1.13	0.78	1	0.88	1.86
Employment	No	0.13	1.68	0.69	1.25	1.21	1.32	1.6	1.1	1.62	0.92	2.19
FOMC	Yes	0.09	0.74	0.28	0.88	0.46	1.6	0.41	0.53	0.48	0.47	1.33
FOMC	No	0.12	1.63	0.69	1.25	1.17	1.33	1.61	1.11	1.58	0.97	2.2
GDP	Yes	0.25	1.15	0.21	1.94	0.71	1.54	1.32	0.61	0.92	0.87	1.67
GDP	No	0.08	1.59	0.7	1.25	1.25	1.32	1.56	1.09	1.62	0.93	2.24
MCSI	Yes	-0.24	0.59	0.45	2.1	0.49	1.57	1.44	1.03	0.86	0.77	1.2
MCSI	No	0.3	1.53	0.69	1.25	1.27	1.32	1.57	1.1	1.62	0.95	2.33
Fed Minutes	Yes	0.33	0.24	-0.24	1.86	-0.22	0.86	0.29	0.51	0.05	-0.12	1.33
Fed Minutes	No	0.09	1.58	0.72	1.26	1.27	1.34	1.81	1.11	1.68	0.94	2.19
Any Macro	Yes	-0.17	1.01	0.36	2.19	1.05	1.66	0.97	1.35	1.56	0.91	2.26
Any Macro	No	0.28	1.59	0.69	1.03	1.18	1.08	1.84	0.93	1.5	0.86	2.16

Table 12: Firm-Specific News at the 1-min Horizon

This table compares the performance of MED across firm-specific news categories for individual stocks at the 1-min horizon, reporting R^2_{OS} values (in percent) and annualized Sharpe ratios (SR) after transaction costs. Panels A, B, and C report results based on the Boudoukh, Feldman, Kogan, and Richardson (2019) dataset from 2005 to 2015, which classifies each firm-day observation into three categories: no news, unidentified news (days on which none of the news coverage is identified as relevant for firm valuation), and identified news (days on which at least one article is identified as relevant). Panels A, B, and C correspond to news coverage during trading hours, overnight, and both, respectively. Panel D reports results based on the previous month's analyst forecast revision, classifying firm-month observations into negative revisions, no revision, and positive revisions. Panel E reports results for firm-day observations partitioned into earnings announcement days and non-announcement days.

Panel A: Trading Hours News			
Metric	No News	Unidentified News	Identified News
R^2_{OS}	0.51	0.38	0.29
SR	2.98	2.42	0.72
Panel B: Overnight News			
Metric	No News	Unidentified News	Identified News
R^2_{OS}	0.47	0.35	0.29
SR	2.74	1.91	0.81
Panel C: Trading Hours or Overnight News			
Metric	No News	Unidentified News	Identified News
R^2_{OS}	0.48	0.35	0.25
SR	3.03	2.19	0.17
Panel D: Analyst Revision			
Metric	Negative Revision	No Revision	Positive Revision
R^2_{OS}	0.39	0.50	0.49
SR	1.64	2.26	2.11
Panel E: Earnings Release			
Metric	No Release	Release	
R^2_{OS}	0.52	0.44	
SR	2.23	2.12	

Table 13: Knight Capital Natural Experiment

This table compares portfolio performance of the linear ensemble (L-MED), nonlinear ensemble (NL-MED), and total ensemble (MED) models on August 1, 2012 (event day) versus the remainder of August 2012 (August 2 to 31). Panel A reports R^2_{OS} (in percent) and Panel B reports Sharpe ratios after transaction costs. Sharpe ratios in this table are computed from 1-min returns and are not annualized, as they are not directly comparable to the annualized Sharpe ratios in Table 3, which are based on monthly excess returns.

Panel A: R^2_{OS} (in percent)					
Time Period	Glitch Affected	L-MED	NL-MED	MED	
August 1	Yes	-1.19	0.02	-0.04	
August 1	No	0.94	0.32	0.49	
Rest of August	Yes	2.28	1.13	1.39	
Rest of August	No	1.98	0.52	0.83	
Panel B: Sharpe Ratio After Transaction Costs					
Time Period	Glitch Affected	L-MED	NL-MED	MED	Intraday Buy and Hold
August 1	Yes	-0.07	0.04	0	-0.03
August 1	No	0.05	0.16	0.19	-0.06
Rest of August	Yes	0.09	0.19	0.18	0.01
Rest of August	No	0.07	0.18	0.19	0.01

Figure 1: Portfolio Predictive Performance at the 1-min Horizon (Percentage R^2_{OS})

This figure presents the out-of-sample R^2 (percentage R^2_{OS}) of the market and sector portfolios for the median ensembles of all linear models (L-MED), all nonlinear models (NL-MED), and all models (MED) at the 1-min horizon from January 2005 to December 2021.

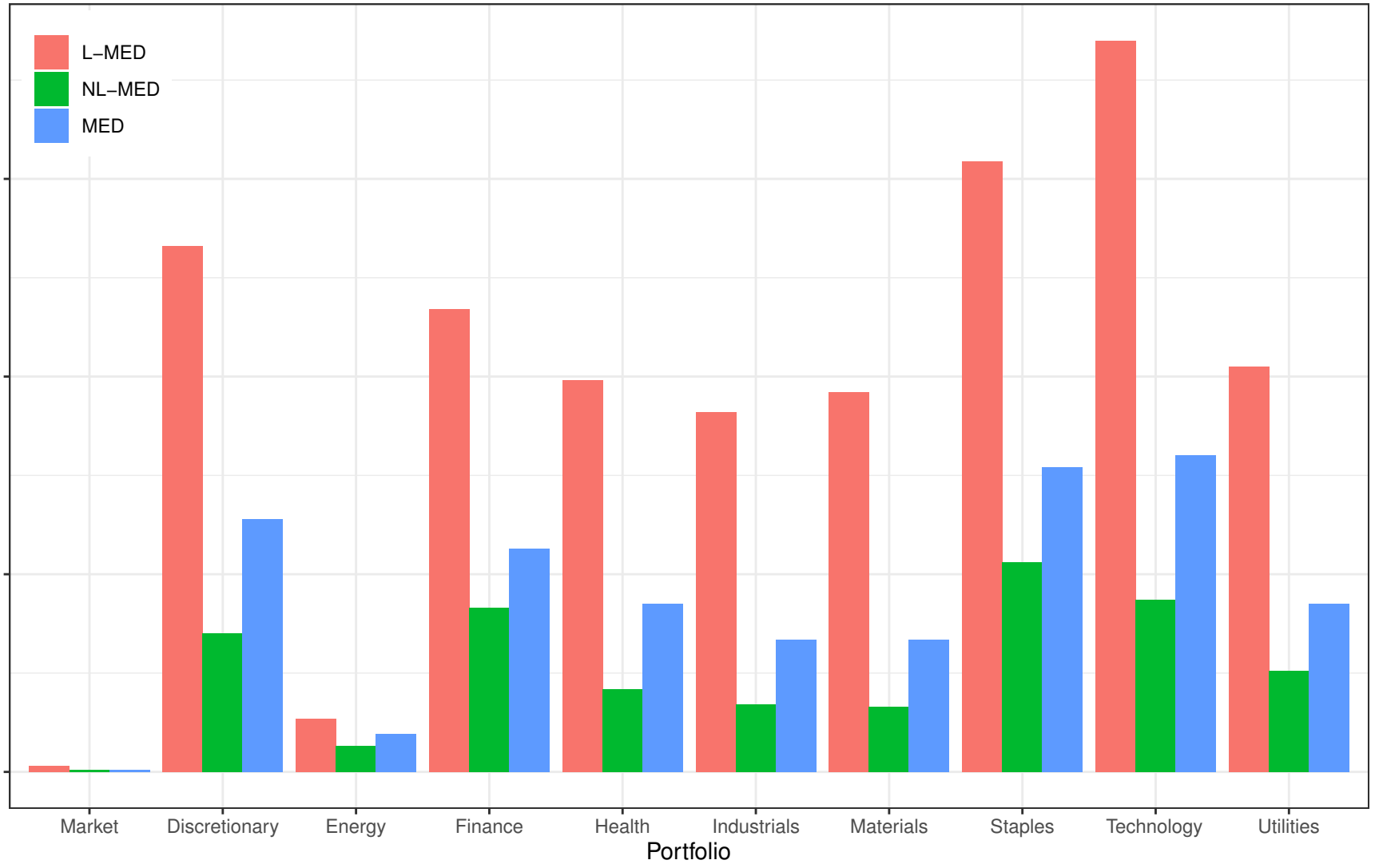


Figure 2: Portfolio Predictive Performance by Time Horizon (Percentage R^2_{OS})

This figure presents the out-of-sample R^2 (percentage R^2_{OS}) of the market and sector portfolios for the total median ensemble (MED) at the 1-min, 5-min, 10-min, 15-min, and 30-min horizons from January 2005 to December 2021.

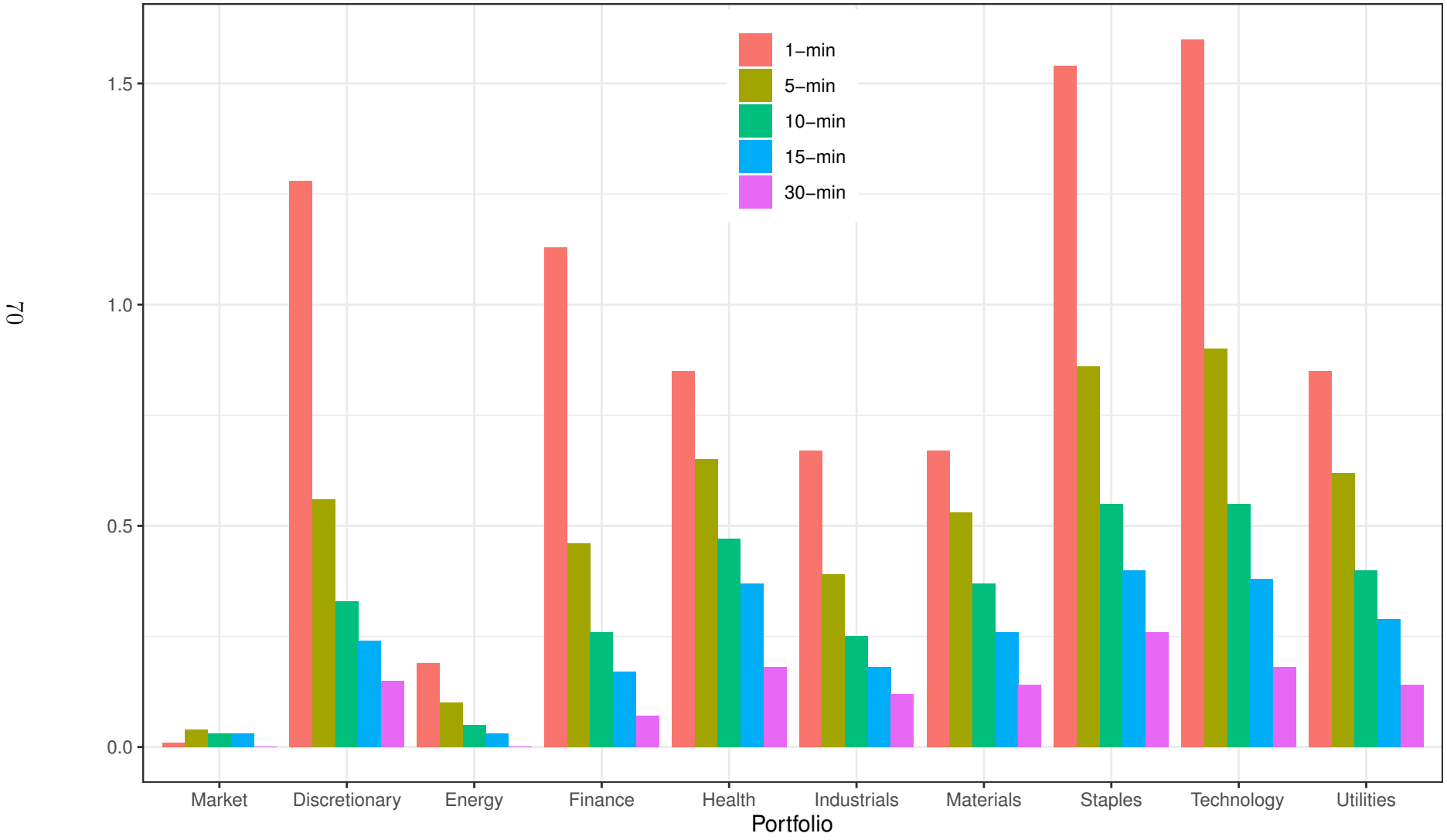


Figure 3: Portfolio Profitability Before Transaction Costs at the 1-min Horizon (Annualized Sharpe Ratio)

This figure presents annualized Sharpe ratios of strategies based on model forecasts for the market and sector portfolios at the 1-min horizon from January 2005 to December 2021, comparing the median ensembles of all linear models (L-MED), all nonlinear models (NL-MED), and all models (MED). We test the equality of Sharpe ratios between L-MED and NL-MED using the two-sided test of Leung and Wong (2008). *, **, and *** indicate the Sharpe ratios are statistically different at the 10%, 5%, and 1% significance levels, respectively.

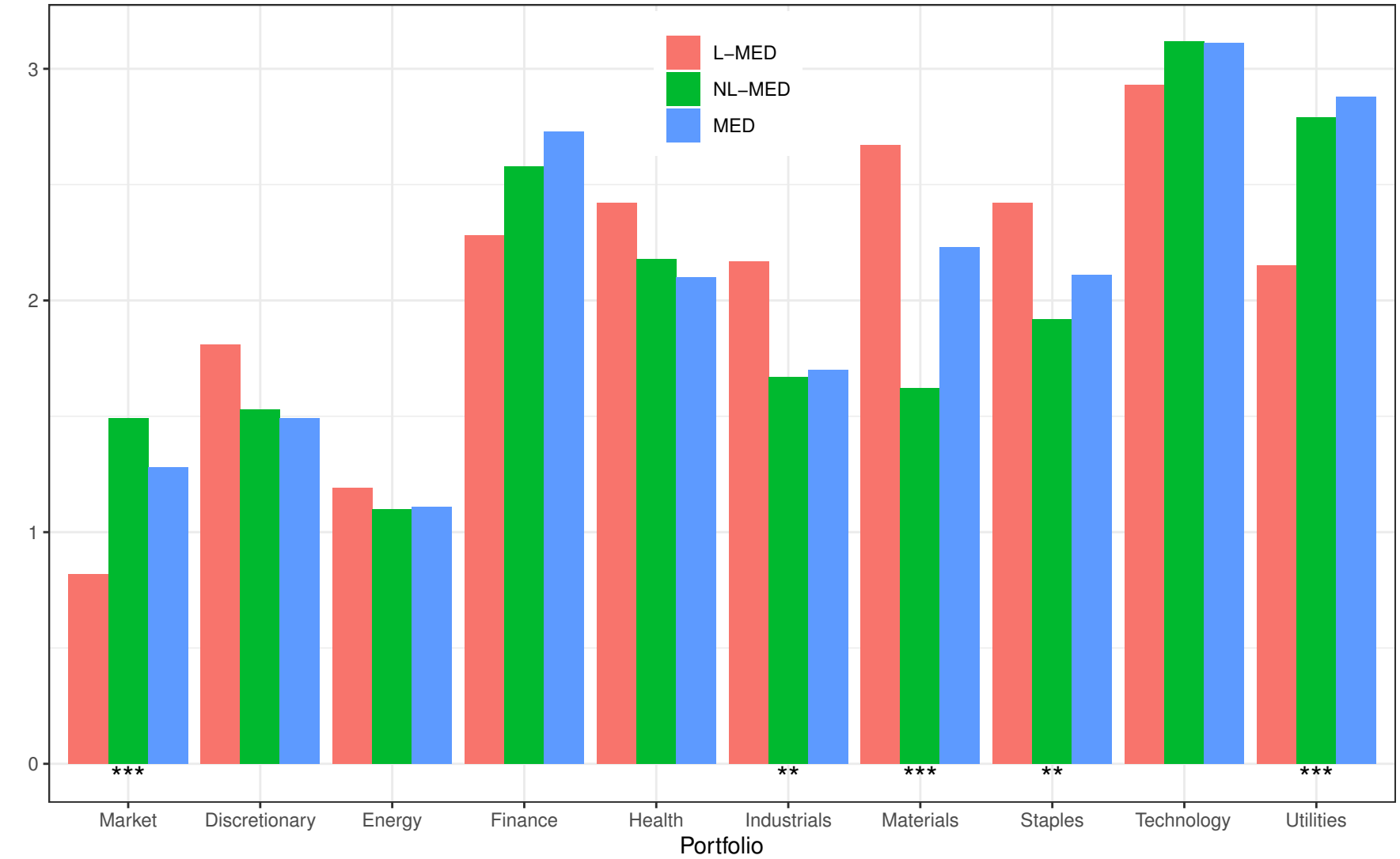


Figure 4: Portfolio Strategy After Transaction Costs at the 1-min Horizon (Annualized Sharpe Ratio)

This figure presents annualized Sharpe ratios after transaction costs for the market and sector portfolios at the 1-min horizon from January 2005 to December 2021, comparing the median ensembles of all linear models (L-MED), all nonlinear models (NL-MED), and all models (MED). We test the equality of Sharpe ratios between L-MED and NL-MED using the two-sided test of Leung and Wong (2008). *, **, and *** indicate the Sharpe ratios are statistically different at the 10%, 5%, and 1% significance levels, respectively.

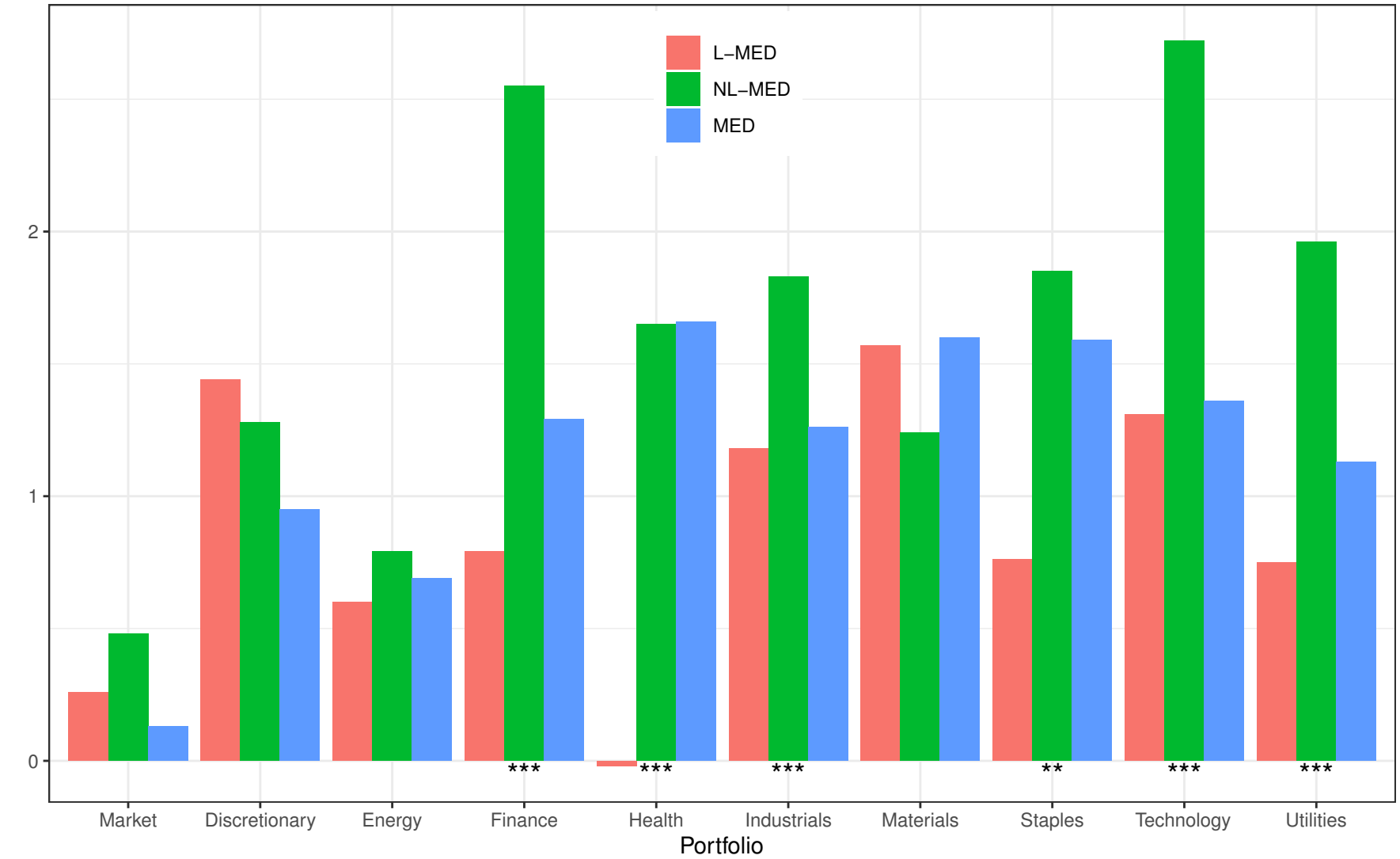


Figure 5: Portfolio Profitability After Transaction Costs by Time Horizon (Annualized Sharpe Ratio)

This figure presents the annualized Sharpe ratio of a strategy based on model forecasts after transaction costs for the market and sector portfolios for the nonlinear median ensemble (NL-MED) for the 1-min, 5-min, 10-min, 15-min, and 30-min horizons from January 2005 to December 2021.

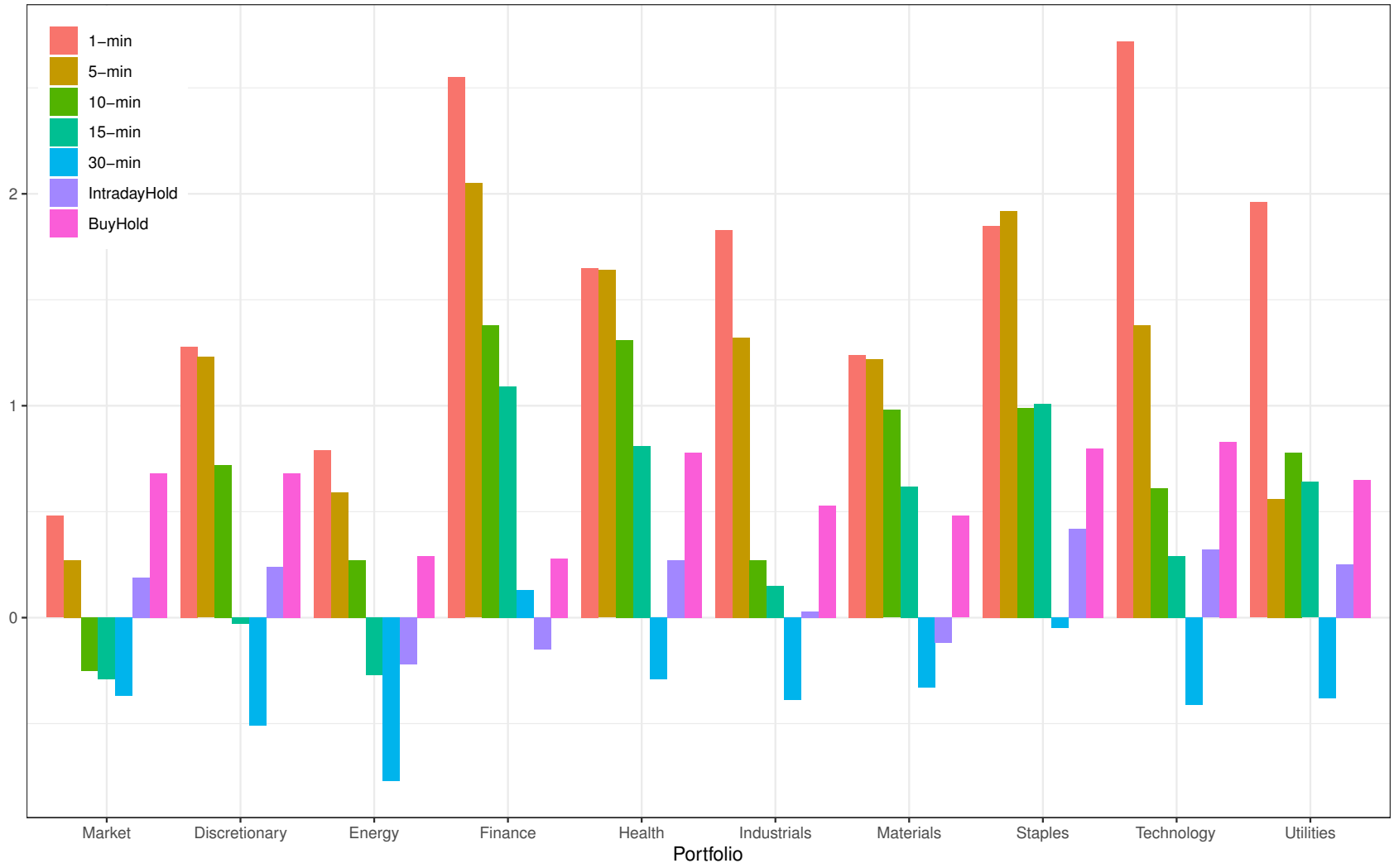


Figure 6: Individual Stock Performance

This figure presents R^2_{OS} (in percent, left bars) and annualized Sharpe ratios (SR) after transaction costs (right bars) for individual stocks at the 1-min horizon from January 2005 to December 2021, where excess returns are equally weighted across stocks in each month.

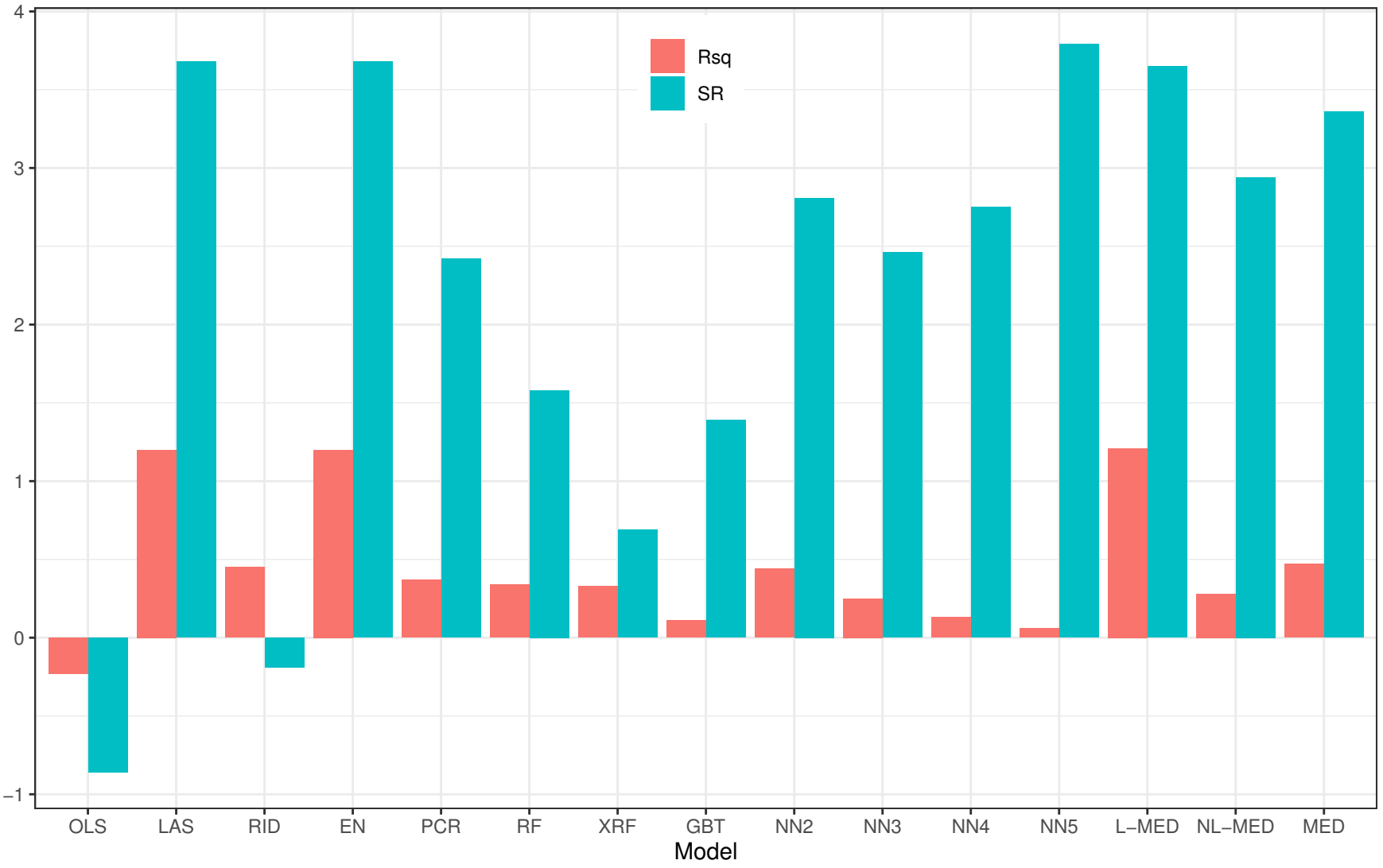


Figure 7: Portfolio Time-of-Day Predictability at the 1-min Horizon (R^2_{OS})

This figure presents, for each portfolio, the R^2_{OS} (in percent) in each half-hour interval from 9:30am to 4:00pm, achieved by the total ensemble (MED) at the 1-min horizon.

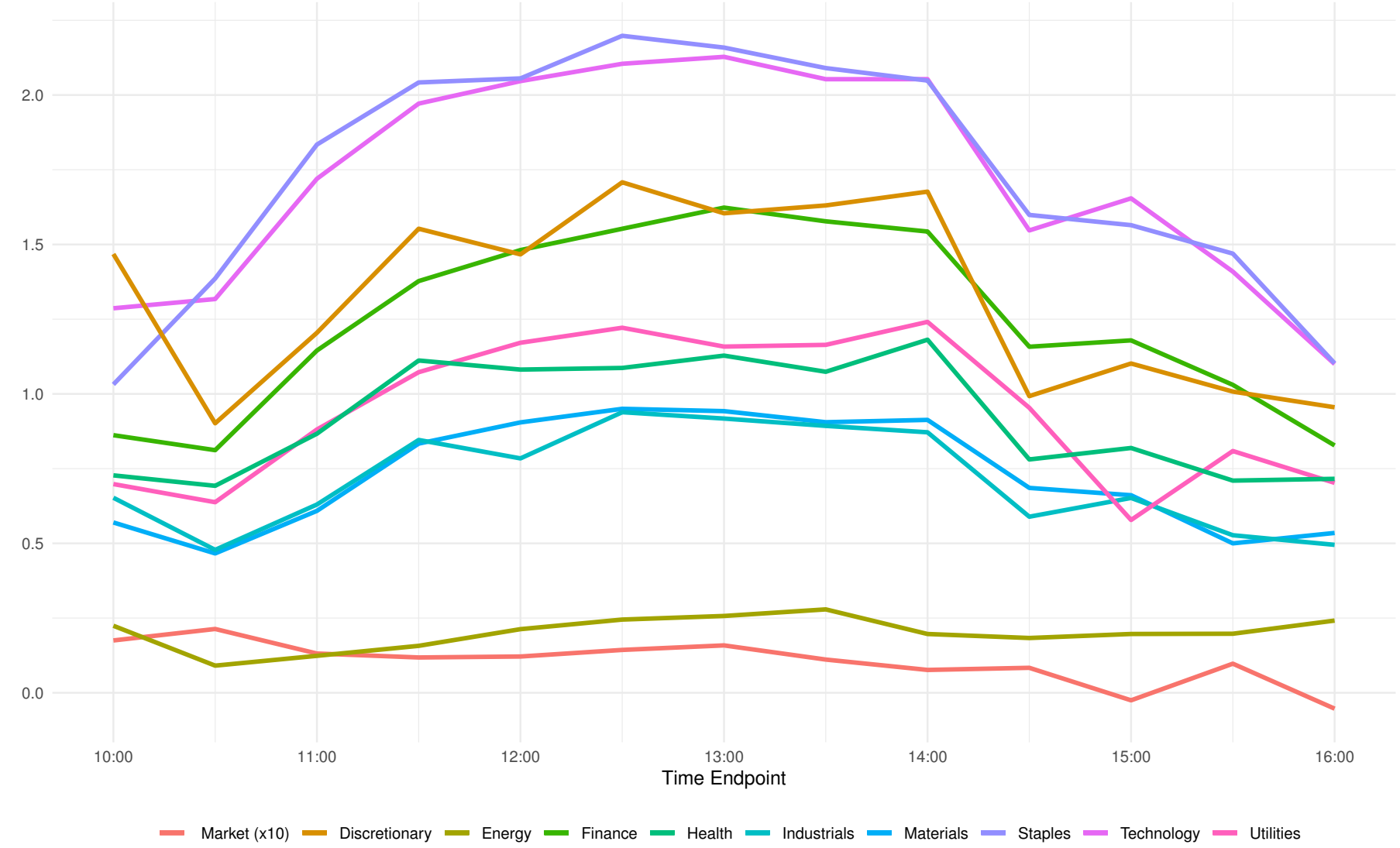


Figure 8: Characteristic Importances at the 1-min Horizon

This figure displays the characteristic importances for the total ensemble model (MED) at the 1-min horizon. Following Gu, Kelly, and Xiu (2020), the variable importance of the j^{th} predictor each month is measured by the reduction in R^2_{OS} from setting all values of predictor j to zero, while holding the remaining model estimates fixed. We then conduct a monthly Fama-MacBeth regression of the variable importance on each contemporaneous firm characteristic. We standardize both the variable importance and characteristics cross-sectionally each month to have a mean of zero and variance of one, allowing for the interpretation of the coefficient as the impact of a one standard deviation change in the characteristic on the standard deviation change in the variable importance. Plotted are the Fama-MacBeth coefficients, ranked by their absolute values, and shaded to indicate their Newey-West t-statistic value.

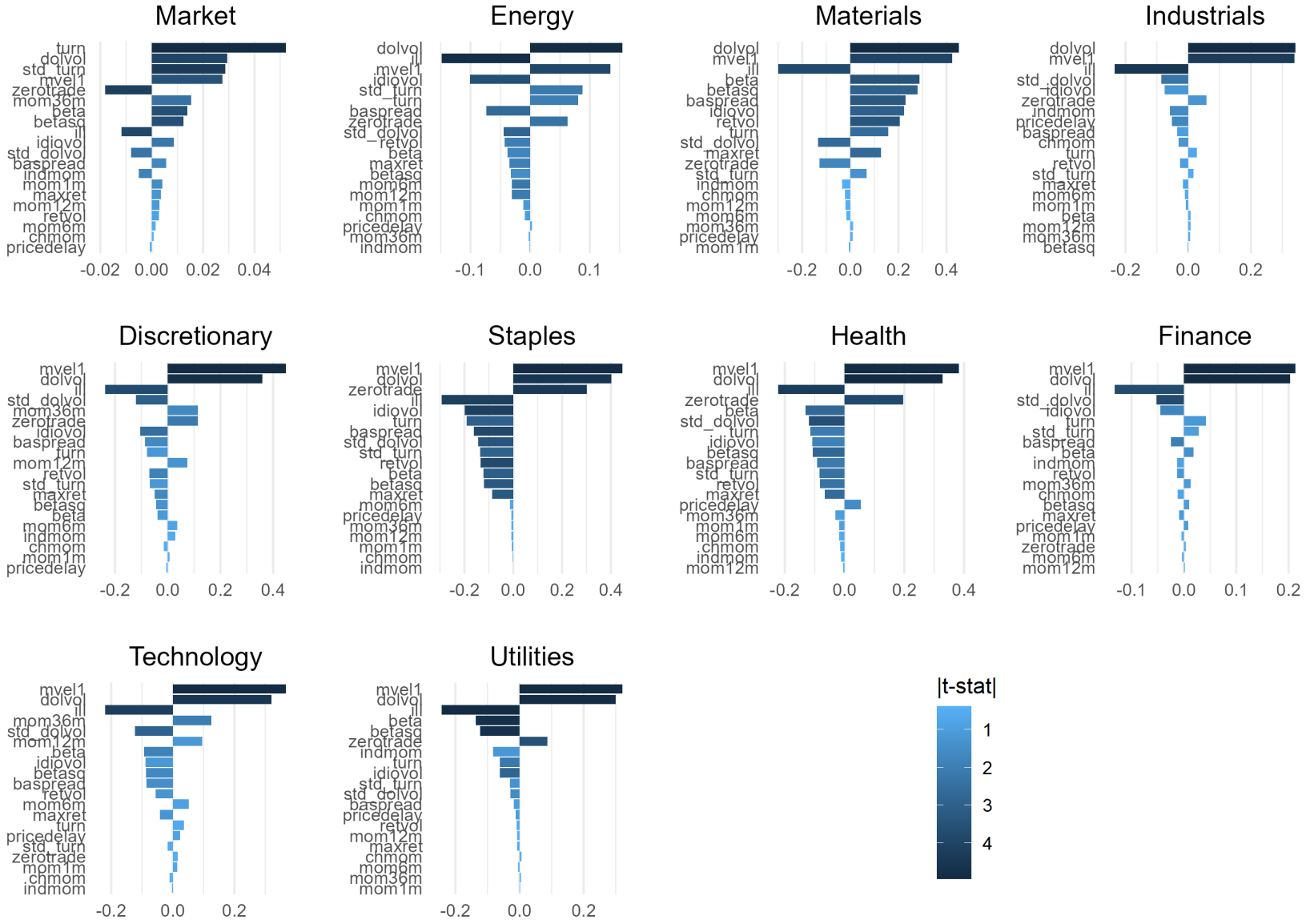
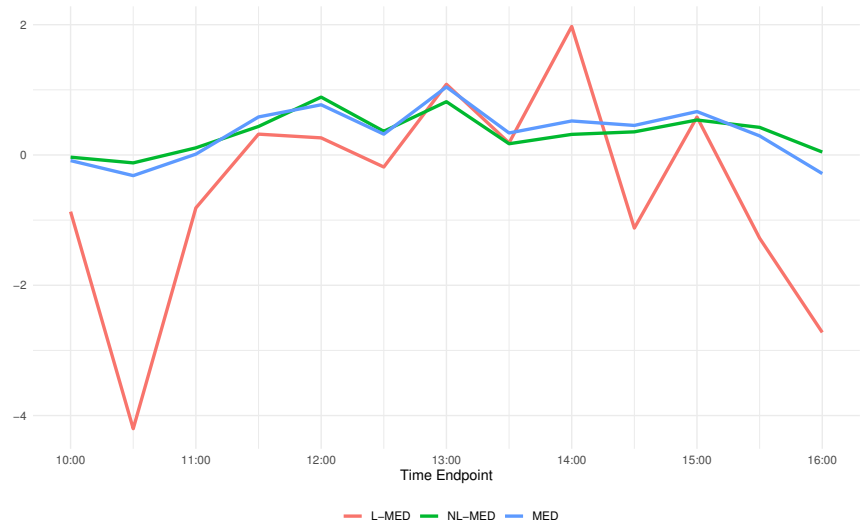
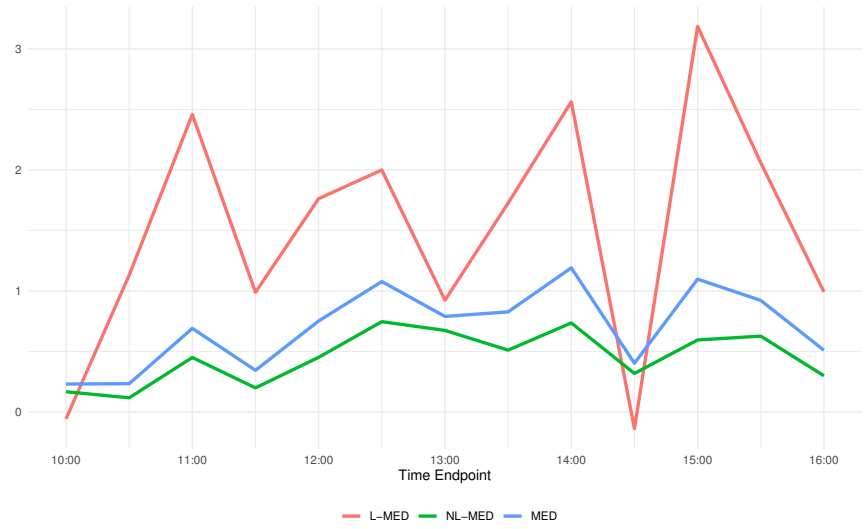


Figure 9: Intraday Dynamics of Knight Capital Natural Experiment

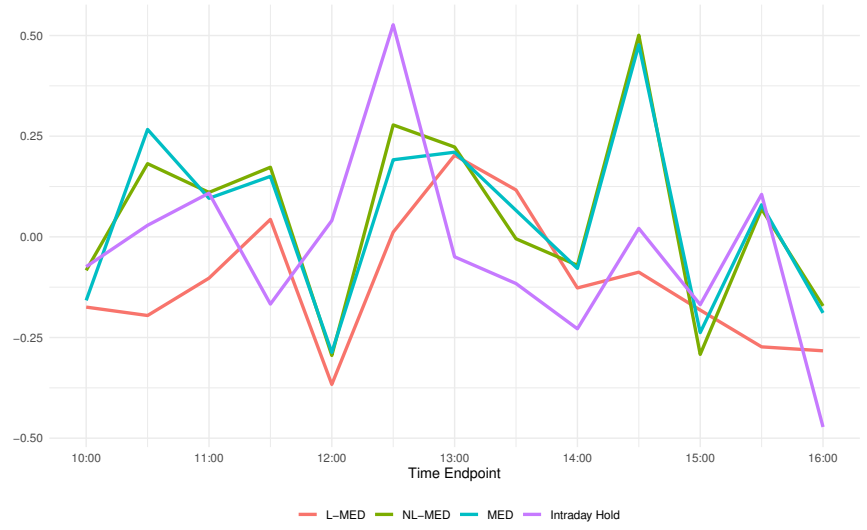
This figure presents model performance for stocks affected and unaffected by the Knight Capital glitch for the linear ensemble (L-MED), nonlinear ensemble (NL-MED), and total ensemble (MED) at the 1-min horizon in each half-hour interval from 9:30am to 4:00pm on the event day (August 1, 2012). Panels (a) and (c) report R^2_{OS} values and Sharpe ratios (SR) after transaction costs for affected firms, and Panels (b) and (d) report the same for unaffected firms. Sharpe ratios after transaction costs are computed from 1-min returns and are not annualized.



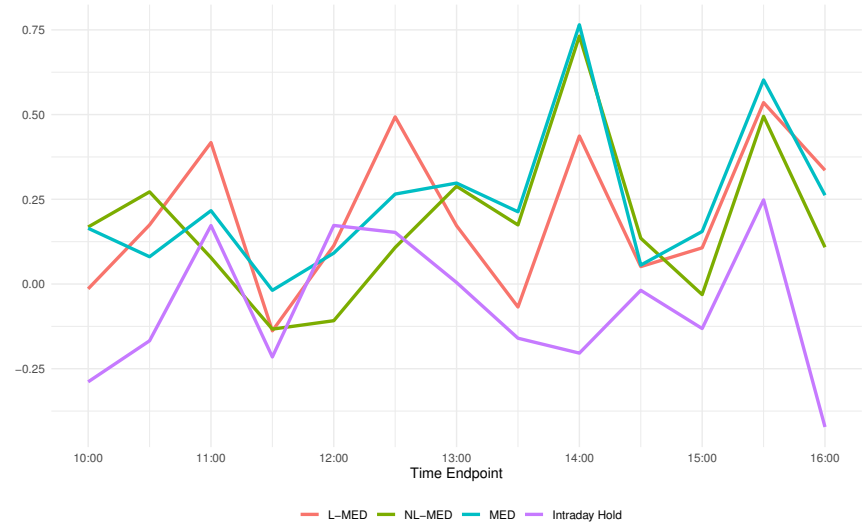
(a) R^2_{OS} of Affected Firms



(b) R^2_{OS} of Unaffected Firms



(c) SR of Affected Firms



(d) SR of Unaffected Firms

Intraday Stock Predictability Everywhere

Internet Appendix

Table of Contents:

- Section IA1 describes the data cleaning and variable construction procedures in detail.
- Section IA2 describes the hyperparameter optimization framework.
- Section IA3 contains additional tables.
- Section IA4 contains additional figures.

IA1 Trade and Quote (TAQ) Cleaning and Variable Construction

Trades and Quotes. This section provides further details on our TAQ data cleaning procedures. The TAQ data requires substantial cleaning due to market microstructure noise. We use the Monthly TAQ Second database for 1999 to 2003 and the Daily TAQ Millisecond/Microsecond database for 2004 to 2021, where the Monthly TAQ database provides coverage for the pre-2004 training and validation years used under yearly re-estimation. We filter noisy observations following procedures based on Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). We calculate returns from prices in the trades database and use bid and ask observations from the quotes database to clean trades and calculate transaction costs. We remove entries with timestamps outside of the trading day (9:30 AM to 4:00 PM) and entries with a zero or negative bid, ask, or price. For each stock, we keep only entries from the exchange with the highest volume in the month.

For the trades data, we delete entries with corrected trades (CORR $\neq$ 0) and entries with an abnormal sale condition, retaining only entries with COND equal to E, F, @, *, @E, @F, *E, and *F.We replace multiple trades with the same second timestamp with a single entry reflecting the median price. We also use quotes to filter trade prices, removing entries with a price above the ask plus the spread or below the bid minus the spread. We remove all entries with a negative spread (ask minus bid) and entries with a spread larger than 50 times the median spread on that day.

We aggregate the cleaned second-by-second price data into 1-min intervals, using the end-of-interval price. Returns at the 5-min, 10-min, 15-min, and 30-min horizons are obtained by summing the corresponding 1-min log returns. We exclude the first return of each day (e.g., 9:30 to 9:31 at the 1-min horizon) and the overnight return, and merge with the daily CRSP data file by CUSIP. We use CUSIPs rather than tickers to merge the TAQ and CRSP databases, because tickers often differ across databases and change over time due to mergers,

acquisitions, and other corporate events. We obtain the CUSIP identifier of each stock by merging trades with the TAQ master files.

Effective Bid-Ask Spread. To calculate the effective bid-ask spread for the test set from 2005 to 2021, we follow procedures based on Holden and Jacobsen (2014) using their publicly available SAS code.[23] We augment their code with the additional filters on the trades data described above. We then calculate the national best bid and national best ask prices from the TAQ NBBO dataset and compute the midquote price. Using the midquote, we calculate the percent effective bid-ask spread as $2|\log(\text{price}) - \log(\text{midpoint})|$.[24] We then aggregate to the second level using the share-weighted average of percent effective bid-ask spreads within each second. Finally, we use the end-of-minute effective bid-ask spread as the transaction cost measure, consistent with how we calculate prices and returns.

Intraday Factors. To construct 1-min versions of the Fama and French (2015) and Carhart (1997) factors, we first replicate their monthly factor construction on a sample consistent with their original papers (i.e., NYSE, NASDAQ, and AMEX) using data from CRSP and Compustat on WRDS.[25] The replication yields correlations of 99.99%, 97.86%, 98.17%, 98.26%, 98.12%, and 99.90% with the mkt, smb, hml, rmw, cma, and umd factors from Kenneth French's website, respectively. We then merge the component portfolios from the replication with the TAQ dataset to construct the six-factor model at the 1-min interval.

To construct the four factors from Pelger (2020) at the 1-min interval, we follow his procedures, restricting the sample to S&P 500 constituents. For the oil, finance, and electricity industries, we use the SIC code groupings from Pelger (2020) to identify firms in each in-

[23] We thank the authors for making their SAS code publicly available.

[24] We use the log definition of the percent effective bid-ask spread to ensure consistency with our use of log returns, as subtracting log spreads from log returns correctly accounts for transaction costs in our setting. This definition is commonly used in the microstructure literature (e.g., Bogousslavsky, Collin-Dufresne, and Sağlam (2021)). We also convert the intraday risk-free rates to log in the excess return formula. Log and arithmetic measures are nearly identical at 1-min intervals and produce very similar results.

[25] We use code from Wayne Chang's website and thank him for making the replication code publicly available.

dustry and calculate equal-weighted 1-min returns to obtain the three industry factors. We also use the equal-weighted market portfolio, consistent with Pelger (2020).

IA2 Hyperparameters

This section details the hyperparameter tuning framework used to estimate our models. Starting with penalized linear models (LAS, RID, and EN), the tuning parameter λ is the main hyperparameter that determines the level of regularization. We select λ using coordinate descent (see Friedman, Hastie, and Tibshirani (2010)). In the case of EN, the convex combination coefficient $\alpha \in [0, 1]$ is a hyperparameter to be optimized, with the endpoints $\{0, 1\}$ corresponding to the discrete choices of only ridge or only lasso regularization, respectively. In PCR, the number of principal components, $1 \leq \kappa \leq K$, is the main hyperparameter.

The tree-based RF, XRF, and GBT models rely on two essential hyperparameters: (1) the number of decision trees in the forest, and (2) the maximum depth of each of these trees, i.e., the maximum path length from any tree root to any of its leaves. More trees in the forest yield greater prediction variance reduction at increased computational cost. Restricting maximum tree depth limits the complexity, both saving computational time and avoiding over-fitting. In GBT models, we also consider the learning rate, ν, where a smaller rate often improves test performance but requires more training steps or trees. For the RF and XRF models, we also adjust the number of predictors used to determine node splits. More predictors permit greater complexity, resulting in bias reduction at the expense of increased prediction variance and computational cost. Additionally, we tune the minimum number of training samples per leaf, where a lower minimum results in more complex trees. For GBT, we use early stopping to halt training when the model stops improving on the validation set, which also helps prevent overfitting.

We adopt the feed-forward neural network architectures from Gu, Kelly, and Xiu (2020),

where NN1 has a single layer containing 32 units and NN5 has five layers containing 32, 16, 8, 4, and 2 units, respectively. All networks use the Rectified Linear Unit (ReLU) activation function $\phi(x) = \max\{0, x\}$. The loss function used for training is critical, since neural networks are highly sensitive to outliers. We use the Huber loss function, a combination of the ℓ^1 and ℓ^2 norms that permits control of the sensitivity to outliers based on a threshold (Huber loss is also used for GBT). We guard against model overfitting by adding ℓ^1 and ℓ^2 regularization to the loss function, and by averaging forecasts from the ten best-performing networks from the search, each with a different hyperparameter configuration. Other important factors involved in training include the number of iterations, or epochs, and the level of *dropout*, the percentage of network units randomly omitted in each training iteration to avoid over-fitting and regularize the optimization problem (see, e.g., Srivastava, Hinton, Krizhevsky, Sutskever, and Salakhutdinov (2014) for a discussion). Models are estimated using stochastic gradient descent optimization with the ADADELTA (Zeiler (2012)) adaptive learning rate method for faster estimation and improved generalization. Lastly, we use early stopping as a regularization technique to prevent overfitting.

Table IA1 summarizes the set of hyperparameters that are tuned for each model in this paper. All models are trained using the rolling window approach detailed in Section 2.3. We use grid search for PCR and random search with 50 models for all other models. Random search is more efficient than grid search (Bergstra and Bengio (2012)).

IA3 Additional tables

This section contains additional tables.

Table IA1: Model Hyperparameters

RID and LAS	λ
EN	λ, $\alpha = \{0, .25, .5, .75, 1\}$
PCR	κ = 1-99 by 2
RF and XRF	#trees = 100-700 by 100, tree depth = 1-40 by 5, #minimum rows per leaf = {1, 25, 50}, columns randomly selected = floor(#features * {.05, .15, .25, .333, .4})
GBT	#trees = 100-700 by 100, Huber threshold = 0.999, tree depth = {1, 2}, learning rate = {0.001, 0.01, 0.1}
NN1-NN5	Activation Function = ReLU with dropout, epochs = {50,100}, Huber threshold = 0.999, input dropout ratio = {0, 0.1, 0.2}, max w^2 = {10, 100, 1000, 3.4028235e+38}, ℓ^1 = {0, 0.00001, 0.0001}, ℓ^2 = {0, 0.00001, 0.0001}, rho (rate time decay) = {0.9, 0.95, 0.99, 0.999}, epsilon (rate time smoothing) = {1e-10, 1e-8, 1e-6, 1e-4}

This table describes the set of hyperparameters that are tuned for each individual model considered.

Table IA2: Autoencoder Portfolio Performance at the 1-min Horizon

Following Gu, Kelly, and Xiu (2021), we estimate latent factors using an autoencoder neural network with one hidden layer containing 6 hidden units. Each hidden unit uses the tanh activation function to capture nonlinearities, and we train the autoencoder for 100 epochs, where the inputs are the 1-min lagged returns z_t from Equation (3). We then use the nonlinear latent factors produced by the autoencoder to estimate the OLS, EN, NN1, NN2, and NN3 models with the same hyperparameters as in Table IA1. Panel A reports R^2_{OS} values and Panel B reports annualized Sharpe ratios after transaction costs.

Panel A: R^2_{OS} (in percent)					
Portfolio	AE_OLS	AE_EN	AE_NN1	AE_NN2	AE_NN3
Market	-7451253.58	0.00	-30691.25	-2485.92	-726.99
Materials	0.80	0.70	0.57	0.42	0.12
Energy	-98.56	0.13	-9.47	0.08	0.04
Finance	-2.00	0.30	0.23	0.27	0.07
Industrials	-0.08	0.53	0.32	0.29	0.08
Technology	0.44	0.84	0.65	0.73	0.27
Staples	-1.70	0.98	0.78	0.76	0.26
Utilities	0.92	0.86	0.62	0.47	0.15
Health	-158.41	0.79	-4.24	-0.74	0.11
Discretionary	0.88	0.83	0.69	0.48	0.18
Panel B: Annualized Sharpe Ratio After Transaction Costs					
Portfolio	AE_OLS	AE_EN	AE_NN1	AE_NN2	AE_NN3
Market	-0.33	0.38	0.16	0.32	-0.24
Materials	0.10	1.15	1.93	2.05	1.92
Energy	-0.26	0.57	-0.27	0.97	0.81
Finance	-0.29	1.07	0.21	1.23	1.23
Industrials	-0.24	0.40	-0.15	0.53	1.98
Technology	-0.17	-0.04	0.48	3.27	2.60
Staples	0.05	1.23	0.21	0.30	0.28
Utilities	0.16	0.67	0.60	1.80	1.24
Health	-0.23	1.31	-0.28	-0.16	-0.10
Discretionary	0.34	1.08	0.47	1.63	1.58

Table IA3: Yearly Re-estimation Portfolio Performance at the 1-min Horizon

This table presents results under yearly re-estimation for the 1-min horizon from 2005 to 2021. Under this timescale, we initially train models on five years of data (1999 to 2003), validate on one year (2004), and evaluate on a one-year test set (2005), then roll the window forward yearly. Panel A reports R^2_{OS} values and Panel B reports annualized Sharpe ratios after transaction costs.

Panel A: R^2_{OS} (in percent)																	
Portfolio	AR1	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	L-MED	NL-MED	MED
Market	-0.03	-0.88	-0.03	-0.31	-0.03	0.00	-0.01	0.00	0.00	-10.07	0.01	0.01	0.01	0.00	-0.01	0.02	0.02
Materials	-0.04	0.75	0.69	0.67	0.69	0.10	0.39	0.24	0.01	0.39	0.35	0.08	0.03	0.00	0.72	0.13	0.26
Energy	-0.04	0.08	0.03	-0.02	0.03	-0.14	0.03	0.04	0.00	0.04	0.02	0.01	0.00	0.00	0.02	0.01	0.04
Finance	-0.10	0.88	1.21	0.72	1.22	0.19	0.47	0.88	0.46	1.13	1.00	0.51	0.23	0.02	1.24	0.64	0.88
Industrials	-0.08	1.00	0.82	1.08	0.82	0.53	0.34	0.40	0.01	0.43	0.25	0.06	0.02	0.02	1.04	0.13	0.29
Technology	-0.13	2.24	2.54	2.20	2.58	0.87	1.29	1.22	0.08	-225.35	1.11	0.43	0.22	0.14	2.59	0.69	1.10
Staples	0.02	2.85	1.79	2.46	1.79	0.60	0.86	0.99	0.01	1.07	0.91	0.40	0.15	0.02	1.97	0.50	0.70
Utilities	-0.66	1.31	1.63	1.16	1.63	0.52	0.44	0.66	0.00	0.18	0.56	0.17	0.06	0.02	1.55	0.26	0.50
Health	-0.11	0.95	1.12	1.18	1.12	0.59	0.45	0.31	0.01	0.44	0.58	0.15	0.07	0.02	1.21	0.25	0.38
Discretionary	-0.03	1.64	0.78	1.38	0.84	0.74	0.54	0.47	0.05	0.57	0.33	0.06	0.03	0.02	1.08	0.19	0.32
Panel B: Annualized Sharpe Ratio After Transaction Costs																	
Portfolio	AR1	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	L-MED	NL-MED	MED
Market	-0.54	-1.93	-0.93	-1.73	-0.93	0.08	0.20	-0.33	-0.21	-0.60	-0.27	0.32	0.08	-0.39	-0.92	0.44	0.68
Materials	-0.43	-0.58	-0.30	-0.54	-0.30	-0.57	1.01	3.64	2.83	0.29	3.21	2.33	2.26	0.43	-0.35	3.74	3.29
Energy	0.04	-0.30	-0.39	-0.35	-0.39	-0.32	0.55	0.68	0.68	-0.37	0.53	0.99	0.48	-0.17	-0.22	1.10	1.29
Finance	0.00	0.00	0.40	-0.14	0.40	-0.02	2.17	1.79	2.03	0.77	1.52	2.96	2.18	1.75	0.39	3.07	1.36
Industrials	-0.23	0.03	1.17	0.32	1.17	-0.05	1.43	1.52	2.67	0.85	2.50	2.22	2.40	1.29	1.24	2.65	2.49
Technology	0.20	-0.06	0.15	-0.07	0.19	0.13	1.94	1.96	0.72	0.25	4.12	3.11	3.37	3.27	0.15	3.02	0.98
Staples	0.20	0.84	1.05	0.59	1.05	-0.21	1.11	1.10	3.68	0.70	3.51	2.54	2.63	1.49	1.20	2.37	2.45
Utilities	-0.43	0.03	0.53	-0.03	0.53	-0.38	1.31	1.56	0.24	-0.01	3.00	2.16	2.09	0.85	0.41	2.46	2.09
Health	0.12	-0.33	0.28	0.20	0.28	-0.27	1.07	0.00	2.51	-0.04	3.08	2.79	2.43	0.93	0.29	2.00	1.34
Discretionary	-0.22	0.89	0.59	0.80	0.78	0.42	1.31	1.23	0.36	0.59	2.64	2.37	1.93	0.60	1.11	1.89	2.25

Table IA4: Daily Re-estimation Portfolio Performance at the 1-min Horizon

This table presents results under daily re-estimation for the 1-min horizon from 2005 to 2021. Under this timescale, we initially train models on 20 days of data (December 2, 2004 to December 30, 2004), validate on one day (December 31, 2004), and evaluate on a one-day test set (January 3, 2005); we then roll the window forward daily. Panel A reports R^2_{OS} values and Panel B reports annualized Sharpe ratios after transaction costs.

Panel A: R^2_{OS} (in percent)

Portfolio	AR1	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	L-MED	NL-MED	MED
Market	-0.05	-12.83	-0.18	-4.05	-0.40	-0.38	-0.51	-0.43	-0.25	-2495932.54	-0.21	-0.13	-0.15	-0.18	-0.21	-0.06	-0.03
Materials	-0.07	1.94	1.69	1.61	1.68	0.62	0.29	0.37	0.01	-32.46	0.88	0.34	0.10	-0.02	1.78	0.45	0.83
Energy	-0.12	-0.53	0.24	-0.48	0.22	-0.04	-0.34	-0.29	-0.24	-15.13	0.09	-0.03	-0.10	-0.16	0.26	0.10	0.21
Finance	0.21	1.19	2.18	1.36	2.13	0.45	0.64	0.65	0.32	-16394.59	1.36	0.65	0.24	0.02	2.20	0.82	1.28
Industrials	-0.06	0.62	1.53	0.80	1.51	0.48	0.21	0.23	0.00	-129.03	0.73	0.25	0.09	0.00	1.61	0.39	0.71
Technology	-0.06	2.65	3.14	2.56	3.04	1.38	1.15	1.15	0.22	-41.98	1.65	0.74	0.32	0.09	3.23	1.02	1.64
Staples	-0.18	2.52	2.69	2.19	2.82	1.31	1.00	0.99	0.19	-8.87	2.05	1.00	0.44	0.11	2.98	1.13	1.67
Utilities	-0.06	1.43	1.81	1.40	1.78	0.90	0.51	0.52	0.02	-110259.20	1.20	0.52	0.20	0.01	1.89	0.66	1.01
Health	-0.12	0.72	1.91	1.00	1.84	0.77	0.17	0.19	-0.04	-5.88	1.17	0.39	0.20	0.06	2.06	0.56	0.98
Discretionary	-0.11	1.70	1.98	1.52	1.92	0.76	0.54	0.58	0.02	-63306534.38	1.15	0.47	0.21	0.02	2.12	0.65	1.09

Panel B: Annualized Sharpe Ratio After Transaction Costs

Portfolio	AR1	OLS	LAS	RID	EN	PCR	RF	XRF	GBT	NN1	NN2	NN3	NN4	NN5	L-MED	NL-MED	MED
Market	-0.41	-1.85	-0.50	-1.17	-0.95	-0.88	-1.02	-1.03	-1.03	0.11	-0.53	-0.45	-0.61	-0.68	-0.72	-0.42	-0.45
Materials	-0.37	0.91	1.09	0.82	0.90	0.01	-0.40	-0.27	-0.46	0.26	1.14	0.46	0.64	-0.20	0.98	1.11	1.13
Energy	-0.43	-0.29	0.39	-0.29	0.35	-0.41	-1.01	-0.87	-0.68	-0.39	0.20	-0.29	-0.32	-0.42	0.27	-0.02	0.29
Finance	0.29	0.02	0.62	0.13	0.58	0.22	-0.03	-0.04	-0.13	0.22	1.00	2.65	1.32	0.28	0.56	2.69	0.92
Industrials	-0.47	-0.17	1.10	-0.06	1.02	0.01	-0.74	-0.62	-0.47	-0.27	0.55	-0.01	-0.21	-0.38	1.07	0.78	0.70
Technology	-0.40	0.61	1.70	0.41	1.37	0.35	0.53	0.83	0.01	-0.05	1.31	0.61	0.64	-0.23	1.70	2.05	1.40
Staples	-0.24	0.05	0.14	-0.08	0.12	-0.16	-0.08	-0.11	-0.24	0.01	1.02	1.11	2.20	0.70	0.13	1.59	0.55
Utilities	-0.12	0.38	0.67	0.51	0.64	0.30	-0.31	-0.52	-0.49	0.02	0.38	1.33	0.14	-0.10	0.79	1.28	1.06
Health	-0.27	-0.23	0.09	-0.29	-0.01	-0.25	-0.57	-0.47	-0.54	-0.48	0.66	0.38	-0.13	-0.28	0.06	0.77	1.07
Discretionary	-0.18	0.90	1.01	0.40	0.89	-0.05	0.22	0.33	0.03	0.26	1.16	0.54	0.24	-0.22	1.17	1.02	1.35

Table IA5: Minutely Re-estimation Portfolio Performance at the 1-min Horizon

This table presents results under minutely re-estimation for the 1-min horizon from 2005 to 2021. Following Chinco, Clark-Joseph, and Ye (2019), we implement a minutely re-estimation timescale in which we initially train models on 30 minutes of data (9:32 to 10:01), validate via ten-fold cross-validation, and evaluate on a 1-min test set (10:02); we then roll the window forward minutely for 359 minutes each day. Due to computational constraints, we restrict this analysis to portfolios at the 1-min horizon and include only AR1, OLS, RID, LAS, and EN. Panel A reports R^2_{OS} values and Panel B reports annualized Sharpe ratios after transaction costs.

Panel A: R^2_{OS} (in percent)						
Portfolio	AR1	OLS	LAS	RID	EN	L-MED
Market	-9.01	-206.43	-3.91	-91.52	-6.61	-6.28
Materials	-7.61	-592781.97	-3.61	-78.87	-8.28	-7.90
Energy	-7.32	-6256186.44	-3.62	-65.46	-8.58	-8.31
Finance	-7.27	-126.89	-3.66	-89.36	-8.49	-8.31
Industrials	-7.54	-371.54	-3.73	-437.20	-7.82	-7.91
Technology	-7.63	-162.33	-3.52	-92.42	-7.60	-7.76
Staples	-9.58	-2322448.49	-3.34	-77.99	-7.48	-7.64
Utilities	-14.21	-118770857.42	-3.96	-70.21	-11.27	-10.29
Health	-7.90	-19601.15	-3.70	-1506.17	-7.76	-8.33
Discretionary	-9.68	-1479.88	-3.58	-193.89	-7.36	-7.58
Panel B: Annualized Sharpe Ratio After Transaction Costs						
Portfolio	AR1	OLS	LAS	RID	EN	L-MED
Market	-1.80	-1.72	-2.11	-2.57	-1.99	-1.79
Materials	-1.72	-2.48	-1.40	-2.82	-1.63	-1.65
Energy	-2.72	-1.07	-2.33	-3.25	-2.72	-2.69
Finance	-1.22	-1.70	-1.09	-1.87	-1.31	-1.25
Industrials	-1.47	-2.16	-1.26	-1.95	-1.40	-1.42
Technology	-1.25	-2.32	-1.22	-2.32	-1.11	-1.04
Staples	-1.24	-1.82	-0.64	-2.57	-0.87	-0.85
Utilities	-0.55	-3.70	-0.99	-2.52	-0.66	-0.71
Health	-1.57	-3.18	-1.39	-3.21	-1.57	-1.63
Discretionary	-1.67	-2.21	-1.40	-2.47	-1.52	-1.51

IA4 Additional figures

This section contains additional figures.

Figure IA1: Portfolio Time-of-Day Predictability at the 1-min Horizon: R^2_{OS} (in percent) with a 60-Month Training Window

This figure presents, for the market and energy portfolios, the R^2_{OS} of the total ensemble (MED) in each half-hour interval from 9:30am to 4:00pm at the 1-min horizon with monthly re-estimation and a 60-month training window.

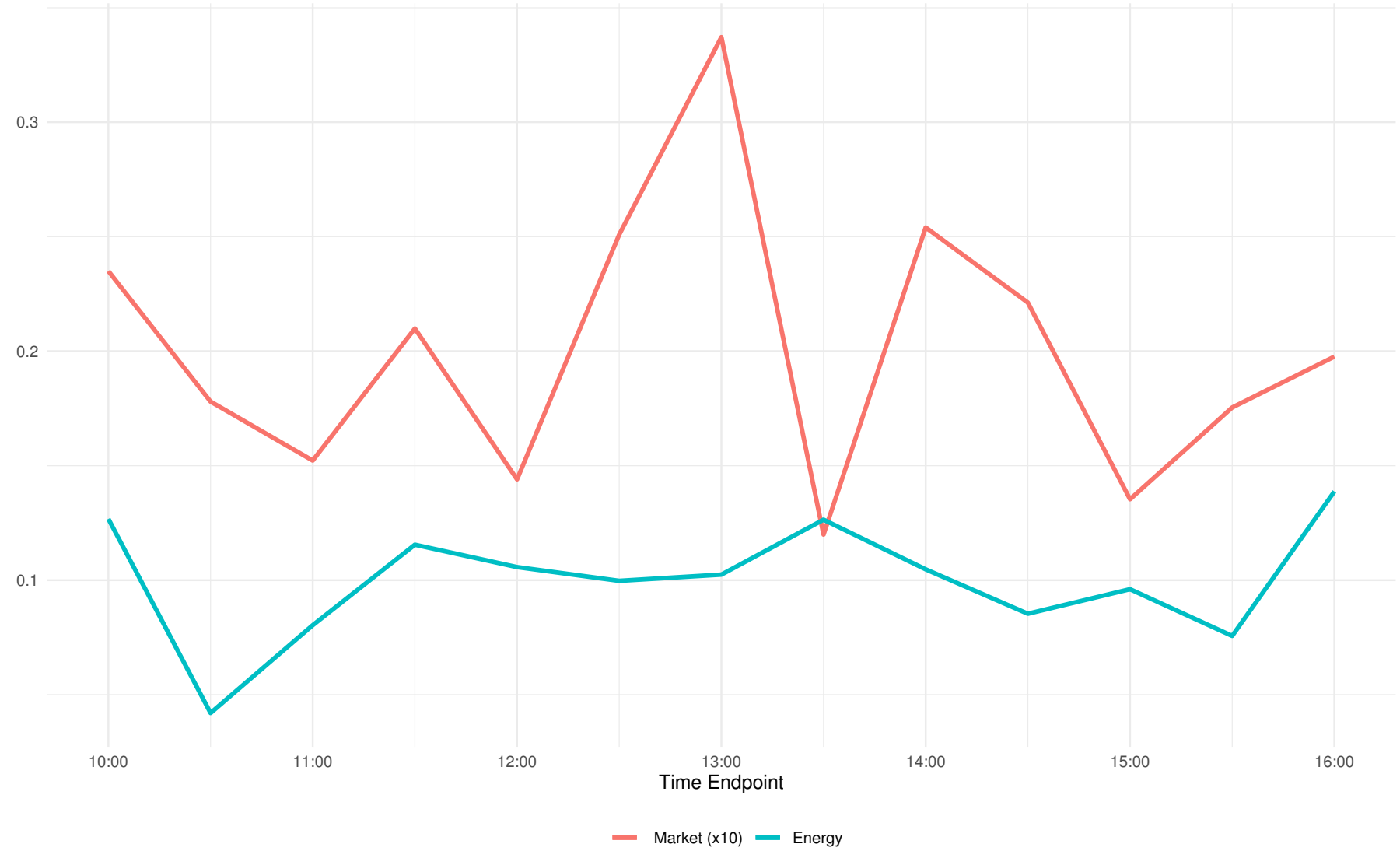


Figure IA2: Portfolio Strategy After Transaction Costs at the 1-min Horizon (Annualized Sharpe Ratio) with Yearly Re-estimation Timescale

This figure presents annualized Sharpe ratios of strategies based on model forecasts after transaction costs for the market and sector portfolios at the 1-min horizon from January 2005 to December 2021 under yearly re-estimation, comparing the median ensembles of all linear models (L-MED), all nonlinear models (NL-MED), and all models (MED). Under this timescale, we initially train models on five years of data (1999 to 2003), validate on one year (2004), and evaluate on a one-year test set (2005), then roll the window forward yearly. We test the equality of Sharpe ratios between L-MED and NL-MED using the two-sided test of Leung and Wong (2008). *, **, and *** indicate the Sharpe ratios are statistically different at the 10%, 5%, and 1% significance levels, respectively.

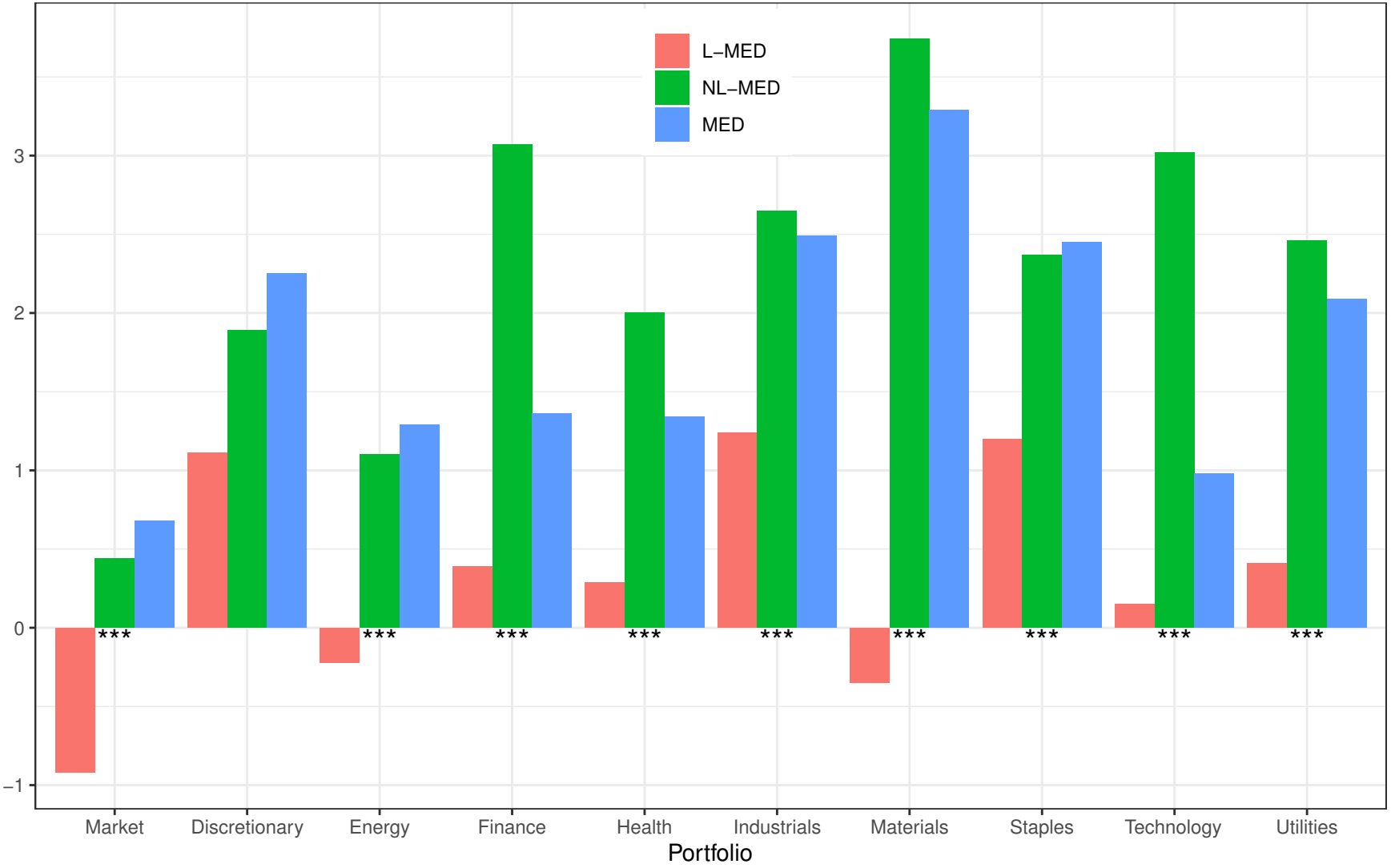


Figure IA3: Portfolio Strategy After Transaction Costs at the 1-min Horizon (Annualized Sharpe Ratio) with Daily Re-estimation Timescale

This figure presents annualized Sharpe ratios of strategies based on model forecasts after transaction costs for the market and sector portfolios at the 1-min horizon from January 2005 to December 2021 under daily re-estimation, comparing the median ensembles of all linear models (L-MED), all nonlinear models (NL-MED), and all models (MED). Under this timescale, we initially train models on 20 days of data (December 2, 2004 to December 30, 2004), validate on one day (December 31, 2004), and evaluate on a one-day test set (January 3, 2005); we then roll the window forward daily. We test the equality of Sharpe ratios between L-MED and NL-MED using the two-sided test of Leung and Wong (2008). *, **, and *** indicate the Sharpe ratios are statistically different at the 10%, 5%, and 1% significance levels, respectively.

