Stock Distribution and Risk Premia Predictability with Quantile Machine Learning

Fred Liu[*†‡]

November 18, 2025

Abstract

We study the predictability of future stock return distributions and risk premia using quantile machine learning models trained on firm characteristics and macroeconomic variables. We introduce a neural network architecture that jointly estimates multiple conditional quantiles while ensuring non-crossing by construction, along with a quantile-based risk premium measure that is robust to outliers and heteroskedasticity. Multi-quantile neural networks outperform linear, tree-based, and feed-forward neural network models, and the robust risk premium forecasts significantly outperform those from conventional machine learning approaches. Finally, we show that return distribution predictability is driven by investor inattention, information frictions, and underreaction to information shocks.

JEL Classification: C21, C45, G0, G1

Keywords: Machine Learning, Robust Risk Premium, Multi-quantile Neural Networks, Quantile Regression, Quantile Spacings

---

[*]We benefited tremendously from discussions with Fahiz Baba-Yara, Leland Bybee (Discussant), Ba Chu (Discussant), Philippe Coulombe, Thibaut Duprey, Robert Engle, Christian Fieberg, Adlai Fisher, Gustavo Freire, Christian Gourieroux, Roxana Halbleib, Bryan Kelly, Aaron Kim (Discussant), Suzanne Lee (Discussant), Gerrit Liedtke, Yueliang Lu (Discussant), Semyon Malamud, Alex Maynard, Ilias Tsiakas, Mihail Velikov, Qinfeng Xu, Wenjun Zhou, Yi Zhou, and seminar and conference participants at the 2025 Finance Management Association in Vancouver, Canadian Econometrics Study Group in Ottawa, Bank of Canada, Canadian Quantitative and Macro Investment Conference in Ottawa, Quantitative Finance and Financial Econometrics Conference in Marseille, Eastern Finance Association in Philadelphia, Midwest Finance Association in Chicago, 2024 Society of Financial Econometrics Summer School organized by NYU Shanghai, International Risk Management Conference at Bocconi University, Generative AI in Finance Conference at Concordia University, and Canadian Economics Association at Toronto Metropolitan University. We gratefully acknowledge support from SSHRC Insight Development Grant 430-2022-00399, and the Digital Research Alliance of Canada Resources for Research Groups 2022 competition (ID 4286). We thank Robert Proner for excellent research assistance. This paper was previously circulated under the title "Quantile Machine Learning and the Cross-Section of Stock Returns" (First version July, 2023).

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

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

1 Introduction

Predictions of stock return quantiles are economically important, as they provide a comprehensive view of the entire conditional distribution of a stock's return, which is valuable to industry and academia. Investors use the distribution for risk management and portfolio allocation, while financial economists use it to test asset pricing models (see Bali, Engle, and Murray (2016) for a comprehensive survey) and to estimate key economic measures, notably the risk premium, which is defined as the expectation over the excess return distribution. However, predicting stock return quantiles is notoriously challenging, because returns exhibit heavy tails, distributional heterogeneity; and depend on many covariates and their complex nonlinear interactions.

Since the seminal work of Koenker and Bassett Jr (1978), financial economists have used quantile regression models to capture heavy tails and heterogeneity in stock return distributions (e.g., Engle and Manganelli (2004)). However, most quantile regression models in the traditional econometrics literature are linear and unable to accommodate a large number of predictors with complex nonlinear interactions, limiting their predictive accuracy. The asset pricing literature has demonstrated that stock returns are a function of hundreds of firm characteristics and macroeconomic variables (Cochrane (2011)). Furthermore, theoretical economic models indicate that return quantiles behave nonlinearly, particularly under certain macroeconomic conditions (e.g., Brunnermeier and Sannikov (2014), He and Krishnamurthy (2013), and Pohl, Schmedders, and Wilms (2018)).

Given the high dimensionality and nonlinearities inherent in predicting return quantiles, machine learning is well suited for such challenging problems through modeling nonlinear relationships while guarding against overfitting on a high-dimensional set of predictors. A seminal contribution to off-the-shelf machine learning models, estimated under mean squared error (MSE) loss, for forecasting stock risk premia is Gu, Kelly, and Xiu (2020) (henceforth GKX).[1] However, it is still little known whether machine learning can significantly improve

[1]For additional examples, see Aleti, Bollerslev, and Siggaard (2025), Azevedo, Hoegner, and Velikov

forecasts of stock return distributions, whether such improvements enhance predictions of risk premia and their economic significance, and if so, what the economic sources of these gains are.

This paper conducts, to our knowledge, the first comprehensive study of the cross-sectional predictability of stock return quantiles estimated using a variety of machine learning techniques trained on a large set of firm characteristics and macroeconomic variables, and clearly demonstrates their advantages over traditional techniques. Traditional econometric techniques typically rely on linear specifications, which lack the flexibility needed to capture the complex functional forms of conditional quantiles, resulting in inaccurate out-of-sample predictions. In contrast, this paper applies machine learning techniques to estimate quantiles with more flexible models without imposing strong ex-ante restrictions. This enhanced flexibility is essential for estimating the unknown and complex functional form of return quantiles while guarding against overfitting using regularization approaches embedded in machine learning techniques.

In the same fashion as GKX, we apply a variety of off-the-shelf linear and nonlinear models to predict the stock return quantile at each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$. Linear models include linear quantile regression (LR) and elastic net (EN) quantile regression. Nonlinear models include quantile random forest (RF), quantile gradient boosted regression trees (GBT), and quantile feed-forward neural network (NN) architectures with 1-4 hidden layers (NN1,...,NN4). In terms of estimation, most of these models (except RF) estimate each quantile separately, which implicitly assumes that the functional forms of different quantiles are independent. However, from an economic perspective, this assumption is unlikely to hold, as the conditional return quantiles for a given firm-month are typically highly related and share a common structure.

To incorporate this economic structure, we introduce two economically motivated neural

(2023), Chinco, Clark-Joseph, and Ye (2019), Choi, Jiang, and Zhang (2025), Freyberger, Neuhierl, and Weber (2020), Huddleston, Liu, and Stentoft (2023), Jiang, Kelly, and Xiu (2023), Leippold, Wang, and Zhou (2022), and Liu and Stentoft (2023)).

network architectures. First, we develop a multi-quantile neural network (MQ) architecture that jointly estimates multiple quantiles within a single functional form. This architecture incorporates an economic restriction of a common structure across the quantiles by sharing latent features across quantiles. MQs utilize a technique called multi-task learning, which enhances model generalization by implicitly improving regularization, and data efficiency by concurrently learning several related tasks. Second, we develop a spaced multi-quantile neural network (SMQ) architecture that extends the MQ to resolve violations of monotonicity in estimated conditional quantile functions, a longstanding issue known as the quantile crossing problem. To our best knowledge, this is the first neural network architecture in the economics literature that guarantees non-crossing quantiles by construction. We include MQ and SMQ architectures with one to two shared layers (MQ1, MQ2 and SMQ1, SMQ2).

We conduct an empirical analysis on a large cross-section of individual stocks spanning 64 years, from 1957 to 2020. Quantile predictive power is evaluated using the $R^1_{os}(\tau)$ statistic from Koenker and Machado (1999), a pseudo-R^2 measure that benchmarks model forecasts against a naive historical quantile. We find that traditional LR fails to deliver out-of-sample forecasting power for conditional quantiles, however the penalized linear model, EN, improves performance throughout. The inclusion of nonlinear interactions through RF does not enhance predictability, as it estimates quantiles without the quantile objective function. In contrast, GBT, estimated with the quantile objective function, achieves higher $R^1_{os}(\tau)$ values than RF across most quantiles. NNs further improve predictive accuracy relative to both linear and tree-based models. MQ and SMQ models are the best performers. Pairwise Diebold–Mariano tests indicate that the MQ and SMQ models significantly outperform the off-the-shelf models across most quantiles. Finally, we show that GBT, LR, and EN exhibit the highest incidence of quantile crossings, whereas MQs have few and SMQs have none by construction.

We introduce a robust measure of the risk premium by integrating over quantiles. This measure is less sensitive to outliers and accommodates heteroskedasticity. We first mathe-

matically prove that the robust risk premium converges to the expected value under relatively weak assumptions, establishing it as an estimator of the risk premium. Empirically, SMQ2 achieves the highest R^2_{os} value of 1.001%, followed closely by MQ1 with 0.989%, while LR records the lowest value of 0.131%. Pairwise Diebold–Mariano tests further show that the MQ and SMQ models significantly outperform linear and tree-based models in forecasting risk premia.

We proceed to investigate the economic significance of the quantile machine learning techniques by forming value-weighted portfolios based on the robust risk premium forecasts, and trading a long-short portfolio that buys stocks with the highest robust risk premium forecasts (decile 10) and sells stocks with the lowest (decile 1). These portfolios deliver high annualized Sharpe ratios. The MQ and SMQ models achieve Sharpe ratios between 1.50 and 1.55, substantially outperforming linear and tree-based models, which record values between 0.77 and 1.10, as well as the NN models, whose Sharpe ratios range from 1.26 to 1.44. The MQ and SMQ models also earn monthly alphas between 2.51% (t-stat = 10.06) and 2.72% (t-stat = 9.47), along with information ratios ranging from 1.52 to 1.58, with respect to the Fama–French five-factor model (Fama and French (2015)) augmented with the momentum factor.

Next, we compare the performance of the robust risk premium and MSE-based risk premium forecasts within the same model class. The robust risk premium delivers superior predictive performance, with Diebold–Mariano statistics indicating statistically significant improvements for the NN2, NN3, NN4, MQ1, SMQ1, and SMQ2 models. Spanning tests further show that long–short portfolios formed using the robust risk premium generate significantly higher economic gains than those based on MSE-based forecasts across all models. Furthermore, the robust risk premium portfolios exhibit smaller drawdowns, higher break-even transaction costs, stronger net-of-transaction-cost performance, and superior results after excluding microcap stocks, demonstrating that the outperformance persists under realistic economic constraints (Avramov, Cheng, and Metzker (2022)).

Furthermore, we investigate the economic sources underlying return distribution predictability. Motivated by the theoretical framework of Hong and Stein (1999), we hypothesize that investors tend to underreact to new information, leading to return distribution predictability. The evidence supports this hypothesis: return distribution predictability is more pronounced among stocks receiving less investor attention, facing greater information frictions, or experiencing larger information shocks. For example, a long–short portfolio based on the robust risk premium earns a Sharpe ratio of 1.88 in the tercile of stocks with the least investor attention, which monotonically decreases to 1.01 in the tercile with the greatest attention. Similarly, both $R^1_{os}(\tau)$ and R^2_{os} values decline as investor attention increases. A parallel pattern emerges with information frictions: a long–short portfolio sorted on the robust risk premium earns a Sharpe ratio of 0.90 in the quintile with the lowest frictions, increasing monotonically to 2.94 in the quintile with the highest. Variable importance analysis identifies volatility, illiquidity, and price-trend variables as the most influential predictors, further supporting information frictions as a primary driver of return distribution predictability. Collectively, these results suggest that machine learning models effectively capture misvaluations arising from investor behavioral biases.

Our main findings remain robust to several alternative specifications and extensions. First, we present empirical evidence that the robust risk premium outperforms other quantile-based estimators of the conditional expectation proposed in the literature, achieving higher R^2_{os} values and Sharpe ratios across most models. Second, we find that increasing the number of quantiles used to estimate the robust risk premium yields only marginal, if any, improvements to statistical or economic performance. Third, we introduce a quantile-based measure of variability-adjusted returns and demonstrate that incorporating variability into the portfolio strategy substantially enhances performance, with the MQ and SMQ models achieving Sharpe ratios between 1.99 and 2.13. Taken together, these results confirm that quantile machine learning and the robust risk premium substantially improve forecasts of stock return distributions and risk premia, leading to large economic gains.

The paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the methodological framework, describes the machine learning models, and defines the quantile-based risk premium measure. Section 4 outlines the data and the approach used to evaluate model performance. Section 5 reports the main empirical findings. Section 6 investigates the economic sources of return distribution predictability. Section 7 presents additional robustness checks and analyses to further support the main results. Section 8 concludes.

2 Related Literature

Our paper contributes to at least three existing strands of literature. First, it contributes to the rapidly expanding literature applying off-the-shelf machine learning techniques to forecast stock returns, aligning most closely with the seminal paper by Gu, Kelly, and Xiu (2020), who apply a variety of off-the-shelf machine learning techniques, estimated under MSE loss, to forecast the *conditional expectation* of stock returns. We build directly on their work, by applying machine learning techniques trained on firm characteristics and macroeconomic variables to forecast the *conditional quantiles* of stock returns. This distributional approach allows us to introduce a quantile-based risk premium measure that is robust to outliers and heteroskedasticity, and delivers significantly improved risk premium forecasts relative to MSE-based models.

Second, our paper is related to the papers that incorporate economic restrictions into machine learning models to improve forecast accuracy and interpretability. Aleti and Bollerslev (2025), Bryzgalova, Pelger, and Zhu (2023), Chen, Pelger, and Zhu (2023), Feng, He, Polson, and Xu (2024), Gu, Kelly, and Xiu (2021), Kelly, Pruitt, and Su (2019), and Kozak, Nagel, and Santosh (2020) develop machine learning techniques with the economic restriction of no-arbitrage for forecasting the *conditional expectation* of stock returns. Jensen, Kelly, Malamud, and Pedersen (2022) incorporate trading costs and use machine learning

with economic restrictions to optimize portfolios. Guijarro-Ordonez, Pelger, and Zanotti (2021) develop a deep learning approach for statistical arbitrage. Cotturo, Liu, and Proner (2024) introduce dynamic multi-task neural networks for factor timing. Our paper introduces two multi-quantile neural network architectures that incorporate an economically motivated common structure across return quantiles, including one that guarantees non-crossing by construction, and uses them to predict the *conditional distribution* of stock returns, yielding significant statistical improvements and economic gains over models without these restrictions.

A few contemporaneous papers study distribution forecasts using machine learning techniques. Keilbar and Wang (2022) use quantile neural networks trained on past returns to model systemic risk. The use of machine learning to quantify uncertainty is related to Allena (2025). However, Allena (2025) uses machine learning to predict the *variance of the estimated expectation*, while our quantile-based methodology predicts the *distribution of future returns*. Independently, Yang, Zhu, Li, and Zhu (2024) extend the autoencoder model of Gu, Kelly, and Xiu (2021) to estimate latent factors and quantile-specific betas, which they use to construct quantile-based approximations of the conditional mean and variance. In contrast, our paper compares a range of machine learning methods to examine the cross-sectional predictability of stock return quantiles, their asset pricing implications, and the underlying economic sources.

Third, our paper is related in spirit to the literature on linear quantile regression in financial economics. Beginning with Koenker and Bassett Jr (1978), there has been an extensive literature on applying linear quantile regression in economics (see Koenker, Chernozhukov, He, and Peng (2017) for a survey). Recently, there has been a growing literature on applying linear quantile regression in empirical finance. Giglio, Kelly, and Pruitt (2016) use a linear quantile regression approach to assess how well various systemic risk measures proposed in the literature predict real economic activity. Adrian, Boyarchenko, and Giannone (2019) find that the lower conditional quantiles of future GDP growth in the US are sensitive to

current financial conditions. Menkveld, Dreber, Holzmeister, Huber, Johannesson, Kirchler, Neusüss, Razen, and Weitzel (2023) use linear quantile regression and the interquartile range to estimate non-standard errors. Cenesizoglu and Timmermann (2008), Pedersen (2015), and Schmidt and Zhu (2016) use linear quantile regression to forecast the distribution of stock market returns. Meligkotsidou, Panopoulou, Vrontos, and Vrontos (2014) and Lima and Meng (2017) examine the use of combined quantiles to estimate market risk premia, while Ma and Pohlman (2008) and Yang, Zhu, Li, and Zhu (2024) apply similar approaches to estimate stock risk premia. Ghysels, Plazzi, and Valkanov (2016) develop a quantile-based measure of conditional skewness and investigate its economic gains in international markets. Ando and Bai (2020) report that risk premia are quantile dependent, and that the common factors explaining the upper and lower tails of global asset returns are different after the subprime crisis. Chen, Dolado, and Gonzalo (2021) use a linear quantile factor model to show that the common factors driving the median of stock returns are different from those driving the tails.

The methodological contributions of our paper specifically to the literature on quantile regression in financial economics are threefold. First, we apply quantile machine learning techniques to model cross-sectional stock returns as nonlinear functions of firm characteristics and macroeconomic variables, showing that neural networks significantly outperform linear quantile regression. Second, we introduce the MQ and SMQ architectures, which incorporate economic structure through multi-task learning, while SMQ also resolves the quantile crossing problem, and show that they generally outperform off-the-shelf quantile machine learning models. Third, we develop a robust risk premium measure by integrating across quantiles, leveraging the inherent stability of quantile-based methods to improve robustness against outliers, and demonstrate substantial gains over MSE-based approaches.

3 Methodology

We index stocks by $i = 1, ..., N$ and months by $t = 1, ..., T$. In its most general form, we describe a stock return's conditional quantile given $\tau \in (0, 1)$ as:

$$Q_\tau \equiv Q_{r_{i,t+1}}(\tau|\mathcal{F}_t) = g_\tau(z_{i,t}), \tag{1}$$

where $r_{i,t+1}$ is the excess return of stock i at time $t+1$. Our objective is to estimate a flexible function $g_\tau(z_{i,t})$ that links the conditional quantile of future returns Q_τ to a P-dimensional vector $z_{i,t}$ of predictors.

3.1 Estimation

3.1.1 Quantile Objective Function

Unlike most regression-based machine learning models, which rely on the MSE objective function, quantile machine learning models typically employ the quantile objective function:

$$\mathcal{L}_\tau(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau \left(r_{i,t+1} - g_\tau(z_{i,t}; \theta) \right), \tag{2}$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}\{u \leq 0\})$ is the quantile loss function. For example, for $\tau = 0.50$, the objective function in Equation (2) is the symmetric mean absolute error (MAE), and is minimized by the conditional median. For $\tau > 0.50$ ($\tau < 0.50$), the objective function in Equation (2) is asymmetric and imposes a higher (lower) penalty on realized returns that exceed the predicted quantile, and is minimized by the respective conditional quantile. In our baseline analysis, we focus on $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$, which are commonly studied in the literature (e.g., Adrian, Boyarchenko, and Giannone (2019); Catania and Luati (2023)). To assess robustness of the main results, Section 7.2 examines the sensitivity of results to alternative quantile specifications.

Financial returns are known to exhibit heavy tails (Mandelbrot (1963), Fama (1963),

Fama (1965)). The MSE objective function places a large emphasis on large errors, hence it is sensitive to outliers. In the machine learning literature, the quantile objective function is a common choice to guard against heavy-tailed observations. The objective function in Equation (2) is robust to heavy-tailed observations, since outliers have minimal influence on quantile estimation. Hence, the quantile objective function can improve the stability of models fitted using such a robust objective function.

Financial returns also frequently exhibit volatility that varies both across firms and through time (Andersen, Bollerslev, Christoffersen, and Diebold (2005)).[2] This heteroskedasticity typically cannot be handled well by the MSE objective function and may distort model fitting. In contrast, models using the quantile objective function are designed to deal with such heteroskedasticity (Koenker and Bassett Jr (1982)), making them more suitable for handling financial returns in which heteroskedasticity is severe under certain macroeconomic conditions such as the Global Financial Crisis.

3.1.2 Off-the-shelf Models

Following the growing literature on machine learning algorithms for asset pricing, we compare a variety of linear and nonlinear quantile models with increasing complexity. We use the LR with all covariates from the traditional econometrics literature (Koenker and Bassett Jr (1978)) as the benchmark model. The LR imposes a linear function $g_\tau(z_{i,t};\theta) = z'_{i,t}\theta$ for the conditional quantile. This specification is simple, but suffers from overfitting when the number of covariates is large. We mitigate LR's propensity to overfit by considering a penalized linear model that adds an EN penalty to the objective function in Equation (2). To model nonlinearities and interaction effects, we consider GBT and NN1-NN4. These nonlinear techniques are analogous to models used in GKX, but are optimized with respect

[2]Internet Appendix Figure IA1 shows that the conditional quantiles exhibit substantial time variation, with the average $Q_{0.05}$ and $Q_{0.95}$ forecasts widening substantially during certain macroeconomic conditions, including the 2008–2009 Global Financial Crisis and the 2020 COVID-19 crisis. The figure also reveals considerable cross-sectional dispersion in the conditional quantiles within each month, particularly in the tails.

to the quantile objective function rather than MSE. We also consider RF, which is optimized using MSE as in GKX, but extended to simultaneously predict all quantiles under the same estimated model (Meinshausen (2006)).[3] Internet Appendix IA1 provides a detailed description of these off-the-shelf quantile machine learning models.

3.1.3 Multi-quantile Neural Networks

The models discussed so far assume that the functional forms of different quantiles are independent (except RF), but this may not hold true in an economic context, where the conditional quantiles for a given firm-month are likely highly related and share a common structure. To incorporate this economic structure, MQ architectures, depicted in Figure 1, estimate all targeted quantiles within a single functional form. The "hard sharing" hidden layers within the MQ interact and nonlinearly transform the input predictors into shared latent features, imposing an economic restriction of a common structure across quantiles. The MQ also includes "task-specific" hidden layers to capture each quantile's nonlinear exposures to the shared latent features, before aggregating these into a prediction for each quantile.

MQs utilize a technique called multi-task learning, which enhances model generalization and training speed by concurrently learning several related tasks. Multi-task learning implicitly provides regularization that is particularly effective in low signal-to-noise ratio settings. Intuitively, independently training a model for each quantile could lead to overfitting to the noise specific to that quantile. By forcing the MQ to jointly predict multiple quantiles, the neural network avoids overfitting to quantile-specific noise, improving its generalizability (Ruder (2017)). Furthermore, multi-task learning exhibits a property called inductive transfer learning, which increases data efficiency by utilizing additional information from the training signals of related tasks (Caruana (1997)).

[3]We include RF, despite its non-reliance on the quantile objective function, due to its popularity as a quantile machine learning model.

3.1.4 Spaced Multi-quantile Neural Networks

One longstanding problem related to quantile regression is the violation of monotonicity in estimated conditional quantile functions, commonly known as the quantile crossing problem, which has received considerable attention in the literature (e.g., Catania and Luati (2023), Chernozhukov, Fernández-Val, and Galichon (2010), Gouriéroux and Jasiak (2008), Granger (2010), Schmidt and Zhu (2016)). The quantile function should increase as τ increases, but estimated functions may violate this condition. Unfortunately, none of the models discussed so far, except RF, are guaranteed to satisfy this monotonicity requirement.

To address the quantile crossing problem, we estimate the distances between adjacent quantiles, known as quantile spacings, rather than the quantiles themselves. This approach, introduced by Granger (2010), directly predicts a reference quantile, typically the median, and models the remaining quantiles as positive functions of distances from the reference quantile, thus satisfying the monotonicity requirement by construction. While this method has been applied in linear settings (e.g., Catania and Luati (2023), Schmidt and Zhu (2016)), to the best of our knowledge, it has not been applied to neural networks.

The SMQ architecture, shown in Figure 2, extends the MQ by resolving the quantile crossing problem while preserving its common structure across quantiles. More specifically, for a set of τ, where $0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K < 1$, the SMQ model is recursively defined as:

$$Q_{\tau_k} = \begin{cases} Q_{\tau_{k+1}} - SP(Y_{\tau_k}) & \text{if } \tau_k < \tau_{k^\star}, \\ Q_{\tau_{k^\star}} & \text{if } \tau_k = \tau_{k^\star}, \\ Q_{\tau_{k-1}} + SP(Y_{\tau_k}) & \text{if } \tau_k > \tau_{k^\star}, \end{cases} \tag{3}$$

where $Q_{\tau_{k^\star}}$ denotes the reference quantile and $Y_{\tau_k} \equiv Y_{r_{i,t+1}}(\tau_k|\mathcal{F}_t)$ denotes the neural network output for τ_k. The function $SP(\cdot)$ denotes the softplus function, defined as $SP(x) = \log(1 + \exp(x))$, which is positive, differentiable, and monotonically increasing. In this context, the

softplus is used to model the spacing between adjacent quantiles, ensuring that all estimated quantiles are increasing by construction.

SMQs and MQs jointly estimate quantiles by minimizing the sum of quantile loss functions across τ levels. For K quantiles, the objective function is given by:

$$\mathcal{L}(\theta) = \sum_{k=1}^{K} \mathcal{L}_{\tau_k}(\theta), \tag{4}$$

where quantile objective function $\mathcal{L}_{\tau_k}(\theta)$ is from Equation (2).[4] Since MQ predicts quantiles directly, the Q_τ in Equation (4) correspond to the neural network outputs Y_{τ_k}, when MQ models are fitted. In contrast, for SMQ, the Q_τ are defined recursively as in Equation (3). We estimate both models using stochastic gradient descent to optimize Equation (4). Notably, for the MQ and SMQ models, the parameters are estimated under the economic restriction of a common structure across quantiles. For SMQ, the parameters are further constrained to ensure monotonicity of the estimated quantiles. This restriction distinguishes SMQ from post-estimation procedures such as ex-post rearrangement methods (e.g., Chernozhukov, Fernández-Val, and Galichon (2010)), which enforce monotonicity after estimation and therefore do not estimate parameters under these economic restrictions.

An important distinction between SMQ and existing linear quantile spacing models (e.g., Catania and Luati (2023); Schmidt and Zhu (2016)) is that the distances between adjacent quantiles are modeled using the softplus function rather than the exponential function. While exponential spacing can be effective in simpler settings, it performs poorly in neural network architectures. The composite functional structure of neural networks can cause exponential spacing to produce extremely large gradients and unstable updates during training, a phenomenon known as the exploding gradient problem. In contrast, the softplus function

[4]When $g_\tau(\cdot)$ is linear, the objective function in Equation (4) is referred to as the Hogg function (Koenker (2005)), which is used to estimate linear multi-quantile models (see, e.g., Catania and Luati (2023), White, Kim, and Manganelli (2015)). Related in spirit, Zou and Yuan (2008) theoretically show that penalized linear quantile regression with parameters shared across quantiles and estimated using the objective function in Equation (4), known as Composite Quantile Regression, satisfies the oracle property under heavy-tailed distributions with infinite variance, and Bradic, Fan, and Wang (2011) extend this result to a high-dimensional setting.

provides smoother and more bounded growth, resulting in greater numerical stability and more reliable training updates.

We evaluate MQ (SMQ) architectures featuring up to two hard sharing layers and one task-specific layer for each quantile. For all MQ (SMQ) architectures, we fix each task-specific layer to comprise 8 neurons for each quantile. The shallowest architecture, denoted as MQ1 (SMQ1), contains a single hard sharing layer with 32 neurons. The subsequent architecture, MQ2 (SMQ2), incorporates two hard sharing layers with 32 and 16 neurons, respectively. We find that architectures with more than two hard sharing layers do not improve results, thus we do not extend the discussion to them for brevity.

3.2 Robust Risk Premium and the Bias-Variance Trade-Off

In this section, we introduce a robust measure of the risk premium by integrating over the quantiles, with the aim of reducing forecast errors by improving forecast stability. The risk premium of stock i in month t is defined as the conditional expectation of excess returns in month $t+1$, i.e., $E(r_{i,t+1}|\mathcal{F}_t)$.

Most machine learning papers forecast risk premia using estimated expectations from models fitted with the MSE objective function. However, the estimated expectations of these models can be highly sensitive to outliers, resulting in forecasts with high variance. For example, OLS estimated coefficients can exhibit an extremely large variance in the presence of heavy-tailed stock returns (Mikosch and de Vries (2013)). Additionally, the MSE objective function typically cannot handle heteroskedasticity in stock returns well, which can further increase estimation variance. For example, heteroskedasticity can increase the variance of OLS estimated coefficients, leading to less accurate predictions.

To address these issues, we introduce a quantile-based risk premium measure, referred to as the robust risk premium measure, which is defined as an integral over the quantiles for $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ and computed as follows:

$$\begin{aligned}\bar{Q} &= (0.05 + \frac{0.25-0.05}{2})Q_{0.05} + (\frac{0.25-0.05}{2} + \frac{0.5-0.25}{2})Q_{0.25} + \\ &(\frac{0.5-0.25}{2} + \frac{0.75-0.50}{2})Q_{0.50} + (\frac{0.75-0.50}{2} + \frac{0.95-0.75}{2})Q_{0.75} + (\frac{0.95-0.75}{2} + (1-0.95))Q_{0.95} \\ &= 0.15Q_{0.05} + 0.225Q_{0.25} + 0.25Q_{0.50} + 0.225Q_{0.75} + 0.15Q_{0.95},\end{aligned} \tag{5}$$

where the weights are derived by evenly partitioning the probability between two adjacent quantiles and allocating the partitioned probability to the nearest quantiles; with an additional 5% allocated to $Q_{0.05}$ and $Q_{0.95}$, respectively. Figure 3 provides an illustration of how we derive our weights using an example distribution. More generally, for a given set $0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K < 1$, the generalized robust risk premium measure is defined as

$$\begin{aligned}\bar{Q}_K &= (\tau_1 + \frac{\tau_2-\tau_1}{2})Q_{\tau_1} + (\frac{\tau_2-\tau_1}{2} + \frac{\tau_3-\tau_2}{2})Q_{\tau_2} + \cdots + \\ &(\frac{\tau_{K-1}-\tau_{K-2}}{2} + \frac{\tau_K-\tau_{K-1}}{2})Q_{\tau_{K-1}} + (\frac{\tau_K-\tau_{K-1}}{2} + (1-\tau_K))Q_{\tau_K},\end{aligned} \tag{6}$$

which converges to the conditional expectation of the respective distribution under general assumptions, as mathematically proven in Section IA2 in the Internet Appendix. Therefore, the robust risk premium measure is an appropriate approximation of the conditional expectation of excess returns, i.e., the risk premium.

Mathematically, the robust risk premium measure can be viewed as a discretization approach that approximates the expected value of a probability distribution using the expectation of a discrete distribution where the random variable only takes on quantiles with the assigned weights as the probability mass. This approach naturally leads to a quantile-based risk premium estimator, referred to as the robust risk premium $\bar{Q}$ forecast, when forecasted quantiles are used in Equation (5).[5] Statistically, this approach serves as an ensemble tech-

[5]For brevity, the terms "robust risk premium $\bar{Q}$" and "robust risk premium $\bar{Q}$ forecast" are used interchangeably in this paper.

nique, reducing variance by integrating various distribution predictions. Economically, this approach amalgamates valuable signals from various segments of the conditional return distribution. Other quantile-based estimators with different weighting schemes or number of quantiles have been proposed in the statistics literature, detailed in Internet Appendix Section IA4. Section 7 compares various approaches as part of robustness checks to support the introduction of Equation (5).

The robust risk premium $\bar{Q}$ reduces the variance in estimating the risk premium by excluding extreme returns outside of the 0.05 and 0.95 quantiles, and adopting the inherent stability of quantile-based measures. It is also robust to heteroskedasticity in stock returns.

While the robust risk premium $\bar{Q}$ reduces variance, it is, unlike MSE-based estimated expectations, typically a biased estimator of the conditional expectation, indicating a potential bias-variance trade-off. The optimal balance in this trade-off depends on the specific characteristics of the excess return distribution and should be determined empirically. Our empirical results in Section 5.4 show that the robust risk premium $\bar{Q}$ substantially outperforms the MSE-based approach used in GKX, indicating that the detrimental effects of a high variance outweigh the negative impact of possible bias embedded in the quantile-based estimation of the risk premium.

4 Data and Performance Evaluation

4.1 Data and Variable Definitions

We employ the same sample and information set as GKX, extended by four years. This sample comprises monthly total equity returns from CRSP for all firms listed on the NYSE, AMEX, and NASDAQ from 1957 to 2020, a span of 64 years. We calculate individual excess returns using the Treasury-bill rate as a proxy for the risk-free rate. For our predictors, $z_{i,t}$, we draw upon an extensive suite of 94 firm-level predictive characteristics from GKX and substitute missing values of characteristics with their cross-sectional median. We rank all

firm characteristics cross-sectionally for each period, subsequently mapping these ranks into the $[-1, 1]$ interval. We also include eight macroeconomic predictors from Welch and Goyal (2008), including dividend-price ratio (DP), earnings-price ratio (EP), book-to-market ratio (BM), net equity expansion (NTIS), Treasury-bill rate (TBL), term spread (TMS), default spread (DFY), and stock variance (SVAR).[6]

4.2 Model Performance Evaluation

By convention we divide our full sample into three disjoint time periods. We begin by estimating the model parameters on an 18-year training sample $\mathcal{T}_1$ (1957-1974); then conduct a hyperparameter optimization on a 12-year validation sample $\mathcal{T}_2$ (1975-1986); and afterwards keep the models fixed for one year and assess the predictive performance on an one-year test sample (1987). We repeat this procedure, by increasing the number of years in the training sample by one year, and moving the validation window and test window forward by one year in each iteration of the remaining years for a total of 34 out-of-sample years, representing the pooled test sample $\mathcal{T}_3$ (1987 - 2020). Internet Appendix IA3 details the setup of the hyperparameter optimization framework.

To assess the predictive performance for quantile forecasts, we adopt the pseudo-R^2 measure from Koenker and Machado (1999), defined as

$$R^1_{os}(\tau) = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3} \rho_\tau(r_{i,t+1} - \hat{r}_{\tau,i,t+1})}{\sum_{(i,t)\in\mathcal{T}_3} \rho_\tau(r_{i,t+1} - \bar{r}_\tau)}, \tag{7}$$

where $\hat{r}_{\tau,i,t+1}$ is the predicted τth quantile for stock i at time $t+1$ and $\bar{r}_\tau$ is the historical τth quantile of the expanding window training sample $\mathcal{T}_1$. The $R^1_{os}(\tau)$ is the quantile analogue of the out-of-sample R^2 often used in the machine learning literature. For every τ, the $R^1_{os}(\tau)$ statistic pools prediction errors across stocks and over time into an overall assessment of each model, and it measures the proportional reduction in the quantile forecasting error (2)

[6]The monthly firm-specific characteristics and macroeconomic predictors are available from Dacheng Xiu's website and Amit Goyal's website, respectively. We thank the authors for making the data available.

for each model relative to a naive forecast using the historical quantile.

To conduct pairwise comparisons of model performance, we use the modified Diebold and Mariano (1995) (DM) test, which accounts for potential cross-sectional dependence in stock returns. The DM test-statistic for a comparison of models 1 and 2 is defined as $DM_{12}(\tau) = \bar{d}_{12}(\tau)/\sigma_{\bar{d}_{12}(\tau)}$, where $\bar{d}_{12}(\tau)$ and $\sigma_{\bar{d}_{12}(\tau)}$ are, respectively, the time-series mean and Newey-West standard error of $d_{12,t+1}(\tau)$ over the test sample for quantile τ. $d_{12,t+1}(\tau)$ is the forecasting error between the two models, calculated as the cross-sectional average of quantile forecast error differentials from each model over each period $t+1$,

$$d_{12,t+1}(\tau) = \frac{1}{n_{3,t+1}} \sum_{i=1}^{n_{3,t+1}} \left(\rho_\tau \left(\hat{e}^{(1)}_{\tau,i,t+1} \right) - \rho_\tau \left(\hat{e}^{(2)}_{\tau,i,t+1} \right) \right), \tag{8}$$

where $\hat{e}^{(j)}_{\tau,i,t+1} = r_{i,t+1} - \hat{r}^{(j)}_{\tau,i,t+1}$ denotes the quantile τ prediction error of model j for stock i at time $t+1$, and $n_{3,t+1}$ is the number of stocks in the test sample.

5 Main Empirical Results

5.1 The Cross-sectional Predictability of Conditional Quantiles

Panel A of Table 1 presents the $R^1_{os}(\tau)$ values across models for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ for the pooled test sample that spans from 1987 through 2020. Among linear models, LR consistently delivers lower $R^1_{os}(\tau)$ values across all quantiles than EN, reflecting LR's lack of regularization. In contrast, imposing sparsity through the EN penalty improves performance across all quantiles, with particularly large gains for the 0.05 quantile, where $R^1_{os}(\tau)$ increases from 15.94% to 17.01%.

For tree-based models, Panel A shows that RF underperforms both linear models across all quantiles, likely because it doesn't use the quantile objective function for estimation. In contrast, GBT outperforms RF for all quantiles except the median, demonstrating that employing the quantile objective function generally improves predictive performance in tree-

based models. The gains are particularly pronounced in the tails, with $R^1_{os}(\tau)$ values of 17.60% for the 0.05 quantile and 13.97% for the 0.95 quantile. However, GBT records an $R^1_{os}(\tau)$ value of -0.47% for the median, indicating that it fails to outperform a naive forecast using the historical cross-sectional median.

Panel A further shows that deep learning models exhibit superior performance. NNs surpass the linear and tree-based models in terms of $R^1_{os}(\tau)$ values across all quantiles. Among them, NN1 delivers the best median forecasts with an $R^1_{os}(\tau)$ of 0.84%, while NN3 provides the strongest tail forecasts, reaching 19.90% for the 0.05 quantile and 15.92% for the 0.95 quantile. MQ and SMQ models emerge as the strongest overall performers, delivering broadly similar results. MQ1 offers the highest $R^1_{os}(\tau)$ values for the 0.05, 0.25, 0.50 quantiles, with 20.31%, 7.22%, and 0.89%, respectively; while SMQ2 leads for the 0.75 and 0.95 quantiles with 2.80% and 16.54%. Across all quantiles, MQ1, MQ2, and SMQ2 consistently outperform all off-the-shelf models, highlighting the benefits of incorporating economic structure and jointly estimating quantiles within a unified architecture.

Comparing the results across quantiles, we see that the first and fifth rows of Table 1 report the highest $R^1_{os}(\tau)$ values, demonstrating that the tail 0.05 and 0.95 quantiles are the most predictable. The second and fourth rows yield lower values, signifying that the inner 0.25 and 0.75 quantiles are less predictable, while the third row displays the lowest values, confirming that the median is the least predictable quantile. These patterns are consistent with a well-documented stylized fact of stock returns: volatility tends to be more predictable than the mean. Moreover, downside quantiles (0.05 and 0.25) are more predictable than upside quantiles (0.95 and 0.75), consistent with the view that downside risks are easier to forecast than their upside counterparts (Bollerslev (2022)).

Panel B of Table 1 reports the percentage of observations in which adjacent quantiles cross for each model. Linear models exhibit moderate crossing rates, peaking at 0.1958% for LR. GBT produces the highest crossing rate of 0.6453%. NN models yield lower rates, with a maximum of 0.0423% for NN4. MQ models, benefiting from their economic structure,

generate far fewer crossings, with a maximum of 0.001% for MQ1. RF and SMQ models have no crossings by construction, but the SMQ models are significantly more accurate than RF, making them the preferred choice when monotonicity is required.

We now conduct pairwise comparisons to assess the statistical performance among the models. Internet Appendix Table IA4 reports the Diebold-Mariano test statistics for all pairwise comparisons among the 12 models for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$. For each panel, a positive statistic indicates the column model outperforms the row model, and a bold number denotes statistical significance at the 5% level for each test. Panel A presents results for the 0.05 quantile. First, the MQ and SMQ models produce statistically significant improvements over all off-the-shelf models, highlighting the value of incorporating economic structure in neural networks. Second, NN models significantly outperform linear and RF models, demonstrating the superior predictive performance of deep learning. Third, GBT significantly outperforms RF, highlighting the importance of using the quantile objective function. Panels B–E report the test statistics for the remaining quantiles. While the broad patterns hold, the number of statistically significant comparisons diminishes at the 0.25, 0.50, and 0.75 quantiles, consistent with their lower predictability.

5.2 Predictive Power of the Robust Risk Premium

This section evaluates the predictive power of the robust risk premium from Equation (5) for individual stock returns. We assess the predictive power of each model using the out-of-sample R-squared (R^2_{os}) metric from GKX:

$$R^2_{os} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \bar{Q}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{i,t+1}}. \tag{9}$$

The R^2_{os} statistic pools prediction squared errors across stocks and over time into a grand panel-level assessment of each model, and it measures the proportional reduction in MSE for each model relative to a naive forecast of zero.

Panel A of Table 2 compares the R^2_{os} values of the 12 models, which closely mirror the quantile predictability results in the previous section. Starting with linear models, the robust risk premium forecasts for LR and EN post R^2_{os} values of 0.131% and 0.477%, respectively. Next for tree-based models, RF yields an R^2_{os} of 0.165% due to its imprecise quantile forecasts, whereas GBT achieves 0.364%, more than doubling RF's performance by employing the quantile objective function. The robust risk premium forecasts from NN1 through NN4 yield R^2_{os} values ranging from 0.870% to 0.977%, substantially outperforming both linear and tree-based models. In comparison, the MQ and SMQ models deliver even higher R^2_{os} values ranging from 0.920% to 1.001%.

To conduct pairwise comparisons of statistical performance across models, we employ the Diebold-Mariano test for differences in out-of-sample predictive accuracy as outlined in GKX. We replace Equation (8) with the forecast error differential between two models. This differential is calculated as the cross-sectional average of squared forecast error differentials from each model over each period $t+1$,

$$d_{12,t+1} = \frac{1}{n_{3,t+1}} \sum_{i=1}^{n_{3,t+1}} \left(\left(\hat{e}^{(1)}_{i,t+1} \right)^2 - \left(\hat{e}^{(2)}_{i,t+1} \right)^2 \right), \tag{10}$$

where $\hat{e}^{(j)}_{i,t+1} = r_{i,t+1} - \hat{r}^{(j)}_{i,t+1}$ represents the prediction error of model j for stock i at time $t+1$, and $n_{3,t+1}$ is the number of stocks in the test sample. The DM test-statistic for a comparison of models 1 and 2 is defined as $DM_{12} = \bar{d}_{12}/\sigma_{\bar{d}_{12}}$, where $\bar{d}_{12}$ and $\sigma_{\bar{d}_{12}}$ are, respectively, the time-series mean and Newey-West standard error of $d_{12,t+1}$ over the test sample.

Panel B of Table 2 reports Diebold-Mariano test statistics across the 12 models. A positive statistic indicates that the column model outperforms the row model, and bold entries denote statistical significance at the 5% level. MQ and SMQ models significantly outperform all linear and tree-based models, while the NNs significantly outperform LR, EN, and RF.

5.3 Machine Learning Portfolios Using the Robust Risk Premium

We next investigate whether the predictive accuracy of the robust risk premium translates into economic gains by forming portfolios based on the $\bar{Q}$ forecasts of the 12 models. At the end of each month, we forecast the one-month-ahead $\bar{Q}$, and then sort stocks into deciles based on these forecasts. We reconstitute portfolios each month using value weights. Table 3 reports the monthly performance results, where "Low" denotes the portfolio with the lowest robust risk premium forecasts (decile 1), "High" denotes the portfolio with the highest robust risk premium forecasts (decile 10), and "H-L" denotes the long-short portfolio that buys the highest robust risk premium portfolio (decile 10) and sells the lowest (decile 1).

Panel A of Table 3 shows that out-of-sample portfolio performance largely mirrors the R^2_{os} results in the previous section. Long–short portfolios based on LR and EN robust risk premium forecasts earn average monthly returns of 1.15% and 1.21% with annualized Sharpe ratios of 0.91 and 0.89, respectively. Among tree-based models, RF delivers the lowest Sharpe ratio of 0.77, whereas GBT improves performance with a Sharpe ratio of 1.10. Portfolios formed from NN models achieve average monthly returns between 2.61% and 2.80% and Sharpe ratios ranging from 1.26 to 1.44. Portfolios formed by MQ and SMQ forecasts deliver the strongest economic performance, generating average monthly returns between 2.96% and 3.17% with Sharpe ratios from 1.50 to 1.55. Together, these results underscore the importance of incorporating economic structure to enhance the economic gains from machine learning techniques.

Panel B of Table 3 details the risk-adjusted performance of the value-weighted long-short portfolios from each machine learning model. Reported are the alphas, R^2 values, and information ratios (IR) with respect to the Fama-French five-factor model (Fama and French (2015)) augmented to include the momentum factor.[7] The first row shows that LR and EN deliver the lowest monthly alphas of 0.64% and 0.65%. Next, RF and GBT record alphas of

[7]Using alternative factor models from Stambaugh and Yuan (2017) and Hou, Mo, Xue, and Zhang (2021) yield similar results, which are available upon request.

1.21% and 1.69%. NN models achieve alphas between 2.25% and 2.42% and IRs ranging from 1.2 to 1.54. The MQ and SMQ models dominate all others, with alphas between 2.51% and 2.72% and IRs between 1.52 and 1.58. In summary, these results highlight the superior risk-adjusted performance of robust risk premium forecasts and neural networks that incorporate economic structure.

The main results in Table 3 are visually summarized in Figure 4, which reports the cumulative performance for the long-short portfolios sorted on $\bar{Q}$ forecasts for all models, along with the cumulative market excess return (mktrf) as a benchmark. MQ and SMQ models consistently outperform all other models; and also perform well in recessions and the post-2000 sample. NN models are the next best performers. GBT achieves moderate performance. Linear models and RF are the worst performers, though they still beat the market benchmark.

5.4 Comparison to MSE-Based Approach

This section compares the performance of the robust risk premium ($\bar{Q}$) forecasts with the MSE-based risk premium ($\bar{\mu}^{MSE}$) forecasts. We replicate the MSE-based models of GKX by largely following their empirical setup. Internet Appendix Table IA3 shows that the resulting R^2_{os} values and Sharpe ratios closely match those reported in their paper. Further details on the replication are provided in Section IA3.2. Having replicated the MSE-based models under the same empirical conditions as our quantile models, we can now compare their performance directly while holding the model class fixed. Table 4 reports the results for the pooled test sample and for subsamples evenly partitioned by firm size. We benchmark the MQ and SMQ models against NN3, the most accurate MSE-based model in GKX and Table IA3.

First, we compare the statistical performance. Panel A of Table 4 reports the pairwise correlations between the $\bar{Q}$ and $\bar{\mu}^{MSE}$ forecasts, the R^2_{os} values for both approaches, and the DM statistic comparing the two, as described in Section 5.2. A positive DM value indicates

that $\bar{Q}$ outperforms $\bar{\mu}^{MSE}$, and bold denotes statistical significance at the 5% level.

For the pooled test sample, the first row of Panel A shows that most models have correlations of 60% or less, indicating that $\bar{Q}$ produces forecasts that meaningfully differ from $\bar{\mu}^{MSE}$. Rows 2 and 3 show that the R^2_{os} values for the $\bar{Q}$ forecasts exceed those of the $\bar{\mu}^{MSE}$ for all models except RF. Row 4 demonstrates that this outperformance is statistically significant for NN2, NN3, NN4, MQ1, SMQ1, and SMQ2. These results highlight the superior predictive power of the robust risk premium relative to forecasts generated by MSE-based estimation. The gains in predictive power of the robust risk premium are likely due to reduced model variance in the presence of heavy-tailed and heteroskedastic returns. The remaining rows in Panel A show that much of the statistical outperformance is concentrated among small firms, which exhibit the heaviest tails. However, even among large firms, the $\bar{Q}$ forecasts yield higher R^2_{os} values than $\bar{\mu}^{MSE}$ across all models. For example, MQ1 achieves an R^2_{os} value of 0.65% for large stocks, outperforming the highest R^2_{os} value for $\bar{\mu}^{MSE}$ of 0.39% attained by NN4.

Next, we compare economic significance. Panel B of Table 4 reports the average monthly return of the long–short portfolios formed using $\bar{Q}$ and $\bar{\mu}^{MSE}$ forecasts. Following Moreira and Muir (2017), we evaluate incremental performance with a spanning test by estimating the time-series regression

$$f^{\bar{Q}}_{t+1} = \alpha_{Span} + \beta f^{\bar{\mu}^{MSE}}_{t+1} + \epsilon_{t+1}, \tag{11}$$

where $f^{\bar{Q}}_{t+1}$ and $f^{\bar{\mu}^{MSE}}_{t+1}$ denote the excess return on the long–short portfolio based on the $\bar{Q}$ and $\bar{\mu}^{MSE}$ forecasts, respectively. A positive intercept, α_{Span}, implies that portfolios constructed from $\bar{Q}$ forecasts deliver higher Sharpe ratios than those based on $\bar{\mu}^{MSE}$ (Moreira and Muir (2017)).

First focusing on the entire sample, Rows 1 and 2 of Panel B show that the average returns based on $\bar{Q}$ forecasts exceed those from $\bar{\mu}^{MSE}$ across all models. The α_{Span} values are positive and statistically significant throughout, indicating that the economic gains from $\bar{Q}$ forecasts

are significantly higher than those from $\bar{\mu}^{MSE}$. The subsequent rows show similar results for small and large firms, with the α_{Spar} values remaining positive statistically significant across all models for each firm-size group.

5.5 Sensitivity to Drawdowns, Transaction Costs, and Portfolio Construction Choices

Avramov, Cheng, and Metzker (2022) express concerns that certain deep learning models may yield long-short portfolios which are infeasible due to economic constraints such as drawdowns and transaction costs. Panels A and B of Table 5 report the drawdowns, portfolio turnover, break-even transaction costs, and net-of-transaction-cost performance of the value-weighted long–short portfolios formed using the $\bar{Q}$ and $\bar{\mu}^{MSE}$ forecasts, respectively. Following Avramov, Cheng, and Metzker (2022), the break-even transaction cost is computed as the six-factor alpha divided by turnover. Transaction costs follow Brandt, Santa-Clara, and Valkanov (2009). For stock i at time t, the cost is $c_{i,t}|w_{i,t} - w_{i,t-1}|$, where $w_{i,t}$ denotes the weight of stock i in the portfolio and $c_{i,t}$ is the one-way transaction cost; portfolio-level costs are calculated by summing across stocks. Inspired by DeMiguel, Martín-utrera, and Uppal (2024), we estimate $c_{i,t}$ as one-half of the average monthly bid–ask spread using the CHL estimator of Abdi and Ranaldo (2017).

Table 5 shows that long–short portfolios sorted on $\bar{Q}$ forecasts experience smaller drawdowns than those sorted on $\bar{\mu}^{MSE}$ forecasts across all models except LR. For example, NN models experience drawdowns between 34.36% and 43.21% for $\bar{Q}$ portfolios, compared with 40.72% to 94.77% for $\bar{\mu}^{MSE}$ portfolios. MQ1 and MQ2 record drawdowns of 34.25% and 35.50%, respectively, while SMQ1 and SMQ2 experience somewhat higher drawdowns of 50.58% and 40.84%, respectively. The table further shows that $\bar{Q}$ portfolios exhibit slightly higher turnover than $\bar{\mu}^{MSE}$ portfolios. For NN models, turnover ranges from 140.7% to 146.8% for $\bar{Q}$ portfolios compared with 122.4% to 139.3% for $\bar{\mu}^{MSE}$ portfolios. MQ and SMQ models experience turnover ranging from 145.6% to 149.4%. Despite these higher turnover

levels, Table 5 shows that $\bar{Q}$ portfolios achieve substantially higher break-even transaction costs, the cost required to fully offset their risk-adjusted returns. For NN models, break-even transaction costs range from 1.53% to 1.67% for $\bar{Q}$ portfolios compared with 1.20% to 1.37% for $\bar{\mu}^{MSE}$ portfolios. MQ and SMQ models achieve break-even transaction costs between 1.72% and 1.82%.

Rows 4–6 of Panels A and B report the monthly average returns, Newey–West t-statistics, and Sharpe ratios of the long–short portfolios net of transaction costs. The $\bar{Q}$ portfolios exhibit substantially stronger net performance than the $\bar{\mu}^{MSE}$ portfolios. For the NN models, net Sharpe ratios range from 0.77 to 0.97 for $\bar{Q}$ portfolios compared with 0.52 to 0.89 for $\bar{\mu}^{MSE}$ portfolios. The MQ and SMQ models achieve even higher net Sharpe ratios, between 0.98 and 1.04. These results show that the $\bar{Q}$ portfolios retain robust economic performance after accounting for drawdowns and transaction costs, both in absolute terms and relative to the $\bar{\mu}^{MSE}$ portfolios.

Finally, Avramov, Cheng, and Metzker (2022) recommend excluding hard-to-arbitrage stocks when constructing portfolios. As a robustness check, Internet Appendix Table IA5 shows that using NYSE breakpoints, the MQ1 portfolio achieves a Sharpe ratio of 1.31, outperforming the highest $\bar{\mu}^{MSE}$ portfolio Sharpe ratio of 1.07 obtained by NN1. Similarly, Internet Appendix Table IA6 shows that after excluding microcap stocks, MQ1 attains a Sharpe ratio of 1.05, exceeding the highest $\bar{\mu}^{MSE}$ portfolio Sharpe ratio of 0.80 delivered by NN1. Finally, Internet Appendix Table IA7 shows that using equal-weighted portfolios, MQ1 achieves a Sharpe ratio of 2.99, slightly outperforming the highest $\bar{\mu}^{MSE}$ portfolio Sharpe ratio of 2.92 obtained by NN1. Overall, these results confirm that the economic gains of $\bar{Q}$ over $\bar{\mu}^{MSE}$ forecasts are not driven by downside risks, transaction costs, or portfolio construction choices.

6 Sources of Return Distribution Predictability

The previous section shows that quantile machine learning models accurately forecast the conditional quantiles and risk premia of individual stocks. This section investigates the economic mechanisms underlying this return distribution predictability.

Hong and Stein (1999) propose a theoretical model in which investors, constrained by their information processing capabilities and computational capacities, can only process a subset of available public information. Their model shows that if information diffuses gradually among investors, they are slow to adjust their demand, leading to only partial incorporation of new information into prices. Consequently, prices tend to underreact to new information, generating return distribution predictability.[8]

Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023) show that such behavioral biases help explain return predictability for left-tail momentum and option returns, respectively. Building on this literature, we test whether similar biases drive the return distribution predictability documented in this paper. In particular, we examine whether investor inattention, information frictions, underreaction to information shocks, and earnings announcements help explain the return distribution predictability.

6.1 Investor Inattention

The investor inattention hypothesis posits that investors' limited cognitive capacity hinders their ability to immediately process all new information relevant to stock prices (Barberis (2018)). This cognitive delay implies that investors require time to process new information, leading to the prediction that stocks receiving less attention will exhibit stronger return distribution predictability. A large empirical literature supports this view, showing that investor inattention often leads to underreaction to public information about firm fundamentals and characteristics (see Atilgan, Bali, Demirtas, and Gunaydin (2020) for a survey).

[8]Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, and Subrahmanyam (2001) also propose models of investor behavior in which investors underreact to information, leading to return predictability.

The investor inattention hypothesis implies greater return distribution predictability in stocks receiving less investor attention, typically proxied by analyst coverage (CVRG). Motivated by Atilgan, Bali, Demirtas, and Gunaydin (2020), we test this conjecture by sorting stocks into terciles according to the number of analysts covering each firm in the most recent month.[9] Within each tercile, we further sort stocks into decile portfolios using robust risk premium forecasts from the EN, GBT, NN3, and MQ1 models.[10]

Table 6 presents the findings. Panel D shows that the MQ1 long–short portfolio achieves a Sharpe ratio of 1.88 in the low–coverage tercile, which declines monotonically to 1.01 in the high–coverage tercile. Panels A, B, and C display a similar declining pattern for EN, GBT, and NN3. However, the spread between the low- and high–coverage terciles widens with model complexity, rising from $1.12 - 0.77 = 0.35$ for EN to $1.88 - 1.01 = 0.87$ for MQ1, indicating that the MQ models more effectively capture return distribution predictability attributable to investor inattention.

Figure 5 illustrates the return distribution predictability for terciles sorted by analyst coverage. The $R^1_{os}(\tau)$ and R^2_{os} values all decline monotonically from the low- to high-coverage terciles. Consistent with the investor inattention hypothesis, these results also indicate stronger return distribution predictability in low-attention stocks and weaker predictability in high-attention stocks.

6.2 Information Frictions

Information frictions due to limits-to-arbitrage impede the full incorporation of information into prices (Pontiff (2006), Shleifer and Vishny (1997)), leading to underreaction and subsequent return continuations. We hypothesize that arbitrage costs partially drive our results, implying a testable prediction: return distribution predictability should be more pronounced in stocks with higher arbitrage costs.

[9] We assign zero coverage to firms with missing observations to maintain a comprehensive sample, though results are similar when these observations are excluded.

[10] Unreported results show that SMQ models produce similar results as MQ models, available upon request.

Rather than relying on a single proxy for information frictions, we follow Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023) and construct an arbitrage index (AI) by incorporating multiple indicators of limits-to-arbitrage. These indicators include firm size, firm age, idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang (2006)), illiquidity (Amihud (2002)), and analyst coverage (Zhang (2006)). To construct the AI, we sort stocks in increasing order by their idiosyncratic volatility and illiquidity and in decreasing order by their size, age, and analyst coverage. Each firm receives a decile rank for each indicator, and the firm-level AI is defined as the sum of these ranks, ranging from 5 to 50, with higher values reflecting stricter limits-to-arbitrage.

Similar to Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023), we sort stocks into quintiles based on their AI. Within each quintile, we form decile portfolios by ranking stocks on their robust risk premium forecasts from the EN, GBT, NN3, and MQ1 models. Table 7 reports the results. Panel D shows that the MQ1 long–short portfolio earns a Sharpe ratio of 0.90 in the lowest AI quintile, which monotonically rises to 2.94 in the highest quintile. Panels A, B, and C display a similar increasing trend for EN, GBT, and NN3. These results indicate that return distribution predictability is stronger in firms with higher arbitrage costs. Moreover, the spread between the lowest and highest AI quintiles widens with model complexity, ranging from $2.20-0.79 = 1.41$ for EN to $2.94 - 0.90 = 2.04$ for MQ1, suggesting that MQ models more effectively extract return distribution predictability associated with information frictions.

Figure 6 illustrates the return distribution predictability for quintiles sorted by AI. The figure reveals a nonlinear "U-shaped" relationship between AI and predictive power. Both $R^1_{os}(\tau)$ and R^2_{os} generally rise from AI2 to AI5, consistent with the hypothesis that higher arbitrage costs increase predictability. Interestingly, the $R^1_{os}(\tau)$ and R^2_{os} values are also relatively high for the lowest AI quintile, AI1, which faces the least information frictions. This likely reflects the fact that AI1 consists of large, liquid, and stable firms that are inherently more predictable. However, this predictability does not translate into profitability. As Table

7 shows, Sharpe ratios for AI1 remain low because low arbitrage costs allow return patterns to be quickly exploited and arbitraged away. By contrast, AI5 exhibits both high predictability and profitability, consistent with higher arbitrage costs impeding the full incorporation of information into prices.

6.3 Underreaction to Information Shocks

Machine learning forecasts can rapidly integrate new information about firm fundamentals and characteristics. If investors underreact to new information, the strongest return distribution predictability should occur among firms experiencing large information shocks, as reflected by large forecast changes. Motivated by Atilgan, Bali, Demirtas, and Gunaydin (2020), we test this conjecture by calculating, for each firm, the change in robust risk premium forecasts between months t and $t-1$, and use these forecast changes in bivariate portfolio sorts to examine their implications for excess returns in month $t+1$. Specifically, for each firm i, we define $\Delta\bar{Q}$ as the difference between the robust risk premium forecast at the end of month t and the forecast at the end of month $t-1$, capturing the innovation in $\bar{Q}$ forecasts. Firms undergoing large positive or negative information shocks from month $t-1$ to t will exhibit correspondingly large positive or negative $\Delta\bar{Q}$ values, while firms with small information shocks will exhibit moderate $\Delta\bar{Q}$ values.

We begin by sorting stocks into terciles based on $\Delta\bar{Q}$ for the EN, GBT, NN3, and MQ1 models. Within each $\Delta\bar{Q}$ tercile, we further sort stocks into decile portfolios by their one-month-ahead robust risk premium forecasts, following the methodology in Section 5.3. Table 8 reports the results. For MQ1, Panel D shows that stocks in the $\Delta\bar{Q}1$, those experiencing large negative information shocks, earn a Sharpe ratio of 1.41. Stocks in $\Delta\bar{Q}3$, those with large positive shocks, earn a Sharpe ratio of 1.57. In contrast, stocks in $\Delta\bar{Q}2$, with relatively stable robust risk premium forecasts, generate a smaller Sharpe ratio of 0.97.

These results align with the hypothesis that investors only partially react to new information regarding changes to firm fundamentals and characteristics. Consequently, prices

underreact to new information, leading to stronger return distribution predictability in stocks with significant information shocks, as predicted by the theoretical framework in Hong and Stein (1999).

Panels B and C exhibit a similar U-shaped pattern in Sharpe ratios for GBT and NN3. In contrast, Panel A shows that EN's highest $\Delta\bar{Q}$ tercile records a Sharpe ratio of 0.71, slightly below the middle tercile's 0.74, indicating that this linear model fails to capture the return distribution predictability associated with large positive information shocks.

6.4 Earnings Announcements

Engelberg, McLean, and Pontiff (2018) report that stock market anomalies are six times more pronounced on earnings announcement days, suggesting that these anomalies stem from mispricing, which are partially corrected when new information arrives. To evaluate the impact of earnings announcements, we follow the subsample approach from Zhan, Han, Cao, and Tong (2022). In each month, we partition our stocks into two subsamples: one containing stocks that had an earnings announcement during the month, and another consisting of all other stocks. We then sort each subsample into decile portfolios based on their $\bar{Q}$ forecasts from the EN, GBT, NN3, and MQ1 models.

Internet Appendix Table IA8 reports the results. Panel A reveals that the EN long-short portfolio achieves a Sharpe ratio of 0.89 for stocks without earnings announcements, compared to only 0.68 for stocks with earnings announcements. Similar trends are evident for GBT, NN3, and MQ1 in Panels B, C, and D, respectively. These patterns contrast with the one documented by Engelberg, McLean, and Pontiff (2018), suggesting that earnings announcements are unlikely to be a source of return distribution predictability.

6.5 Sources of Return Predictability for MSE-Based Models

Can behavioral biases also help explain stock return predictability in MSE-based models, and if so, to what extent? To our best knowledge, this question has not been explicitly

examined in the existing literature. Internet Appendix Tables IA9, IA10, and IA11 replicate the earlier bivariate portfolio sorts based on investor attention, information frictions, and information shocks, respectively, for the MSE-based models (EN, RF, GBT, and NN3).

The results indicate that the return predictability of MSE-based models is generally more pronounced for stocks with lower investor attention, higher information frictions, and larger information shocks, confirming that these models are likewise influenced by behavioral biases. However, their overall economic performance is considerably weaker. For example, Table IA9 shows that, for the MSE-based NN3 model, the analyst-coverage terciles have Sharpe ratios of 1.68, 1.55, and 0.79, respectively, whereas for the MQ1 model in Table 6, the corresponding Sharpe ratios are substantially higher at 1.88, 1.73, and 1.01. These findings demonstrate that, conditional on investor attention, long–short portfolios based on robust risk premium forecasts consistently outperform those derived from MSE-based forecasts. The comparisons based on information frictions and information shocks yield similar conclusions.

6.6 Which Covariates Matter for Each Quantile?

We further investigate the sources of return distribution predictability by identifying which covariates are important determinants of each stock return quantile, while simultaneously controlling for the many other predictor variables. Motivated by the approach in GKX, we discover influential variables for each quantile by setting all values of the predictor in question to zero, while holding the remaining model estimates fixed. Specifically, for every $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$, the variable importance of the j^{th} predictor is measured by calculating the reduction in $R^1_{os}(\tau)$ from setting all values of the j^{th} predictor to zero within each test sample, and then averaging the $R^1_{os}(\tau)$ reductions across test samples as a single importance measure. The variable importance within each model is normalized to sum to one, allowing for the interpretation of relative importance for the model.[11]

Internet Appendix Figure IA2 presents the variable importance of the top ten firm-level

[11] Popular Shapley value packages are not compatible with MQ and SMQ models, so we employ this variable importance approach, inspired by GKX, as a computationally efficient and easily interpretable alternative.

predictors for each model, and Internet Appendix Figure IA3 reports the corresponding variable rankings across models.[12] Panel A in both figures show that, for the 0.05 quantile, the models largely agree on the most influential predictors. These are primarily risk measures such as total and idiosyncratic return volatility (retvol, idiovol); followed by illiquidity measures including the bid–ask spread (baspread), share turnover (turn), and firm size (mvel1); and price-trend variables such as momentum (mom6m, mom12m). This set of predictors is consistent with Chen, Hong, and Stein (2001), who show that risk, illiquidity, and price-trend variables are strong predictors of crashes.[13] Notably, these influential characteristics provide further support for information frictions as a source of return distribution predictability, as they are closely linked to limits to arbitrage. Although few accounting variables are influential, secured debt indicator (securedind) ranks highly, reflecting its relation to default risk. Among macroeconomic variables, the aggregate dividend–price (DP) and earnings–price (EP) ratios are the most influential. The other panels show that the significant predictors for the remaining quantiles are similar to those for the 0.05 quantile.

7 Robustness Checks and Additional Analyses

This section presents robustness checks and additional analyses that further support the main results. Section 7.1 compares the robust risk premium to other quantile-based estimators of the conditional expectation. Section 7.2 studies the sensitivity of the results to the number of quantiles. Section 7.3 examines the implications of incorporating variability forecasts in portfolio strategies.

[12] See Welch and Goyal (2008) and Table A.6 of the Internet Appendix in GKX for detailed definitions of the macroeconomic variables and firm-level characteristics, respectively.

[13] These findings are also consistent with the Value-at-Risk literature (e.g., Liu and Stentoft (2021)) that relies on volatility and price-trend predictors.

7.1 Alternative Quantile-Based Estimators

The use and evaluation of quantile-based estimators of the conditional expectation in finance remains relatively underexplored. This section compares the robust risk premium with other quantile-based estimators proposed in the literature. These estimators differ primarily in the weights assigned to each quantile. They can also be broadly grouped into those using five quantiles ($\bar{Q}_{YZLZ}$, $\bar{Q}_{FQ}$) and those using three ($\bar{Q}_{WWLT}$, $\bar{Q}_{TM}$, $\bar{Q}_{MP}$). Section IA4 in the Internet Appendix details these measures and their weighting schemes. Because it's unlikely to find a universal weighting scheme that is optimal across all distributions, the appropriate weights for stocks depend on the characteristics of their excess return distributions and should be determined empirically. Internet Appendix Table IA12 empirically compares the statistical and economic performance of the alternative estimators with the robust risk premium $\bar{Q}$. The bold entries indicate the best-performing weighting scheme for each model.

Panel A shows that $\bar{Q}$ achieves the highest R^2_{os} values for all models except RF, indicating that it delivers the greatest statistical performance among quantile-based estimators of the conditional expectation. It is observed that the five-quantile estimators consistently outperform the three-quantile estimators, suggesting that increasing the number of quantiles from three to five may improve predictive performance. Panel B demonstrates that long–short portfolios formed using $\bar{Q}$ forecasts achieve the highest Sharpe ratios for all models except LR, EN, RF, and NN4, highlighting that $\bar{Q}$ empirically offers the strongest economic significance. Again, five-quantile estimators generally outperform their three-quantile counterparts, although the magnitude of improvement is modest. Overall, the results indicate that the robust risk premium generally provides the best statistical and economic performance among these quantile-based estimators.

7.2 Sensitivity to the Number of Quantiles

The previous section suggests that five-quantile estimators outperform three-quantile estimators. A natural question is whether incorporating additional quantiles can consistently

enhance performance. Analogous to choosing the number of parameters in a model, increasing the number of quantiles may improve the precision of the robust risk premium estimate by reducing bias, but it may also increase model complexity and variance. To investigate this bias-variance trade-off, we extend the MQ1 model to predict K quantiles ($K \in \{21, 25, 33, 51, 99\}$), corresponding to approximately uniform τ-intervals of 0.05, 0.04, 0.03, 0.02, and 0.01, respectively.[14] $K = 5$ is the baseline specification. We then compute the generalized robust risk premium, $\bar{Q}_K$, using all K quantiles according to Equation (6).

Internet Appendix Table IA13 presents the results. Panel A shows that the baseline model's R^2_{os} of 0.99% is only marginally exceeded by the model with $K = 25$ that attains an R^2_{os} of 1.02%. The remaining models yield R^2_{os} values between 0.87% and 0.91%. Panel B reports that the baseline model's Sharpe ratio of 1.53 is slightly outperformed by the models with 21, 25, and 33 quantiles, which achieve Sharpe ratios of 1.54, 1.57, and 1.54, respectively, while the models with $K = 51$ and $K = 99$ produce Sharpe ratios of 1.51. Overall, these results indicate that increasing the number of quantiles beyond five yields negligible, if any, gains in statistical or economic performance.

7.3 Robust Risk Premium to Variability Managed Portfolios

Moreira and Muir (2017) show that incorporating variability can enhance the economic performance of aggregate portfolios. Building on this insight, we examine whether incorporating variability as a secondary portfolio criterion generates additional economic gains in the cross-section. To do so, we combine the robust risk premium with the interquartile range, denoted as $IQR = Q_{0.75} - Q_{0.25}$, to construct a quantile-based measure of variability-adjusted returns that is robust to outliers. We refer to this measure as the robust risk premium to variability (RRPV) ratio:

[14] We focus on MQ1 because it is both computationally efficient and highly accurate. Off-the-shelf models (except RF) estimate each quantile separately, substantially increasing computation time. RF is excluded because of its considerably lower predictive accuracy relative to MQ1. SMQ models are slower to compute than MQ due to their more complex structure.

$$RRPVratio = \frac{\bar{Q}}{IQR}. \tag{12}$$

This section examines whether portfolios based on the robust risk premium can be enhanced by incorporating the $RRPV$ ratio and whether it contains predictive information for Sharpe ratios.[15]

Allena (2025) demonstrates that a double-sorting portfolio strategy combining machine learning forecasts of the risk premium and its standard error can enhance performance. Inspired by their approach, we design a similar strategy that leverages machine learning forecasts of the $RRPV$ ratio to refine the portfolios detailed in Section 5.3 and briefed below. At the end of each month, we sort stocks into deciles based on their one-month-ahead robust risk premium forecasts. The portfolios with the lowest and highest robust risk premium forecasts are labeled as "Low" and "High", respectively.

To form RRPV-Managed portfolios, we further divide each of the above portfolios into deciles based on their $RRPV$ ratio forecasts, rebalancing them monthly using value weights. Within each robust risk premium decile, D1 (D10) denotes the portfolio with the lowest (highest) $RRPV$ ratio forecasts. This procedure generates subportfolios $\{L_{D1}, ..., L_{D10}\}$ and $\{H_{D1}, ..., H_{D10}\}$. Our RRPV-Managed strategy, denoted as "H_{D10}-L_{D1}", takes a long position in H_{D10}, consisting of stocks with the highest $\bar{Q}$ and $RRPV$ ratio forecasts, and a short position in L_{D1}, consisting of stocks with the lowest $\bar{Q}$ and $RRPV$ ratio forecasts.

Panel A of Internet Appendix Table IA14 shows that the RRPV-Managed strategy substantially enhances Sharpe ratios relative to the baseline results in Section 5.3 (detailed in

[15]We use $\bar{Q}$ for the numerator because it is insensitive to outliers, unlike expected returns estimated from models fitted with the MSE objective function. For the denominator, we employ the interquartile range instead of the standard deviation, since the former is a robust measure of variability while the latter is sensitive to extreme values. Under a Gaussian distribution, the $RRPV$ ratio is approximately equal to the Sharpe ratio divided by 1.35, making it a scaled version of the moment-based Sharpe ratio under normality. For heavy-tailed distributions, the $RRPV$ ratio provides a robust alternative measure of risk-adjusted returns: for example, in a Cauchy distribution, the Sharpe ratio is undefined, whereas $RRPV$ always exists as long as $Vol > 0$. While there is no theoretical equivalence between the $RRPV$ ratio and the Sharpe ratio without strong distributional assumptions such as normality, the practical usefulness of the $RRPV$ ratio as a measure of risk-adjusted performance is ultimately an empirical question examined in this section.

Table 3) across all nonlinear models. For example, NN models achieve Sharpe ratios ranging from 1.53 to 1.99, while MQ and SMQ models attain even higher values between 1.99 and 2.13. The Sharpe ratio of 2.13 obtained by SMQ1 substantially exceeds the peak value of 1.51 documented in Allena (2025), indicating that strategies incorporating the $RRPV$ ratio as a secondary portfolio criterion outperform those that instead use the variance of the risk premium, as in Allena (2025). Panel B further shows that alphas also substantially improve across all nonlinear models. Panel C reports the Sharpe ratios of the RRPV-Managed portfolios. Across nonlinear models, Sharpe ratios for D1 are generally lower than those for D10, implying that the $RRPV$ ratio contains predictive information for Sharpe ratios, thereby validating its usefulness as a measure of variability-adjusted returns.

Internet Appendix Table IA15 displays the results for the RRPV-Managed equal-weighted portfolios. Panel A also demonstrates that the RRPV-Managed strategies improve the Sharpe ratios of the equal-weighted long-short portfolios relative to Table IA7 for all nonlinear models except NN1. The MQ and SMQ models achieve high Sharpe ratios ranging from 3.05 to 3.25.

8 Conclusion

Building on the success of a rapidly growing literature that applies machine learning techniques to model asset returns, this paper studies the predictability of future stock return distributions and risk premia using quantile machine learning models. Forecasts of the future return distribution provide investors with a means to better measure risk and manage portfolios, and offer financial economists a powerful tool to test asset pricing models and estimate risk premia.

This paper leverages a set of 94 firm-level predictors and 8 macroeconomic variables from 1957 to 2020 to forecast stock return quantiles. It demonstrates several results that enhance our knowledge of the cross-section of stock returns. First, it introduces a neural

network architecture that incorporates an economic restriction by jointly estimating multiple conditional quantiles, and is also designed to produce non-crossing quantile forecasts. These models generally outperform linear, tree-based, and feed-forward neural network models across quantiles. Second, this paper introduces a quantile-based risk premium measure that is robust to outliers and heteroskedasticity. The robust risk premium forecasts achieve significantly higher out-of-sample statistical and economic performance than MSE-based forecasts. Third, we present compelling evidence that return distribution predictability is driven by investor behavioral biases, including investor inattention, information frictions, and underreaction to information shocks.

The quantile machine learning approaches studied in this paper will also lead to a promising direction for future work. While this paper demonstrates the success of quantile machine learning techniques using stocks as a widely studied testing ground, there are many recalcitrant asset classes where robustness can have a greater impact. For example, bonds (Bianchi, Büchner, and Tamoni (2021)), options (Bali, Beckmeyer, Mörke, and Weigert (2023)), international assets (Leippold, Wang, and Zhou (2022)), intraday stocks (Huddleston, Liu, and Stentoft (2023), Liu and Stentoft (2023)), currencies (Filippou, Rapach, Taylor, and Zhou (2021)), and mutual funds (Kaniel, Lin, Pelger, and Van Nieuwerburgh (2023)), can exhibit heavy-tailed returns, and may benefit from applying the quantile machine learning approaches explored in this paper. Finally, our paper contributes a distributional approach to the machine learning techniques applied in the rapidly growing fintech industry.

References

Abdi, F., and A. Ranaldo (2017): "A simple estimation of bid-ask spreads from daily close, high, and low prices," *The Review of Financial Studies*, 30(12), 4437–4480.

Adrian, T., N. Boyarchenko, and D. Giannone (2019): "Vulnerable growth," *American Economic Review*, 109(4), 1263–89.

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*.

Allena, R. (2025): "Confident risk premiums and investments using machine learning uncertainties," *The Review of Financial Studies*, p. hhaf087.

Amihud, Y. (2002): "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 5(1), 31–56.

Andersen, T. G., T. Bollerslev, P. Christoffersen, and F. X. Diebold (2005): "Volatility forecasting," *NBER Working Paper No. w11188*.

Ando, T., and J. Bai (2020): "Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity," *Journal of the American Statistical Association*, 115(529), 266–279.

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.

Avramov, D., S. Cheng, and L. Metzker (2022): "Machine learning versus economic restrictions: Evidence from stock return predictability," *Management Science*.

Azevedo, V., C. Hoegner, and M. Velikov (2023): "The expected returns on machine-learning strategies," *Available at SSRN 4702406*.

Bali, T. G., H. Beckmeyer, M. Mörke, and F. Weigert (2023): "Option Return Predictability with Machine Learning and Big Data," *The Review of Financial Studies*, hhad017.

Bali, T. G., R. F. Engle, and S. Murray (2016): *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

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.

Barberis, N., A. Shleifer, and R. Vishny (1998): "A model of investor sentiment," *Journal of financial economics*, 49(3), 307–343.

Bianchi, D., M. Büchner, and A. Tamoni (2021): "Bond risk premiums with machine learning," *The Review of Financial Studies*, 34(2), 1046–1089.

Bollerslev, T. (2022): "Realized semi (co) variation: Signs that all volatilities are not created equal," *Journal of Financial Econometrics*, 20(2), 219–252.

Bradic, J., J. Fan, and W. Wang (2011): "Penalized composite quasi-likelihood for ultrahigh dimensional variable selection," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(3), 325–349.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009): "Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns," *The Review of Financial Studies*, 22(9), 3411–3447.

Brunnermeier, M. K., and Y. Sannikov (2014): "A macroeconomic model with a financial sector," *American Economic Review*, 104(2), 379–421.

Bryzgalova, S., M. Pelger, and J. Zhu (2023): "Forest through the trees: Building cross-sections of stock returns," *Journal of Finance, forthcoming*.

Caruana, R. (1997): "Multitask learning," *Machine Learning*, 28, 41–75.

Catania, L., and A. Luati (2023): "Semiparametric modeling of multiple quantiles," *Journal of Econometrics*, 237(2), 105365.

Cenesizoglu, T., and A. Timmermann (2008): "Is the distribution of stock returns predictable?," *Available at SSRN 1107185*.

Chen, J., H. Hong, and J. C. Stein (2001): "Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices," *Journal of Financial Economics*, 61(3), 345–381.

Chen, L., J. J. Dolado, and J. Gonzalo (2021): "Quantile factor models," *Econometrica*, 89(2), 875–910.

Chen, L., M. Pelger, and J. Zhu (2023): "Deep learning in asset pricing," *Management Science*.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and probability curves without crossing," *Econometrica*, 78(3), 1093–1125.

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.

Choi, D., W. Jiang, and C. Zhang (2025): "Alpha go everywhere: Machine learning and international stock returns," *Review of Asset Pricing Studies*, p. raaf005.

Cochrane, J. H. (2011): "Presidential address: Discount rates," *The Journal of Finance*, 66(4), 1047–1108.

Cotturo, P., F. Liu, and R. Proner (2024): "Multi-Factor Timing with Deep Learning," *Available at SSRN*.

Dabney, W., G. Ostrovski, D. Silver, and R. Munos (2018): "Implicit quantile networks for distributional reinforcement learning," in *International Conference on Machine Learning*, pp. 1096–1105. PMLR.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam (2001): "Overconfidence, arbitrage, and equilibrium asset pricing," *The Journal of Finance*, 56(3), 921–965.

DeMiguel, V., A. Martín-utrera, and R. Uppal (2024): "A Multifactor Perspective on Volatility-Managed Portfolios," *The Journal of Finance*, 79(6), 3859–3891.

Diebold, F., and R. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–63.

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

Engle, R. F., and S. Manganelli (2004): "CAViaR: Conditional autoregressive value at risk by regression quantiles," *Journal of business & economic statistics*, 22(4), 367–381.

Fama, E. F. (1963): "Mandelbrot and the stable Paretian hypothesis," *The Journal of Business*, 36(4), 420–429.

——— (1965): "The behavior of stock-market prices," *The Journal of Business*, 38(1), 34–105.

Fama, E. F., and K. R. French (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

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.

Filippou, I., D. Rapach, M. P. Taylor, and G. Zhou (2021): "Exchange rate prediction with machine learning and a smart carry portfolio," *Available at SSRN 3455713*.

Freyberger, J., A. Neuhierl, and M. Weber (2020): "Dissecting characteristics nonparametrically," *The Review of Financial Studies*, 33(5), 2326–2377.

Friedman, J. H. (2001): "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, pp. 1189–1232.

Gallant, A. R., and H. White (1992): "On learning the derivatives of an unknown mapping with multilayer feedforward networks," *Neural Networks*, 5(1), 129–138.

Ghysels, E., A. Plazzi, and R. Valkanov (2016): "Why invest in emerging markets? The role of conditional return asymmetry," *The Journal of Finance*, 71(5), 2145–2192.

Giglio, S., B. Kelly, and S. Pruitt (2016): "Systemic risk and the macroeconomy: An empirical evaluation," *Journal of Financial Economics*, 119(3), 457–471.

Gouriéroux, C., and J. Jasiak (2008): "Dynamic quantile models," *Journal of Econometrics*, 147(1), 198–205.

Granger, C. W. (2010): "Some thoughts on the development of cointegration," *Journal of Econometrics*, 158(1), 3–6.

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.

Guijarro-Ordonez, J., M. Pelger, and G. Zanotti (2021): "Deep learning statistical arbitrage," *arXiv preprint arXiv:2106.04028*.

Hansen, B. E. (1994): "Autoregressive conditional density estimation," *International Economic Review*, pp. 705–730.

He, Z., and A. Krishnamurthy (2013): "Intermediary asset pricing," *American Economic Review*, 103(2), 732–70.

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.

Hornik, K., M. Stinchcombe, and H. White (1989): "Multilayer feedforward networks are universal approximators," *Neural Networks*, 2(5), 359–366.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021): "An augmented q-factor model with expected growth," *Review of Finance*, 25(1), 1–41.

Huddleston, D., F. Liu, and L. Stentoft (2023): "Intraday market predictability: A machine learning approach," *Journal of Financial Econometrics*, 21(2), 485–527.

Jensen, T. I., B. T. Kelly, S. Malamud, and L. H. Pedersen (2022): "Machine learning and the implementable efficient frontier," *Swiss Finance Institute Research Paper*, 1(22-63).

Jiang, C., E. Maasoumi, and Z. Xiao (2020): "Quantile aggregation and combination for stock return prediction," *Econometric Reviews*, 39(7), 715–743.

Jiang, J., B. Kelly, and D. Xiu (2023): "(Re-) Imag (in) ing price trends," *The Journal of Finance*, 78(6), 3193–3249.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1991): *The theory and practice of econometrics*. John wiley & sons.

Kaniel, R., Z. Lin, M. Pelger, and S. Van Nieuwerburgh (2023): "Machine-learning the skill of mutual fund managers," *Journal of Financial Economics*, 150(1), 94–138.

Keilbar, G., and W. Wang (2022): "Modelling systemic risk using neural network quantile regression," *Empirical Economics*, 62(1), 93–118.

Kelly, B. T., S. Pruitt, and Y. Su (2019): "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, 134(3), 501–524.

Kingma, D. P., and J. Ba (2017): "Adam: A Method for Stochastic Optimization," .

Koenker, R. (2005): *Quantile regression*, vol. 38. Cambridge university press.

Koenker, R., and G. Bassett Jr (1978): "Regression quantiles," *Econometrica: Journal of the Econometric Society*, pp. 33–50.

——— (1982): "Robust tests for heteroscedasticity based on regression quantiles," *Econometrica: Journal of the Econometric Society*, pp. 43–61.

Koenker, R., V. Chernozhukov, X. He, and L. Peng (2017): *Handbook of Quantile Regression*. CRC press.

Koenker, R., and J. A. Machado (1999): "Goodness of fit and related inference processes for quantile regression," *Journal of the American Statistical Association*, 94(448), 1296–1310.

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

Leippold, M., Q. Wang, and W. Zhou (2022): "Machine learning in the Chinese stock market," *Journal of Financial Economics*, 145(2), 64–82.

Lima, L. R., and F. Meng (2017): "Out-of-sample return predictability: A quantile combination approach," *Journal of Applied Econometrics*, 32(4), 877–895.

Liu, F., and L. Stentoft (2021): "Regulatory capital and incentives for risk model choice under Basel 3," *Journal of Financial Econometrics*, 19(1), 53–96.

——— (2023): "Intraday Stock Predictability Everywhere," *Available at SSRN 4496917*.

Ma, L., and L. Pohlman (2008): "Return forecasts and optimal portfolio construction: a quantile regression approach," *The European Journal of Finance*, 14(5), 409–425.

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *The Journal of Business*, 36(4), 394–419.

Meinshausen, N. (2006): "Quantile regression forests.," *Journal of Machine Learning Research*, 7(6).

Meligkotsidou, L., E. Panopoulou, I. D. Vrontos, and S. D. Vrontos (2014): "A quantile regression approach to equity premium prediction," *Journal of Forecasting*, 33(7), 558–576.

Menkveld, A. J., A. Dreber, F. Holzmeister, J. Huber, M. Johannesson, M. Kirchler, S. Neusüss, M. Razen, and U. Weitzel (2023): "Non-standard errors," *forthcoming in The Journal of Finance*.

Mikosch, T., and C. G. de Vries (2013): "Heavy tails of OLS," *Journal of Econometrics*, 172(2), 205–221.

Moreira, A., and T. Muir (2017): "Volatility-managed portfolios," *The Journal of Finance*, 72(4), 1611–1644.

Nair, V., and G. E. Hinton (2010): "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*.

Pedersen, T. Q. (2015): "Predictable return distributions," *Journal of Forecasting*, 34(2), 114–132.

Pohl, W., K. Schmedders, and O. Wilms (2018): "Higher order effects in asset pricing models with long-run risks," *The Journal of Finance*, 73(3), 1061–1111.

Pontiff, J. (2006): "Costly arbitrage and the myth of idiosyncratic risk," *Journal of Accounting and Economics*, 42(1-2), 35–52.

Ruder, S. (2017): "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*.

Schmidt, L., and Y. Zhu (2016): "Quantile spacings: A simple method for the joint estimation of multiple quantiles without crossing," *Available at SSRN 2220901*.

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

Stambaugh, R. F., and Y. Yuan (2017): "Mispricing factors," *The Review of Financial Studies*, 30(4), 1270–1315.

Tukey, J. W., et al. (1977): *Exploratory data analysis*, vol. 2. Springer.

Wan, X., W. Wang, J. Liu, and T. Tong (2014): "Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range," *BMC Medical Research Methodology*, 14(1), 1–13.

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.

White, H., T.-H. Kim, and S. Manganelli (2015): "VAR for VaR: Measuring tail dependence using multivariate regression quantiles," *Journal of econometrics*, 187(1), 169–188.

Wright, M. N., and A. Ziegler (2015): "ranger: A fast implementation of random forests for high dimensional data in C++ and R," *arXiv preprint arXiv:1508.04409*.

Yang, X., Z. Zhu, D. Li, and K. Zhu (2024): "Asset pricing via the conditional quantile variational autoencoder," *Journal of Business & Economic Statistics*, 42(2), 681–694.

Zhan, X., B. Han, J. Cao, and Q. Tong (2022): "Option return predictability," *The Review of Financial Studies*, 35(3), 1394–1442.

Zhang, X. F. (2006): "Information uncertainty and stock returns," *The Journal of Finance*, 61(1), 105–137.

Zhao, Z., and Z. Xiao (2014): "Efficient regressions via optimally combining quantile information," *Econometric Theory*, 30(6), 1272–1314.

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

Zou, H., and M. Yuan (2008): "Composite Quantile Regression and the Oracle Model Selection Theory," *The Annals of Statistics*, pp. 1108–1126.

Table 1: Monthly Out-of-Sample Quantile Prediction Performance (Percentage $R^1_{os}(\tau)$)

Panel A of this table reports the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ for the pooled test sample of individual stock excess returns from 1987-2020. We compare 12 models, including linear quantile regression with all covariates (LR), elastic net quantile regression (EN), quantile random forest (RF), quantile gradient boosted regression trees (GBT), quantile feed-forward neural networks with 1-4 layers (NN1,...,NN4), multi-quantile neural networks with 1-2 shared layers (MQ1, MQ2), and spaced multi-quantile neural networks with 1-2 shared layers (SMQ1, SMQ2). The $R^1_{os}(\tau)$ pools quantile prediction errors across stocks and over time into a grand panel-level assessment of each model for a given τ, and is defined as

$$R^1_{os}(\tau) = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3} \rho_\tau(r_{i,t+1} - \hat{r}_{\tau,i,t+1})}{\sum_{(i,t)\in\mathcal{T}_3} \rho_\tau(r_{i,t+1} - \bar{r}_\tau)}.$$

Panel B reports the percentage of observations where adjacent quantiles cross for each model.

Panel A: Monthly Out-of-Sample Quantile Prediction Performance (Percentage $R^1_{os}(\tau)$)												
τ	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
0.05	15.94	17.01	12.60	17.60	19.46	19.85	19.90	18.79	20.31	20.15	20.31	20.30
0.25	5.65	6.09	4.67	5.81	6.99	7.11	7.10	6.32	7.22	7.16	7.18	7.18
0.50	0.36	0.60	0.27	-0.47	0.84	0.71	0.75	0.62	0.89	0.85	0.79	0.86
0.75	1.92	2.15	1.37	2.10	2.69	2.57	2.65	2.22	2.74	2.70	2.71	2.80
0.95	12.72	12.87	10.79	13.97	15.69	15.86	15.92	15.19	16.26	16.36	16.22	16.54
Panel B: Adjacent Quantile Crossings in Percentages												
% Crossings	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
$Q_{0.05} > Q_{0.25}$	0.0768	0.0005	0	0.4870	0.0000	0.0002	0.0001	0.0052	0.0000	0.0001	0	0
$Q_{0.25} > Q_{0.50}$	0.1958	0.1220	0	0.5052	0.0011	0.0014	0.0004	0.0423	0.0000	0.0005	0	0
$Q_{0.50} > Q_{0.75}$	0.0714	0.0290	0	0.6453	0.0028	0.0039	0.0004	0.0008	0.0010	0.0001	0	0
$Q_{0.75} > Q_{0.95}$	0.1481	0.0137	0	0.0001	0.0000	0.0000	0.0003	0.0005	0.0009	0.0000	0	0

Table 2: Out-of-Sample Robust Risk Premium Prediction Performance (Percentage R^2_{os})

Panel A reports the out-of-sample R^2_{os} values (in percentages) of robust risk premium forecasts for the pooled test sample of individual stock excess returns from 1987-2020. The robust risk premium is defined as $\bar{Q} = 0.15Q_{0.05} + 0.225Q_{0.25} + 0.25Q_{0.50} + 0.225Q_{0.75} + 0.15Q_{0.95}$. We compare 12 models in total including linear quantile regression with all covariates (LR), elastic net quantile regression (EN), quantile random forest (RF), quantile gradient boosted regression trees (GBT), quantile neural networks with 1-4 layers (NN1,...,NN4), multi-quantile neural networks with 1-2 shared layers (MQ1 ,MQ2), and spaced multi-quantile neural networks with 1-2 shared layers (SMQ1, SMQ2). The R^2_{os} statistic pools prediction squared errors across stocks and over time into a grand panel-level assessment of each model, and it measures the proportional reduction in mean squared error for each model relative to a naive forecast of zero. The R^2_{os} is defined as

$$R^2_{os} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \bar{Q}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{i,t+1}}. \tag{13}$$

Panel B reports pairwise Diebold-Mariano test statistics, comparing the out-of-sample firm-level $\bar{Q}$ prediction performance among the 12 models. A positive number indicates that the column model outperforms the row model, and bold font signifies that the difference is statistically significant at the 5% level.

Panel A: Out-of-sample $\bar{Q}$ Prediction Performance (Percentage R^2_{OS})											
LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
0.131	0.477	0.165	0.364	0.889	0.874	0.977	0.870	0.989	0.920	0.965	1.001
Panel B: Comparison of $\bar{Q}$ Prediction Performance Using Diebold-Mariano Tests											
	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
LR	1.65	0.06	0.93	**2.12**	**2.29**	**2.25**	**2.16**	**3.00**	**2.63**	**2.80**	**2.80**
EN		-0.99	-0.51	**2.30**	**2.39**	**2.33**	**2.27**	**2.77**	**2.06**	**2.60**	**2.85**
RF			0.47	**2.24**	**2.10**	**2.40**	**2.22**	**2.58**	**2.15**	**2.47**	**2.59**
GBT				1.69	1.80	1.95	1.70	**2.31**	**2.14**	**2.11**	**2.11**
NN1					-0.17	1.22	0.04	0.90	0.22	0.67	0.95
NN2						1.48	0.23	1.06	0.40	0.88	1.26
NN3							-1.38	0.03	-0.51	-0.11	0.16
NN4								0.73	0.19	0.58	0.86
MQ1									-0.88	-0.47	0.24
MQ2										0.65	0.94
SMQ1											0.75

Table 3: Performance of Value-Weighted Decile Portfolios Sorted on Robust Risk Premium Forecasts

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on robust risk premium forecasts. At the end of each month, we predict the one-month-ahead robust risk premium, and then sort stocks into deciles based on their forecasts. We reconstitute portfolios each month using value weights. "Low" denotes the portfolio with the lowest robust risk premium forecasts (decile 1), "High" denotes the portfolio with the highest robust risk premium forecasts(decile 10), "H-L" denotes the long-short portfolio that buys the highest robust risk premium portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Panel B presents the alphas (α), Newey-West t-statistics (t(α)), R^2 values, and information ratios (IR) for the value-weighted long-short portfolios, with respect to the Fama-French five-factor model augmented to include the momentum factor.

Panel A: Performance of Portfolios Sorted on $\bar{Q}$ Forecasts

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
Low (L)	0.01	0.01	0.02	-0.89	-0.90	-0.86	-0.87	-0.67	-1.09	-1.20	-1.03	-1.18
2	0.45	0.39	0.44	-0.12	0.11	0.05	-0.01	0.24	-0.21	-0.15	-0.22	-0.18
3	0.55	0.63	0.55	0.16	0.33	0.41	0.41	0.49	0.37	0.27	0.30	0.29
4	0.64	0.64	0.65	0.48	0.57	0.65	0.64	0.71	0.64	0.55	0.56	0.53
5	0.70	0.70	0.73	0.64	0.69	0.78	0.77	0.86	0.73	0.73	0.74	0.77
6	0.81	0.80	0.75	0.69	0.81	0.84	0.88	0.93	0.85	0.89	0.93	0.88
7	0.91	0.88	0.94	0.92	0.97	1.02	1.03	1.09	1.08	1.00	1.06	1.08
8	0.91	0.95	0.90	1.01	1.07	1.18	1.28	1.26	1.13	1.12	1.23	1.21
9	1.03	0.95	1.06	1.11	1.33	1.40	1.42	1.43	1.43	1.43	1.53	1.58
High (H)	1.16	1.23	1.20	1.30	1.82	1.85	1.93	1.94	1.88	1.91	1.98	1.98
H-L	1.15	1.21	1.18	2.19	2.72	2.71	2.80	2.61	2.96	3.11	3.01	3.17
t	4.86	4.71	4.34	6.21	8.36	8.32	8.72	7.47	9.49	9.74	9.10	9.23
SR	0.91	0.89	0.77	1.10	1.44	1.41	1.42	1.26	1.53	1.55	1.50	1.53

Panel B: Risk-Adjusted Performance of Portfolios Sorted on $\bar{Q}$ Forecasts

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
α	0.64	0.65	1.21	1.69	2.26	2.26	2.42	2.25	2.51	2.61	2.62	2.72
t(α)	4.29	3.93	4.93	6.12	8.42	8.24	8.36	6.1	10.06	9.91	9.82	9.47
R^2	46	52	18	47	40	33	24	18	33	29	31	26
IR	0.7	0.69	0.88	1.17	1.54	1.43	1.41	1.2	1.58	1.54	1.57	1.52

Table 4: Comparison to MSE-Based Approach

This table compares the performance of the robust risk premium ($\bar{Q}$) and MSE-based ($\bar{\mu}^{MSE}$) forecasts for the pooled test sample and for subsamples evenly partitioned by firm size. Panel A reports the pairwise correlations between the $\bar{Q}$ and $\bar{\mu}^{MSE}$ forecasts, the R^2_{os} values for both approaches, and the Diebold–Mariano (DM) statistics described in Section 5.2. A positive DM statistic indicates that $\bar{Q}$ outperforms $\bar{\mu}^{MSE}$; bold values denote statistical significance at the 5% level. Panel B reports the average monthly excess returns of the long–short portfolios formed using $\bar{Q}$ and $\bar{\mu}^{MSE}$ forecasts. It also reports the alphas (α_{Span}) and Newey-West t-statistics (t(α)) from the time-series regression

$$f^{\bar{Q}}_{t+1} = \alpha_{Span} + \beta f^{\bar{\mu}^{MSE}}_{t+1} + \epsilon_{t+1}, \tag{14}$$

where $f^{\bar{Q}}_{t+1}$ and $f^{\bar{\mu}^{MSE}}_{t+1}$ denote the excess returns of the long–short portfolios based on the $\bar{Q}$ and $\bar{\mu}^{MSE}$ forecasts, respectively. We benchmark the MQ and SMQ models against the MSE-based NN3 model.

Panel A: Statistical Performance Comparison Between $\bar{Q}$ and $\bar{\mu}^{MSE}$ Forecasts

Sample	Metric	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
All	Correlation	0.33	0.80	0.39	0.57	0.83	0.60	0.60	0.53	0.60	0.59	0.60	0.60
	R^2_{os} for $\bar{Q}$	0.13	0.48	0.17	0.36	0.89	0.87	0.98	0.87	0.99	0.92	0.96	1.00
	R^2_{os} for $\bar{\mu}^{MSE}$	-0.23	0.33	0.22	0.15	0.50	0.37	0.56	0.39	0.56	0.56	0.56	0.56
	DM	0.87	1.41	-0.15	0.71	1.92	**3.46**	**2.18**	**2.70**	**2.35**	1.65	**2.10**	**2.29**
Small	Correlation	0.34	0.80	0.44	0.57	0.86	0.61	0.66	0.55	0.66	0.65	0.66	0.66
	R^2_{os} for $\bar{Q}$	0.12	0.48	0.21	0.47	1.02	1.00	1.11	0.97	1.11	1.05	1.10	1.14
	R^2_{os} for $\bar{\mu}^{MSE}$	0.17	0.41	0.31	0.24	0.70	0.44	0.66	0.39	0.66	0.66	0.66	0.66
	DM	-0.12	0.88	-0.29	1.03	1.90	**3.47**	**2.41**	**2.99**	**2.40**	**1.96**	**2.20**	**2.39**
Large	Correlation	0.33	0.81	0.32	0.58	0.76	0.57	0.46	0.47	0.45	0.45	0.46	0.45
	R^2_{os} for $\bar{Q}$	0.17	0.47	0.05	0.07	0.50	0.50	0.59	0.57	0.65	0.53	0.57	0.59
	R^2_{os} for $\bar{\mu}^{MSE}$	-1.39	0.11	-0.03	-0.10	-0.10	0.17	0.26	0.39	0.26	0.26	0.26	0.26
	DM	1.78	1.93	0.06	0.28	1.77	1.77	1.11	0.86	1.49	0.74	1.22	1.31

Panel B: Economic Performance Comparison Between $\bar{Q}$ and $\bar{\mu}^{MSE}$ Forecasts

Sample	Metric	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
All	H-L for $\bar{Q}$	1.15	1.21	1.18	2.19	2.72	2.71	2.80	2.61	2.96	3.11	3.01	3.17
	H-L for $\bar{\mu}^{MSE}$	0.86	0.96	0.95	1.14	2.49	1.99	2.23	1.82	2.23	2.23	2.23	2.23
	α_{Span}	0.63	0.39	0.87	1.33	0.53	1.21	1.13	1.28	1.42	1.47	1.28	1.42
	t(α_{Span})	2.82	2.52	2.99	4.20	2.74	4.29	4.17	3.94	5.01	5.05	4.20	4.95
Small	H-L for $\bar{Q}$	2.67	2.67	2.19	3.82	4.43	4.38	4.73	4.49	4.62	4.77	4.68	4.72
	H-L for $\bar{\mu}^{MSE}$	2.05	2.02	2.22	1.68	4.15	3.49	3.78	2.78	3.78	3.78	3.78	3.78
	α_{Span}	1.61	0.96	1.76	2.60	0.74	1.17	1.42	3.32	1.56	1.43	1.39	1.37
	t(α_{Span})	4.59	3.99	4.04	7.13	2.80	3.62	3.95	5.19	5.45	4.57	4.24	3.99
Large	H-L for $\bar{Q}$	1.02	1.06	0.89	1.58	1.82	1.81	1.95	1.83	2.02	2.03	2.03	2.06
	H-L for $\bar{\mu}^{MSE}$	0.77	0.84	0.61	0.96	1.52	1.20	1.53	1.21	1.53	1.53	1.53	1.53
	α_{Span}	0.56	0.36	0.69	0.87	0.49	0.86	0.74	0.89	0.88	0.75	0.82	0.83
	t(α_{Span})	2.43	2.75	2.80	3.72	3.15	3.47	3.51	3.74	4.08	3.29	3.68	3.52

Table 5: Drawdowns, Turnover, and Transaction Costs Analyses of Value-weighted Long-Short Portfolios

Panel A provides the maximum drawdowns ("Max DD"), average monthly percentage change in holdings ("Turnover"), and break-even transaction costs ("Break-even TC") for value-weighted long-short portfolios sorted on robust risk premium ($\bar{Q}$) forecasts. It also reports average monthly returns net of transaction costs ("Net H–L"), their Newey–West t-statistics ("t(Net)"), and annualized Sharpe ratios ("Net SR"). Panel B presents the corresponding results for MSE-based ($\bar{\mu}^{MSE}$) forecasts. Following Avramov, Cheng, and Metzker (2022), the break-even transaction cost is computed as the six-factor alpha divided by turnover. Transaction costs follow Brandt, Santa-Clara, and Valkanov (2009). For stock i at time t, the cost is $c_{i,t}|w_{i,t} - w_{i,t-1}|$, where $w_{i,t}$ denotes the weight of stock i in the portfolio and $c_{i,t}$ is the one-way transaction cost; portfolio-level costs are calculated by summing across stocks. Inspired by DeMiguel, Martín-utrera, and Uppal (2024), we estimate $c_{i,t}$ as one-half of the average monthly bid–ask spread using the CHL estimator of Abdi and Ranaldo (2017).

Panel A: Value-Weighted Portfolios Sorted on $\bar{Q}$ Forecasts

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
Max DD(%)	51.16	55.3	65.36	60.46	36.48	34.36	39.02	43.21	34.25	35.50	50.58	40.84
Turnover(%)	133.2	134.0	176.4	153.6	140.7	144.2	145.1	146.8	145.6	147.9	147.9	149.4
Breakeven TC(%)	0.48	0.49	0.69	1.10	1.61	1.57	1.67	1.53	1.72	1.76	1.77	1.82
Net H-L	0.55	0.61	0.17	1.24	1.82	1.75	1.81	1.56	1.97	2.06	1.94	2.06
t(Net)	2.34	2.39	0.66	3.56	5.71	5.50	5.84	4.62	6.52	6.72	6.07	6.27
Net SR	0.44	0.45	0.12	0.62	0.97	0.91	0.93	0.77	1.03	1.04	0.98	1.01

Panel B: Value-Weighted Portfolios Sorted on $\bar{\mu}^{MSE}$ Forecasts

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4
Max DD(%)	49.67	60.91	87.33	65.99	49.59	40.72	44.46	94.77
Turnover(%)	125.2	124.8	132.9	98.2	139.3	130.9	131.1	122.4
Breakeven TC(%)	0.45	0.42	0.71	0.50	1.37	1.26	1.37	1.20
Net H-L	0.31	0.40	0.23	0.63	1.62	1.21	1.35	0.97
t(Net)	1.56	1.48	0.72	2.55	4.97	3.69	3.89	3.02
Net SR	0.30	0.30	0.13	0.42	0.89	0.66	0.75	0.52

Table 6: Value-Weighted Bivariate Portfolios Sorted on Analyst Coverage and Robust Risk Premium Forecasts

This table reports the monthly performance of value-weighted portfolios, first sorted by analyst coverage (CVRG) and then by the robust risk premium $\bar{Q}$ forecasts for the EN, GBT, NN3, and MQ1 models. We sort stocks into terciles based on the number of analysts covering the firm in the most recent month, then we divide each CVRG tercile into deciles based on their $\bar{Q}$ forecasts. Within each CVRG tercile, "Low" denotes the portfolio with the lowest $\bar{Q}$ forecasts (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ forecasts (decile 10), "H-L" denotes the long-short portfolio that buys the highest $\bar{Q}$ portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Alphas (α) and Newey-West t-statistics (t(α)) for the long-short portfolios are with respect to the Fama-French five-factor model augmented to include the momentum factor.

	Panel A: EN			Panel B: GBT			Panel C: NN3			Panel D: MQ1		
	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3
Low (L)	-0.58	-0.43	0.08	-1.98	-1.42	-0.22	-1.40	-1.35	-0.46	-1.52	-1.47	-0.52
2	0.19	0.22	0.49	-0.32	-0.25	0.17	-0.04	-0.32	0.03	-0.14	-0.35	-0.06
3	0.56	0.47	0.73	0.19	0.00	0.34	0.45	0.15	0.45	0.39	-0.03	0.46
4	0.45	0.53	0.60	0.37	0.44	0.60	0.45	0.57	0.68	0.33	0.36	0.71
5	0.57	0.72	0.76	0.60	0.60	0.69	0.62	0.71	0.75	0.61	0.76	0.75
6	0.55	0.76	0.80	0.63	0.76	0.79	0.68	1.02	0.88	0.61	0.88	0.92
7	0.63	0.90	0.86	0.54	0.95	0.87	0.98	1.08	0.95	0.86	1.09	1.01
8	0.75	0.86	0.91	0.90	1.01	0.98	1.20	1.28	1.05	1.11	1.21	1.04
9	1.04	0.96	0.98	1.09	1.19	1.11	1.60	1.51	1.29	1.69	1.50	1.25
High (H)	1.20	1.19	1.16	1.44	1.33	1.32	2.44	2.16	1.55	2.35	2.06	1.48
H-L	1.78	1.62	1.08	3.42	2.75	1.54	3.84	3.51	2.00	3.87	3.53	2.00
t	6.12	5.58	4.00	9.80	7.22	4.39	9.60	10.12	6.08	10.43	10.22	6.13
SR	1.12	1.05	0.77	1.49	1.30	0.78	1.70	1.72	1.00	1.88	1.73	1.01
α	1.50	0.99	0.53	3.18	2.33	1.07	3.95	3.19	1.48	3.94	3.09	1.45
t(α)	6.22	4.52	3.07	9.23	6.77	3.75	8.58	8.54	5.24	8.93	10.10	5.81

Table 7: Value-Weighted Bivariate Portfolios Sorted on Arbitrage Index and Robust Risk Premium Forecasts

This table reports the monthly performance of value-weighted portfolios, first sorted by an index of arbitrage costs and then by the robust risk premium $\bar{Q}$ forecasts for the EN, GBT, NN3, and MQ1 models. Arbitrage Index (AI) construction follows Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023), and includes idiosyncratic volatility, illiquidity, size, age, and analyst coverage. We sort stocks into quintiles based on their AI, then we divide each AI quintile into deciles based on their $\bar{Q}$ forecasts. Within each AI quintile, "Low" denotes the portfolio with the lowest $\bar{Q}$ forecasts (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ forecasts (decile 10), "H-L" denotes the long-short portfolio that buys the highest $\bar{Q}$ portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Alphas (α) and Newey-West t-statistics (t(α)) for the long-short portfolios are with respect to the Fama-French five-factor model augmented to include the momentum factor.

	Panel A: EN					Panel B: GBT					Panel C: NN3					Panel D: MQ1				
	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5
Low (L)	0.21	0.09	-0.51	-1.18	-2.15	0.17	-0.02	-0.67	-1.67	-3.41	0.11	-0.16	-0.77	-1.74	-3.04	0.15	-0.19	-0.83	-1.73	-2.98
2	0.60	0.64	0.16	-0.46	-0.84	0.43	0.33	-0.10	-0.39	-0.91	0.42	0.46	-0.06	-0.38	-0.90	0.46	0.46	0.00	-0.36	-1.11
3	0.62	0.71	0.56	0.28	-0.44	0.63	0.74	0.38	0.04	-0.70	0.69	0.78	0.38	0.19	-0.51	0.51	0.57	0.37	-0.08	-0.37
4	0.77	0.77	0.67	0.68	-0.19	0.70	0.68	0.41	0.58	0.04	0.70	0.79	0.76	0.59	0.32	0.67	0.84	0.66	0.48	-0.10
5	0.85	0.83	0.82	0.68	0.43	0.74	0.84	0.87	0.53	0.31	0.81	1.01	0.96	0.59	0.48	0.90	0.93	0.94	0.72	0.33
6	0.79	0.81	0.82	0.81	0.76	0.80	0.83	0.99	0.95	0.61	0.95	1.01	1.16	1.11	1.01	0.96	0.97	0.87	1.08	0.95
7	0.90	0.97	0.99	1.03	0.62	0.86	1.16	1.12	1.00	1.02	0.90	1.21	1.13	1.38	1.41	1.02	1.14	1.22	1.21	1.35
8	0.93	1.17	1.14	1.17	1.03	0.92	1.24	1.30	1.37	1.37	0.91	1.10	1.31	1.58	1.79	0.92	1.24	1.34	1.61	1.93
9	0.90	1.20	1.17	1.32	1.52	1.03	1.34	1.32	1.47	1.78	1.16	1.43	1.46	1.73	2.52	0.99	1.35	1.45	1.64	2.29
High (H)	1.14	1.46	1.49	1.64	1.99	1.24	1.33	1.52	2.00	2.95	1.27	1.66	1.79	2.58	4.23	1.29	1.60	1.72	2.50	4.10
H-L	0.94	1.37	2.00	2.82	4.14	1.08	1.34	2.19	3.66	6.36	1.16	1.82	2.55	4.32	7.27	1.14	1.79	2.55	4.23	7.08
t	4.50	4.05	5.96	8.51	11.63	4.77	4.12	6.30	9.61	15.75	5.38	5.80	7.53	11.04	13.16	5.37	5.40	7.54	11.13	14.06
SR	0.79	0.82	1.14	1.52	2.20	0.85	0.71	1.17	1.67	2.82	0.88	0.93	1.28	2.13	2.84	0.90	0.94	1.36	2.22	2.94
α	0.49	0.72	1.38	2.23	3.56	0.87	0.86	1.87	3.36	5.98	0.87	1.38	2.19	3.97	7.29	0.75	1.16	2.11	3.85	7.00
t(α)	3.34	2.72	6.65	8.67	10.76	3.95	3.21	6.80	10.07	14.31	4.53	5.32	7.61	10.91	12.57	4.04	5.08	8.13	11.30	13.67

Table 8: Value-Weighted Bivariate Portfolios Sorted on Changes and Levels of Robust Risk Premium Forecasts
This table reports the monthly performance of value-weighted portfolios, first sorted by the monthly change in the robust risk premium ($\Delta\bar{Q}$) forecasts and then by the robust risk premium ($\bar{Q}$) forecasts for the EN, GBT, NN3, and MQ1 models. For each firm i, we define $\Delta\bar{Q}$ as the robust risk premium forecast at the end of month t minus the robust risk premium forecast at the end of month $t-1$. We sort stocks into terciles based on $\Delta\bar{Q}$, then we divide each $\Delta\bar{Q}$ tercile into deciles based on their $\bar{Q}$ forecasts. Within each $\Delta\bar{Q}$ tercile, "Low" denotes the portfolio with the lowest $\bar{Q}$ forecasts (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ forecasts (decile 10), "H-L" denotes the long-short portfolio that buys the highest $\bar{Q}$ portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Alphas (α) and Newey-West t-statistics (t(α)) for the long-short portfolios are with respect to the Fama-French five-factor model augmented to include the momentum factor.

Performance of Bivariate Portfolio Sorts on $\Delta\bar{Q}$ and $\bar{Q}$ Forecasts												
	Panel A: EN			Panel B: GBT			Panel C: NN3			Panel D: MQ1		
	$\Delta\bar{Q}1$	$\Delta\bar{Q}2$	$\Delta\bar{Q}3$	$\Delta\bar{Q}1$	$\Delta\bar{Q}2$	$\Delta\bar{Q}3$	$\Delta\bar{Q}1$	$\Delta\bar{Q}2$	$\Delta\bar{Q}3$	$\Delta\bar{Q}1$	$\Delta\bar{Q}2$	$\Delta\bar{Q}3$
Low (L)	-0.31	0.12	0.14	-1.87	-0.67	-0.70	-1.74	-0.18	-0.39	-1.86	-0.22	-0.73
2	0.16	0.51	0.74	-0.32	0.11	-0.09	-0.67	0.34	0.19	-0.67	0.22	0.46
3	0.34	0.64	0.67	-0.13	0.56	0.44	-0.06	0.59	0.85	-0.29	0.52	0.68
4	0.56	0.57	0.75	0.13	0.58	0.64	0.14	0.64	1.07	0.01	0.79	1.14
5	0.48	0.75	0.84	0.18	0.64	0.80	0.13	0.74	1.07	0.18	0.73	1.18
6	0.60	0.70	0.87	0.43	0.85	1.08	0.57	0.95	1.24	0.31	0.79	1.38
7	0.60	0.85	0.91	0.56	0.93	1.11	0.70	0.97	1.35	0.69	1.01	1.39
8	0.84	1.01	0.98	0.69	1.05	1.14	0.87	1.12	1.52	0.85	1.27	1.51
9	1.01	1.04	1.17	0.87	1.14	1.35	1.08	1.49	1.80	0.97	1.12	1.94
High (H)	1.12	1.13	1.18	1.17	1.26	1.66	1.25	1.54	2.85	1.14	1.44	2.95
H-L	1.42	1.01	1.04	3.04	1.93	2.36	2.98	1.72	3.23	3.00	1.67	3.67
t	4.44	3.90	3.81	7.36	5.81	6.52	8.05	6.04	6.30	8.47	5.14	7.35
SR	0.86	0.74	0.71	1.34	1.07	1.14	1.36	1.07	1.28	1.41	0.97	1.57
α	0.87	0.57	0.61	2.45	1.60	1.90	2.44	1.31	3.03	2.49	1.20	3.35
t(α)	3.70	2.78	2.98	7.94	5.71	5.46	7.93	5.00	5.11	8.48	4.47	6.19

Figure 1: Multi-quantile Neural Network Architecture

This figure illustrates an example of a multi-quantile neural network architecture used to forecast quantiles for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$. The "hard sharing" hidden layers within the MQ interact and nonlinearly transform the input predictors into common latent features, imposing an economically motivated restriction of a common structure across quantiles. The MQ includes "task-specific" hidden layers that capture each quantile's nonlinear exposures to the shared latent features, subsequently aggregating these into a final prediction for each quantile.

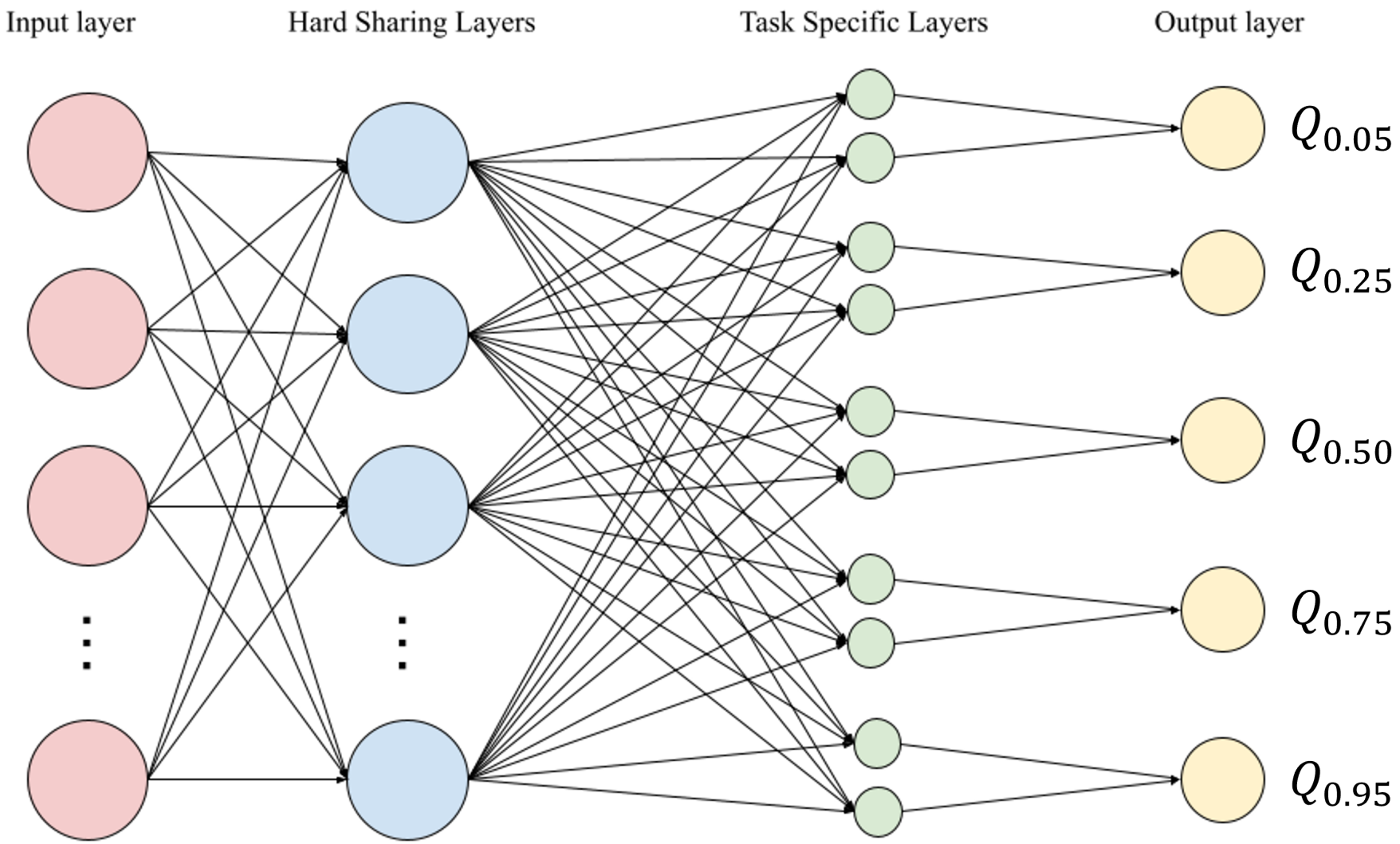


Figure 2: Spaced Multi-quantile Neural Network Architecture

This figure illustrates an example of a spaced multi-quantile neural network architecture used to forecast quantiles for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$. The "hard sharing" hidden layers within the SMQ interact and nonlinearly transform the input predictors into common latent features, imposing an economically motivated restriction of a common structure across quantiles. The SMQ includes "task-specific" hidden layers that capture each quantile's nonlinear exposures to the shared latent features, subsequently aggregating these into a final prediction for each quantile. For a set of τ, where $0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K < 1$, the SMQ model is recursively defined as:

$$Q_{\tau_k} = \begin{cases} Q_{\tau_{k+1}} - SP(Y_{\tau_k}) & \text{if } \tau_k < \tau_{k^\star}, \\ Q_{\tau_{k^\star}} & \text{if } \tau_k = \tau_{k^\star}, \\ Q_{\tau_{k-1}} + SP(Y_{\tau_k}) & \text{if } \tau_k > \tau_{k^\star}, \end{cases}$$

where $Q_{\tau_{k^\star}}$ denotes the reference quantile and $Y_{\tau_k} \equiv Y_{\tau_{i,t+1}}(\tau_k|\mathcal{F}_t)$ denotes the neural network output for τ_k. The function $SP(\cdot)$ denotes the softplus function, defined as $SP(x) = \log(1 + \exp(x))$, which is positive, differentiable, and monotonically increasing. In this context, the softplus is used to model the spacing between adjacent quantiles, ensuring that all estimated quantiles are increasing by construction.

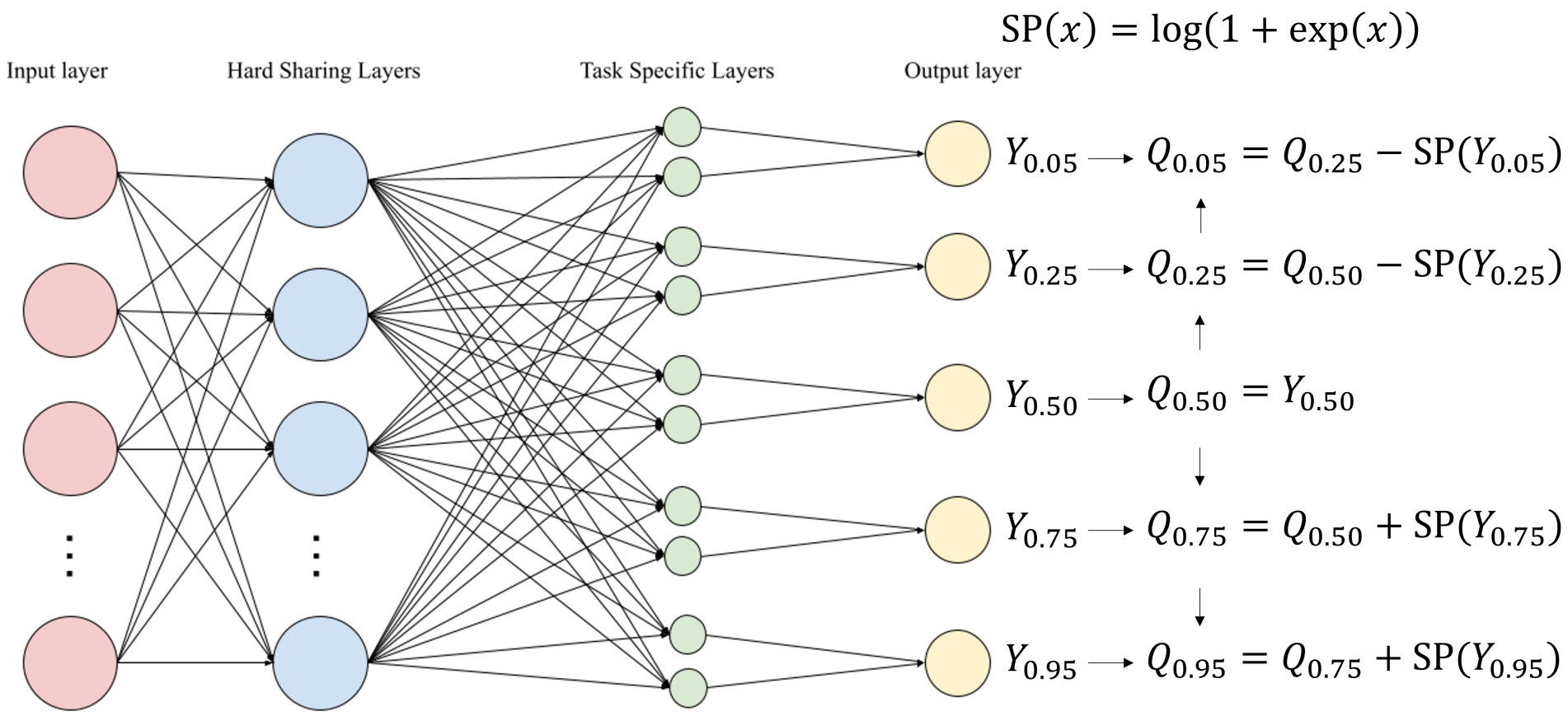


Figure 3: Robust Risk Premium Weights on Example Distribution

This figure illustrates how we derive the weights in Equation (5) using an example distribution, by evenly partitioning the probability between two adjacent quantiles and allocating the partitioned probability to the nearest quantiles; with an additional 5% allocated to $Q_{0.05}$ and $Q_{0.95}$. Specifically, our weights are calculated as $\{((\tau_1-0)+\frac{\tau_2-\tau_1}{2}),(\frac{\tau_2-\tau_1}{2}+\frac{\tau_3-\tau_2}{2}),(\frac{\tau_3-\tau_2}{2}+\frac{\tau_4-\tau_3}{2}),(\frac{\tau_4-\tau_3}{2}+\frac{\tau_5-\tau_4}{2}),(\frac{\tau_5-\tau_4}{2}+(1-\tau_5))\}$ or $\{((0.05-0)+\frac{0.25-0.05}{2}),(\frac{0.25-0.05}{2}+\frac{0.5-0.25}{2}),(\frac{0.5-0.25}{2}+\frac{0.75-0.5}{2}),(\frac{0.75-0.5}{2}+\frac{0.95-0.75}{2}),(\frac{0.95-0.75}{2}+(1-0.95))\}$. More generally, for a given set $0<\tau_1<\cdots<\tau_{K-1}<\tau_K<1$, the measure can be expressed as Equation (6), which converges to the conditional expectation of the respective distribution under general assumptions, as mathematically proven in Section IA2 in the Internet Appendix. Therefore, the robust risk premium measure is an estimator of the conditional expectation of excess returns, i.e., the risk premium. This example distribution is a skewed Student's t (Hansen (1994)) with 15 degrees of freedom and a skewness parameter of 0.2, 1 variance, and 0 expected value.

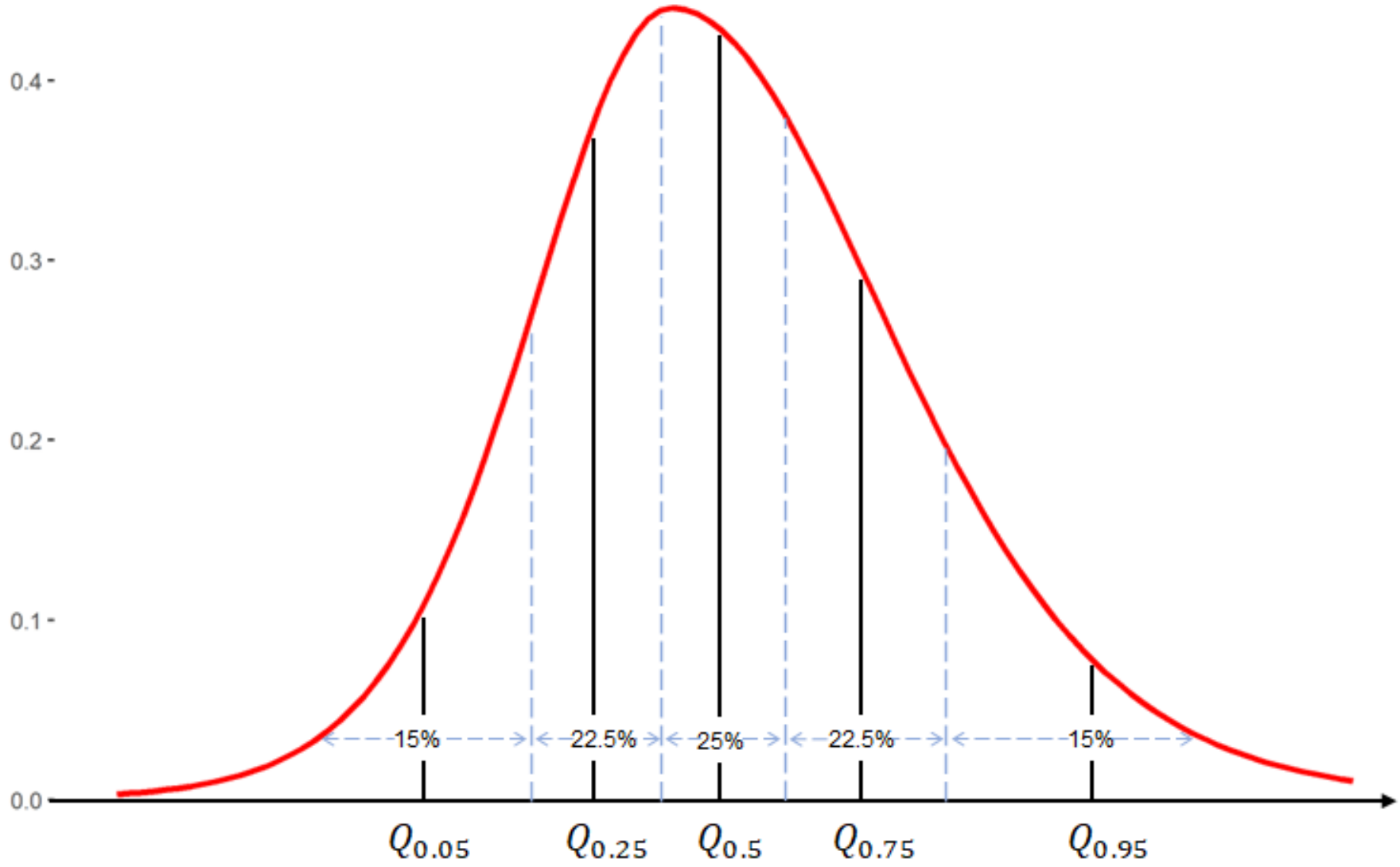


Figure 4: Cumulative Return of Long-Short Portfolios Sorted on Robust Risk Premium Forecasts

This figure displays the cumulative log returns of long-short portfolios (value-weighted) sorted on robust risk premium $\bar{Q}$ forecasts for the 12 models. At the end of each month, we calculate the one-month-ahead forecasted conditional robust risk premium measure $\bar{Q} = 0.15Q_{0.05} + 0.225Q_{0.25} + 0.25Q_{0.50} + 0.225Q_{0.75} + 0.15Q_{0.95}$, then sort stocks into deciles based on $\bar{Q}$ forecasts into value-weighted portfolios. The High-Low portfolio corresponds to the long-short portfolio that buys stocks with the highest robust risk premium measure $\bar{Q}$ forecasts (decile 10) and sells stocks with the lowest (decile 1). mktrf denotes the cumulative log return of the CRSP value-weighted market portfolio. The shaded periods show NBER recession dates.

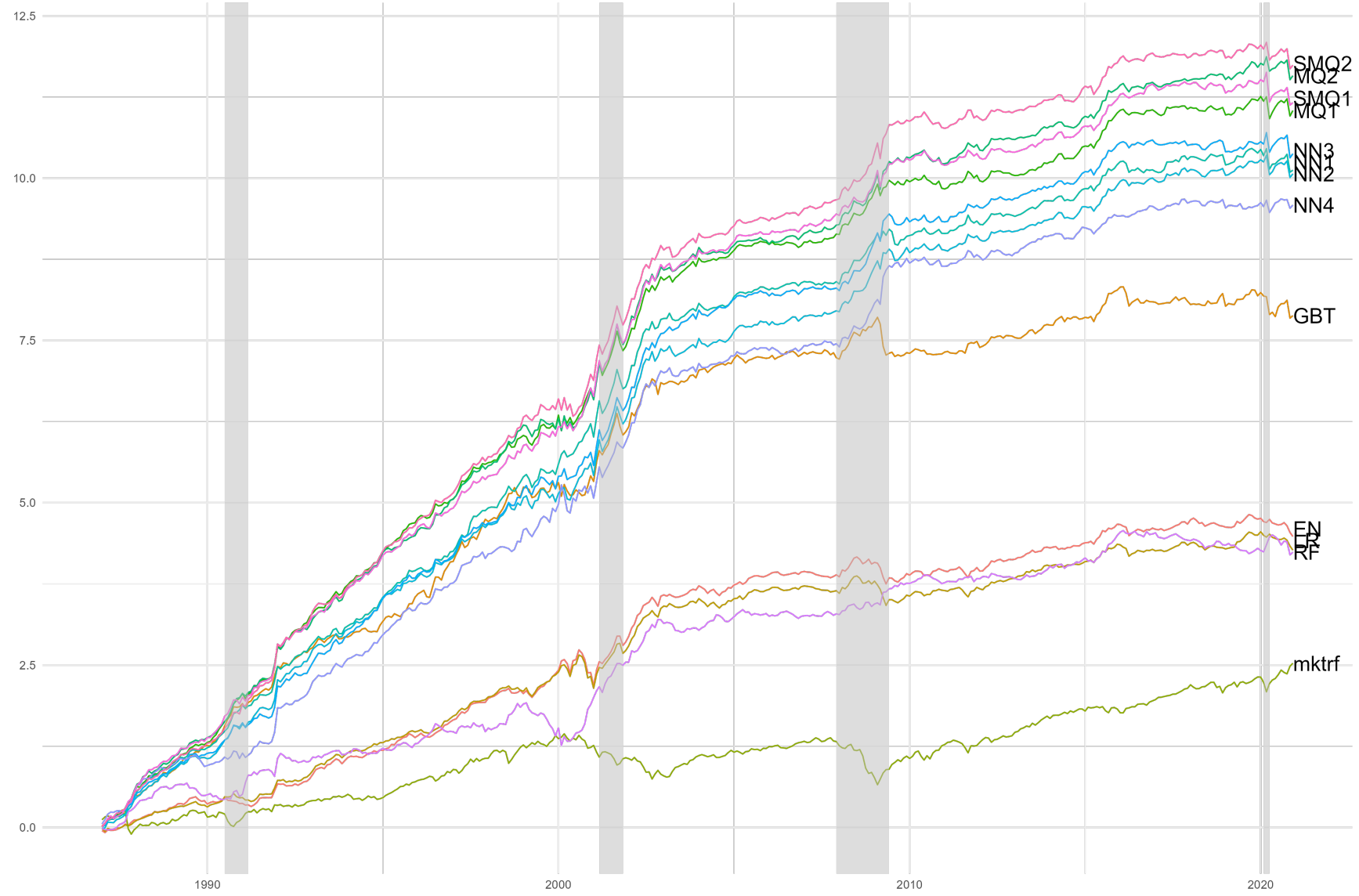


Figure 5: Return Distribution Predictability Conditional on Investor Inattention

This figure shows the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ and R^2_{os} values (in percentages) of the MQ1 model for stocks sorted by a proxy for investor attention. We sort stocks into terciles based on the number of analysts covering the firm in the most recent month. Stocks with lower analyst coverage (CVRG1) receive less investor attention, while stocks with higher analyst coverage (CVRG3) receive more investor attention.

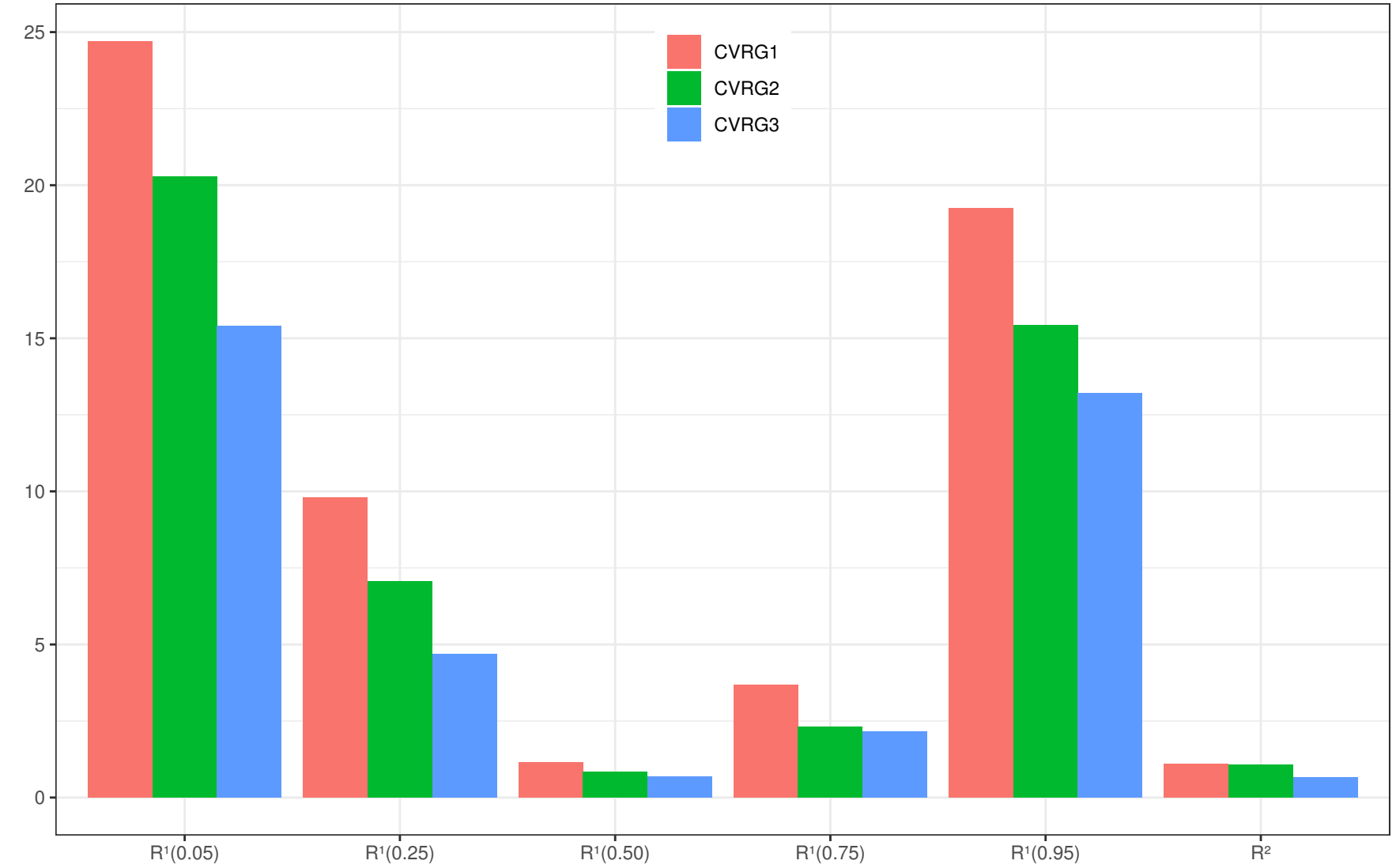
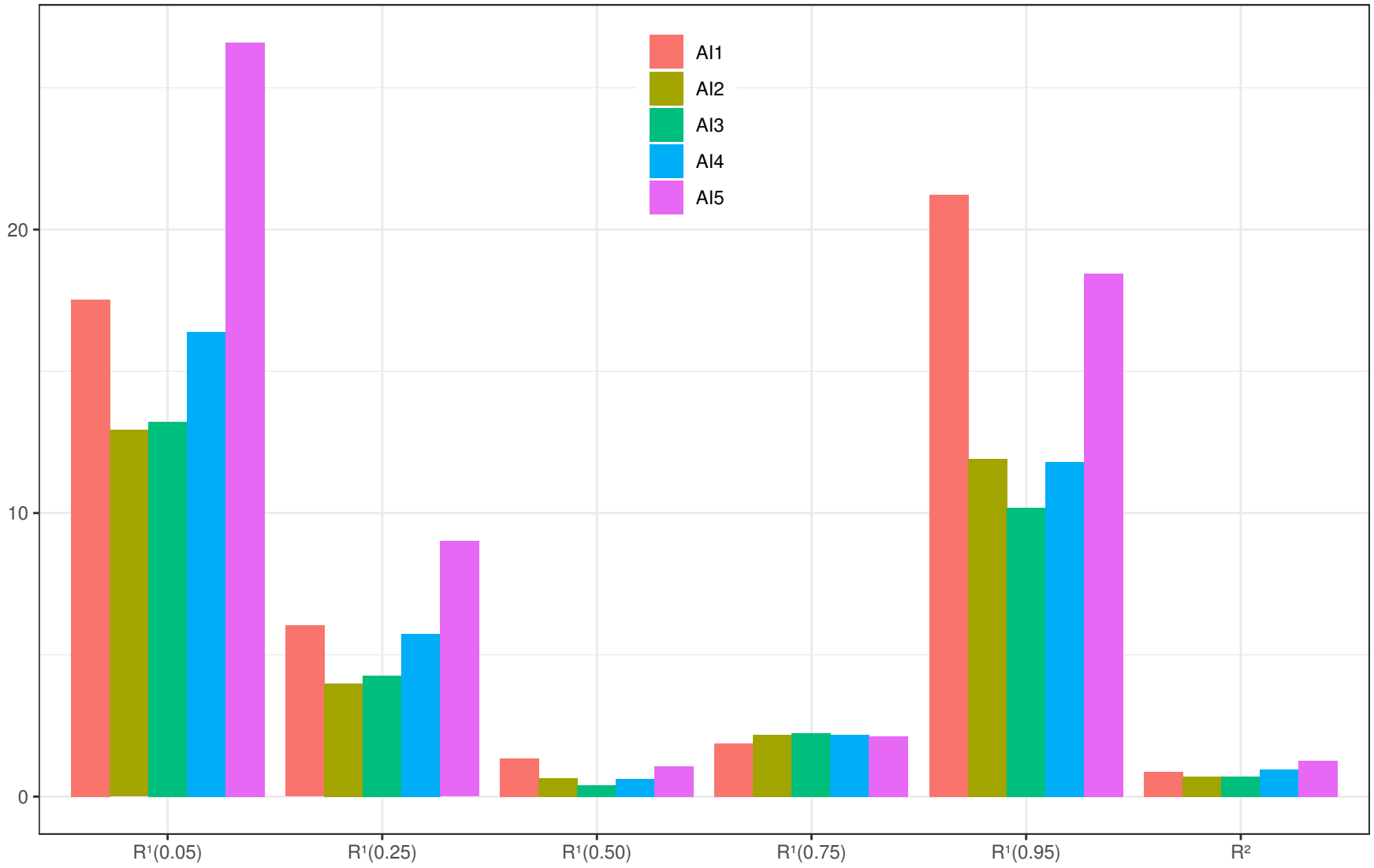


Figure 6: Return Distribution Predictability Conditional on Information Frictions

This figure shows the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ and R^2_{os} values (in percentages) of the MQ1 model for stocks sorted by an index of arbitrage costs. Arbitrage Index (AI) construction follows Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023), and includes idiosyncratic volatility, illiquidity, size, age, and analyst coverage. We sort stocks into quintiles based on their AI. Stocks with the lowest Arbitrage Index (AI1) face the least information frictions, whereas stocks with the highest Arbitrage Index (AI5) face the greatest information frictions.



Stock Distribution and Risk Premia Predictability with Quantile Machine Learning

Internet Appendix

Table of Contents:

- Section IA1 provides details on the **off-the-shelf quantile machine learning models** used in this paper.
- Section IA2 provides a **mathematical proof** that the robust risk premium convergences to the expected value.
- Section IA3 provides details on the **hyperparameter optimization framework** for the quantile and MSE-based models.
- Table IA1 presents the **set of hyperparameters** used for the quantile models.
- Table IA2 presents the **set of hyperparameters** used for the MSE-based models.
- Table IA3 presents the **replication results** for the MSE-based models.
- Table IA4 reports the **pairwise Diebold-Mariano test statistics** for the quantile model forecasts.
- Table IA5 compares the performance of portfolios sorted on the robust risk premium and MSE-based forecasts using **NYSE breakpoints.**
- Table IA6 compares the performance of portfolios sorted on the robust risk premium and MSE-based forecasts, **excluding microcap stocks.**
- Table IA7 compares the performance of **equal-weighted portfolios** sorted on the robust risk premium and MSE-based forecasts.
- Table IA8 presents **bivariate portfolios sorted on earnings events and robust risk premium forecasts.**

- Table IA9 presents **bivariate portfolios sorted on analyst coverage and MSE-based forecasts.**
- Table IA10 presents **bivariate portfolios sorted on information frictions and MSE-based forecasts.**
- Table IA11 presents **bivariate portfolios sorted on changes and levels of MSE-Based Forecasts.**
- Table IA12 provides a **comparison of quantile-based estimators** of the conditional expectation.
- Table IA13 presents the performance of the **generalized robust risk premium with different quantile specifications**.
- Table IA14 provides an analysis of **value-weighted** RRPV-managed portfolios.
- Table IA15 provides an analysis of **equal-weighted** RRPV-managed portfolios.
- Figure IA1 shows a **time series plot of the cross-sectional mean and dispersion of MQ1 quantile forecasts**.
- Figure IA2 shows **variable importance plots** by model for each quantile.
- Figure IA3 displays **variable importance rankings** by model for each quantile.

IA1 Off-the-shelf Quantile Machine Learning Models

This section introduces several methods for estimating the function g_τ in Equation (1). Since machine learning models using the MSE objective function have been discussed extensively in e.g., Gu, Kelly, and Xiu (2020), we focus on the application of these machine learning methods using the quantile objective function. In this paper we consider specifically 1) linear models, 2) nonlinear models.

IA1.1 Linear Models

IA1.1.1 Linear Quantile Regression

The linear quantile regression (Koenker and Bassett Jr (1978)) imposes that the conditional quantile $g_\tau(\cdot)$ can be approximated by a linear function

$$g_\tau(z_{i,t};\theta) = z'_{i,t}\theta. \tag{15}$$

In the ordinary least squares case, the conditional expectation is often estimated by minimizing the mean squared error. Analogously, the conditional quantile can be estimated by minimizing the quantile objective function

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau\left(r_{i,t+1} - z'_{i,t}\theta\right), \tag{16}$$

where $\rho(u) = u(\tau - \mathbf{1}\{u \leq 0\})$ is the quantile loss function.

IA1.1.2 Penalized Linear Quantile Regression

The linear quantile regression can overfit in the presence of a large number of predictors. We mitigate overfitting by imposing a penalized objective function

$$\mathcal{L}(\theta;\cdot) = \mathcal{L}(\theta) + \phi(\theta;\cdot), \tag{17}$$

where we append to the original objective function in (16) a penalty function $\phi(\theta;\cdot)$. We set $\phi(\theta;\cdot)$ to be the popular elastic net (Zou and Hastie (2005)) penalty defined as

$$\phi(\theta;\lambda,\alpha) = \lambda(1-\alpha)\sum_{j=1}^{P}|\theta_j| + \frac{1}{2}\lambda\alpha\sum_{j=1}^{P}\theta_j^2, \tag{18}$$

which involves two non-negative hyperparameters λ and α. Tuning parameter λ controls the amount of shrinkage, where larger values of λ correspond to more shrinkage. When $\alpha = 0$, (18) corresponds to the lasso penalty. The lasso's absolute value penalty imposes sparsity since a subset of the coefficients can be set to exactly 0, hence lasso can be used for variable selection. When $\alpha = 1$, (18) corresponds to the ridge penalty which shrinks coefficient estimates toward zero but does not impose exact zeros anywhere. When $0 < \alpha < 1$, we have the elastic net penalty that combines the sparsity of the lasso penalty with the shrinkage of the ridge penalty. One advantage of the elastic net penalty is an improved ability to handle highly correlated predictors (Zou and Hastie (2005)), which is a property that is inherited from the ridge penalty. Tuning parameter λ is chosen using hyperparameter tuning, which is described in further detail in Section IA3.

IA1.2 Nonlinear Models

Linear quantile regression is robust against heavy tails that are attributed to financial returns. However, it does not allow for nonlinearities and interaction effects between predictors. We next consider nonlinear tree-based and neural network methods.

IA1.2.1 Gradient Boosted Trees

Gradient boosted trees (GBT) is an ensemble method that sequentially combines many decision trees into a single predictive model. Unlike single decision trees that are likely to overfit, GBT learns slowly from a sequence of shallow decision trees. The general GBT algorithm for an arbitrary differentiable loss function (similar to Friedman (2001)) is given

by Algorithm 1.

Algorithm 1 Gradient Boosted Regression Trees

1. Initialize $g_0(\cdot) = argmin_\gamma \sum_{i=1}^{N} \mathcal{L}(r_{i,t+1}, \gamma)$
2. **For** *$m = 1$ to M:*
   (a) **For** *$i = 1, 2, ..., N$ and $t = 1, ..., T$ compute:*
   $$\epsilon_{i,t+1} = -\left[\frac{\partial \mathcal{L}(r_{i,t+1}, g)}{\partial g}\right]_{g=g_{m-1}(z_{i,t})}$$
   (b) Fit a regression tree to the targets $\epsilon_{i,t+1}$ giving terminal regions $R_{jm}, j = 1, 2, ..., J_m$.
   (c) **For** *$j = 1, 2, ..., J_m$ compute*
   $$\gamma_{jm} = argmin_\gamma \sum_{z_{i,t} \in R_{jm}} \mathcal{L}(r_{i,t+1}, g_{m-1}(z_{i,t}) + \gamma).$$
   (d) Update $g_m(\cdot) = g_{m-1}(\cdot) + \lambda \sum_{j=1}^{J_m} \gamma_{jm} \mathbf{1}(\cdot \in R_{jm})$.
3. Output $\hat{g}(z_{i,t}) = \sum_{m=1}^{M} \lambda g_m(\cdot)$.

For GBT, the ensemble prediction is a sum of individual decision tree predictions that are shrunken by a shrinkage parameter $\lambda \in (0, 1)$ that prevents the model from overfitting. At each new step m, a shallow tree is fitted to the residuals from the model with $m-1$ trees, and its residual forecast is added to the total with a weight of λ. This is iterated until there are a total of M trees in the ensemble.

When the loss function L is MSE, then the negative gradient is the residual $r_{i,t+1} - g_{m-1}(z_{i,t})$ and the terminal regions predict the mean of the residuals in the corresponding region. In contrast, when using quantile gradient boosting given τ, the negative gradient calculated using the quantile loss in (2) is

$$-\left[\frac{\partial \mathcal{L}(r_{i,t+1}, g)}{\partial g}\right]_{g=g_{m-1}(z_{i,t})} = \tau \mathbf{1}(r_{i,t+1} \geq g_{m-1}(z_{i,t})) + (\tau - 1)\mathbf{1}(r_{i,t+1} < g_{m-1}(z_{i,t})), \quad (19)$$

and the terminal regions predict the τ quantile of the residuals in the corresponding region. For example, when $\tau = 0.5$, then the negative gradient is the sign of the residual, $\tau sign[r_{i,t+1} - g_{m-1}(z_{i,t})]$, and the terminal regions predict the median of the residuals (Friedman (2001)).

The main GBT hyperparameters are the shrinkage parameter λ, the depth of each in-

dividual tree, and the number of trees. To further mitigate overfitting, we implement a stochastic variation of gradient boosting which samples a fraction of the training observations without replacement. We also incorporate early stopping that stops adding trees when the validation set's quantile error starts to increase. Section IA3 provides further details on the hyperparameters.

IA1.2.2 Quantile Random Forest

Quantile random forest, an extension of the random forest algorithm, focuses on estimating conditional quantiles, not just conditional expectations. In a manner analogous to random forest's aggregation of conditional expectation predictions from multiple decision trees, quantile random forest combines conditional distribution functions from numerous decision trees to predict the conditional distribution function, $P_t(r_{i,t+1} \leq y)$. Quantile random forest grows decision trees like random forest, splitting on a random subset of predictors using the mean squared error criterion. However, unlike random forest, which stores only the mean of excess returns in the leaf node, quantile random forest retains all excess returns. This allows for the calculation of the conditional distribution function of excess returns

$$P_t(r_{i,t+1} \leq y) = E_t[\mathbf{1}(r_{i,t+1} \leq y)], \tag{20}$$

from which quantile predictions can be obtained for every given τ as $g_\tau(z_{i,t}) = inf\{y : P_t(r_{i,t+1} \leq y) \geq \tau\}$.

Formally, we adopt the notation from Meinshausen (2006) and call θ_m the random parameter vector that determines how tree $m = 1, ..., M$ is grown. Define leaf $\ell = 1, ..., L$ corresponding to a terminal region R_ℓ. For any firm in the test set with predictor z, there is one unique leaf ℓ such that $z \in R_\ell$, denoted $\ell(z, \theta_m)$. Let the weight vector $w_{i,t}(z, \theta_m) = \frac{\mathbf{1}(z_{i,t} \in R_{\ell(z,\theta_m)})}{\#\{(i',t'): z_{(i',t')} \in R_{\ell(z,\theta_m)}\}}$ if observation $z_{i,t}$ is part of leaf $\ell(z, \theta_m)$ and 0 if it is not. For a collection of M decision trees, let $w_{i,t}(z)$ be the average of $w_{i,t}(z, \theta_m)$,

$$w_{i,t}(z) = \frac{1}{M}\sum_{m=1}^{M} w_{i,t}(z, \theta_m). \tag{21}$$

The prediction of random forest is then $\hat{\mu}(z) = \sum_{(i,t)\in\mathcal{T}_1} w_{i,t}(z) r_{i,t+1}$, which approximates the conditional expectation of excess returns. By (20), the conditional distribution of stock returns is estimated by

$$\hat{P}_t(r_{i,t+1} \leq y) = \sum_{(i,t)\in\mathcal{T}_1} w_{i,t}(z)\mathbf{1}(r_{i,t+1} \leq y), \tag{22}$$

using the same weights as random forest defined in equation (21). Estimates of a stock return's conditional quantile given τ are then

$$\hat{g}_\tau(z) = inf\{y : \hat{P}_t(r_{i,t+1} \leq y) \geq \tau\}. \tag{23}$$

The main hyperparameter in quantile random forest is its maximum depth, which controls the complexity of each decision tree.

IA1.2.3 Feed-Forward Neural Networks

Neural networks are powerful machine learning models that have been theoretical demonstrated to be "universal approximators" for any smooth predictive association (Gallant and White (1992), Hornik, Stinchcombe, and White (1989)). In this paper, we use traditional feed-forward neural networks (NN) to approximate $g_\tau(z_{i,t}; \theta)$, which consists of an input layer of predictors, hidden layers that interact and nonlinearly transform predictors, and an output layer that aggregates the hidden layers into a final prediction. Most neural network regression models are estimated by minimizing the MSE of prediction errors, which results in a model that predicts the conditional expectation. In contrast, we estimate the neural network model by minimizing the quantile objective function (2), which results in a model that predicts the conditional quantile. The non-differentiability of (2) at exactly zero is not an issue in practice, because it occurs with low probability and can be handled by setting the

gradient to zero. We utilize stochastic gradient descent (SGD) to train the neural network, where at each iteration of the optimization, SGD evaluates the gradient on a small random subset of the data, significantly increasing the speed of the optimization.

The hidden units in neural networks are important for improving the flexibility of the model. Each hidden unit has an activation function that takes the outputs from the previous layer and nonlinearly transforms its aggregated signal before sending it to the next layer. In this paper, we use the popular Rectified linear unit (ReLU) activation function defined as $\phi(x) = \max\{0, x\}$ (Nair and Hinton (2010)). Following Gu, Kelly, and Xiu (2020), we use architectures with up to four hidden layers, where the shallowest network has a single layer with 32 units and the deepest has four hidden layers with 32, 16, 8, and 4 units, respectively. All architectures are fully connected so each unit receives an input from all layers below.

Due to their complexity, neural networks are subject to overfitting. We address this issue in several ways. First, we add a penalty term to the objective function similar to (17), which shrinks coefficients toward zero using the elastic net penalty (18). Second, we utilize early stopping, which stops the optimization when the validation set's quantile error begins to increase, allowing the model to generalize better and reducing the computational cost. Third, we adopt an ensemble approach from Gu, Kelly, and Xiu (2020) by averaging across ten neural network configurations to reduce the variance of predictions. Neural networks use hyperparameter tuning to find the appropriate level of regularization to mitigate overfitting. Section IA3 provides further details on these hyperparameters.

IA2 Theorem of Robust Risk Premium Convergence to Expected Value

Theorem 1. *Let $F_t(x) = P(r_{i,t+1} \leq x|\mathcal{F}_t)$ be the cumulative distribution function of stock excess return $r_{i,t+1}$ and conditional on information set $\mathcal{F}_t$.*

For $0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K < 1$, let

$$\begin{aligned}
&\bar{Q}_K = (\tau_1 + \frac{\tau_2 - \tau_1}{2})Q_{\tau_1} + (\frac{\tau_2 - \tau_1}{2} + \frac{\tau_3 - \tau_2}{2})Q_{\tau_2} + \cdots + \\
&(\frac{\tau_{K-1} - \tau_{K-2}}{2} + \frac{\tau_K - \tau_{K-1}}{2})Q_{\tau_{K-1}} + (\frac{\tau_K - \tau_{K-1}}{2} + (1 - \tau_K))Q_{\tau_K} = \\
&= (\frac{\tau_2 + \tau_1}{2})Q_{\tau_1} + (\frac{\tau_3 - \tau_1}{2})Q_{\tau_2} + \cdots + (\frac{\tau_K - \tau_{K-2}}{2})Q_{\tau_{K-1}} + (1 - \frac{\tau_K + \tau_{K-1}}{2})Q_{\tau_K}
\end{aligned} \tag{24}$$

where $Q_{\tau_k} \equiv Q_{r_{i,t+1}}(\tau_k|\mathcal{F}_t)$ are quantiles of $r_{i,t+1}$, $k \in \{1, ..., K\}$, and let $\Delta_1 = max\{\tau_2 + \tau_1, 1 - (\tau_K + \tau_{K-1})\}$ and $\Delta_2 = \max\limits_{2 \leq k \leq K-1} \{Q_{\frac{\tau_{k+1}+\tau_k}{2}} - Q_{\frac{\tau_k+\tau_{k-1}}{2}}\}$.

Assume that $F_t(x)$ is continuous and strictly monotonically increasing, $E_t(r_{i,t+1})$ exists, and

$$E_t(r_{i,t+1}) = \int_{\mathbb{R}} x d\mu_{F_t}, \tag{25}$$

where $\int_{\mathbb{R}} x d\mu_{F_t}$ is a Lebesgue–Stieltjes integral, μ_{F_t} is the Lebesgue–Stieltjes measure associated with F_t, and $\mathbb{R}$ represents the set of real numbers, then

$$\lim_{\substack{\Delta_1 \to 0 \\ \Delta_2 \to 0}} \bar{Q}_K = E_t(r_{i,t+1}). \tag{26}$$

Proof. Without loss of generality, fix firm i and time t. Since $\int_{\mathbb{R}} x d\mu_{F_t}$ exists, then Riemann-Stieltjes integral $\int_{\mathbb{R}} x dF_t(x)$ exists and $E_t(r_{i,t+1}) = \int_{\mathbb{R}} x dF_t(x)$.

Step 1: First assume that $r_{i,t+1}$ is bounded on interval $[a, b]$, where a and b are real numbers, and $a < b$, then

$$E_t(r_{i,t+1}) = \int_a^b x dF_t(x). \tag{27}$$

Let us partition $[a, b]$ into disjoint subintervals; $a = x_0 < x_1 < \cdots < x_{K-1} < x_K = b$; and let $\delta_K = \max_{1 \le k \le K}(x_k - x_{k-1})$, and $RS_t(K)$ be the Riemann-Stieltjes sum defined as

$$RS_t(K) = \sum_{k=1}^{K} c_k(F_t(x_k) - F_t(x_{k-1})), \tag{28}$$

where $c_k \in (x_{k-1}, x_k]$.

Let ϵ be any positive number. Since x is continuous and $F_t(x)$ is increasing, then

$$\lim_{\delta_K \to 0} RS_t(K) = \int_a^b x dF_t(x) = E_t(r_{i,t+1}), \tag{29}$$

and there is a $\delta_K > 0$ such that,

$$|RS_t(K) - E_t(r_{i,t+1})| < \epsilon. \tag{30}$$

Let $x_0 = a, x_K = b, x_k = Q_{\frac{\tau_{k+1}+\tau_k}{2}}$ for $k \in \{1, ..., K-1\}$; and let $c_k = Q_{\tau_k}$ so that $c_k \in (x_{k-1}, x_k]$ for $k \in \{1, ..., K\}$.

Select $0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K < 1$, such that $\Delta_1 = max\{\tau_2 + \tau_1, 1 - (\tau_K + \tau_{K-1})\}$ is sufficiently small, $\max\{(Q_{\frac{\tau_2+\tau_1}{2}} - a), (b - Q_{\frac{\tau_K+\tau_{K-1}}{2}})\} < \delta_K$, and $\Delta_2 = \max_{2 \le k \le K-1}\{Q_{\frac{\tau_{k+1}+\tau_k}{2}} - Q_{\frac{\tau_k+\tau_{k-1}}{2}}\} < \delta_K$, then

$$
\begin{aligned}
RS_t(K) &= \sum_{k=1}^{K} c_k(F_t(x_k) - F_t(x_{k-1})) \\
&= (F_t(Q_{(\frac{\tau_2+\tau_1}{2})}) - F_t(a))Q_{\tau_1} + (F_t(Q_{(\frac{\tau_3+\tau_2}{2})}) - F_t(Q_{(\frac{\tau_2+\tau_1}{2})}))Q_{\tau_2} + \cdots \\
&+ (F_t(Q_{(\frac{\tau_K+\tau_{K-1}}{2})}) - F_t(Q_{(\frac{\tau_{K-1}+\tau_{K-2}}{2})}))Q_{\tau_{K-1}} + (F_t(b) - F_t(Q_{(\frac{\tau_K+\tau_{K-1}}{2})})Q_{\tau_K} \\
&= (\frac{\tau_2+\tau_1}{2} - F_t(a))Q_{\tau_1} + (\frac{\tau_3+\tau_2}{2} - \frac{\tau_2+\tau_1}{2})Q_{\tau_2} + \cdots \\
&+ (\frac{\tau_K+\tau_{K-1}}{2} - \frac{\tau_{K-1}+\tau_{K-2}}{2})Q_{\tau_{K-1}} + (F_t(b) - \frac{\tau_K+\tau_{K-1}}{2})Q_{\tau_K} \\
&= (\frac{\tau_2+\tau_1}{2} - F_t(a))Q_{\tau_1} + (\frac{\tau_3-\tau_1}{2})Q_{\tau_2} + \cdots + (\frac{\tau_K-\tau_{K-2}}{2})Q_{\tau_{K-1}} + (F_t(b) - \frac{\tau_K+\tau_{K-1}}{2})Q_{\tau_K} \\
&= \bar{Q}_K - F_t(a)Q_{\tau_1} - (1 - F_t(b))Q_{\tau_K},
\end{aligned}
\tag{31}
$$

where the third equality is guaranteed because $F_t(x)$ is continuous and strictly monotonically increasing.

Since $F_t(a) = 0$ and $F_t(b) = 1$, then $RS_t(K) = \bar{Q}_K$. Therefore

$$
|\bar{Q}_K - E_t(r_{i,t+1})| = |RS_t(K) - E_t(r_{i,t+1})| < \epsilon, \tag{32}
$$

and $\bar{Q}_K$ converges to $E_t(r_{i,t+1})$.

Step 2. Next assume $r_{i,t+1}$ is not bounded, and without loss of generality assume $r_{i,t+1} \in (-\infty, \infty)$.[16]

Let ϵ be any positive number. Since $\int_{\mathbb{R}} x dF_t(x)$ exists, then there is a sufficiently large interval $[c, d]$, such that for any $a < c < 0$ and any $b > d > 0$,

[16]The proof of one side unbounded is similar.

$$
\begin{aligned}
&|\int_a^b xdF_t(x) - E_t(r_{i,t+1})| < \frac{\epsilon}{4}, \\
&|\int_{-\infty}^a xdF_t(x)| < \frac{\epsilon}{4}, \\
&|\int_b^{\infty} xdF_t(x)| < \frac{\epsilon}{4}.
\end{aligned} \tag{33}
$$

Also, we can select $0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K < 1$ such that Δ_1 and Δ_2 are sufficiently small, $a < Q_{\tau_1} < 0$, $0 < Q_{\tau_K} < b$, and

$$
|RS_t(K) - \int_a^b xdF_t(x)| < \frac{\epsilon}{4}, \tag{34}
$$

where $RS_t(K)$, Δ_1, and Δ_2 are defined the same as in Step 1.

Since $a < Q_{\tau_1} < 0$, then $\int_{-\infty}^a xdF_t(x) \leq \int_{-\infty}^a Q_{\tau_1} dF_t(x) = Q_{\tau_1} F_t(a) < 0$, therefore, $|Q_{\tau_1} F_t(a)| \leq |\int_{-\infty}^a xdF_t(x)| < \frac{\epsilon}{4}$.

Also, since $0 < Q_{\tau_K} < b$, then $\int_b^{\infty} xdF_t(x) \geq \int_b^{\infty} Q_{\tau_K} dF_t(x) = Q_{\tau_K}(1 - F_t(b)) \geq 0$, therefore, $|Q_{\tau_K}(1 - F_t(b))| \leq |\int_b^{\infty} xdF_t(x)| < \frac{\epsilon}{4}$. Hence,

$$
|\bar{Q}_K - RS_t(K)| = |-Q_{\tau_1} F_t(a) - Q_{\tau_K}(1 - F_t(b))| \leq |Q_{\tau_1} F_t(a)| + |Q_{\tau_K}(1 - F_t(b))| < \frac{\epsilon}{4} + \frac{\epsilon}{4} = \frac{\epsilon}{2}.
$$

Therefore,

$$
|\bar{Q}_K - E_t(r_{i,t+1})| \leq |\bar{Q}_K - RS_t(K)| + |RS_t(K) - \int_a^b xdF_t(x)| + |\int_a^b xdF_t(x) - E_t(r_{i,t+1})| < \epsilon,
$$

and $\bar{Q}_K$ converges to $E_t(r_{i,t+1})$. □

IA3 Hyperparameters and Computational Framework

IA3.1 Quantile Models

This section discusses in more detail the hyperparameters that are tuned for each of the quantile models used in this paper using the validation set approach discussed in Section 4.

For the EN, tuning parameter λ is the main hyperparameter that determines the level of shrinkage. For GBT, the main hyperparameters include the number of trees M, the maximum depth of each tree, and the shrinkage parameter λ, and regularization parameter ℓ^2. The shrinkage parameter controls the rate at which boosting learns, where a smaller rate typically leads to better performance but requires more trees in the ensemble. The depth controls the complexity of the boosted ensemble, where a greater depth allows for more interactions among variables. Regularization parameter ℓ^2 penalizes terminal region outputs, effectively shrinking them towards zero. For RF, the main hyperparameter is the depth of each tree.

NNs require choosing many hyperparameters to find the best model flexibility. We mitigate overfitting by adding ℓ^1 and ℓ^2 regularization to the loss function, using batch normalization between layers to stabilize activation variances, and employing early stopping with optimized epoch counts. We further reduce prediction variance by averaging forecasts across ten independently trained networks, following Gu, Kelly, and Xiu (2020). All neural network models are estimated using the Adaptive Moment Estimation (ADAM) optimizer from Kingma and Ba (2017).

For MQ and SMQ, we use ReLU activations in all hidden layers and estimate weights using ADAM. The regularization design mirrors that of NNs, incorporating ℓ^1 and ℓ^2 penalties, batch normalization, early stopping, and ensemble averaging across ten independently initialized models. For SMQ, we additionally apply gradient clipping, which prevents exploding gradients by limiting excessively large updates during training.

Table IA1 summarizes the set of hyperparameters that are tuned for each of the quantile

Table IA1: Hyperparameters for Quantile Models

LR	Learning rate $\in \{0.001, 0.01\}$ Epochs = $\{100\}$ Patience = 5
EN	$\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}\}$ $\alpha = 0.5$ Learning rate $\in \{0.001, 0.01\}$ Epochs = $\{100\}$ Patience = 5
GBT	#Trees = 1000 Tree depth 1∼6 Learning rate $\in \{0.01, 0.1\}$ $\ell^2 \in \{10^2, 10^3, 10^4, 10^5\}$ Sample rate = 0.5 Patience = 5
RF	#Trees = 300 #Features in each split = floor(#Predictors/3) = 34 Tree depth = 20∼50
NN1-NN4 and MQ1-MQ2	Activation Functions = Relu Epochs = $\{200\}$ $\ell^1 \in \{10^{-6}, 10^{-5}, 10^{-4}\}$ $\ell^2 \in \{10^{-6}, 10^{-5}, 10^{-4}\}$ Learning rate $\in \{0.001, 0.01\}$ Patience = 5
SMQ1-SMQ2	Same hyperparameters as NN and MQ models Gradient clipping = 30

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

models considered in this paper. To tune these at a particular time, the sample of data is partitioned into training, testing, and validation sets described in Section 4. The set of hyperparameters with the lowest quantile loss in the validation set is selected as the best model through grid search.

LR and EN are estimated using TensorFlow by a neural network with 0 hidden layers and only an output layer; we find that this approach outperforms popular packages for estimating high-dimensional quantile regressions such as a semismooth Newton coordinate descent algorithm used in the previous version of the paper. NNs are also estimated using Keras Sequential API and Tensorflow. MQ and SMQ are implemented using the Keras Functional API and TensorFlow. RF is estimated using a fast implementation of random forests for high-dimensional data implemented in the ranger package (Wright and Ziegler

(2015)). GBT is estimated using the LightGBM package.

IA3.2 MSE-Based Models

The MSE-based models are estimated following the same procedures as the quantile models to facilitate a fair comparison. GKX estimate several models using the Huber loss, a hybrid objective function that combines MSE with MAE to downweight outliers; for brevity, we refer to both MSE and Huber specifications as MSE-based models. The statistics literature shows that the Huber loss function performs well under contaminated normal distributions, whereas the quantile loss function is more suitable for heavy-tailed distributions, and is most efficient under double-exponential distributions (Bradic, Fan, and Wang (2011)). Because each loss function performs best for a particular class of distributions, it remains an empirical question which one is most appropriate for the excess stock return distribution. In this paper, all MSE-based models (except for RF) are estimated using the Huber loss, with the threshold set at the 99.9th percentile of the training sample. The RF models are estimated using the MSE loss in both the MSE-based and quantile-based approaches.

Table IA1 summarizes the hyperparameters tuned for each MSE-based model. These closely mirror those used for the quantile-based models, with deviations introduced only to improve the performance of the MSE-based specifications. Hyperparameters are tuned using the validation set approach described in Section 4. The combination of hyperparameters that yield the lowest MSE loss in the validation set is selected as the optimal configuration through grid search.

Regularization techniques are also consistent across the quantile and MSE-based models. For example, feed-forward neural networks for the MSE-based models also employ ℓ^1 and ℓ^2 regularization, early stopping, batch normalization, and ensemble averaging across ten independently trained configurations.

For the MSE-based models, LR and EN are estimated using the SGDRegressor package, which implements stochastic gradient descent similar to TensorFlow's optimization routine.

Table IA2: Hyperparameters for MSE-Based Models

LR	Loss = Huber
EN	$\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	$\alpha = 0.5$
	Loss = Huber
GBT	#Trees = 1000
	Tree depth 1∼6
	Learning rate $\in \{0.001, 0.01\}$
	$\ell^2 \in \{10^5, 10^6, 10^7, 10^8\}$
	Sample rate = 0.5
	Loss = Huber
	Patience = 5
RF	#Trees = 300
	#Features in each split = floor($\sqrt{\#Predictors}$) = 10
	Tree depth = 1∼6
NN1-NN4	Activation Functions = Relu
	Epochs = {200}
	Loss = Huber
	$\ell^1 \in \{10^{-6}, 10^{-5}, 10^{-4}\}$
	$\ell^2 \in \{10^{-6}, 10^{-5}, 10^{-4}\}$
	Learning rate $\in \{0.001, 0.01\}$
	Epochs = {200}
	Patience = 5

This table describes the set of hyperparameters that are tuned for each individual MSE-based model considered in the paper.

NNs are estimated using Keras Sequential API and TensorFlow. RF is estimated using the RandomForestRegressor package from scikit-learn, and GBT model is estimated using the LightGBM package.

Table IA3 reports the R^2_{os} values and portfolio sort results for the MSE-based models. Due to computational constraints, we exclude industry dummies and omit interactions between macroeconomic variables and firm characteristics that would expand the predictor set to 920, as in GKX, for both the quantile and MSE-based models. Despite these simplifications, the results in Table IA3 closely match those in GKX, confirming that the replication is successful for the extended 1987–2020 sample. Notably, the replicated NN1 and NN3 models achieve higher R^2_{os} values than in GKX because our MSE-based NNs are estimated using the Huber loss, whereas GKX employ the MSE loss for NNs.

Table IA3: Performance of the MSE-Based Approach Replication

This table presents the results for the MSE-based models. Panel A reports the out-of-sample R^2_{os} values (in percentages) of MSE-based model forecasts for the entire panel of individual stock excess returns. Panel B reports the monthly performance of value-weighted decile portfolios, sorted on MSE-based model forecasts. The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios.

Panel A: Percentage R^2_{OS} of $\bar{\mu}^{MSE}$								
	LR	EN	RF	GBT	NN1	NN2	NN3	NN4
	-0.23	0.33	0.22	0.15	0.5	0.37	0.56	0.39
Performance of Value-Weighted Portfolios sorted on $\bar{\mu}^{MSE}$								
	LR	EN	RF	GBT	NN1	NN2	NN3	NN4
Low (L)	0.36	0.19	0.32	-0.09	-0.80	-0.36	-0.57	-0.36
2	0.51	0.56	0.55	0.35	-0.09	0.06	-0.19	0.11
3	0.71	0.61	0.56	0.64	0.23	0.39	0.29	0.30
4	0.78	0.70	0.72	0.61	0.52	0.46	0.49	0.45
5	0.87	0.70	0.74	0.72	0.70	0.62	0.76	0.67
6	0.82	0.82	0.74	0.71	0.74	0.81	0.83	0.80
7	0.89	0.83	0.83	0.77	0.91	0.92	0.96	0.99
8	0.90	0.84	0.93	0.78	0.97	1.02	0.95	1.06
9	0.99	0.91	0.96	0.84	1.21	1.23	1.13	1.12
High (H)	1.22	1.15	1.27	1.05	1.69	1.62	1.66	1.46
H-L	0.86	0.96	0.95	1.14	2.49	1.99	2.23	1.82
t	4.30	3.56	2.96	4.52	7.42	5.91	6.25	5.52
SR	0.82	0.72	0.52	0.76	1.35	1.08	1.22	0.97

IA4 Related Literature on Quantile-Based Estimation of the Conditional Expectation

The literature offers three main approaches of combining quantiles to estimate the conditional expectation. The first is to compute a weighted average of coefficient estimates in linear quantile regression (e.g., Jiang, Maasoumi, and Xiao (2020); Zhao and Xiao (2014)), however, this approach is generally unsuitable for machine learning applications, where many models, including RF and GBT, are nonparametric. The second averages quantile estimates using least-squares weights (e.g., Lima and Meng (2017); Meligkotsidou, Panopoulou, Vrontos, and Vrontos (2014)), but this introduces an additional layer of estimation and potential model error, so we do not pursue it further, leaving it for possible future exploration.

The third approach applies fixed weights and can be broadly categorized into estimators that use three or five quantiles. Among three-quantile estimators, a classic example from the statistics literature is Tukey et al. (1977)'s trimean, which assigns weights $\{0.25, 0.50, 0.25\}$, yielding $\bar{Q}_{TM} = 0.25Q_{0.25} + 0.5Q_{0.50} + 0.25Q_{0.75}$. Another estimator from the statistics literature (see Wan, Wang, Liu, and Tong (2014) for a comprehensive survey) uses equal weights, $\{1/3, 1/3, 1/3\}$, producing $\bar{Q}_{WWLT} = (Q_{0.25} + Q_{0.50} + Q_{0.75})/3$. However, these estimators developed under strong distributional assumptions such as normality are often inappropriate for the heavy-tailed distributions typical of stock returns. In contrast, Equation (6)'s convergence to the conditional expectation relies on much weaker distributional assumptions, making it more robust to model misspecification. Finally, Ma and Pohlman (2008) propose a probability-based method that assigns weights proportional to the probability mass of each quantile interval, allocating most of the weight to the median. Specifically, by partitioning the probability space into $(0, 0.05]$, $(0.05, 0.95)$, and $[0.95, 1)$, the implied weights are $\{0.05, 0.90, 0.05\}$, yielding $\bar{Q}_{MP} = 0.05Q_{0.05} + 0.9Q_{0.50} + 0.05Q_{0.95}$.

Among five-quantile estimators, a classic example from the statistics literature is provided by Judge, Griffiths, Hill, Lütkepohl, and Lee (1991)'s five-quantile estimator. This estimator

assigns weights $\{0.05, 0.25, 0.4, 0.25, 0.05\}$, yielding $\bar{Q}_{FQ} = 0.05Q_{0.05} + 0.25Q_{0.25} + 0.4Q_{0.50} + 0.25Q_{0.75} + 0.05Q_{0.95}$. Closest to this paper is the data discretization approach in Yang, Zhu, Li, and Zhu (2024), which uses the minimized 1-Wasserstein metric (Dabney, Ostrovski, Silver, and Munos (2018)) to determine weights. For our chosen quantiles, their method implies weights $\{0.1, 0.3, 0.2, 0.3, 0.1\}$, yielding $\bar{Q}_{YZLZ} = 0.1Q_{0.05} + 0.3Q_{0.25} + 0.2Q_{0.50} + 0.3Q_{0.75} + 0.1Q_{0.95}$. As a simple numerical comparison, consider the Skewed Student's t distribution (Hansen (1994)), depicted in Figure 3, with 15 degrees of freedom, 0.2 skewness parameter, 1 variance, and 0 expected value. Using the theoretical quantiles, Yang, Zhu, Li, and Zhu (2024)'s weights would produce -0.012 (calculated by $0.1 \times (-1.51) + 0.3 \times (-0.68) + 0.2 \times (-0.07) + 0.3 \times 0.61 + 0.1 \times 1.74$) as the estimated expected value, while our weights would yield 0.00125 (calculated by $0.15 \times (-1.51) + 0.225 \times (-0.68) + 0.25 \times (-0.07) + 0.225 \times 0.61 + 0.15 \times 1.74$), which is a better estimate of the expected value of the example distribution. This is likely because their approach assigns too little weight to the median for certain probability distributions, in particular for distributions that have a high concentration around the median such as stock return distributions.

A subset of these estimators have been employed by Ma and Pohlman (2008), Meligkotsidou, Panopoulou, Vrontos, and Vrontos (2014), Lima and Meng (2017), and Yang, Zhu, Li, and Zhu (2024). To our best knowledge, the weights used to construct the robust risk premium $\bar{Q}$ in Equation (5) are new to this literature.

Table IA4: Comparisons of Out-of-Sample Quantile Prediction Performance using Diebold-Mariano Tests

This table reports the pairwise Diebold-Mariano test statistics, comparing the out-of-sample quantile prediction performance among the machine learning models for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$. We compare 12 models in total including linear quantile regression with all covariates (LR), elastic net quantile regression (EN), quantile random forest (RF), quantile gradient boosted regression trees (GBT), quantile feed-forward neural networks with 1-4 layers (NN1,...,NN4), multi-quantile neural networks with 1-2 shared layers (MQ1, MQ2), and spaced multi-quantile neural networks with 1-2 shared layers (SMQ1, SMQ2). A positive number indicates that the column model outperforms the row model, and bold font signifies that the difference is statistically significant at the 5% level.

Panel A: Diebold-Mariano Tests for the $\tau = 0.05$

	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
LR	1.08	**-2.79**	1.17	**3.24**	**3.75**	**3.71**	**2.86**	**4.32**	**4.07**	**4.10**	**4.11**
EN		-7.65	0.54	**8.90**	**9.45**	**9.62**	**6.30**	**10.61**	**9.66**	**10.35**	**9.91**
RF			**4.31**	**8.81**	**9.31**	**9.35**	**8.49**	**10.57**	**10.00**	**9.99**	**9.97**
GBT				**2.07**	**2.47**	**2.57**	1.30	**2.91**	**2.78**	**2.86**	**2.77**
NN1					**2.11**	**2.59**	**-3.13**	**4.15**	**3.31**	**5.21**	**4.14**
NN2						0.45	**-5.28**	**3.14**	**2.45**	**2.91**	**3.20**
NN3							**-5.52**	**2.76**	**2.22**	**2.83**	**2.88**
NN4								**7.37**	**5.39**	**6.82**	**6.32**
MQ1									-0.83	-0.24	-0.09
MQ2										0.83	1.29
SMQ1											0.11

Panel B: Diebold-Mariano Tests for the $\tau = 0.25$

	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
LR	1.38	**-2.33**	0.71	**3.67**	**4.12**	**4.21**	**2.10**	**4.60**	**4.37**	**4.21**	**4.17**
EN		**-5.85**	-1.32	**7.16**	**8.27**	**8.51**	1.71	**8.60**	**7.18**	**7.55**	**6.78**
RF			**4.41**	**7.34**	**7.38**	**7.59**	**5.60**	**8.10**	**7.77**	**7.74**	**7.49**
GBT				**5.00**	**5.78**	**5.83**	**2.29**	**5.79**	**5.50**	**5.36**	**5.05**
NN1					1.95	1.81	**-3.76**	**3.57**	1.77	**2.70**	**1.97**
NN2						-0.12	**-5.29**	1.36	0.51	0.65	0.58
NN3							**-5.47**	1.46	0.59	0.72	0.65
NN4								**4.87**	**4.61**	**4.32**	**4.17**
MQ1									-0.70	-1.02	-0.57
MQ2										0.12	0.24
SMQ1											0.08

Panel C: Diebold-Mariano Tests for the $\tau = 0.5$											
	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
LR	1.61	-0.37	**-2.52**	**2.69**	**2.26**	**2.49**	1.66	**3.19**	**2.98**	**2.83**	**3.06**
EN		**-2.04**	**-3.18**	**3.23**	**2.35**	**3.48**	0.53	**3.13**	**2.35**	**2.97**	**3.04**
RF			**-2.16**	**3.37**	**2.56**	**2.82**	**2.12**	**3.61**	**3.09**	**2.94**	**3.26**
GBT				**3.86**	**3.53**	**3.61**	**3.22**	**4.14**	**4.10**	**3.79**	**4.01**
NN1					**-2.08**	-1.22	**-2.97**	0.64	0.04	-0.67	0.23
NN2						1.05	-1.95	**2.11**	1.35	1.53	**2.08**
NN3							**-3.86**	1.39	0.80	0.59	1.21
NN4								**2.47**	1.84	**2.21**	**2.42**
MQ1									-0.73	-1.77	-0.65
MQ2										-0.85	0.31
SMQ1											1.90
Panel D: Diebold-Mariano Tests for the $\tau = 0.75$											
	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
LR	1.58	**-2.00**	0.99	**2.97**	**2.59**	**2.75**	1.93	**3.34**	**3.31**	**3.06**	**3.68**
EN		**-3.31**	-0.15	**3.47**	**2.73**	**2.93**	0.79	**3.72**	**3.49**	**3.28**	**4.57**
RF			**3.44**	**5.41**	**4.75**	**4.90**	**3.44**	**5.72**	**5.18**	**5.53**	**6.40**
GBT				**2.76**	**2.10**	**2.37**	0.50	**3.06**	**2.75**	**2.90**	**3.71**
NN1					**-2.56**	-0.76	**-2.26**	0.65	0.04	0.17	1.53
NN2						**2.02**	-1.69	**2.33**	1.54	1.65	**2.49**
NN3							-1.95	1.19	0.60	0.74	1.69
NN4								**2.51**	**2.34**	**2.24**	**3.21**
MQ1									-0.68	-0.78	1.25
MQ2										0.15	1.43
SMQ1											1.95
Panel E: Diebold-Mariano Tests for the $\tau = 0.95$											
	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
LR	1.03	**-5.44**	**5.62**	**7.12**	**7.39**	**6.97**	**5.53**	**7.75**	**7.70**	**7.61**	**8.03**
EN		**-5.49**	**5.62**	**6.70**	**6.94**	**6.57**	**5.23**	**7.24**	**7.17**	**7.11**	**7.62**
RF			**9.44**	**10.60**	**10.95**	**10.77**	**9.15**	**11.99**	**11.50**	**12.14**	**12.87**
GBT				**3.44**	**3.76**	**3.73**	**2.33**	**4.30**	**4.31**	**4.19**	**4.80**
NN1					**3.16**	**2.30**	**-2.68**	**5.26**	**6.09**	**4.03**	**5.19**
NN2						0.77	**-3.52**	**4.04**	**4.95**	**2.91**	**4.43**
NN3							**-5.15**	**3.31**	**4.09**	**2.21**	**4.01**
NN4								**5.86**	**6.19**	**4.93**	**5.79**
MQ1									1.70	-0.45	**2.90**
MQ2										-1.37	1.55
SMQ1											**4.59**

Table IA5: Performance of Portfolios with NYSE Breakpoints Sorted on Robust Risk Premium Forecasts

Panel A reports the monthly performance of value-weighted decile portfolios, formed using NYSE breakpoints, sorted on robust risk premium forecasts. At the end of each month, we predict the one-month-ahead robust risk premium, and then sort stocks into deciles based on their forecasts. We reconstitute portfolios each month using value weights. "Low" denotes the portfolio with the lowest robust risk premium forecasts (decile 1), "High" denotes the portfolio with the highest robust risk premium forecasts (decile 10), "H-L" denotes the long-short portfolio that buys the highest robust risk premium portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Panel B reports the monthly performance of value-weighted decile portfolios, formed using NYSE breakpoints, which are sorted on MSE-based forecasts.

Panel A: Performance of Value-Weighted Portfolios Sorted on $\bar{Q}$ Forecasts

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
Low (L)	0.20	0.18	0.26	-0.24	-0.20	-0.24	-0.23	-0.24	-0.37	-0.37	-0.33	-0.36
2	0.56	0.58	0.48	0.36	0.42	0.54	0.44	0.48	0.47	0.29	0.32	0.34
3	0.62	0.60	0.61	0.49	0.55	0.57	0.49	0.44	0.49	0.51	0.56	0.54
4	0.70	0.77	0.67	0.55	0.71	0.59	0.73	0.65	0.62	0.73	0.70	0.61
5	0.79	0.79	0.67	0.68	0.73	0.83	0.75	0.76	0.78	0.73	0.72	0.78
6	0.85	0.81	0.80	0.77	0.85	0.90	0.89	0.92	0.89	0.88	0.81	0.88
7	0.91	0.90	0.82	0.89	0.88	0.96	0.95	0.99	0.92	0.90	1.00	0.92
8	0.93	0.96	0.88	0.95	0.96	0.96	0.97	1.01	1.06	1.03	1.02	1.05
9	1.00	0.96	0.87	1.08	1.11	1.19	1.26	1.24	1.18	1.20	1.20	1.26
High (H)	1.16	1.18	1.14	1.26	1.40	1.41	1.52	1.47	1.49	1.49	1.56	1.49
H-L	0.96	1.01	0.88	1.50	1.60	1.65	1.75	1.71	1.86	1.86	1.88	1.85
t	4.66	4.46	4.55	5.71	7.06	7.05	7.39	6.96	8.41	7.93	8.48	8.10
SR	0.84	0.82	0.79	1.00	1.16	1.13	1.16	1.17	1.31	1.20	1.29	1.23

Panel B: Performance of Value-Weighted Portfolios Sorted on $\bar{\mu}^{MSE}$

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4
Low (L)	0.33	0.25	0.35	-0.01	-0.18	0.22	-0.09	0.03
2	0.54	0.60	0.57	0.47	0.38	0.44	0.54	0.46
3	0.75	0.69	0.62	0.53	0.60	0.52	0.63	0.39
4	0.79	0.74	0.66	0.74	0.63	0.68	0.66	0.65
5	0.83	0.74	0.72	0.92	0.74	0.75	0.76	0.72
6	0.80	0.83	0.81	0.88	0.76	0.76	0.86	0.65
7	0.88	0.88	0.77	0.73	0.89	0.86	0.81	0.84
8	0.86	0.97	0.85	0.65	0.95	0.95	0.82	0.70
9	0.95	0.96	0.86	0.77	0.98	1.14	1.03	0.92
High (H)	1.20	1.17	0.98	1.00	1.34	1.34	1.41	1.28
H-L	0.86	0.92	0.63	1.08	1.51	1.07	1.47	1.35
t	4.99	3.29	2.82	4.04	5.99	4.68	5.49	3.55
SR	0.90	0.70	0.45	0.77	1.07	0.80	0.91	0.70

Table IA6: Performance of Portfolios Sorted on Robust Risk Premium Forecasts Excluding Microcaps

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on robust risk premium forecasts, for all stocks above the 20th percentile of NYSE cap weights (i.e., microcaps below the 20th percentile NYSE cap are excluded). At the end of each month, we predict the one-month-ahead robust risk premium, and then sort stocks into deciles based on their forecasts. We reconstitute portfolios each month using value weights. "Low" denotes the portfolio with the lowest robust risk premium forecasts (decile 1), "High" denotes the portfolio with the highest robust risk premium forecasts (decile 10), "H-L" denotes the long-short portfolio that buys the highest robust risk premium portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Panel B reports the monthly performance of value-weighted decile portfolios, which are sorted on MSE-based forecasts, for all stocks above the 20th percentile of NYSE cap weights.

Panel A: Performance of Value-Weighted Portfolios Sorted on $\bar{Q}$ Forecasts

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
Low (L)	0.16	0.12	0.24	-0.10	-0.25	-0.31	-0.38	-0.26	-0.37	-0.41	-0.38	-0.42
2	0.50	0.52	0.51	0.28	0.27	0.31	0.27	0.34	0.24	0.14	0.13	0.21
3	0.63	0.64	0.66	0.51	0.49	0.51	0.52	0.51	0.55	0.53	0.55	0.46
4	0.67	0.64	0.68	0.55	0.64	0.69	0.73	0.71	0.66	0.68	0.65	0.63
5	0.73	0.80	0.72	0.68	0.82	0.79	0.75	0.82	0.73	0.79	0.75	0.80
6	0.84	0.79	0.81	0.80	0.85	0.86	0.87	0.92	0.93	0.84	0.84	0.86
7	0.93	0.88	0.74	0.89	0.87	0.93	0.95	0.95	0.97	0.96	0.96	0.95
8	0.89	0.90	0.87	0.95	1.00	0.95	0.96	1.00	0.97	1.03	1.06	1.02
9	1.00	0.95	0.93	1.10	1.01	1.16	1.18	1.14	1.15	1.15	1.14	1.23
High (H)	1.09	1.14	1.06	1.25	1.35	1.34	1.42	1.40	1.42	1.37	1.46	1.37
H-L	0.93	1.02	0.82	1.35	1.60	1.65	1.80	1.66	1.79	1.78	1.84	1.79
t	4.22	4.32	3.57	4.84	5.72	5.71	6.50	5.83	6.37	6.14	6.43	6.20
SR	0.78	0.79	0.66	0.84	0.97	0.95	1.08	0.98	1.05	0.98	1.08	1.00

Panel B: Performance of Value-Weighted Portfolios Sorted on $\bar{\mu}^{MSE}$

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4
Low (L)	0.34	0.27	0.39	0.13	-0.12	0.26	-0.09	0.07
2	0.48	0.61	0.65	0.50	0.28	0.39	0.33	0.48
3	0.64	0.62	0.68	0.71	0.55	0.51	0.54	0.55
4	0.78	0.67	0.80	0.69	0.59	0.65	0.70	0.61
5	0.83	0.76	0.68	0.84	0.67	0.62	0.82	0.74
6	0.82	0.78	0.73	0.73	0.85	0.78	0.88	0.82
7	0.85	0.84	0.78	0.74	0.80	0.83	0.84	0.86
8	0.85	0.85	0.79	0.75	0.94	0.91	0.93	0.92
9	0.92	0.87	0.84	0.92	0.98	1.02	0.98	0.92
High (H)	1.12	1.08	0.93	0.98	1.21	1.23	1.14	1.13
H-L	0.78	0.81	0.54	0.86	1.34	0.98	1.23	1.06
t	4.30	3.13	2.16	3.70	4.51	3.49	4.35	3.77
SR	0.77	0.61	0.36	0.63	0.80	0.59	0.76	0.66

Table IA7: Performance of Equal-Weighted Portfolios Sorted on Robust Risk Premium Forecasts

Panel A reports the monthly performance of equal-weighted decile portfolios, sorted on robust risk premium forecasts. At the end of each month, we predict the one-month-ahead robust risk premium, and then sort stocks into deciles based on their forecasts. We reconstitute portfolios each month using equal weights. "Low" denotes the portfolio with the lowest robust risk premium forecasts (decile 1), "High" denotes the portfolio with the highest robust risk premium forecasts (decile 10), "H-L" denotes the long-short portfolio that buys the highest robust risk premium portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Panel B reports the monthly performance of equal-weighted decile portfolios, which are sorted on MSE-based forecasts.

Panel A: Performance of Equal-Weighted Portfolios Sorted on $\bar{Q}$ Forecasts

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
Low (L)	-0.67	-0.75	-0.43	-1.16	-1.37	-1.43	-1.47	-1.40	-1.50	-1.55	-1.47	-1.57
2	0.19	0.26	0.26	0.10	-0.11	-0.15	-0.19	-0.01	-0.20	-0.19	-0.21	-0.19
3	0.62	0.59	0.49	0.38	0.28	0.29	0.30	0.33	0.27	0.24	0.28	0.23
4	0.73	0.75	0.58	0.61	0.56	0.61	0.56	0.50	0.54	0.53	0.52	0.51
5	0.79	0.87	0.68	0.76	0.69	0.76	0.67	0.68	0.69	0.72	0.70	0.69
6	0.96	0.92	0.79	0.89	0.84	0.83	0.87	0.80	0.87	0.88	0.85	0.85
7	1.03	1.01	0.91	1.04	1.04	1.01	1.06	1.03	1.11	1.06	1.07	1.08
8	1.11	1.07	0.99	1.29	1.27	1.24	1.23	1.23	1.26	1.29	1.27	1.33
9	1.28	1.30	1.23	1.45	1.50	1.54	1.63	1.55	1.59	1.61	1.64	1.64
High (H)	1.57	1.58	2.11	2.25	2.90	2.92	2.95	2.90	2.97	3.01	2.96	3.04
H-L	2.24	2.33	2.54	3.41	4.27	4.35	4.42	4.29	4.47	4.56	4.42	4.61
t	8.47	8.60	8.79	11.67	13.50	13.53	13.39	11.49	14.33	14.41	13.94	14.10
SR	1.76	1.70	1.68	2.11	2.98	2.83	2.75	2.56	2.99	2.87	2.78	2.87

Panel B: Performance of Equal-Weighted Portfolios Sorted on $\bar{\mu}^{MSE}$

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4
Low (L)	-0.23	-0.32	-0.27	-0.18	-1.32	-0.73	-1.08	-0.62
2	0.24	0.35	0.20	0.42	-0.03	-0.03	-0.05	0.21
3	0.52	0.63	0.33	0.56	0.34	0.21	0.37	0.50
4	0.63	0.78	0.56	0.63	0.55	0.48	0.54	0.66
5	0.75	0.80	0.63	0.83	0.74	0.66	0.75	0.73
6	0.76	0.81	0.75	0.89	0.87	0.76	0.90	0.89
7	0.89	0.97	0.93	0.97	1.00	0.97	0.97	0.95
8	1.06	0.96	1.05	1.03	1.20	1.23	1.18	1.06
9	1.28	1.16	1.25	1.03	1.52	1.51	1.47	1.32
High (H)	1.71	1.46	2.18	1.43	2.74	2.53	2.56	1.91
H-L	1.94	1.79	2.46	1.61	4.06	3.26	3.64	2.52
t	7.01	6.41	7.99	5.99	12.92	10.59	11.66	8.98
SR	1.50	1.36	1.43	1.17	2.92	2.37	2.63	1.71

Table IA8: Value-Weighted Bivariate Portfolios Sorted on Earnings Events and Robust Risk Premium Forecasts

This table reports the monthly performance of value-weighted portfolios, first sorted by earnings announcement status and then by the robust risk premium $\bar{Q}$ forecasts for the EN, GBT, NN3, and MQ1 models. In each month, we partition our stocks into two subsamples: one subsample contains stocks that experienced an earnings announcement during that month, while all other stocks belong to the second subsample. We sort stocks into two subsamples based on their earnings announcement status, then we divide each earnings announcement subsample into deciles based on their $\bar{Q}$ forecasts. Within each earnings announcement subsample, "Low" denotes the portfolio with the lowest $\bar{Q}$ forecasts (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ forecasts (decile 10), "H-L" denotes the long-short portfolio that buys the highest $\bar{Q}$ portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Alphas (α) and Newey-West t-statistics (t(α)) for the long-short portfolios are with respect to the Fama-French five-factor model augmented to include the momentum factor.

	Panel A: EN		Panel B: GBT		Panel C: NN3		Panel D: MQ1	
	Stocks Without Earnings Events	Stocks With Earnings Events	Stocks Without Earnings Events	Stocks With Earnings Events	Stocks Without Earnings Events	Stocks With Earnings Events	Stocks Without Earnings Events	Stocks With Earnings Events
Low (L)	-0.26	0.19	-1.19	-0.33	-1.10	-0.34	-1.23	-0.36
2	0.28	0.58	-0.21	0.30	-0.22	0.17	-0.34	0.04
3	0.50	0.93	0.11	0.28	0.32	0.70	0.26	0.59
4	0.55	0.93	0.31	0.55	0.52	0.92	0.47	0.85
5	0.56	1.07	0.52	0.82	0.64	1.17	0.57	0.96
6	0.68	1.17	0.59	1.06	0.77	1.00	0.72	1.28
7	0.74	1.16	0.72	1.37	0.93	1.47	0.91	1.56
8	0.83	1.41	0.91	1.33	1.17	1.35	1.06	1.43
9	0.79	1.41	0.93	1.62	1.25	1.77	1.21	1.81
High (H)	1.03	1.59	1.16	2.14	1.77	2.47	1.76	2.16
H-L	1.29	1.40	2.35	2.48	2.87	2.81	2.99	2.52
t	4.90	3.43	6.83	5.31	8.85	5.92	9.30	5.59
SR	0.89	0.68	1.14	0.93	1.44	1.09	1.50	0.94
α	0.73	0.81	1.82	1.76	2.65	2.12	2.63	1.76
t(α)	4.01	2.71	6.62	4.16	8.94	4.73	9.64	4.61

Table IA9: Value-Weighted Bivariate Portfolios Sorted on Analyst Coverage and MSE-Based Forecasts

This table reports the monthly performance of value-weighted portfolios, first sorted by analyst coverage (CVRG) and then by the MSE-based forecast $\bar{\mu}^{MSE}$ for the EN, RF, GBT, and NN3 models. We sort stocks into terciles based on the number of analysts following the firm in the most recent month, then we divide each CVRG tercile into deciles based on their $\bar{\mu}^{MSE}$. Within each CVRG tercile, "Low" denotes the portfolio with the lowest $\bar{\mu}^{MSE}$ (decile 1), "High" denotes the portfolio with the highest $\bar{\mu}^{MSE}$ (decile 10), "H-L" denotes the long-short portfolio that buys the highest $\bar{\mu}^{MSE}$ portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Alphas (α) and Newey-West t-statistics (t(α)) for the long-short portfolios are with respect to the Fama-French five-factor model augmented to include the momentum factor.

Performance of Bivariate Portfolio Sorts on Analyst Coverage and MSE-Based Forecasts												
	Panel A: EN			Panel B: RF			Panel C: GBT			Panel D: NN3		
	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3
Low (L)	-0.05	-0.18	0.25	-0.19	-0.06	0.38	-0.13	-0.07	-0.01	-1.25	-1.23	-0.32
2	0.40	0.55	0.58	0.31	0.54	0.67	0.30	0.40	0.45	-0.33	-0.06	0.20
3	0.46	0.47	0.69	0.36	0.55	0.78	0.30	0.46	0.65	0.10	-0.04	0.47
4	0.50	0.65	0.62	0.56	0.48	0.76	0.37	0.62	0.77	0.44	0.24	0.58
5	0.54	0.54	0.77	0.59	0.70	0.80	0.49	0.57	0.81	0.42	0.72	0.77
6	0.57	0.74	0.75	0.56	0.73	0.66	0.55	0.49	0.71	0.55	0.89	0.92
7	0.49	0.88	0.85	0.73	0.80	0.76	0.78	0.74	0.76	0.87	0.93	0.94
8	0.68	0.87	0.85	0.98	0.97	0.80	0.65	0.83	0.84	1.15	0.94	0.85
9	0.72	0.88	0.94	1.17	1.01	0.92	0.70	0.79	0.93	1.53	1.21	1.04
High (H)	1.06	1.05	1.10	1.18	1.28	0.99	0.91	0.81	1.05	1.96	1.71	1.22
H-L	1.11	1.23	0.84	1.37	1.35	0.61	1.04	0.89	1.06	3.21	2.94	1.54
t	3.59	4.24	3.08	3.85	3.71	2.15	4.21	2.92	3.93	8.31	7.67	4.41
SR	0.76	0.84	0.60	0.66	0.65	0.35	0.69	0.50	0.68	1.68	1.55	0.79
α	0.96	0.76	0.36	1.57	1.47	0.41	0.79	0.35	0.38	3.17	2.60	0.94
t(α)	3.53	2.83	1.68	3.75	3.62	1.62	2.93	1.41	2.13	7.42	6.35	3.08

Table IA10: Value-Weighted Bivariate Portfolios Sorted on Arbitrage Index and MSE-Based Forecasts

This table reports the monthly performance of value-weighted portfolios, first sorted by an index of arbitrage costs and then by the MSE-based forecasts $\bar{\mu}^{MSE}$ for the EN, RF, GBT, and NN3 models. Arbitrage Index (AI) construction follows Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023), and includes idiosyncratic volatility, illiquidity, size, age, and analyst coverage. We sort stocks into quintiles based on their AI, then we divide each AI quintile into deciles based on their $\bar{\mu}^{MSE}$. Within each AI quintile, "Low" denotes the portfolio with the lowest $\bar{\mu}^{MSE}$ (decile 1), "High" denotes the portfolio with the highest $\bar{\mu}^{MSE}$ (decile 10), "H-L" denotes the long-short portfolio that buys the highest $\bar{\mu}^{MSE}$ portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Alphas (α) and Newey-West t-statistics (t(α)) for the long-short portfolios are with respect to the Fama-French five-factor model augmented to include the momentum factor.

Performance of Bivariate Portfolio Sorts on Arbitrage Index and MSE-Based Forecasts																				
	Panel A: EN					Panel B: RF					Panel C: GBT					Panel D: NN3				
	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5
Low (L)	0.31	0.28	-0.11	-0.73	-1.76	0.55	0.58	-0.10	-0.58	-1.12	0.37	0.23	-0.07	-0.75	-1.38	0.21	-0.11	-0.54	-1.45	-2.63
2	0.62	0.61	0.29	-0.18	-0.59	0.70	0.87	0.34	0.05	-0.44	0.60	0.69	0.28	0.20	-0.29	0.48	0.52	-0.08	-0.22	-1.17
3	0.61	0.73	0.55	0.31	-0.05	0.69	0.74	0.55	0.19	-0.18	0.68	0.72	0.33	0.29	-0.33	0.68	0.68	0.50	-0.13	-0.59
4	0.67	0.80	0.57	0.44	-0.07	0.79	0.79	0.65	0.48	-0.19	0.80	0.95	0.70	0.23	0.19	0.78	0.85	0.66	0.34	0.16
5	0.81	0.76	0.70	0.61	0.31	0.76	0.85	0.77	0.94	0.30	0.79	0.79	1.03	0.64	0.45	0.77	0.93	0.89	0.71	0.06
6	0.73	0.97	0.83	0.82	0.42	0.73	0.87	0.89	0.95	0.88	0.77	0.86	0.91	0.65	0.49	0.93	0.94	0.92	0.97	0.91
7	0.91	0.91	0.92	0.97	0.79	0.81	0.81	1.12	0.70	0.81	0.79	0.82	0.98	0.99	0.38	0.88	0.95	0.90	1.14	1.23
8	0.98	1.02	0.85	0.97	1.13	0.78	0.98	1.04	0.83	1.22	0.87	1.01	0.90	0.99	0.76	0.88	1.04	1.23	1.41	1.56
9	0.90	1.25	1.19	1.25	1.20	0.76	0.97	1.11	1.20	1.26	0.87	0.98	0.75	0.93	1.03	1.01	1.24	1.29	1.56	1.91
High (H)	1.05	1.24	1.39	1.23	1.50	0.91	1.11	1.08	1.58	3.22	0.95	1.14	1.13	1.21	1.24	0.99	1.50	1.47	1.87	3.03
H-L	0.73	0.96	1.50	1.96	3.26	0.36	0.53	1.18	2.17	4.34	0.58	0.90	1.19	1.96	2.62	0.77	1.60	2.00	3.32	5.66
t	3.17	2.94	4.23	5.75	8.72	2.02	1.75	3.93	6.73	7.63	3.23	2.81	3.61	5.55	6.50	3.61	5.12	6.03	8.43	11.70
SR	0.59	0.60	0.83	1.08	1.80	0.31	0.29	0.64	1.08	1.42	0.53	0.51	0.68	1.10	1.39	0.55	0.84	1.12	1.73	2.77
α	0.30	0.48	1.03	1.45	2.70	0.16	0.31	1.00	2.21	4.41	0.18	0.26	0.59	1.48	2.19	0.39	1.17	1.56	2.98	5.44
t(α)	1.64	1.72	3.46	4.60	6.82	0.87	1.14	3.53	6.50	7.49	1.18	1.06	2.42	4.79	5.47	2.09	4.30	5.94	8.40	11.44

Table IA11: Value-Weighted Bivariate Portfolios Sorted on Changes and Levels of MSE-Based Forecasts

This table reports the monthly performance of value-weighted portfolios, first sorted by the monthly change in the MSE-based forecast $\bar{\mu}^{MSE}$ ($\Delta\bar{\mu}$) and then by the MSE-based forecast level ($\bar{\mu}^{MSE}$) for the EN, RF, GBT, and NN3 models. For each firm i, we define $\Delta\bar{\mu}$ as the MSE-based forecast at the end of month t minus the MSE-based forecast at the end of month $t-1$. We sort stocks into terciles based on $\Delta\bar{\mu}$, then we divide each $\Delta\bar{\mu}$ tercile into deciles based on their $\bar{\mu}^{MSE}$. Within each $\Delta\bar{\mu}$ tercile, "Low" denotes the portfolio with the lowest $\bar{\mu}^{MSE}$ (decile 1), "High" denotes the portfolio with the highest $\bar{\mu}^{MSE}$ (decile 10), "H-L" denotes the long-short portfolio that buys the highest $\bar{\mu}^{MSE}$ portfolio (decile 10) and sells the lowest (decile 1). The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Alphas (α) and Newey-West t-statistics (t(α)) for the long-short portfolios are with respect to the Fama-French five-factor model augmented to include the momentum factor.

Performance of Bivariate Portfolio Sorts on $\Delta\bar{\mu}$ and $\bar{\mu}^{MSE}$ Forecasts												
	Panel A: EN			Panel B: RF			Panel C: GBT			Panel D: NN3		
	$\Delta\bar{\mu}1$	$\Delta\bar{\mu}2$	$\Delta\bar{\mu}3$	$\Delta\bar{\mu}1$	$\Delta\bar{\mu}2$	$\Delta\bar{\mu}3$	$\Delta\bar{\mu}1$	$\Delta\bar{\mu}2$	$\Delta\bar{\mu}3$	$\Delta\bar{\mu}1$	$\Delta\bar{\mu}2$	$\Delta\bar{\mu}3$
Low (L)	-0.19	0.36	0.42	0.02	0.15	0.44	-0.30	0.13	0.36	-1.19	-0.08	-0.29
2	0.41	0.51	0.58	0.49	0.52	0.61	0.04	0.37	0.49	-0.56	0.43	0.20
3	0.40	0.56	0.66	0.58	0.50	0.62	0.23	0.71	0.82	-0.19	0.53	0.73
4	0.60	0.69	0.84	0.54	0.56	0.81	0.46	0.68	0.80	-0.07	0.84	0.80
5	0.51	0.71	0.76	0.80	0.72	0.87	0.47	0.67	0.69	0.18	0.82	0.98
6	0.75	0.71	0.82	0.89	0.85	0.86	0.74	0.58	0.67	0.36	0.77	1.05
7	0.73	0.81	0.83	0.95	0.78	0.96	0.82	0.69	0.57	0.47	0.89	1.24
8	0.75	0.81	0.92	0.86	0.90	1.17	0.86	0.64	1.11	0.81	1.01	1.46
9	0.86	1.03	1.05	0.86	0.87	1.18	0.84	0.86	0.94	0.83	1.04	1.47
High (H)	0.99	1.17	1.13	0.83	1.04	1.81	0.86	0.88	1.24	0.94	1.20	2.25
H-L	1.18	0.81	0.70	0.81	0.89	1.38	1.16	0.75	0.88	2.13	1.28	2.54
t	4.14	3.01	2.47	2.27	3.27	3.15	3.93	2.72	2.49	5.89	4.50	5.96
SR	0.81	0.58	0.48	0.44	0.59	0.56	0.75	0.47	0.51	1.08	0.78	1.20
α	0.76	0.44	0.34	0.58	0.56	1.70	0.61	0.33	0.23	1.57	0.76	2.21
t(α)	3.15	1.88	1.42	1.85	2.16	3.22	2.82	1.40	0.88	4.68	2.96	4.87

Table IA12: Comparison of Quantile-Based Estimators of the Conditional Expectation

This table reports results for different quantile-based estimators of the conditional expectation. Panel A reports the out-of-sample R^2_{os} values (in percentages) for the pooled test sample of individual stock excess returns. Panel B reports the monthly performance of value-weighted decile portfolios, sorted on quantile-based estimator forecasts. The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. The quantile-based estimators are defined as:

$$\bar{Q} = 0.15Q_{0.05} + 0.225Q_{0.25} + 0.25Q_{0.50} + 0.225Q_{0.75} + 0.15Q_{0.95}$$
$$\bar{Q}_{YZLZ} = 0.1Q_{0.05} + 0.3Q_{0.25} + 0.2Q_{0.50} + 0.3Q_{0.75} + 0.1Q_{0.95}$$
$$\bar{Q}_{FQ} = 0.05Q_{0.05} + 0.25Q_{0.25} + 0.4Q_{0.50} + 0.25Q_{0.75} + 0.05Q_{0.95}$$
$$\bar{Q}_{WWLT} = (Q_{0.25} + Q_{0.50} + Q_{0.75})/3$$
$$\bar{Q}_{TM} = 0.25Q_{0.25} + 0.5Q_{0.50} + 0.25Q_{0.75}$$
$$\bar{Q}_{MP} = 0.05Q_{0.05} + 0.9Q_{0.50} + 0.05Q_{0.95}.$$

Detailed descriptions of these estimators are provided in Internet Appendix Section IA4.

Panel A: Monthly Out-of-sample Percentage R^2_{OS} of Quantile-Based Estimators

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
$\bar{Q}$	**0.13**	**0.48**	0.17	**0.36**	**0.89**	**0.87**	**0.98**	**0.87**	**0.99**	**0.92**	**0.96**	**1.00**
$\bar{Q}_{YZLZ}$	0.03	0.38	0.35	0.27	0.77	0.75	0.87	0.81	0.86	0.79	0.83	0.87
$\bar{Q}_{FQ}$	-0.19	0.16	**0.43**	0.01	0.43	0.43	0.57	0.54	0.49	0.43	0.51	0.51
$\bar{Q}_{WWLT}$	-0.46	-0.11	0.30	-0.26	0.02	0.01	0.14	0.17	0.07	0.01	0.08	0.09
$\bar{Q}_{TM}$	-0.50	-0.15	0.27	-0.37	-0.06	-0.04	0.09	0.11	-0.03	-0.09	0.01	-0.01
$\bar{Q}_{MP}$	-0.30	0.01	0.36	-0.61	0.18	0.20	0.36	0.32	0.22	0.17	0.24	0.22

Panel B: Sharpe ratios of Value-Weighted Portfolios Sorted on Quantile-Based Estimators

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
$\bar{Q}$	0.91	0.89	0.77	**1.10**	**1.44**	**1.41**	**1.42**	1.26	**1.53**	**1.55**	**1.50**	**1.53**
$\bar{Q}_{YZLZ}$	**0.96**	0.95	0.79	1.08	1.33	1.32	1.36	**1.45**	1.40	1.48	1.46	1.47
$\bar{Q}_{FQ}$	0.94	**1.00**	**0.81**	1.08	1.18	1.20	1.28	1.32	1.23	1.31	1.37	1.35
$\bar{Q}_{WWLT}$	0.89	**1.00**	0.71	0.99	1.12	1.14	1.16	1.15	1.10	1.22	1.23	1.22
$\bar{Q}_{TM}$	0.86	0.94	0.67	1.01	1.07	1.11	1.13	1.17	1.10	1.17	1.20	1.20
$\bar{Q}_{MP}$	0.84	0.83	0.71	1.00	0.99	0.99	1.20	1.19	1.08	1.14	1.16	1.28

Table IA13: Performance of Generalized Robust Risk Premium with Different Numbers of Quantiles

This table presents the results for the generalized robust risk premium computed using MQ1 with varying numbers of quantiles, K. Panel A reports the out-of-sample R^2_{os} values (in percentages) for the pooled test sample of individual stock excess returns. Panel B reports the monthly performance of value-weighted decile portfolios, sorted on the generalized robust risk premium forecasts. The average realized returns are in monthly percentage. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the long-short portfolios. Reported are values for:

$$K = 5, (\tau_k \in \{0.05, 0.25, 0.5, 0.75, 0.95\});$$
$$K = 21, (\tau_k \in \{0.01, 0.05, 0.10, 0.15, ..., 0.85, 0.90, 0.95, 0.99\});$$
$$K = 25, (\tau_k = \{0.02, 0.06, 0.10, 0.14, ..., 0.86, 0.90, 0.94, 0.98\});$$
$$K = 33, (\tau_k = \{0.02, 0.05, 0.08, 0.11, ..., 0.89, 0.92, 0.95, 0.98\});$$
$$K = 51, (\tau_k = \{0.01, 0.03, 0.05, ..., 0.47, 0.49, 0.5, 0.51, 0.53, ...0.95, 0.97, 0.99\});$$
$$K = 99, (\tau_k = \{0.01, 0.02, 0.03, ...0.97, 0.98, 0.99\}).$$

The generalized robust risk premium, $\bar{Q}_K$, is computed using all K quantiles according to Equation (6).

	$K=5$	$K=21$	$K=25$	$K=33$	$K=51$	$K=99$
Panel A: Monthly Percentage R^2_{OS} of $\bar{Q}_K$ Forecasts						
R^2_{OS}	0.99	0.87	1.02	0.87	0.91	0.86
Panel B: Performance of Portfolios Sorted on $\bar{Q}_K$ Forecasts						
Low (L)	-1.09	-1.17	-1.20	-1.20	-1.15	-1.22
2	-0.21	-0.12	-0.20	-0.20	-0.16	-0.05
3	0.37	0.35	0.20	0.28	0.22	0.25
4	0.64	0.56	0.53	0.51	0.52	0.51
5	0.73	0.72	0.75	0.78	0.71	0.75
6	0.85	0.94	0.79	0.82	0.91	0.81
7	1.08	1.04	1.06	1.00	1.05	1.03
8	1.13	1.21	1.16	1.13	1.21	1.15
9	1.43	1.44	1.31	1.42	1.39	1.42
High (H)	1.88	1.88	1.92	1.90	1.96	1.84
H-L	2.96	3.06	3.12	3.10	3.11	3.07
t	9.49	9.86	9.46	9.73	9.74	9.37
SR	1.53	1.54	1.57	1.54	1.51	1.51
α	2.51	2.57	2.70	2.65	2.69	2.68
t(α)	10.06	9.89	9.54	9.70	9.57	9.48
R^2	33	30	28	31	31	31

Table IA14: Performance of RRPV-Managed Value-Weighted Portfolios

This reports the Sharpe ratios of the RRPV-Managed value-weighted portfolios. At the end of each month, we sort stocks into deciles based on their one-month-ahead robust risk premium forecasts. The portfolios with the lowest and highest robust risk premium forecasts are labeled as "Low" and "High", respectively. To form RRPV-Managed portfolios, we further divide each of the above portfolios into deciles based on their $RRPV$ ratio forecasts, rebalancing them monthly using value weights. Within each robust risk premium decile, D1 (D10) denotes the portfolio with the lowest (highest) $RRPV$ ratio forecasts. This procedure generates subportfolios $\{L_{D1}, ..., L_{D10}\}$ and $\{H_{D1}, ..., H_{D10}\}$. Panel A displays the monthly average returns of our RRPV-Managed strategy, denoted as "H_{D10}-L_{D1}", which takes a long position in H_{D10}, consisting of stocks with the highest $\bar{Q}$ and $RRPV$ ratio forecasts, and a short position in L_{D1}, consisting of stocks with the lowest $\bar{Q}$ and $RRPV$ ratio forecasts. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the H_{D10}-L_{D1} long-short portfolios. Panel B presents the alphas (α), Newey-West t-statistics (t(α)), and R^2 values for the value-weighted long-short portfolios, with respect to the Fama-French five-factor model augmented to include the momentum factor. Panel C compares the Sharpe ratios of the RRPV-Managed portfolios between D1 and D10 across deciles of robust risk premium forecasts.

Panel A: Performance of RRPV-Managed Value-Weighted Long-Short Portfolios

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
H_{D10}-L_{D1}	0.81	0.83	2.29	5.05	4.28	5.32	5.91	6.56	5.17	6	5.95	6.28
t	3.13	2.82	4.19	9.54	8.91	9.94	9.67	8.95	10.42	10.7	11.05	10.44
SR	0.53	0.52	0.8	1.62	1.53	1.99	1.77	1.77	1.99	2.03	2.13	2.1

Panel B: Risk-Adjusted Performance of RRPV-Managed Value-Weighted Long-Short Portfolios

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
α	0.36	0.36	1.92	4.12	3.51	4.84	5.74	6.44	4.85	5.4	5.49	5.98
t(α)	1.65	1.54	4.53	9.55	7.98	8.64	8.34	7.7	9.39	10.02	10.61	9.71
R^2	27	37	31	37	29	21	8	6	18	23	20	18

Panel C: Sharpe Ratios of RRPV-Managed Value-Weighted Portfolios												
	LR		EN		RF		GBT		NN1		NN2	
	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10
Low (L)	0.2	0.0	0.1	-0.1	-0.2	0.1	-1.1	-0.2	-0.8	0.0	-0.9	-0.1
2	0.4	0.0	0.4	0.1	0.1	0.1	-0.1	0.1	-0.2	0.1	-0.1	0.2
3	0.2	0.1	0.4	0.2	0.2	0.3	0.1	0.2	0.0	0.2	0.1	0.2
4	0.2	0.2	0.2	0.3	0.1	0.2	0.1	0.3	0.4	0.4	0.3	0.4
5	0.3	0.2	0.3	0.4	0.3	0.4	0.3	0.3	0.4	0.5	0.2	0.6
6	0.3	0.4	0.2	0.4	0.2	0.4	0.4	0.4	0.4	0.7	0.4	0.5
7	0.2	0.2	0.2	0.7	0.4	0.4	0.5	0.4	0.3	0.9	0.5	0.9
8	0.3	0.5	0.3	0.7	0.2	0.5	0.6	0.7	0.5	0.9	0.4	1.0
9	0.3	0.7	0.2	0.9	0.1	0.8	0.6	0.8	0.3	1.3	0.3	1.2
High (H)	0.3	0.8	0.3	0.8	0.0	1.0	0.8	0.8	0.5	1.1	0.5	1.4
	NN3		NN4		MT1		MT2		SMQ1		SMQ2	
	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10
Low (L)	-0.9	-0.1	-0.9	-0.2	-0.9	0.0	-0.9	-0.1	-0.9	-0.1	-1.0	-0.1
2	-0.1	0.1	0.0	0.2	-0.1	0.0	-0.3	0.1	-0.1	0.0	-0.1	0.1
3	0.2	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3
4	0.4	0.3	0.4	0.6	0.2	0.4	0.2	0.3	0.3	0.3	0.3	0.3
5	0.3	0.6	0.4	0.5	0.3	0.3	0.4	0.4	0.5	0.4	0.4	0.6
6	0.6	0.7	0.4	0.7	0.4	0.6	0.4	0.7	0.5	0.7	0.5	0.6
7	0.5	0.9	0.4	0.8	0.5	0.9	0.5	0.7	0.6	0.9	0.5	0.8
8	0.5	1.2	0.5	0.8	0.5	1.0	0.5	1.0	0.4	1.2	0.6	1.1
9	0.4	1.3	0.3	1.1	0.6	1.2	0.7	1.1	0.5	1.3	0.6	1.3
High (H)	0.5	1.1	0.5	1.1	0.5	1.3	0.4	1.6	0.4	1.5	0.5	1.5

Table IA15: Performance of RRPV-Managed Equal-Weighted Portfolios

This reports the Sharpe ratios of the RRPV-Managed equal-weighted portfolios. At the end of each month, we sort stocks into deciles based on their one-month-ahead robust risk premium forecasts. The portfolios with the lowest and highest robust risk premium forecasts are labeled as "Low" and "High", respectively. To form RRPV-Managed portfolios, we further divide each of the above portfolios into deciles based on their $RRPV$ ratio forecasts, rebalancing them monthly using equal weights. Within each robust risk premium decile, D1 (D10) denotes the portfolio with the lowest (highest) $RRPV$ ratio forecasts. This procedure generates subportfolios $\{L_{D1}, ..., L_{D10}\}$ and $\{H_{D1}, ..., H_{D10}\}$. Panel A displays the monthly average returns of our RRPV-Managed strategy, denoted as "H_{D10}-L_{D1}", which takes a long position in H_{D10}, consisting of stocks with the highest $\bar{Q}$ and $RRPV$ ratio forecasts, and a short position in L_{D1}, consisting of stocks with the lowest $\bar{Q}$ and $RRPV$ ratio forecasts. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the H_{D10}-L_{D1} long-short portfolios. Panel B presents the alphas (α), Newey-West t-statistics (t(α)), and R^2 values for the equal-weighted long-short portfolios, with respect to the Fama-French five-factor model augmented to include the momentum factor. Panel C compares the Sharpe ratios of the RRPV-Managed portfolios between D1 and D10 across deciles of robust risk premium forecasts.

Panel A: Performance of RRPV-Managed Equal-Weighted Long-Short Portfolios

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
H_{D10}-L_{D1}	1.8	2.1	4.98	5.81	7.09	7.49	8.52	9.3	7.57	8.42	8.06	8.54
t	6.28	7.16	11.45	12.61	14.59	14.84	14.1	13.2	14.85	14.97	14.48	16.03
SR	1.12	1.38	2.04	2.2	2.85	2.94	2.85	2.64	3.05	3.11	3.07	3.25

Panel B: Performance of RRPV-Managed Equal-Weighted Long-Short Portfolios

	LR	EN	RF	GBT	NN1	NN2	NN3	NN4	MQ1	MQ2	SMQ1	SMQ2
α	1.19	1.66	4.79	5.1	6.67	7.27	8.54	9.3	7.34	8.2	7.65	8.37
t(α)	4.56	8.09	11.14	13.95	12.25	12.6	12.05	11.52	13.14	13.56	12.9	13.27
R^2	32	40	14	40	8	6	1	6	4	2	5	4

Panel C: Sharpe Ratios of RRPV-Managed Equal-Weighted Portfolios												
	LR		EN		RF		GBT		NN1		NN2	
	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10
Low (L)	-0.3	-0.1	-0.5	-0.1	-0.5	-0.2	-1.4	0.0	-1.1	-0.2	-1.1	-0.2
2	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	-0.1	-0.1	-0.3	0.0
3	0.4	0.3	0.4	0.1	0.1	0.3	0.3	0.4	0.1	0.2	0.2	0.1
4	0.3	0.4	0.4	0.2	0.1	0.2	0.4	0.4	0.4	0.3	0.3	0.2
5	0.3	0.3	0.6	0.4	0.2	0.3	0.5	0.4	0.4	0.4	0.4	0.4
6	0.6	0.4	0.6	0.4	0.3	0.4	0.6	0.5	0.4	0.6	0.4	0.5
7	0.5	0.4	0.6	0.6	0.3	0.5	0.6	0.6	0.4	0.9	0.4	0.8
8	0.6	0.5	0.6	0.7	0.3	0.6	0.8	0.8	0.5	1.0	0.4	1.1
9	0.7	0.8	0.7	0.9	0.3	0.8	0.9	1.0	0.4	1.4	0.4	1.3
High (H)	0.7	1.0	0.6	1.1	0.3	1.4	1.2	1.2	0.5	1.7	0.5	1.6
	NN3		NN4		MT1		MT2		SMQ1		SMQ2	
	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10
Low (L)	-1.4	-0.2	-1.2	-0.2	-1.2	-0.3	-1.3	-0.2	-1.2	-0.1	-1.3	-0.2
2	-0.3	0.0	0.1	0.1	-0.2	0.0	-0.3	0.1	-0.2	0.1	-0.3	0.1
3	0.2	0.2	0.1	0.2	0.1	0.2	0.1	0.1	0.2	0.2	0.0	0.3
4	0.3	0.3	0.3	0.3	0.1	0.3	0.4	0.2	0.2	0.3	0.2	0.3
5	0.4	0.4	0.3	0.4	0.5	0.2	0.5	0.4	0.4	0.4	0.4	0.4
6	0.4	0.7	0.3	0.5	0.5	0.6	0.5	0.7	0.4	0.5	0.5	0.7
7	0.5	0.8	0.3	0.8	0.5	0.8	0.5	0.8	0.4	0.8	0.4	0.8
8	0.4	1.2	0.3	0.9	0.5	1.1	0.6	1.0	0.4	1.1	0.6	1.0
9	0.5	1.4	0.3	1.2	0.5	1.4	0.5	1.3	0.6	1.4	0.5	1.5
High (H)	0.5	1.5	0.4	1.6	0.6	1.7	0.6	1.7	0.4	1.8	0.6	1.8

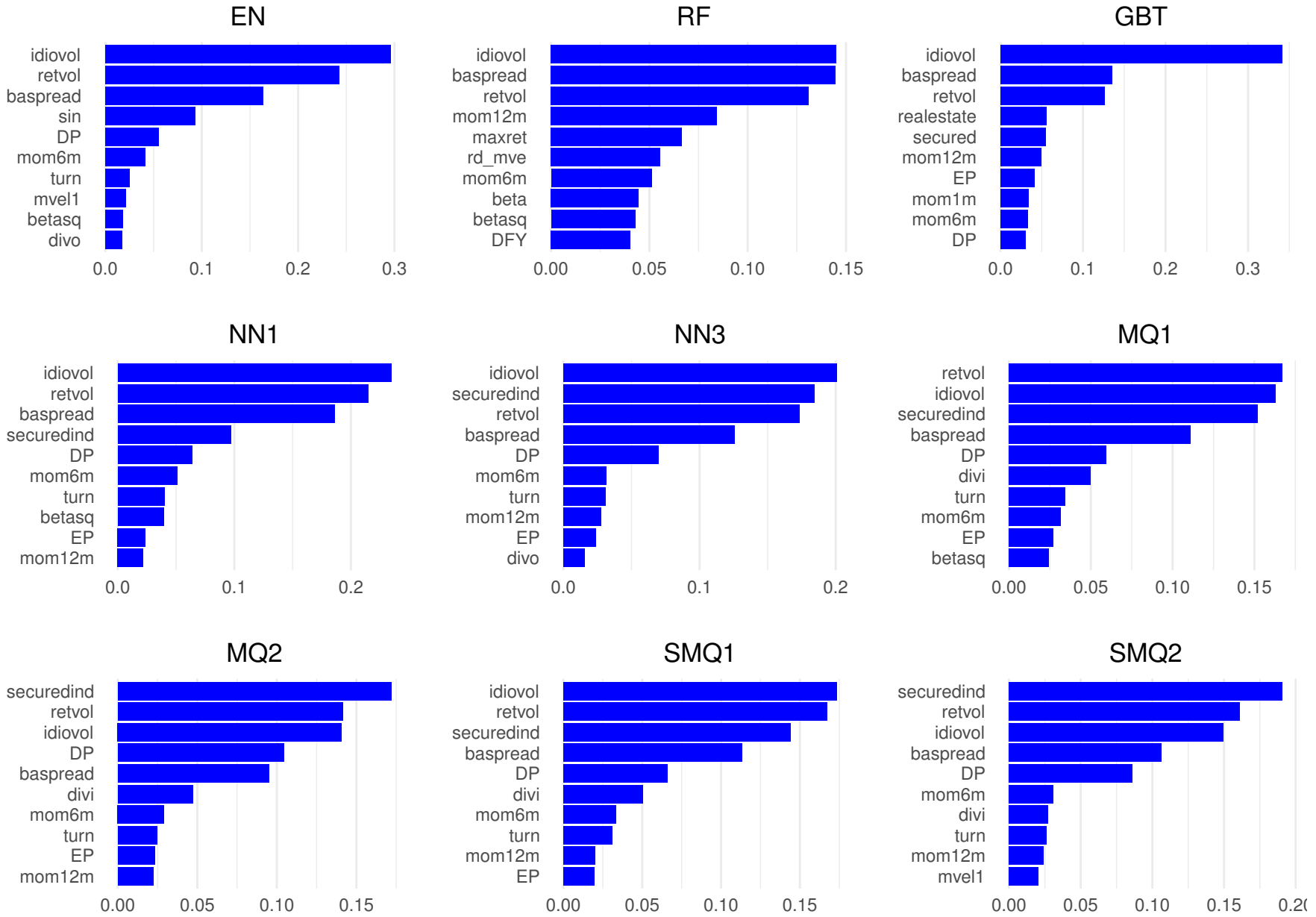
Figure IA1: Time Series of the Cross-Sectional Mean and Dispersion of MQ1 Quantile Forecasts

This figure shows the time-series of the cross-sectional mean (solid line) and the range between the 25th and 75th percentiles (shaded area) of the one-month-ahead MQ1 quantile forecasts across all stocks in each month, for each $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$.

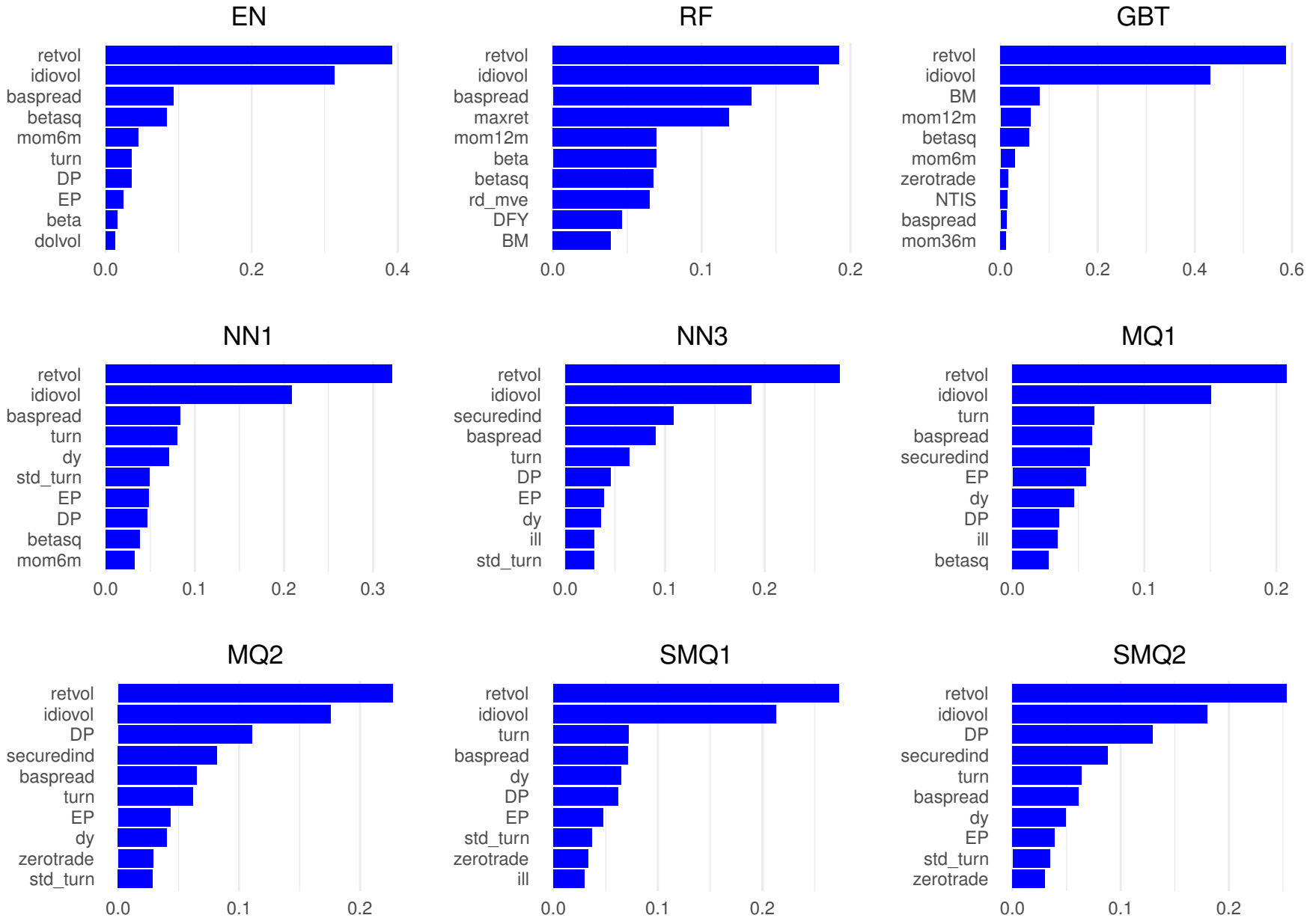


Figure IA2: Variable Importance by Model for each Quantile

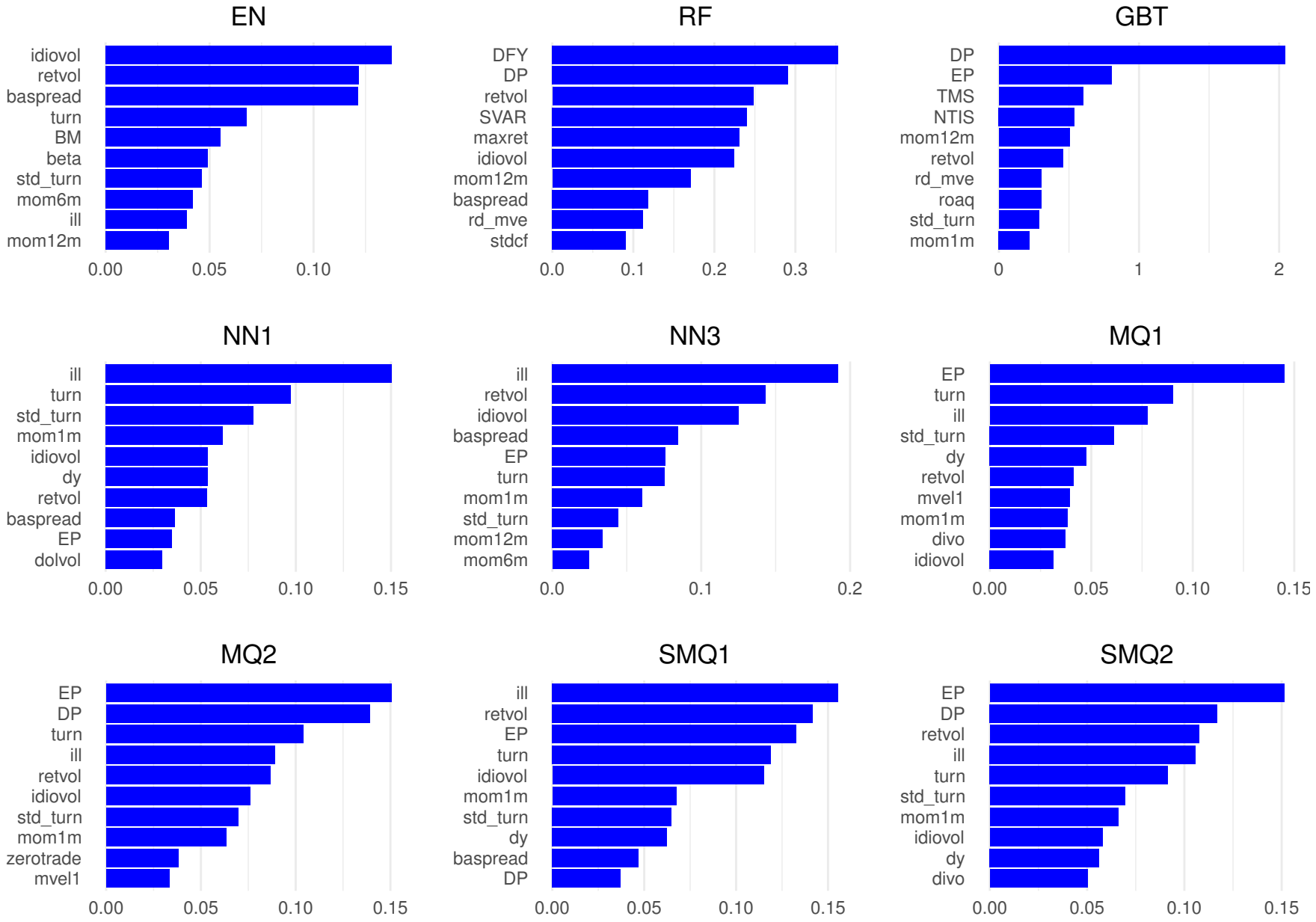
This figure reports the variable importance for the top 10 most influential variables of each model. For each model and each τ ($\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$), the variable importance of the j^{th} predictor is measured by calculating the reduction in $R^1_{os}(\tau)$ from setting all values of the j^{th} predictor to zero within each test sample, and then averaging the $R^1_{os}(\tau)$ reductions across test samples as a single importance measure. For every τ, the variable importance within each model is normalized to sum to one. As predictors, we use 94 firm-level characteristics from Gu, Kelly, and Xiu (2020) and eight macroeconomic variables from Welch and Goyal (2008).



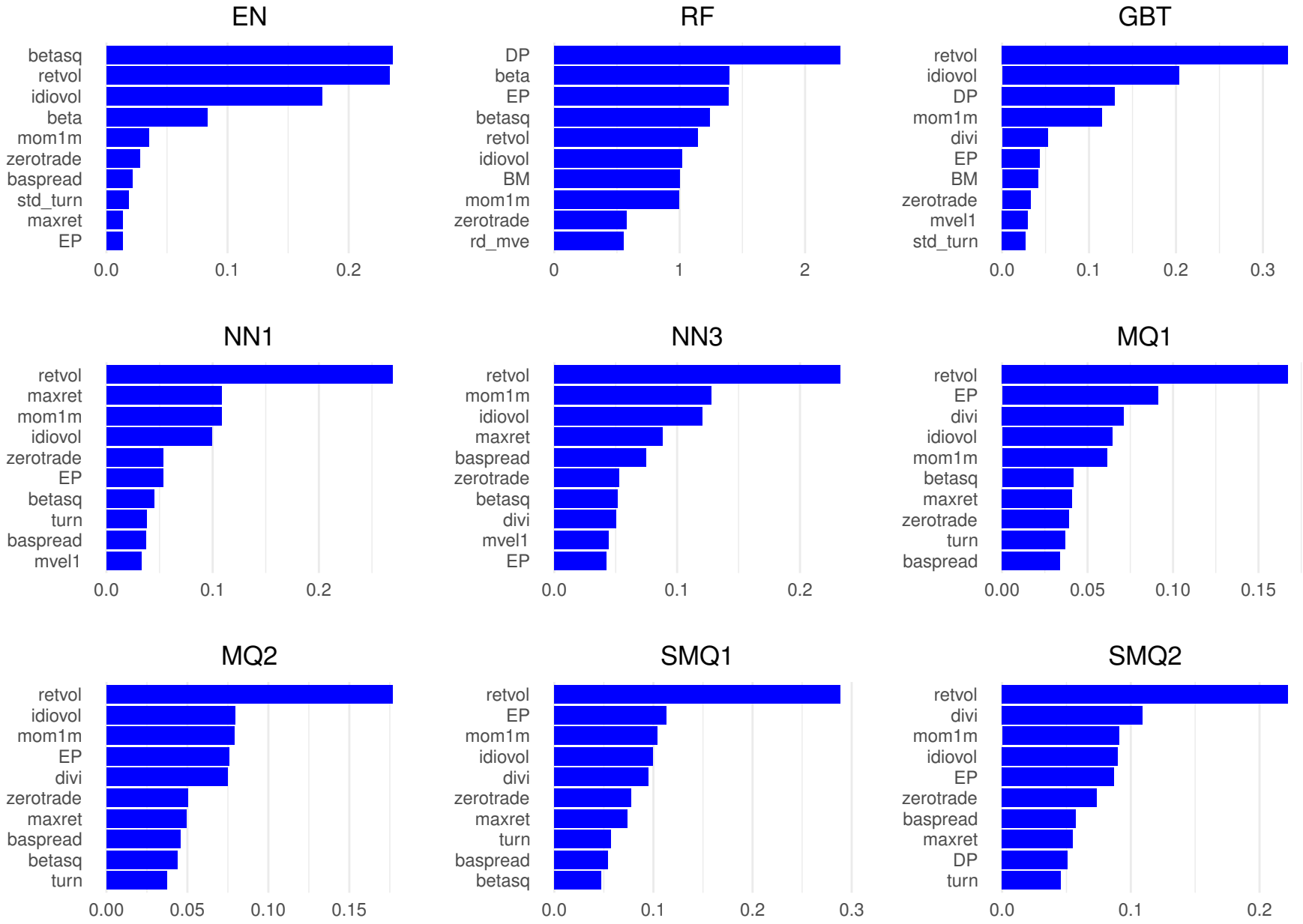
Panel A: Variable Importance Plot for $\tau = 0.05$



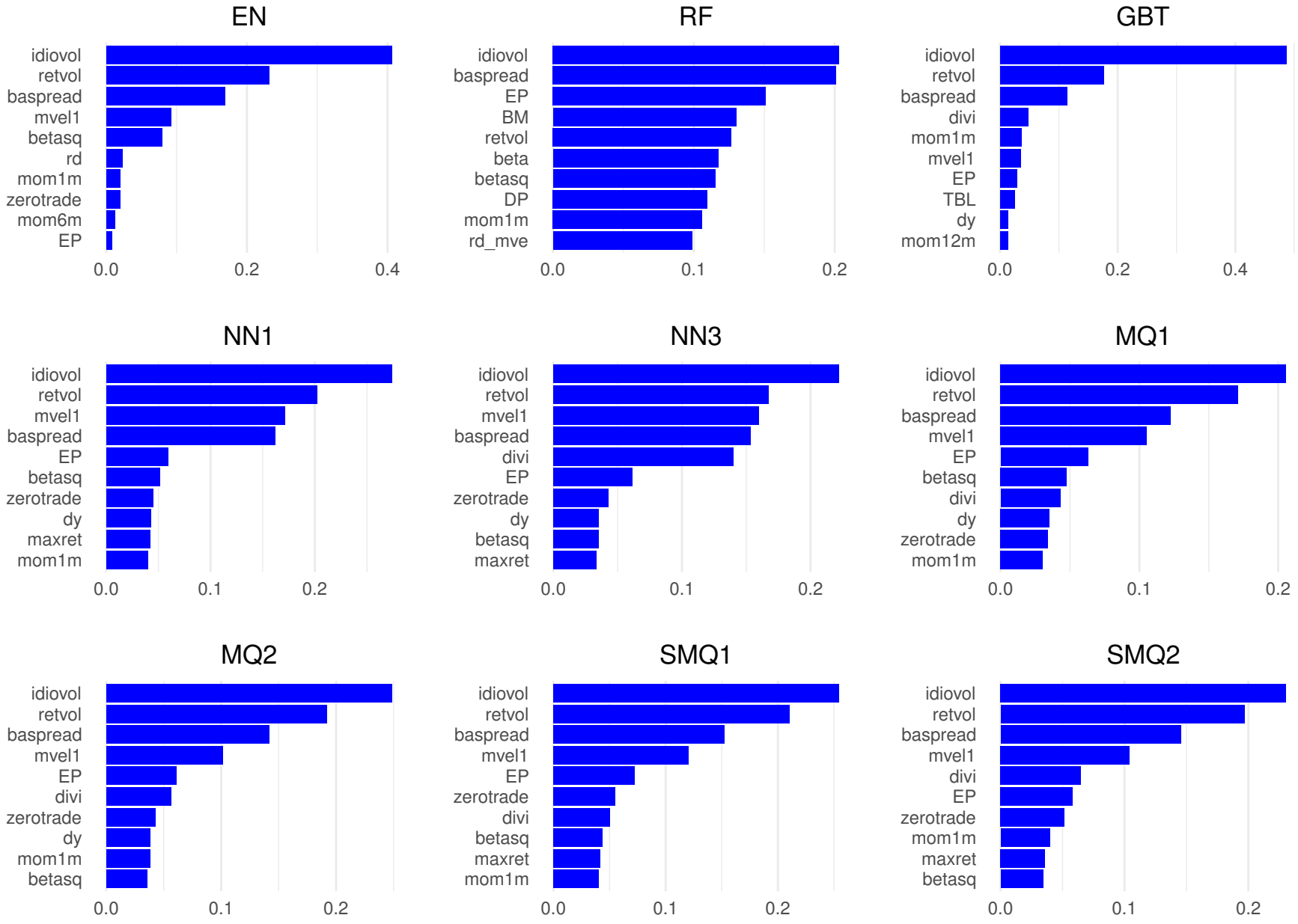
Panel B: Variable Importance Plot for $\tau = 0.25$



Panel C: Variable Importance Plot for $\tau = 0.50$



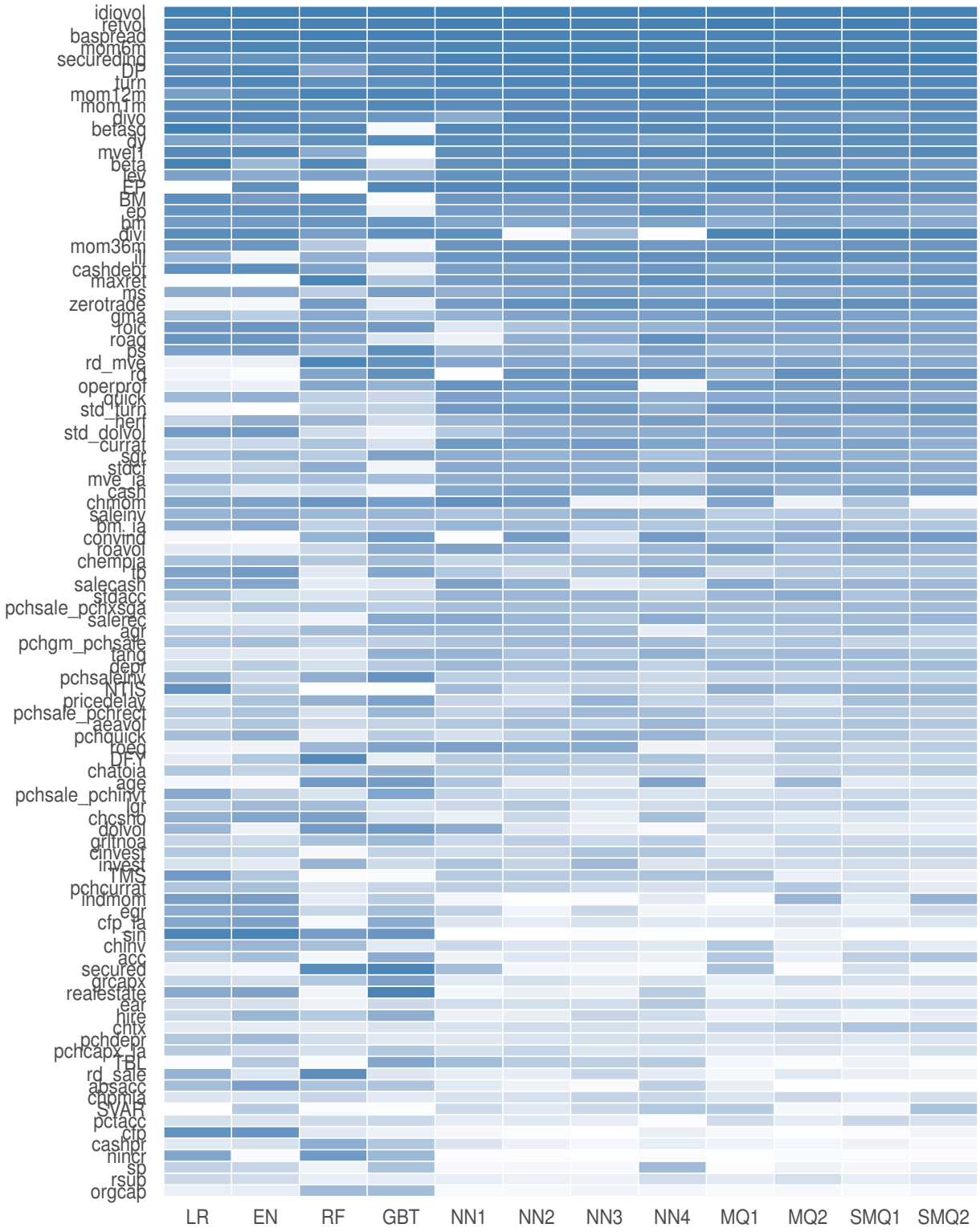
Panel D: Variable Importance Plot for $\tau = 0.75$



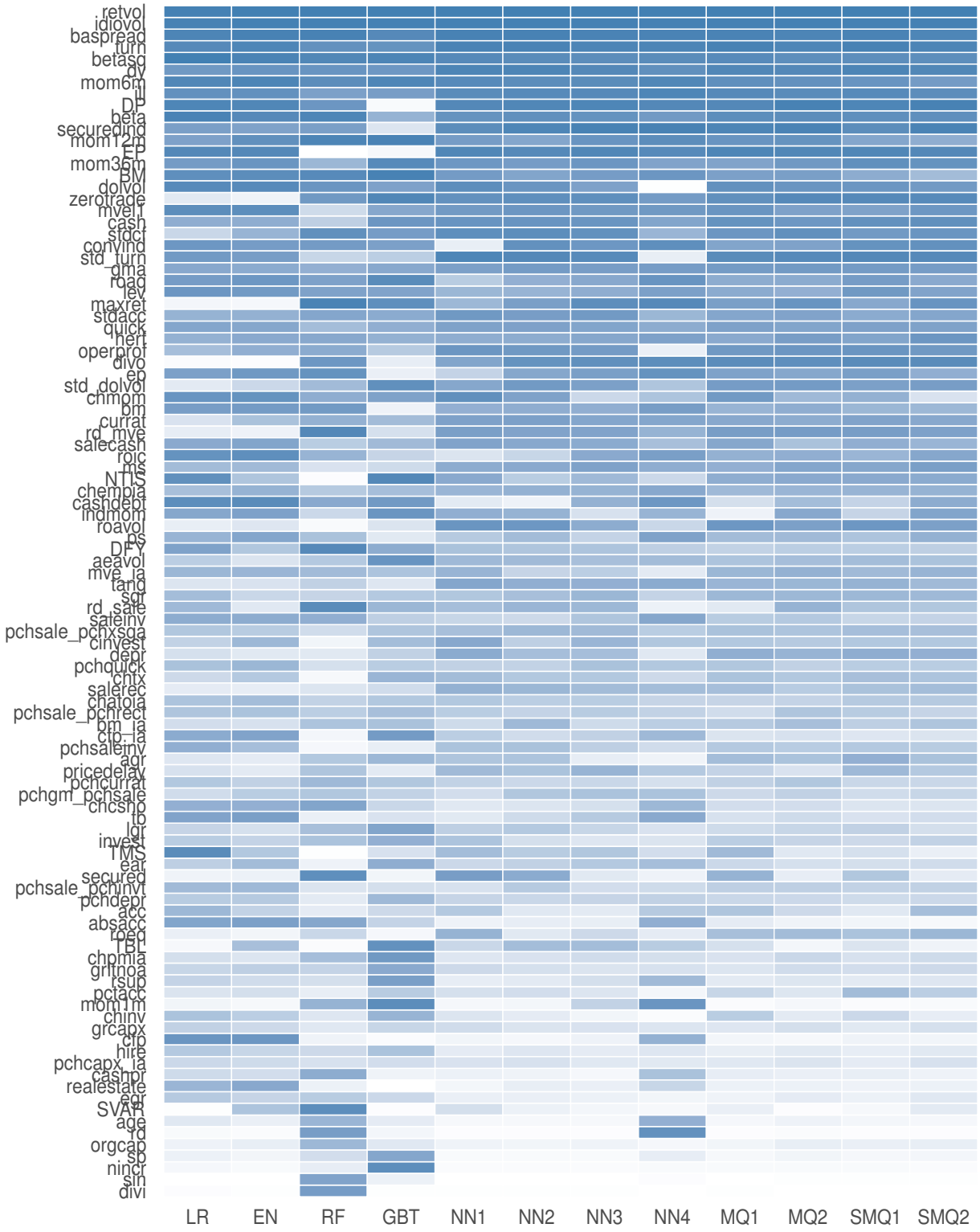
Panel E: Variable Importance Plot for $\tau = 0.95$

Figure IA3: Variable Importance Ranking by Model for each Quantile

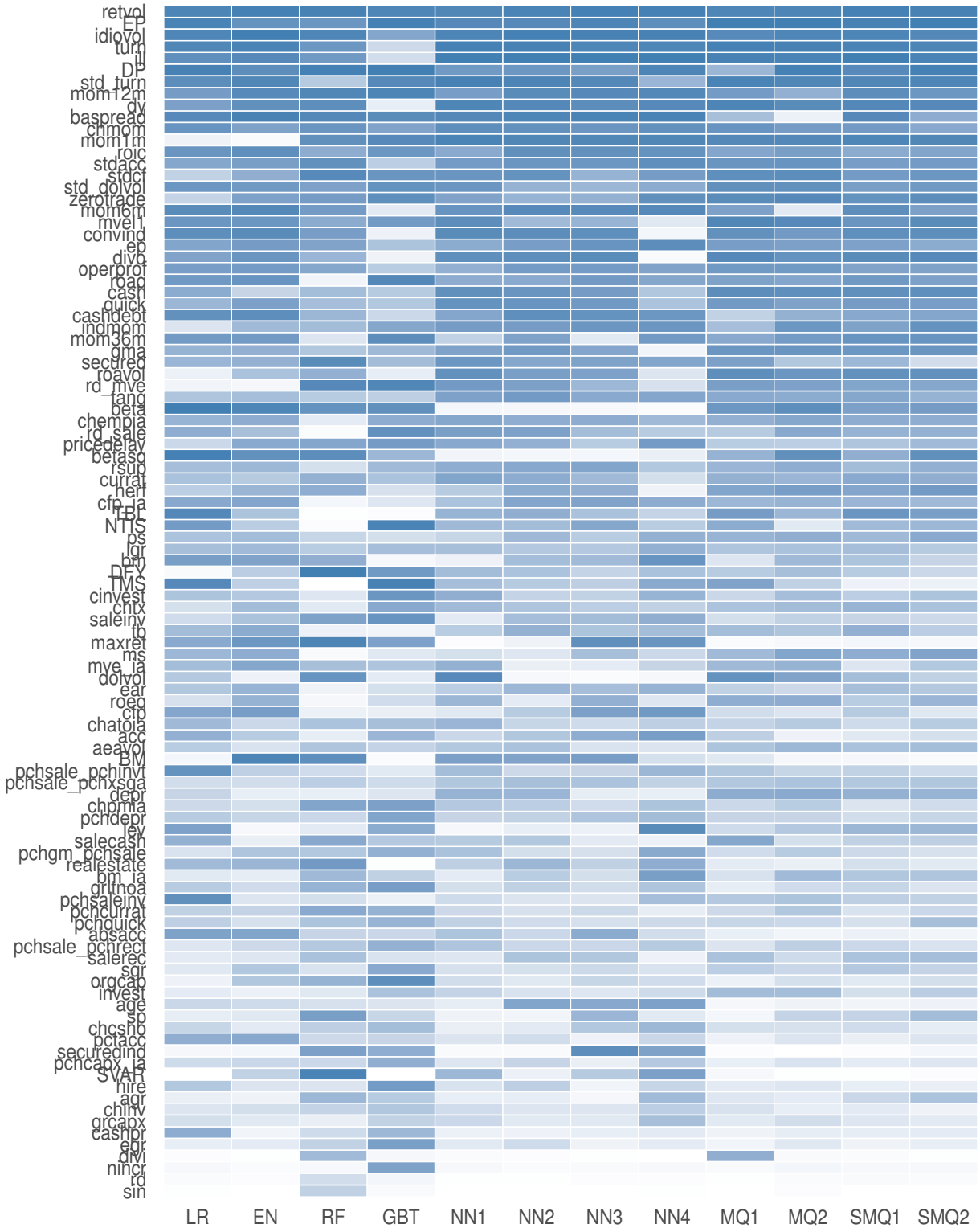
This figure reports the overall rankings of predictors for all models. For each model and each τ ($\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$), the variable importance of the j^{th} predictor is measured by calculating the reduction in $R^1_{os}(\tau)$ from setting all values of the j^{th} predictor to zero within each test sample, and then averaging the $R^1_{os}(\tau)$ reductions across test samples as a single importance measure. We then rank the importance of each variable for each model, and then sum their ranks. Variables are ordered so that the highest ranks are on top and lowest are on the bottom. The color gradient within each column indicates the most influential (dark) to least influential (white) variables.



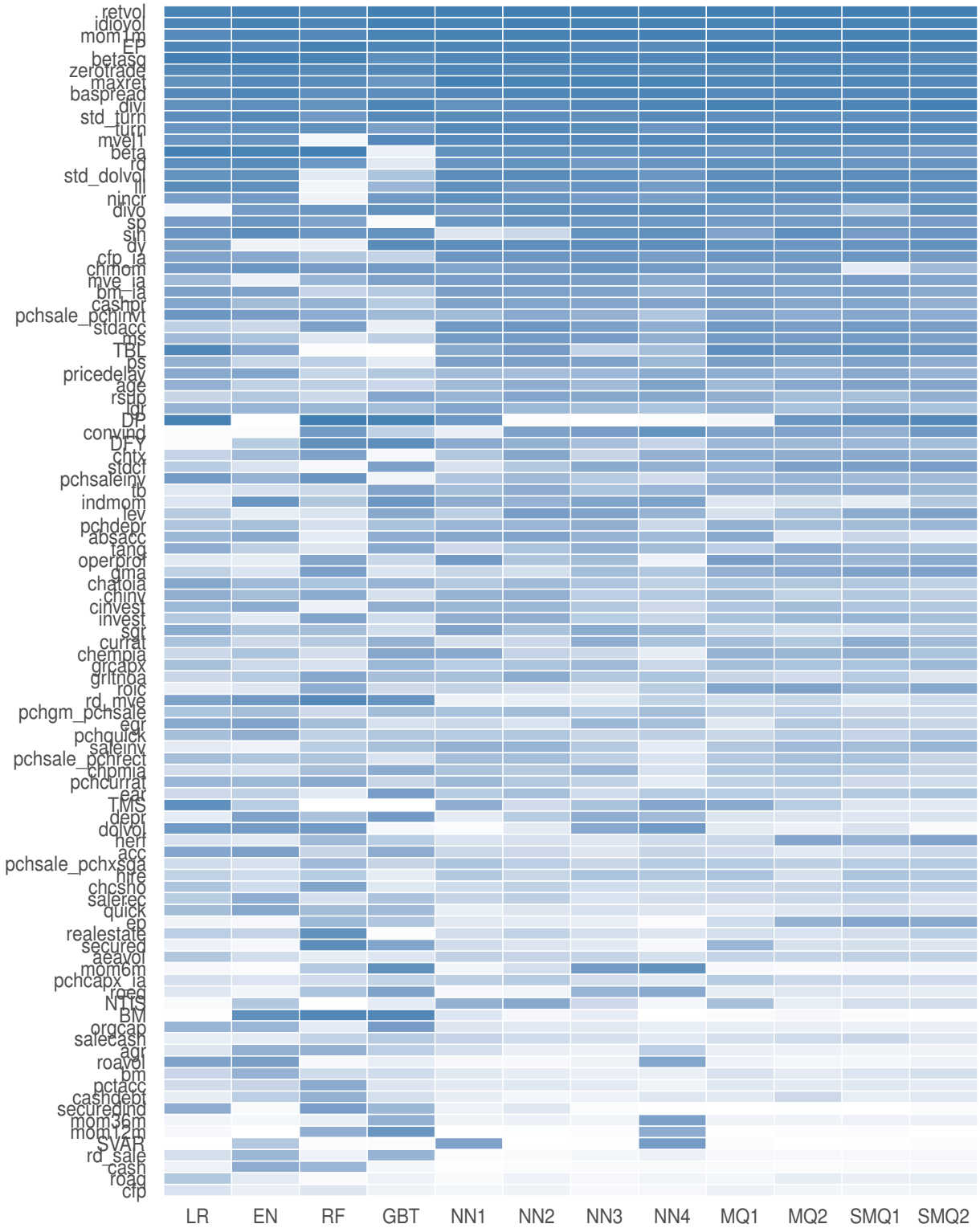
Panel A: Variable Importance Ranking for $\tau = 0.05$



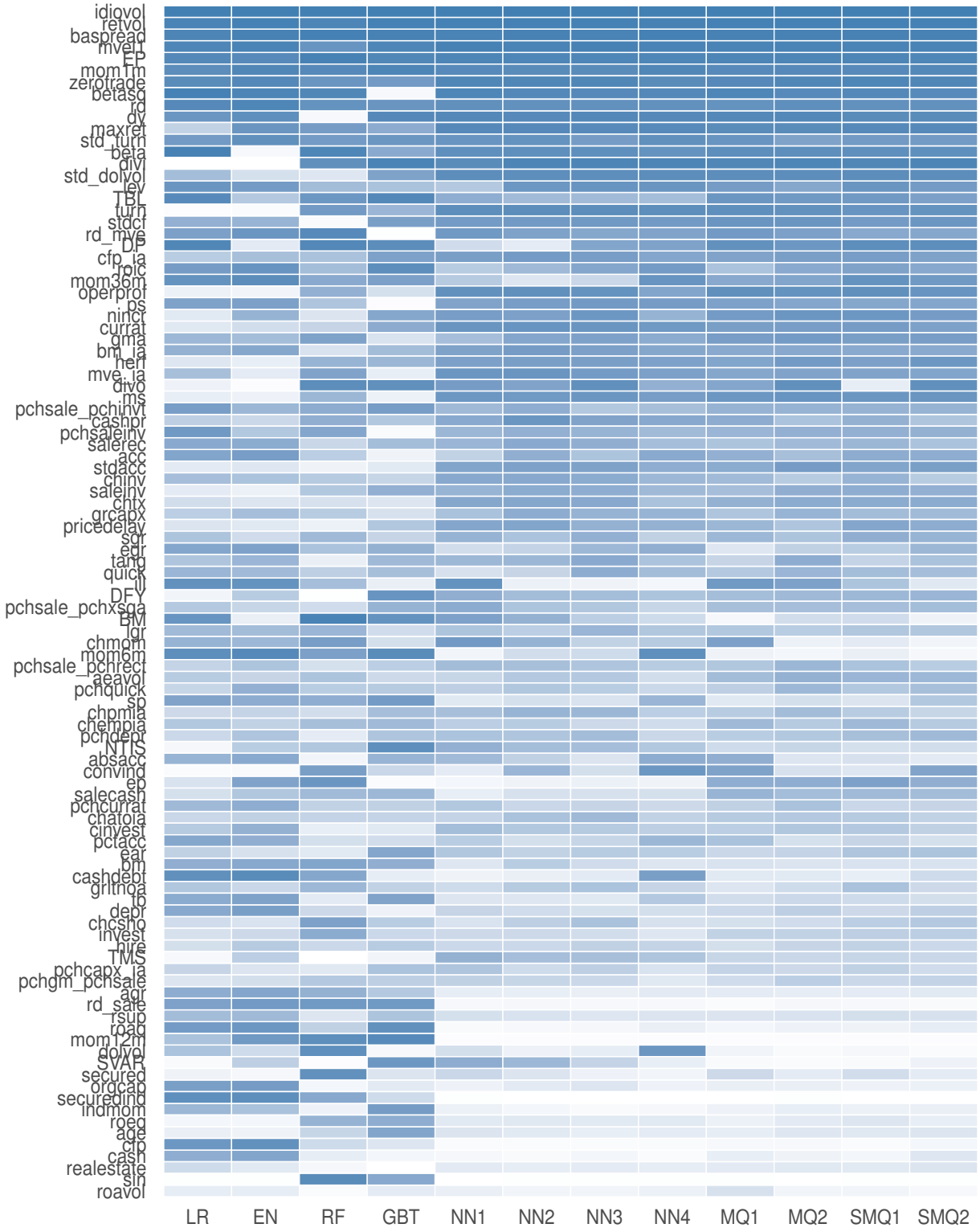
Panel B: Variable Importance Ranking for $\tau = 0.25$



Panel C: Variable Importance Ranking for $\tau = 0.50$



Panel D: Variable Importance Ranking for $\tau = 0.75$



Panel E: Variable Importance Ranking for $\tau = 0.95$