Quantile Machine Learning and the Cross-Section of Stock Returns: Robust Risk Premia and Machine Learning Risk Forecasts

Fred Liu[*†‡]

August 30, 2024

Abstract

We investigate the predictability of future stock return quantiles using machine learning models trained on firm characteristics and macroeconomic variables, and find that multi-task neural networks dominate linear, tree-based, and feed-forward neural network models. We introduce a quantile-based risk premium measure, robust to outliers and heteroskedasticity, and demonstrate that it delivers significant predictive and economic gains for investors, outperforming leading machine learning techniques. Our approach extends to predicting volatility, skewness, and Value-at-Risk, where we uncover distinct asymmetries between downside risk versus upside potential in their relation to future returns. We also demonstrate that a portfolio strategy incorporating machine learning Sharpe ratio forecasts leads to significant economic gains. Finally, we provide compelling evidence that return distribution predictability is driven by underreaction to information shocks, investor inattention, and information frictions.

JEL Classification: C21, C45, C52, C55, C58, G0, G1, G17

Keywords: Machine Learning, Quantile Regression, Robust Risk Premia, Risk Forecasting, Multi-task Neural Networks

[*]We benefited tremendously from discussions with Thibaut Duprey, Bryan Kelly, Semyon Malamud, Alex Maynard, Ilias Tsiakas, Qinfeng Xu, and seminar and conference participants at the 2024 Society of Financial Econometrics Summer School organized by NYU Shanghai, International Risk Management Conference at Bocconi 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.

[†]Assistant Professor of Finance, Department of Economics and Finance, University of Guelph, 50 Stone Rd E, Guelph, Ontario, N1G2W1, 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, since they provide valuable information about a stock return's conditional distribution. The conditional distribution is a cornerstone in numerous asset pricing theories, including risk aversion (Markowitz (1952), Sharpe (1965)), downside risk (Markowitz (1959), Roy (1952)), and tail risk (Barro (2006), Gabaix (2012)). Furthermore, the predicted quantiles can be leveraged to forecast important economic measures, including volatility, skewness, tail risk, and the risk premium, all benefiting from the inherent stability of quantile-based measures, making them robust to outliers.[1] However, predicting stock return quantiles is notoriously difficult, because financial returns exhibit heavy tails, heteroskedasticity; and may depend on many covariates and their complex nonlinear interactions. This paper underscores the statistical and economic significance of using quantile machine learning to predict the conditional distribution of stock excess returns, and investigates its economic source, paving the way for a deeper dive into the economic mechanisms of asset pricing.

Our primary contributions are fivefold. First, we provide a new set of benchmarks for the predictive accuracy of machine learning models in forecasting the conditional quantiles of individual stocks. Second, we introduce a robust risk premium measure by integrating over quantiles, and demonstrate substantial improvements to the statistical and economic significance of machine learning predictions. A multi-task neural network that incorporates economic structure by jointly estimating quantiles produces robust risk premium forecasts that enjoy an out-of-sample predictive R^2 of 1.16% and a value-weighted long-short decile spread with an out-of-sample Sharpe ratio of 1.71, outperforming the best methods in Gu, Kelly, and Xiu (2020) (henceforth GKX) by 190% and 27%, respectively. Third, we extend our quantile machine learning approach to predict risk measures such as volatility, skewness, and Value-at-Risk (VaR); and find an asymmetric relationship between risks and

[1] In alignment with the academic finance literature, we characterize the risk premium of a stock as its conditional expected return in excess of the risk-free rate.

future returns, with downside risk negatively related to future returns and upside potential positively related, which challenges conventional economic theories. Fourth, by incorporating risk aversion into our portfolio strategy through Sharpe ratio forecasts, we elevate the value-weighted long-short Sharpe ratio up to 2.51. Fifth, we provide a behavioral explanation for our results by showing that return distribution predictability is more pronounced in firms that experience larger information shocks, receive less attention, or encounter greater information frictions.

Most quantile regression models in the traditional econometrics literature are linear and unable to accommodate a large number of predictors with complex nonlinear interactions. This limitation hinders those models' predictive accuracy. The finance 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 can 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 nonlinearity of return quantiles and the high-dimensional problems related to using many firm characteristics and macroeconomic variables, machine learning is well suited for such challenging prediction problems through modeling nonlinear relationships with interactions while guarding against overfitting on a high-dimensional set of predictors, and with a capability for variable selection.

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, in general, rely on unregularized linear specifications, which can underfit the true functional form of the 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 flex-

ibility is essential for approximating 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 linear and nonlinear models to predict quantile τ ($\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$) of stock returns. 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 networks with 1-4 hidden layers (NN1,...,NN4). We also introduce an economically motivated multi-task neural network (MT) architecture. Unlike feed-forward neural networks that independently estimate models for each quantile, our MT approach jointly estimates multiple quantiles within a single functional form, incorporating economic structure by sharing common latent features across quantiles. This multi-task learning approach not only acts as a regularizer to prevent overfitting to quantile-specific noise, but also increases data efficiency through simultaneous learning of related tasks. We include MT architectures with one to three shared layers (MT1,...,MT3).

We assess the predictive power of the different models using the $R^1_{os}(\tau)$ statistic from Koenker and Machado (1999), which is a pseudo-R^2 measure that benchmarks the model predictions against a naive forecast of the historical quantile. First, we show that the traditional unregularized LR fails to deliver out-of-sample forecasting power for conditional quantiles, generally producing the lowest $R^1_{os}(\tau)$ values across all quantiles, however, the penalized linear model, EN, improves the $R^1_{os}(\tau)$ values across all quantiles. We further demonstrate that the inclusion of nonlinear interactions with RF does not enhance predictive power, as it estimates quantiles without the quantile objective function. In contrast, GBT, which is estimated using the quantile objective function, improves the out-of-sample $R^1_{os}(\tau)$ values across all quantiles over RF. Feed-forward neural networks increase the $R^1_{os}(\tau)$ values across all quantiles relative to tree-based and linear models. Multi-task neural networks are the best performers, and the MT1 and MT2 models achieve the highest $R^1_{os}(\tau)$

values across all quantiles. Finally, we make pairwise comparisons of the predictive accuracy of different models using the Diebold-Mariano test statistic. We show that MT1 and MT2 statistically significantly outperform all other models (with the exception of MT3) across all quantiles.

We introduce a robust measure of the risk premium by integrating over quantiles. The robust risk premium is less sensitive to outliers and accommodates heteroskedasticity. We first mathematically prove that the robust risk premium converges to the expected value under relatively weak assumptions. We then empirically demonstrate that the robust risk premium improves the R^2_{OS} relative to GKX for most models, reaching R^2_{OS} values of 1.16% and 1.1% for MT1 and MT2, respectively. 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 (decile 10) and sells stocks with the lowest (decile 1). We find that our models generally earn high Sharpe ratios, up to 1.71 for MT1. Feed-forward neural networks are the next best models with Sharpe ratios of up to 1.36 for NN4. Tree-based and linear models underperform with a maximum Sharpe ratio of 0.91 for both the EN and RF models. Furthermore, MT1 earns a monthly alpha of 2.65% (t-stat of 9.38) and information ratio of 1.66 with respect to the Fama-French five-factor model (Fama and French (2015)) augmented to include the momentum factor, and a maximum drawdown of only 27.92%. Additionally, MT1 earns a notable break-even transaction cost of 1.83% and maintains its economic significance after the exclusion of microcap stocks; which shows that the model can deliver substantial risk-adjusted returns in the presence of realistic economic constraints (Avramov, Cheng, and Metzker (2022)).

Our quantile machine learning approaches shed new light on the cross-sectional relationship between conditional risk measures and expected returns, as we extend our quantile machine learning approach to predict conditional volatility, skewness, and tail risk. Unlike moment-based risk measures, which are sensitive to outliers, our robust risk measures

are insensitive to the heavy tails of stock returns (Fama (1963), Fama (1965), Mandelbrot (1963)). Moving beyond prevailing risk measures in the existing literature that typically only condition on just historical returns or option prices, our machine learning risk measures condition on a large set of firm characteristics and macroeconomic variables, which better mirrors the information set available to investors. These machine learning risk measures also incorporate complex functional forms, reflecting the sophisticated strategies and intuitions that investors exercise in risk management

We form portfolios based on the robust risk measure forecasts and find no consistent relationship between conditional volatility and average returns. Inspired by the asymmetric investor preferences towards downside risk versus upside potential (Kahneman and Tversky (1979)), we decompose volatility into its downside and upside components, corresponding to the negative and positive regions of the conditional distribution of returns, respectively. We show that upside volatility has a significant positive relationship with average returns. In contrast, downside volatility has a significant negative relationship with average returns, which is consistent with the empirical evidence in Ang, Hodrick, Xing, and Zhang (2006). Notably, the long-short portfolio sorted on upside volatility delivers an alpha of 1.42% (t-stat of 3.42) for MT1, while the long-short portfolio sorted on downside volatility earns an alpha of -1.49% (t-stat of -4.51). These results suggest that asymmetries are necessary for understanding the relationship between volatility and expected returns. Both observed relationships contradict the traditional risk-return trade-off, which posits that stocks with higher upside potential should yield lower expected returns, whereas those with higher downside risk should generate higher expected returns.

We further study the role of asymmetries by forming portfolios based on conditional skewness—a measure of return asymmetry. Across nearly all models, we find a significant positive relationship between conditional skewness and average returns. Notably, a long-short portfolio sorted on conditional skewness delivers an alpha of 0.96% (t-stat of 3.62) for MT1. These results challenge several economic theories, such as the risk aversion (Harvey

and Siddique (2000), Kraus and Litzenberger (1976)) and lottery preferences (Barberis and Huang (2008), Brunnermeier, Gollier, and Parker (2007), Mitton and Vorkink (2007)), which predict a negative relationship.

Next, we explore the role of asymmetries for tails risks by forming portfolios sorted on VaR, a common measure of tail risk. Across all models, we uncover a significant negative relationship between conditional tail risk and average returns, contrasting with the risk-return tradeoff. Notably, the long-short portfolio sorted on VaR earns an alpha of -1.18% (t-stat of -3.54) for MT1. These results align with the left-tail momentum documented in Atilgan, Bali, Demirtas, and Gunaydin (2020) and support the economic theory suggesting that investors underreact to adverse information (Hong and Stein (1999)). In contrast, the right tail has no significant relationship with average returns, further spotlighting the asymmetries between upside potential versus downside risk.

Moreover, motivated by investor risk aversion, we consider a portfolio strategy that incorporates conditional volatility forecasts. First, we introduce a quantile-based measure of risk-adjusted returns, denoted the robust Sharpe ratio, which is designed to be insensitive to outliers. Second, we empirically demonstrate that the robust Sharpe ratio predicts the moment-based Sharpe ratio, validating its effectiveness as a robust measure of risk-adjusted returns. Third, we implement a Sharpe-Managed portfolio strategy, inspired by Allena (2021), that buys stocks with the highest robust risk premium and highest robust Sharpe ratio forecasts, and sells those with the lowest robust risk premium and lowest robust Sharpe ratio forecasts. Implementing this portfolio strategy enhances the realized Sharpe ratio across most models, and the MT models achieve Sharpe ratios between 2.25 and 2.51, significantly surpassing the peak Sharpe ratio of 1.51 earned by a comparable volatility-based portfolio strategy in Allena (2021). These results underscore the substantial economic gains derived from incorporating risk aversion (Markowitz (1952), Sharpe (1965)) as a secondary criterion in portfolio strategies.

Finally, we investigate potential sources of return distribution predictability. Given that

empirical observations diverge from traditional risk-return theories, we explore behavioral explanations. Inspired by Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023), we hypothesize that investors tend to underreact to new information, leading to short-run return continuations and return distribution predictability, as implied by the theoretical framework in Hong and Stein (1999). Consistent with our hypothesis, we find that return distribution predictability is more pronounced for stocks that experience larger information shocks, receive less attention, or encounter greater information frictions. For example, a long-short portfolio sorted on the robust risk premium earns a Sharpe ratio of 0.98 in the quintile with the lowest firm-level information frictions, which monotonically increases to 3.1 in the quintile experiencing the highest information frictions. These results align with the notion that machine learning models are effective in identifying misvaluations influenced by investor behavioral biases.

We further study the sources of return distribution predictability by identifying the most influential variables for each stock return quantile. The set of important variables varies across quantiles, but is in close agreement between models for each quantile. Volatility, liquidity, and price trends are the most influential characteristics for the 0.05, 0.25, 0.75, and 0.95 quantiles, which further supports information frictions as a key driver of return distribution predictability. For the median, highly influential variables include momentum, valuation ratios, and fundamental signals. The aggregate dividend-price ratio is the most influential macroeconomic variable for downside quantiles, and the earnings-price ratio is the most influential macroeconomic variable for upside quantiles. These results indicate that different predictors are important for forecasting different parts of the future return distribution.

Our paper develops quantile machine learning techniques for the cross-section of stock returns, and contributes to at least three existing strands of literature. First, our paper contributes to the rapidly expanding literature applying machine learning techniques to forecast stock returns, and it is closest to the seminal paper by Gu, Kelly, and Xiu (2020), who

apply machine learning techniques to forecast the *conditional expectation* of stock returns (see also, e.g., Aleti, Bollerslev, and Siggaard (2023), Chinco, Clark-Joseph, and Ye (2019), Choi, Jiang, and Zhang (2023), Freyberger, Neuhierl, and Weber (2020), Huddleston, Liu, and Stentoft (2023), Leippold, Wang, and Zhou (2022), and Liu and Stentoft (2023)). In contrast, we apply 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, robust to outliers and heteroskedasticity, and use machine learning to predict volatility, skewness, and tail risk.

Furthermore, several recent papers introduce machine learning models with economic restrictions to improve forecasts. Aleti and Bollerslev (2024), Bryzgalova, Pelger, and Zhu (2023), Chen, Pelger, and Zhu (2023), Feng, He, Polson, and Xu (2023), Gu, Kelly, and Xiu (2021), 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. This paper develops multi-task neural networks that incorporate an economic restriction based on a common structure across return quantiles, which are used to predict the *conditional distribution* of stock returns, leading to significant statistical improvements and economic gains compared with models without economic 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 (2021). However, Allena (2021) 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 (2023) extend the autoencoder model

of Gu, Kelly, and Xiu (2021) to estimate latent factors and quantile-specific betas. Their paper focuses on using autoencoders to predict quantile-based approximations of the conditional expectation and variance, while this paper compares a variety of machine learning techniques to study the cross-sectional predictability of stock return quantiles. In addition, their paper only conducts equal-weighted portfolio sorts with fixed transaction costs, while this paper demonstrates the economic gains of quantile machine learning on both equal- and value-weighted portfolios, accounting for economic constraints as advocated by Avramov, Cheng, and Metzker (2022). Furthermore, we also study the relationship between machine learning risk measures and expected returns, and demonstrate that these relationships are influenced by behavioral biases.

Second, our empirical study contributes to the literature on conditional risk measures and the cross-section of expected returns (see Bali, Engle, and Murray (2016) for a survey). Our robust volatility and skewness measures closely resemble the quantile-based risk measures in Ghysels (2014) and Ghysels, Plazzi, and Valkanov (2016). Moreover, our tail risk measure is similar to the VaR measure in Atilgan, Bali, Demirtas, and Gunaydin (2020). However, rather than forecasting quantiles solely based on past daily returns, our machine learning methodology conditions on an extensive set of firm characteristics and macroeconomic variables, mirroring the comprehensive information set available to investors. This enhanced approach uncovers marked asymmetries between downside risk and upside potential in relation to future returns. Building on the findings of Atilgan, Bali, Demirtas, and Gunaydin (2020), we observe that the anomalous relationship between VaR and expected returns has a behavioral explanation. However, we extend their analysis to further elucidate the anomalous relationship between downside volatility, skewness, and the robust risk premium with expected returns.

Third, our paper is related in spirit to the literature on quantile regression in empirical finance. Beginning with Koenker and Bassett Jr (1978), there has been an extensive literature on applying quantile linear regression in economics (see Koenker, Chernozhukov,

He, and Peng (2017) for a survey). Recently, there has been a growing literature on applying quantile linear regression in empirical finance. Giglio, Kelly, and Pruitt (2016) use a quantile regression approach to assess how well various systemic risk measures proposed in the literature predict real economic activity. Menkveld, Dreber, Holzmeister, Huber, Johannesson, Kirchler, Neusüss, Razen, and Weitzel (2023) use quantile regression and the interquartile range to estimate non-standard errors. On the Macro-Finance side, Adrian, Boyarchenko, and Giannone (2019) find that the lower conditional quantiles of future GDP growth in the US are sensitive to current financial conditions. Chen, Dolado, and Gonzalo (2021) use a quantile factor model to show that the common factors driving the median of stock returns are different from those driving the tails. 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. Our paper complements the quantile regression literature by using quantile machine learning techniques to model cross-sectional stock returns as a nonlinear function of firm characteristics and macroeconomic variables, including introducing a neural network model that incorporates economic structure through multi-task learning.

The paper is organized as follows: Section 2 introduces the methodological framework, machine learning models, and quantile-based measures, Section 3 discusses the approach to evaluate model performance and the data, Section 4 presents the main empirical results, Section 5 investigates the sources of return distribution predictability, and Section 6 summarizes our conclusions.

2 Methodology

We index stocks as $i = 1, ..., N$ 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:

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

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 $Q_{r_{i,t+1}}$ to a P-dimensional vector $z_{i,t}$ of predictors.

2.1 Estimation

Unlike most regression-based machine learning models, which rely on the mean squared error (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 $\tau = 0.5$, the objective function (2) is the symmetric mean absolute error (MAE), and is minimized by the conditional median. For $\tau > 0.5$ ($\tau < 0.5$), objective function (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.

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. Objective function (2) has a built-in robustness to heavy-tailed observations, since outliers have minimal influence on quantiles. 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)). This heteroskedas-

ticity is not captured by the MSE objective function. In contrast, models using the quantile objective function are designed to account for such heteroskedasticity (Koenker and Bassett Jr (1982)), making them more suitable for handling the heterogeneous volatility of financial returns.

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 linear quantile regression (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 elastic net penalty (EN) to objective function (2). To model nonlinearities and interaction effects, we consider quantile gradient boosted regression trees (GBT) and deep quantile feed-forward neural networks with 1-4 layers (NN1,...,NN4). These nonlinear techniques are analogous to models used in GKX, but are optimized to the quantile objective function rather than MSE.[2] We also consider quantile random forest (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 models.

The models discussed so far assume that the functional forms of each quantile 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. Multi-task neural network (MT) architectures, depicted in Figure 1, are designed to estimate all targeted quantiles within a single functional form. The "hard sharing" hidden layers within the MT interact and nonlinearly transform the input predictors into shared latent features, imposing an economic restriction of a common structure across quantiles. The MT also includes "task-specific"

[2]GKX estimates some models using the Huber loss, a hybrid loss function that augments MSE with MAE for outliers.

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

hidden layers to capture each quantile's nonlinear exposures to the shared latent features, before aggregating these into a prediction for each quantile. MTs enhance generalization by implicitly improving regularization and data efficiency. We consider MT architectures featuring up to three hard sharing layers (MT1,...,MT3). A detailed description of these methods is available in Internet Appendix IA1.2.4.

2.2 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 premiums 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 face of heavy-tailed stock returns (Mikosch and de Vries (2013)). Additionally, models estimated with MSE are incapable of accommodating the heteroskedasticity in stock returns, which can further increase 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 that is an integral over the quantiles, defined as

$$\begin{aligned}\bar{Q} &= (0.05 + \frac{0.25-0.05}{2})Q_{0.05} + (\frac{0.25-0.05}{2} + \frac{0.5-0.25}{2})Q_{0.25} + \\ &(\frac{0.5-0.25}{2} + \frac{0.75-0.5}{2})Q_{0.5} + (\frac{0.75-0.5}{2} + \frac{0.95-0.75}{2})Q_{0.75} + (\frac{0.95-0.75}{2} + (1-0.95))Q_{0.95} \\ &= 0.15Q_{0.05} + 0.225Q_{0.25} + 0.25Q_{0.5} + 0.225Q_{0.75} + 0.15Q_{0.95},\end{aligned} \tag{3}$$

where $Q_\tau \equiv Q_{r_{i,t+1}}(\tau|\mathcal{F}_t)$ is from Equation (1), and 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 2 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 measure can be expressed as

$$\begin{aligned}\bar{Q}_K = &(\tau_1 + \frac{\tau_2 - \tau_1}{2})Q_{\tau_1} + (\frac{\tau_2 - \tau_1}{2} + \frac{\tau_3 - \tau_2}{2})Q_{\tau_2} + \cdots + \\ &(\frac{\tau_{K-1} - \tau_{K-2}}{2} + \frac{\tau_K - \tau_{K-1}}{2})Q_{\tau_{K-1}} + (\frac{\tau_K - \tau_{K-1}}{2} + (1 - \tau_K))Q_{\tau_K},\end{aligned} \tag{4}$$

which converges to the conditional expectation of the respective distribution under general assumptions, as mathematically proven in Section IA2 in the Internet Appendix.[4] Therefore, the robust risk premium measure is also an estimator of the conditional expectation of excess returns, i.e., the risk premium.

Mathematically, the robust risk premium $\bar{Q}$ can be viewed as a data discretization approach that estimates 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.[5] Statistically, this measure serves as an ensemble

[4]The robust statistics literature typically uses quantile-based estimates of the expected value that rely on strong distributional assumptions such as normality (see Wan, Wang, Liu, and Tong (2014) for a comprehensive survey), which is inappropriate for the heavy-tailed distribution of stock returns. In contrast, Equation (4)'s convergence to the conditional expectation relies on much weaker distributional assumptions, thus the measure is more robust to model misspecification. In unreported results, we found that our weighting scheme significantly outperforms those in Wan, Wang, Liu, and Tong (2014), both statistically and economically, indicating that their distributional assumption of normality is inappropriate for stock returns. These results are available upon request.

[5]Our approach estimates the expected value of a probability distribution using a discrete distribution function, which is similar in spirit to the data discretization approach in Yang, Zhu, Li, and Zhu (2023). However, using their discretization method, based on the minimized 1-Wasserstein metric (Dabney, Ostrovski, Silver, and Munos (2018)), the set of weights for our chosen quantiles would be $\{0.1, 0.3, 0.2, 0.3, 0.1\}$. As a simple numerical comparison, consider the Skewed Student's t distribution (Hansen (1994)), depicted in Figure 2, with 15 degrees of freedom, 0.2 skewness parameter, 1 variance, and 0 expected value. Using the true quantiles, Yang, Zhu, Li, and Zhu (2023)'s weights would produce -0.012 (calculated by $0.1\times(-1.51)+0.3\times(-0.68)+0.2\times(-0.07)+0.3\times0.61+0.1\times1.74$) as the estimated expected value, while our weights would yield 0.00125 (calculated by $0.15\times(-1.51)+0.225\times(-0.68)+0.25\times(-0.07)+0.225\times0.61+0.15\times1.74$), which

technique, reducing variance by integrating various distribution predictions. Economically, this approach amalgamates valuable signals from various segments of the conditional return distribution.

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. The robust risk premium also accounts for heteroskedasticity in stock returns.

While the robust risk premium measure reduces variance, it is, unlike 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. Our empirical results in Section 4.2 show that the robust risk premium substantially outperforms the MSE objective function 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.

2.3 Conditional Volatility, Skewness, and Value-at-Risk Measures

The predictive relationship between stock returns and risk measures such as volatility, skewness, and tail risk remains a central research area in financial economics. While recent studies affirm a connection between these risk measures and future stock returns, empirical evidence offers divergent views regarding the direction of these relationships.

Prominently, Ang, Hodrick, Xing, and Zhang (2006) discover that stocks with higher idiosyncratic volatilities exhibit lower average returns, which implies a negative trade-off between risk and returns. However, economic theory recognizes that investors care differently about downside losses than upside gains due to loss aversion (Kahneman and Tversky

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. We also empirically find that our weighting scheme considerably outperforms theirs both statistically and economically (see Table IA2).

(1979)) and disappointment aversion (Gul (1991)). Bolstering this viewpoint, recent literature underscores that upside and downside volatilities bear distinct associations with future returns (Bollerslev, Li, and Zhao (2020), Farago and Tédongap (2018)).

The empirical relationship between skewness and average stock returns has elicited divergent findings in academic literature. On one hand, studies by Amaya, Christoffersen, Jacobs, and Vasquez (2015), Boyer, Mitton, and Vorkink (2010), Conrad, Dittmar, and Ghysels (2013), and Harvey and Siddique (2000) identify a negative relationship, a finding that aligns with economic theories such as risk aversion and lottery preferences. In contrast, papers by Bali, Hu, and Murray (2019), Ghysels, Plazzi, and Valkanov (2016), Rehman and Vilkov (2012), and Xing, Zhang, and Zhao (2010) reveal a positive relationship.

Recent papers provide conflicting empirical evidence about the relationship between tail risk and average returns. Bégin, Dorion, and Gauthier (2020), Huang, Liu, Rhee, and Wu (2012), Kapadia and Zekhnini (2019), Kelly and Jiang (2014), and Liu (2020) show that tail risk is positively associated with average returns, consistent with the risk-return tradeoff. However, Atilgan, Bali, Demirtas, and Gunaydin (2020) show that tail risk is negatively associated with average returns, which can be explained by investor underreaction.

One potential reason for the conflicting empirical results is that the literature typically constructs risk measures using linear models and conditions on a limited set of variables, such as past stock returns or option prices, which limits their predictive power. This section constructs machine learning risk measures, which condition on a broader set of firm-level characteristics and macroeconomic variables, mirroring the rich information accessible to investors. These machine learning risk measures also incorporate complex functional forms, capturing the nuanced thought processes and intuitions that investors exercise throughout their risk management processes.

To address the significant challenge of measuring higher order moments due to their sensitivity to outliers (Kim and White (2004)), we construct firm-level quantile-based risk measures that are insensitive to outliers. For a robust measure of conditional volatility, we

employ the widely used interquartile range as follows:

$$Vol_{i,t+1} = Q_{0.75} - Q_{0.25}, \tag{5}$$

where $Q_\tau \equiv Q_{r_{i,t+1}}(\tau|\mathcal{F}_t)$. The interquartile range assesses the conditional dispersion of returns, excluding the tails, with higher values indicating larger fluctuations of returns. We use the interquartile range instead of the standard deviation, since the latter tends to be dominated by the size of extreme outliers.[6] Furthermore, the interquartile range can be viewed as a scaled version of the standard deviation under certain conditions. For example, in a Gaussian distribution, the interquartile range approximately equals to 1.35 times the standard deviation.

We also decompose the volatility measure into separate components, associated with the positive and negative regions of the conditional distribution of returns:

$$Vol^+_{i,t+1} = Q_{0.75}, \qquad Vol^-_{i,t+1} = -Q_{0.25}, \tag{6}$$

which correspond to upside and downside volatility, respectively. By defining $Vol^-_{i,t+1} = -Q_{0.25}$, we multiply $Q_{0.25}$ by -1 so that higher values of $Vol^-_{i,t+1}$ correspond to higher levels of downside volatility. Evidently, the two volatility components add up to the total volatility, $Vol^+_{i,t+1} + Vol^-_{i,t+1} = Vol_{i,t+1}$.

Furthermore, we employ the median and inner 0.25 and 0.75 quantiles to construct a robust measure of conditional skewness (Bowley (1920)), as defined below:

$$Skew_{i,t+1} = \frac{(Q_{0.75} - Q_{0.5}) - (Q_{0.5} - Q_{0.25})}{Q_{0.75} - Q_{0.25}}. \tag{7}$$

Conditional skewness measures return asymmetry, where a positive skewness indicates that a stock's upside potential exceeds its downside risk. We use a robust skewness measure

[6]For certain heavy-tailed distributions, such as a Student's t distribution with 2 degrees of freedom, the standard deviation is infinite, whereas the interquartile range remains finite.

instead of a moment-based skewness measure, since the latter is particularly sensitive to outliers, more so than the first two moments (Kim and White (2004)). Furthermore, the robust skewness measure approximates a scaled version of the conditional third moment under certain conditions (Ghysels, Plazzi, and Valkanov (2016)).

Finally, we construct a robust measure of conditional tail risk using Value-at-Risk at the 95% confidence level, defined as

$$VaR_{i,t+1} = -Q_{0.05}, \tag{8}$$

where we multiply $Q_{0.05}$ by -1 so that higher values of $VaR_{i,t+1}$ correspond to higher levels of left-tail risk. VaR is a common regulatory measure that quantifies how much returns could decline over a given time period with a given probability. For example, losses, i.e., negative returns, greater than $VaR_{i,t+1}$ are expected to occur less than 5% of the time during the next month for stock i.

For forecasting firm-level robust conditional risk measures, we replace each quantile with their corresponding machine learning forecasts, and then study the empirical relationship between these machine learning risk measure forecasts and future returns.

3 Data and Performance Evaluation

3.1 Data and Variable Definitions

We employ our quantile machine learning techniques on U.S. stocks, utilizing the same sample as GKX to enable a direct comparison. This sample comprises monthly total equity returns from CRSP for all firms listed on the NYSE, AMEX, and NASDAQ from 1957 to 2016, a span of 60 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 eight macroeconomic predictors from Welch and

Goyal (2008).[7] We substitute missing characteristics with their cross-sectional median, and rank all firm characteristics cross-sectionally for each period, subsequently mapping these ranks into the $[-1, 1]$ interval. The macroeconomic predictors include: 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).

3.2 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 $\mathcal{T}_3$ (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 30 out-of-sample years from 1987 - 2016. 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

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

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 over time into an overall assessment of each model, and it measures the proportional reduction in the quantile forecasting error (2) for

[7]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.

each model relative to a naive forecast of the historical quantile.

To make pairwise comparisons of models, 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 $t+1$,

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

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.

4 Empirical Results

4.1 The Cross-sectional Predictability of Conditional Quantiles

Table 1 presents the $R^1_{os}(\tau)$ values of quantile τ ($\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$) for the pooled test sample that spans from 1987 through 2016. We compare twelve models in total including LR with all covariates, LR-4 (which only includes size, book-to-market, momentum, and return volatility as covariates), EN, RF, GBT, feed-forward neural networks with one to four layers (NN1,...,NN4), and multi-task neural networks with one to three shared layers (MT1,...,MT3).

Beginning with linear models, Table 1 reports that the LR model with all covariates generally produces the lowest $R^1_{os}(\tau)$ values across all five quantiles. For instance, LR's $R^1_{os}(\tau)$ of -0.86% for the median implies that a naive forecast of the historical median across all stocks outperforms it. This outcome is not surprising given LR's susceptibility to

overfitting the training sample due to lack of regularization. However, limiting LR to a sparse parameterization by incorporating only four covariates (size, book-to-market, momentum, and return volatility) substantially boosts the $R^1_{os}(\tau)$ of the median to 0.36%. Alternatively, restricting LR to a sparse parameterization with the elastic net penalty considerably enhances the $R^1_{os}(\tau)$ values across all quantiles compared to LR, particularly for the 0.05 and 0.95 quantiles, which achieve $R^1_{os}(\tau)$ values of 12.87% and 10.29%, respectively.

For tree-based models, Table 1 reports that RF underperforms EN across most quantiles, likely because RF doesn't use the quantile objective function for estimation. GBT outperforms RF in every quantile, demonstrating that employing the quantile objective function improves predictions in tree-based models. This improvement is particularly notable at the 0.05 and 0.95 quantiles with $R^1_{os}(\tau)$ values of 13.69% and 12.78%, respectively.

Table 1 also reveals that feed-forward neural networks surpass both the linear and tree-based models in terms of $R^1_{os}(\tau)$ values across all quantiles. As observed in GKX, shallower networks perform better. The two-layered NN2 achieves an $R^1_{os}(\tau)$ of 0.61% for the median, more than doubling the $R^1_{os}(\tau)$ of 0.29% observed with the GBT model. Among feed-forward networks, NN1 has the highest $R^1_{os}(\tau)$ values for the 0.05 and 0.95 quantiles of 16.2% and 13.52%, respectively.

Multi-task neural networks emerge as the top performers. In particular, MT2 is the best model in three quantiles, with respective $R^1_{os}(\tau)$ values of 20.24%, 7.18%, 0.92%, 2.7%, and 16.21% for the 0.05, 0.25, 0.5, 0.75, and 0.95 quantiles. Furthermore, MT1 emerges as the top model in two quantiles, boasting respective $R^1_{os}(\tau)$ values of 20.37%, 7.16%, 0.91%, 2.79%, and 16.2%. This performance emphasizes the importance of incorporating economic structure and jointly estimating all quantiles within the same model. However, consistent with GKX, the benefits of deep learning seem to be limited, as MT3 underperforms MT1 and MT2. Nonetheless, MT3 outperforms the feed-forward neural networks for all quantiles, with the exception of the 0.75 quantile.

Comparing the results across quantiles, we see that the first and fifth rows of Table 1 have

the highest $R^1_{os}(\tau)$ values, demonstrating that the outer 0.05 and 0.95 quantiles are the most predictable. The second and fourth rows have lower $R^1_{os}(\tau)$ values, signifying that the inner 0.25 and 0.75 quantiles are less predictable than the outer quantiles. The third row displays the lowest $R^1_{os}(\tau)$ values, indicating that the median is the least predictable quantile. These observations are in line with a widely accepted feature of stock returns: volatility is typically more predictable than the mean. Moreover, downside quantiles (0.05 and 0.25) are more predictable than upside quantiles (0.95 and 0.75), which supports the notion that downside risks are more predictable than their upside counterparts (Bollerslev (2022)).

We now conduct pairwise comparisons of the statistical differences across models. Table 2 presents the Diebold-Mariano test statistics across the twelve models for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$. For each panel, a positive statistic indicates the column model outperforms the row model, and a bold number denotes significance at the 5% level for each test.

Panel A of Table 2 displays the Diebold-Mariano test statistics for the 0.05 quantile. First, the MT models produce statistically significant improvements over all other models, underscoring the importance of imposing economic structure in neural networks. Second, feed-forward neural networks significantly outperform the linear and RF models, suggesting superior performance of deep learning approaches for this quantile. Third, GBT significantly outperforms most of the linear and RF models, indicating the importance of modeling nonlinearities using the quantile objective function. The conclusions above are generally confirmed by Panel E of Table 2, which reports the Diebold-Mariano test statistics for the 0.95 quantile.

Panel B of Table 2 displays the Diebold-Mariano test statistics for the 0.25 quantile. Again, the MT models provide significant improvements over all other models, and NNs significantly outperform tree-based and linear models. These findings are largely confirmed by Panels C and D of Table 2, which report the Diebold-Mariano test statistics for the 0.5 and 0.75 quantiles. However, these panels contain fewer significant test statistics, reflecting the diminished predictability of these quantiles.

4.2 Predictive Power of the Robust Risk Premium

This section evaluates the predictive power of the robust risk premium from Equation (3) for individual stock returns. We are particularly interested in determining whether using the robust risk premium to forecast returns improves the predictive power relative to the methods in GKX.[8] We assess the predictive power of each model using the out-of-sample R-squared (R^2_{os}) metric from GKX:

$$R^2_{os} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \bar{Q}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{i,t+1}}. \tag{11}$$

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 3 presents a comparison of the R^2_{os} values of the twelve models. Starting with linear models, the robust risk premium forecasts for LR, LR-4, and EN achieve R^2_{os} values of -1.44%, 0.2%, and 0.34%, respectively, surpassing their counterparts in GKX, which have R^2_{os} values of -3.46%, 0.16%, and 0.11%, respectively. This suggests that utilizing the robust risk premium allows linear models to effectively reduce variance when outliers are present, resulting in superior performance compared to the conditional expectation-based models from GKX.

Moving on to tree-based methods, the RF model generates an R^2_{os} of 0.19%, due to its imprecise quantile forecasts. In contrast, GBT records an R^2_{os} of 0.49%, surpassing the R^2_{os} of 0.34% that the GBT model attains in GKX.

The robust risk premiums of neural networks NN1 through NN4 register R^2_{os} values ranging from 0.45% to 0.52%, outperforming the networks reported in GKX, which recorded R^2_{os} values between 0.33% and 0.4%. This is likely because of the detrimental impact of outliers on the performance of feed-forward neural networks estimated using the MSE objective func-

[8] We also replicated the conditional expectation-based models in GKX and found quantitatively similar results, thus we use their reported results as transparent benchmarks.

tion. MT1 and MT2 have the highest R^2_{os} values of 1.16% and 1.1%, respectively, more than doubling the best-performing model in GKX with an R^2_{os} of 0.4%. MT3 also outperforms feed-forward neural networks with an R^2_{os} of 0.68%.

This comparison emphasizes the superior predictive capabilities of robust risk premiums for predicting individual stock returns, compared to forecasts generated by models estimated with MSE. This improved predictive accuracy is likely due to the reduced model variance in the context of heavy-tailed and heteroskedastic stock returns.

Panel A of Table IA2 in the Internet Appendix presents the R^2_{os} values obtained using an alternative quantile-based risk premium measure, $\bar{Q}_{YZLZ} = 0.1Q_{0.05} + 0.3Q_{0.25} + 0.2Q_{0.5} + 0.3Q_{0.75} + 0.1Q_{0.95}$, based on the methodology of Yang, Zhu, Li, and Zhu (2023). Notably, the R^2_{os} values of this alternative approach underperform those reported in Panel A of Table 3 across all models, with the exceptions of LR and RF. Such a comparison illuminates the superior predictive power of the robust risk premium measure in Equation (3), demonstrating that it provides a more accurate estimate of the risk premium.

To conduct pairwise comparisons of statistical differences across models, we employ the Diebold-Mariano test for differences in out-of-sample predictive accuracy as outlined in GKX. We replace Equation (10) 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 $t+1$,

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

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 3 presents the Diebold-Mariano test statistics across the twelve models.

A positive statistic indicates the column model outperforms the row model, and a bold number denotes significance at the 5% level for each test. Notably, MT1 and MT2 statistically significantly outperform all other models except MT3. The NN and GBT models outperform RF and linear models, though the statistical significance is only evident over LR-4.

4.3 Machine Learning Portfolios Using the Robust Risk Premium

We now investigate whether the improved predictive accuracy of the robust risk premium forecasts translate into economic gains by forming portfolios based on the $\bar{Q}$ predictions of the twelve models. At the end of each month, we predict the one-month-ahead robust risk premium, and then sort stocks into deciles based on these forecasts. We reconstitute portfolios each month using value weights. Table 4 reports the monthly performance results, where "Low" denotes the portfolio with the lowest robust risk premium (decile 1), "High" denotes the portfolio with the highest robust risk premium (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 4 demonstrates that the out-of-sample portfolio performance mostly aligns with the R^2_{os} results reported in Section 4.2. Long-short portfolios constructed using linear model forecasts have monthly average returns from 0.95% to 1.25% and annualized Sharpe ratios between 0.5 and 0.91. Tree-based models, RF and GBT, display modest performances with Sharpe ratios of 0.91 and 0.76, respectively. Feed-forward neural networks significantly outperform linear and tree-based models. Long-short portfolios constructed using feed-forward neural network forecasts achieve average returns between 2.12% and 2.48%, and Sharpe ratios from 1.21 to 1.36. We note that the NN1, NN2, NN3, and NN4 models surpass their counterparts in GKX, recording Sharpe ratios of 1.21, 1.29, 1.34, and 1.36, respectively, compared to 1.17, 1.16, 1.2, and 1.35 in GKX. These findings underscore the economic gains of utilizing the robust risk premium to enhance portfolios.

Portfolios based on multi-task neural networks attain the highest Sharpe ratios. MT1

earns a Sharpe ratio of 1.71, marking a 27% improvement over the highest Sharpe ratio of 1.35 in GKX. MT2 and MT3 also achieve high Sharpe ratios of 1.63 and 1.57, respectively. Collectively, these observations underscore the necessity of imposing economic structure to enhance the economic gains of machine learning techniques.

Panel B of Table 4 details the risk-adjusted performance of the value-weighted long-short portfolios from each machine learning model. Reported are the alphas, information ratios (IR), and R^2 values with respect to the Fama-French five-factor model (Fama and French (2015)) augmented to include the momentum factor.[9] The first row shows that linear models earn the lowest monthly alphas ranging from 0.24% to 0.54%. This is followed by the tree-based models with RF and GBT recording alphas of 1.33% and 0.79%, respectively. Feed-forward neural networks achieve alphas of 1.4%, 1.83%, 1.95%, and 1.85%, respectively, outperforming their GKX counterpart alphas of 1.2%, 1.33%, 1.52%, and 1.76%. Consequently, the information ratios of our feed-forward neural network models are 1.02, 1.22, 1.27, and 1.26, respectively, beating the GKX information ratios of 0.92, 0.93, 0.96, and 1.18. MT models dominate all others, with alphas spanning 2.65% to 2.77%, and information ratios between 1.55 and 1.66. These results underscore the superior risk-adjusted performance delivered by the robust risk premium.

The main results in Table 4 are visually summarized in Figure 3, which reports the cumulative performance for the long-short portfolios sorted on $\bar{Q}$ for all models, along with the cumulative market excess return as a benchmark. MT models perform consistently well in recessions and the post-2000 sample, and dominate the other models by a substantial margin. Feed-forward neural networks are the next best performing models. Tree-based models achieve a moderate performance. Linear models are the worst performers, though they still beat the market benchmark.

Panel B of Table IA2 in the Internet Appendix showcases value-weighted portfolios sorted on an alternative quantile-based risk premium measure, $\bar{Q}_{YZLZ} = 0.1Q_{0.05} + 0.3Q_{0.25} +$

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

$0.2Q_{0.5} + 0.3Q_{0.75} + 0.1Q_{0.95}$, based on the methodology of Yang, Zhu, Li, and Zhu (2023). Notably, with the exception of the LR-4 model, the Sharpe ratios using this alternative approach are lower than those reported in Table 4. Such a comparison amplifies the enhanced economic significance of the robust risk premium measure defined in Equation (3), reaffirming that it provides a superior estimate of the risk premium.

Table IA3 in the Internet Appendix presents equal-weighted portfolios sorted on the robust risk premium across the twelve models. As expected, equal-weighted portfolio results are typically better than value-weighted ones. MT1 records an average return of 4.92% and a Sharpe ratio of 3.19, greatly exceeding the highest Sharpe ratio of 2.45 in GKX.

Avramov, Cheng, and Metzker (2022) highlight concerns that certain deep learning models may yield long-short portfolios which are infeasible due to economic constraints such as drawdowns and transaction costs. Panel A of Table 5 reports the drawdowns, maximum losses, portfolio turnovers, and break-even transaction costs of the value-weighted long-short portfolios from each machine learning model. MT1 registers the smallest drawdown of 27.92%, while MT2 and MT3 experience modest drawdowns of 39.55% and 48.08%, respectively, indicating the effectiveness of quantile-based measures in improving stability and protecting investors from extreme crashes. In comparison, feed-forward neural networks undergo slightly higher drawdowns ranging from 34.34% to 62.1%, comparable to their GKX counterparts, which vary from 30.84% to 61.6%. RF and GBT endure high drawdowns of 65.36% and 99.78%, respectively. Linear models also face substantial drawdowns, ranging between 36.51% and 90.4%.

The third row of Panel A reveals that MT models register higher turnovers, ranging between 145.06% and 148.29%. Similarly, feed-forward neural networks show turnovers from 136.68% to 138.01%, exceeding the turnovers of GKX counterparts, which lie between 121.02% to 126.81%. The fourth row shows that despite their higher turnover, MT models register substantial break-even transaction costs, i.e., the cost required to completely absorb the risk-adjusted return. Following Avramov, Cheng, and Metzker (2022), we calculate the

break-even transaction cost of each model as the six-factor alpha divided by the turnover. The break-even transaction cost for MT models vary from 1.79% to 1.88%, which significantly exceed the average transaction cost of 0.64% for this sample of stocks, estimated by Avramov, Cheng, and Metzker (2022). Additionally, NN2, NN3, and NN4 exhibit high break-even transaction costs of 1.34%, 1.42%, and 1.34%, respectively. These results significantly exceed the break-even transaction costs of other deep learning models as reported by Avramov, Cheng, and Metzker (2022); which include 0.92% for NN3 in GKX, 1.12% for the generative adversarial network from Chen, Pelger, and Zhu (2023), and 0.48% for the conditional autoencoder from Gu, Kelly, and Xiu (2021). This suggests that incorporating the robust risk premium with deep learning models can be used to achieve higher risk-adjusted returns compared to leading deep learning models, particularly when accounting for transaction costs.

Furthermore, Avramov, Cheng, and Metzker (2022) recommend the exclusion of hard-to-arbitrage stocks when constructing portfolios. As a robustness check, Table IA8 in the Internet Appendix displays similar results for both value-weighted and equal-weighted portfolios when we apply portfolio sorts with NYSE breakpoints. Notably, both value-weighted and equal-weighted MT1 portfolios still achieve significant Sharpe ratios of 1.34 and 3.1, respectively. Additionally, Table IA10 in the Internet Appendix reveals that, after excluding microcap stocks, the results are still economically significant. Value-weighted and equal-weighted MT1 portfolios achieve high Sharpe ratios of 1.03 and 1.16, respectively. These results underscore the capability of the robust risk premium measure to achieve significant risk-adjusted returns in the face of economic constraints.

4.4 Conditional Volatility and Expected Returns

This section studies the empirical relationship between the robust conditional volatility (5) and average returns. Table 6 displays the performance of value-weighted portfolios sorted on robust conditional volatility predictions from our twelve models. Average returns generally

decrease with an increase in Vol predictions across all models, consistent with Ang, Hodrick, Xing, and Zhang (2006). However, only one t-statistic is significant at the 5% confidence level, demonstrating that the relationship between Vol and returns is weak. Panel B reports that none of the alphas are significant at the 5% level, agreeing with Bégin, Dorion, and Gauthier (2020), who argue that volatility under normal market conditions isn't priced. Table IA4 indicates that equal-weighted portfolios, sorted on Vol, deliver positive but insignificant average returns across all models, suggesting there is no robust relationship between Vol and average returns.

Next, motivated by the asymmetric preferences of investors toward upside potential and downside risk (Kahneman and Tversky (1979)), we bifurcate our analysis into upside and downside volatilities. Panel A of Table 7 reports the performance of value-weighted portfolios sorted on upside volatility, Vol^+. Interestingly, average returns increase with an increase in Vol^+ for all models, with LR-4 being the sole exception. While the long-short portfolios are only significant for MT1 and MT2, Panel B reports that the alphas are positive and statistically significant for all nonlinear models except GBT. In contrast, the alphas are not significant for the linear models, suggesting that nonlinearities play a crucial role in capturing the anomalous positive relationship between Vol^+ and future returns. The positive relationship between upside volatility and average returns is inconsistent with the risk-return tradeoff, which predicts a negative relationship since stocks with higher upside potential should yield lower expected returns.

In contrast, Panel A of Table 8 displays a significant negative relationship between downside volatility, Vol^-, and average returns. Both Panels A and B reveal that the average returns and alphas of the long-short portfolios are negative and statistically significant across all models, with the exception of GBT. However, Panel B shows that the alphas of NN and MT are larger in magnitude compared to those of linear and tree-based models, suggesting that neural networks are the most effective models in capturing the anomalous negative relationship between Vol^- and future returns. These results show that the negative relationship

found in Table 6 is concentrated in downside volatility. The negative relationship between downside volatility and average returns is also inconsistent with the risk-return tradeoff, which predicts a positive relationship since stocks with higher downside risk should earn higher expected returns.

In summary, the asymmetry between upside and downside volatility explains the muted relationship of Vol with average returns. This is because Vol^+ and Vol^- exhibit opposite relationships with average returns, effectively neutralizing any relationship when combined. This observed asymmetry between upside and downside volatility is consistent with the arbitrage asymmetry theory postulated by Stambaugh, Yu, and Yuan (2015), where stocks with higher Vol^- (Vol^+) are overpriced (underpriced) and exhibit a negative (positive) relationship with average returns due to limits-to-arbitrage. These results contrast with prospect theory, which predicts the opposite relationships for both Vol^+ and Vol^- (Barberis, Mukherjee, and Wang (2016)).

4.5 Conditional Skewness and Expected Returns

Table 9 reveals a positive relationship between robust conditional skewness and average returns for all twelve models. Panel A of Table 9 reports that long-short portfolios sorted by Skew yield positive and statistically significant returns for nine out of twelve models. Notably, Skew forecasts from MT3 yield a long-short return of 1.51% (t-stat of 4.32) with a Sharpe ratio of 0.83. Panel B of Table 9 reports that alphas for NN2-NN4 and MT1-MT3 are statistically significant at the 5% level, with Skew forecasts from MT1 achieving an alpha of 0.96% (t-stat of 3.62) and an information ratio of 0.72. In contrast, the alphas for linear and tree-based models are not statistically significant, indicating that neural networks are important for uncovering the anomalous positive relationship between Skew and future returns. Table IA5 in the Internet Appendix shows that the equal-weighted long-short portfolios, sorted on Skew, all yield positive average returns and alphas (except LR-4), with the majority of them showing statistical significance at the 1% level. MT1 yields a Sharpe ratio

of 1.37 and six-factor alpha of 2.05% (t-stat of 6.4).

Our results demonstrate that stocks with higher predicted upside potential yield higher average returns. Conversely, stocks with more predicted downside earn lower average returns. These findings contrast with several prevailing economic theories, such as the risk aversion (Harvey and Siddique (2000), Kraus and Litzenberger (1976)) and lottery preferences (Barberis and Huang (2008), Brunnermeier, Gollier, and Parker (2007), Mitton and Vorkink (2007)), which predict a negative relationship between skewness and expected returns.

This significant positive relationship between conditional skewness and average returns corroborates the empirical findings of Bali, Hu, and Murray (2019), Ghysels, Plazzi, and Valkanov (2016), Rehman and Vilkov (2012), and Xing, Zhang, and Zhao (2010). By utilizing nonlinear models and incorporating a broad range of variables, we've solidified the evidence supporting this relationship. Notably, stocks exhibiting positive skewness imply a greater upside potential relative to their downside risk. This insight coherently dovetails with Section 4.4, where we delineate that upside volatility is positively related, while downside volatility is negatively related, to average returns.

Table IA9 in the Internet Appendix further corroborates these findings, reporting similar results when sorting portfolios with NYSE breakpoints. Also, after excluding microcap stocks, long-short portfolios produce positive average returns across all models, with the majority showing statistical significance at the 5% level, as demonstrated in Table IA11.

4.6 Conditional Tail Risk and Expected Returns

Table 10 presents a consistent negative relationship between robust conditional tail risk and average returns. Panel A of Table 10 reports that long-short portfolios sorted by VaR yield negative and statistically significant returns for all models. Notably, VaR forecasts from NN2 register a long-short return of -2.22% (t-stat of -3.3) and a Sharpe ratio of -0.71. Panel B of Table 10 reports that alphas are all negative, and are statistically significant at the 5% level for nine out of twelve models, with NN2 posting an alpha of -1.49% (t-stat

of −5.11). By comparison, Atilgan, Bali, Demirtas, and Gunaydin (2020) report that long-short portfolios sorted on VaR, conditioned on daily returns, yield value-weighted returns of −0.78% (t-stat of −2.34) and alphas of −0.94% (t-stat of −4.42). This underscores the enhanced economic significance of forecasting VaR using machine learning models that condition on a comprehensive set of firm characteristics and macroeconomic variables.

Our findings lend further support to the negative relationship between VaR and average returns as previously demonstrated in Atilgan, Bali, Demirtas, and Gunaydin (2020). In line with their work, we find that this relationship is mostly due to the very negative returns observed in Decile 10. This suggests that investors overprice securities bearing higher tail risk, subsequently facing negative returns in the future. This behavior aligns well with the theoretical framework proposed by Hong and Stein (1999), wherein tail risk events persist over time due to short-term underreaction in asset prices.

While it's understood that the left-tail of returns correlates negatively with average returns, attributed to investor underreaction, an intriguing follow-up question arises: do investors similarly underreact to exceptionally positive returns? If so, we'd anticipate a positive correlation between the right-tail of returns and average returns. Table IA6 provides insights into this question, detailing the performance of value-weighted returns sorted by $Q_{0.95}$—the right-tail counterpart of VaR. As delineated in Panel A of Table IA6, long-short portfolios sorted by $Q_{0.95}$ present negative results, except NN2, albeit statistically insignificant for all models. Delving into Panel B, alphas are predominantly positive across models, except for LR-4. However, statistical significance emerges solely for NN1 and NN2. These results show that there is no consistent relationship between the right-tail of returns and average returns, indicating an asymmetry between left- and right-tail risks. This finding resonates with the empirical observations in Atilgan, Bali, Demirtas, and Gunaydin (2020) and Hong, Lim, and Stein (2000), reinforcing the established asymmetry between upside potential and downside risk, as detailed in Section 4.4.

4.7 Robust Sharpe Ratio and Sharpe-Managed Portfolios

In this section, we study whether using investor risk aversion as a secondary portfolio criteria can lead to economic gains. To do so, we leverage the measures from Equations (3) and (5) to introduce a quantile-based measure of risk-adjusted returns that is robust to outliers, which we refer to as the robust Sharpe ratio:

$$SR_Q = \frac{\bar{Q}}{Vol}. \tag{13}$$

The numerator of (13) is the robust risk premium $\bar{Q}$, which is an estimator of the conditional expectation of excess returns. We choose $\bar{Q}$ because of its insensitivity to outliers, unlike estimated expectations from models fitted with the MSE objective function. For the denominator, we utilize the interquartile range in lieu of the standard deviation. This choice is motivated by the fact that the interquartile range, a common measure of variability, is robust to outliers, whereas the standard deviation is not.

Notably, under the Gaussian distribution, SR_Q is approximately equal to the Sharpe ratio divided by 1.35, and thus can be viewed as a scaled version of the moment-based Sharpe ratio under certain conditions. For heavy-tailed distributions, SR_Q can be viewed as a robust measure of risk-adjusted returns.[10] While there's no theoretical link between the robust Sharpe ratio and the moment-based Sharpe ratio without imposing strict distributional assumptions, its usefulness as a measure of risk-adjusted returns is ultimately an empirical question. This section investigates whether the robust Sharpe ratio contains predictive information for the moment-based Sharpe ratio, and whether portfolios based on the robust risk premium can be improved by incorporating SR_Q.

Allena (2021) demonstrates that employing a strategy based on machine learning forecasts of the risk premium and its standard error can enhance portfolio performance. Inspired by the strategy in Allena (2021), we consider a similar portfolio strategy that uses machine

[10]For certain heavy-tailed distributions, such as a Cauchy distribution, the moment-based Sharpe ratio is undefined, whereas SR_Q always exists as long as $Vol > 0$.

learning SR_Q forecasts to enhance the portfolios in Section 4.3. At the end of each month, we sort stocks into deciles based on their one-month-ahead robust risk premium forecast. The portfolios with the lowest and highest robust risk premium forecasts are labeled as "Low" and "High", respectively. We then form Sharpe-Managed portfolios by subdividing each portfolio into deciles based on their SR_Q forecasts. We reconstitute portfolios monthly using value weights. Within each decile, D1 (D10) denotes the portfolio with the lowest (highest) SR_Q forecasts. This results in deciles $\{Low_{D1}, ..., Low_{D10}\}$ and $\{High_{D1}, ..., High_{D10}\}$ within the Low and High portfolios, respectively. Here, Low_{D10} (Low_{D1}) and $High_{D10}$ ($High_{D1}$) denote subgroups of stocks with the highest (lowest) SR_Q forecasts from the Low and High portfolios, respectively. Our Sharpe-Managed strategy, denoted as "H_{D10}-L_{D1}", involves exclusively taking long positions on $High_{D10}$ and short positions on Low_{D1}.

Panel A of Table 11, displays the realized moment-based Sharpe ratios of the Sharpe-Managed portfolios. Across models, the Sharpe ratios of D1 are generally less than D10, indicating that SR_Q contains predictive information for the realized moment-based Sharpe ratios, validating it as a useful measure of risk-adjusted returns. The Sharpe ratios of D1 do not exceed 0.7, even in the decile with the highest expected returns, suggesting that these portfolios contain highly volatile stocks. Conversely, D10 Sharpe ratios are higher, peaking at 1.5 for MT1 in the "High" decile, indicating that this portfolio features stocks with low volatility and high returns.

Panel B of Table 11 reports that the Sharpe-Managed strategy significantly enhances the Sharpe ratios relative to Section 4.3 for most nonlinear models. The Sharpe ratio for RF and GBT increase to 0.95 and 1.04, respectively. NN2, NN2, and NN4 garner Sharpe ratios of 1.31, 1.65, and 1.42, respectively. MTs achieve impressive Sharpe ratios between 2.25 and 2.51, with MT2 recording the highest Sharpe ratio. However, the linear and NN1 models experience reduced Sharpe ratios compared to Section 4.3. Panel C demonstrates that risk-adjusted returns are substantially improved for every model. MT2 posts the highest six-factor alpha of 5.73% (t-stat of 12.01), reemphasizing the economic gains of accurately

predicting robust risk premiums and conditional volatility through imposing economic structure. Additionally, the Sharpe ratio of 2.51 attained by MT2 significantly surpasses the peak Sharpe ratio of 1.51 documented in Allena (2021), which uses a comparable value-weighted portfolio strategy based on machine learning forecasts of the risk premium and its standard error. This comparison suggests that strategies incorporating the conditional distribution of returns through SR_Q outperform those relying on the distribution of the risk premium as described in Allena (2021). Overall, these results show that considering risk aversion in portfolio strategies can substantially improve portfolios due to SR_Q's ability to predict Sharpe ratios.

Table IA7 in the Internet Appendix displays the results for the Sharpe-Managed equal-weighted portfolios. Panel B demonstrates that the Sharpe-Managed strategies generally improve the Sharpe ratios of the equal-weighted long-short portfolios relative to Table IA3. The MT models secure the highest Sharpe ratios from 3.19 to 3.42.

5 Sources of Return Distribution Predictability

Sections 4.1 and 4.2 demonstrate that quantile machine learning models accurately predict the conditional quantiles and risk premia of individual stocks. Sections 4.3 - 4.6 reveal predictive relationships between forecasts of risk premium, downside risk, skewness, and tail risk with expected returns, which cannot be accounted for by exposures to systematic risk factors and are difficult to reconcile with traditional risk-return theories. Consequently, this section investigates behavioral explanations to uncover the drivers of 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 incor-

poration of new information into prices. Consequently, prices tend to underreact to new information, causing short-run return continuations and return distribution predictability.[11] These short-run continuations help explain our empirical findings that stocks with higher upside potential (i.e., $\bar{Q}$ or $Skew$) earn significantly higher average returns, while stocks with higher downside risk (i.e., Vol^{-} and VaR) earn significantly lower average returns.

Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023) demonstrate that these behavioral biases can explain the return predictability documented in their papers. In this section, we explore potential economic sources driving the return distribution predictability, by adopting tests inspired by Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023). We specifically investigate whether behavioral biases linked to underreaction to news, investor inattention, information frictions, and mispricing due to biased expectations explain the return distribution predictability. Given the consistency of empirical findings across various models, our analysis will concentrate on the MT1 model.

5.1 Underreaction to Information Shocks

Machine learning forecasts rapidly integrate new information on firm fundamentals and characteristics. If investors underreact to this new information, the most pronounced return distribution predictability should manifest in firms experiencing substantial information shocks, as reflected by large forecast changes. Motivated by Atilgan, Bali, Demirtas, and Gunaydin (2020), we test this conjecture by calculating the change in machine learning forecasts for each firm between months t and $t-1$, and use these forecast changes in bivariate portfolio sorts to determine whether they have any implication for excess returns in month $t+1$. 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 $t-1$, representing the innovation

[11] 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 short-run return continuations.

in $\bar{Q}$. Similarly, DeltaSkew, DeltaVol^{-}, and DeltaVaR are defined as the difference in their respective forecasts over the same period. Stocks undergoing large positive or negative information shocks from month $t-1$ to t will exhibit correspondingly high positive or negative Delta values, while stocks with minimal changes in forecasts will have only moderate Delta values.

We begin our analysis by sorting stocks into terciles based on Delta$\bar{Q}$. Within each Delta$\bar{Q}$ tercile, we further sort stocks into decile portfolios based on their one-month-ahead robust risk premium forecasts, following the methodology described in Section 4.3. We repeat this procedure for stocks based on their DeltaSkew, DeltaVol^{-}, and DeltaVaR, sorting them into terciles, and then into decile portfolios according to their Skew, Vol^{-}, and VaR forecasts, respectively.

Table 12 presents the results of our Delta analysis. Panel A illustrates that stocks in the lowest Delta$\bar{Q}$ tercile, which experienced a large negative information shock, achieve a long-short return of 3.65% (t-stat of 9.37) and a Sharpe ratio of 1.78 when sorted on $\bar{Q}$. Additionally, stocks in the highest Delta$\bar{Q}$ tercile, which experienced a large positive information shock, achieve a long-short return of 3.8% (t-stat of 7.05) and a Sharpe ratio of 1.55. Conversely, stocks in the middle Delta$\bar{Q}$ tercile, with relatively unchanged robust risk premium forecasts, post a lower long-short return of 2.15% (t-stat of 5.91) and a Sharpe ratio of 1.18. Panels B, C, and D display similar patterns for DeltaSkew, DeltaVol^{-}, and DeltaVaR, respectively.

Figure 4 displays the quantile and return prediction performance for terciles sorted on DeltaSkew. For each quantile, the lowest and highest terciles achieve higher $R^1_{os}(\tau)$ values than the middle tercile, indicating that stocks experiencing large negative and positive information shocks have more predictable distributions than stocks with relatively unchanged forecasts. Additionally, the figure illustrates that the highest tercile achieves the highest R^2_{os} value, indicating that stocks with large positive information shocks exhibit more predictable returns.

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 this new information, leading to stronger return continuations and return distribution predictability in stocks with significant forecast changes—those experiencing large information shocks—as predicted by the theoretical framework in Hong and Stein (1999).

Notably, the long-short portfolios for DeltaVol^{-} and DeltaVaR in Panels C and D are not statistically significant for stocks in the middle tercile, suggesting that underreaction to new information may explain these downside volatility and tail risk anomalies. In contrast, the long-short portfolios for Delta$\bar{Q}$ and DeltaSkew in Panels A and B remain statistically significant across all terciles, suggesting that these anomalies cannot be solely attributed to underreaction to new information.

5.2 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 a more pronounced underreaction to new information and return distribution predictability. Numerous studies empirically support this, showing that investor inattention often results in underreaction to public information about firm fundamentals and characteristics.[12]

The investor inattention hypothesis predicts stronger return distribution predictability in stocks with less attention, as typically proxied by analyst coverage (CVRG). Motivated by Atilgan, Bali, Demirtas, and Gunaydin (2020), we test this conjecture by sorting stocks into

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

terciles according to the number of analysts covering each firm in the most recent month.[13] Within each tercile, stocks are then sorted into decile portfolios based on their $\bar{Q}$, $Skew$, Vol^{-}, or VaR.

Table 13 presents the findings. Panel A reveals that the long-short portfolio sorted on $\bar{Q}$ achieves a Sharpe ratio of 2.15 in the low analyst coverage tercile, which monotonically decreases to 1.06 in the high analyst coverage tercile. Similarly, Panels B, C, and D demonstrate that Sharpe ratios for the long-short portfolios sorted on $Skew$, Vol^{-}, and VaR, respectively, progressively decrease in magnitude from low to high analyst coverage terciles. Figure 4 illustrates the quantile and return prediction performance for terciles sorted on analyst coverage. Across all quantiles, the $R^1_{os}(\tau)$ values monotonically decrease from the low analyst coverage tercile to the high tercile. Additionally, the R^2_{os} values are lowest in the high tercile. Consistent with the investor inattention hypothesis, these results indicate stronger (weaker) return distribution predictability in stocks receiving less (more) attention.

5.3 Information Frictions

Information frictions due to limits-to-arbitrage can 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 findings, leading to a testable prediction: return distribution predictability should be more pronounced in stocks with higher arbitrage costs.

Rather than relying on a single proxy for information frictions, we adopt the approach of Atilgan, Bali, Demirtas, and Gunaydin (2020) and Bali, Beckmeyer, Mörke, and Weigert (2023) by constructing an arbitrage index (AI) incorporating multiple indicators of limits-to-arbitrage. These indicators include firm size, firm age, idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang (2006)), Amihud illiquidity (Amihud (2002)), and analyst coverage (Zhang (2006)). To construct the AI, we sort stocks in increasing order by their idiosyncratic volatil-

[13] We account for missing coverage observations by assigning a zero value to ensure a comprehensive sample, though our conclusions remain unchanged when these missing observations are excluded.

ity and illiquidity and in decreasing order by their size, age, and analyst coverage. Each firm is given the corresponding score of its decile rank for each indicator. The firm-level AI is then calculated as the sum of these scores, which range from 5 to 50, with higher values indicating 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 AI quintile, stocks are further divided into decile portfolios based on their $\bar{Q}$, $Skew$, Vol^{-}, or VaR. The results, displayed in Table 14, reveal a distinct pattern: Panel A indicates that the long-short portfolio sorted on $\bar{Q}$ earns a Sharpe ratio of 0.98 in the lowest AI quintile, which monotonically increases to 3.1 in the highest AI quintile. Similarly, Panels B, C, and D show an increasing trend in absolute Sharpe ratios from the lowest to the highest AI quintiles for $Skew$, Vol^{-}, and VaR respectively. This pattern indicates that return distribution predictability is pronounced in firms with higher arbitrage costs. Notably, the long-short portfolios for Delta$\bar{Q}$ and DeltaSkew in Panels A and B remain statistically significant for all quintiles, suggesting that information frictions provide only a partial explanation for these anomalies.

Figure 6 illustrates the quantile and return prediction performance for quantiles sorted on AI. Interestingly, the figure reveals a nonlinear "U-shaped" relationship between AI and predictive power. $R^1_{os}(\tau)$ and R^2_{os} values generally increase from AI3 - AI5, aligning with expectations that higher arbitrage costs enhance predictability. However, the $R^1_{os}(\tau)$ and R^2_{os} values are also relatively high for the lowest AI quantile, AI1, which has the least information frictions. This phenomenon likely occurs because AI1, by definition, includes large, liquid, and stable stocks, which are inherently more predicable. However, this higher predictive power does not translate into greater profitability, as indicated by the relatively low Sharpe ratios for AI1 in Table 14, since the stocks in this portfolio have low arbitrage costs, allowing predictable patterns to be quickly arbitraged away. In contrast, AI5 exhibits both high predictability and profitability due to the higher arbitrage costs impeding the full

incorporation of information into prices.

5.4 Biased Expectations

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 due to biased expectations, which are partially corrected when new information arrives. To evaluate the impact of earnings announcements in our analysis, 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 $\bar{Q}$, $Skew$, Vol^{-}, or VaR.

Table 15 reports the results. Panel A of Table 15 reveals that the long-short portfolio sorted on $\bar{Q}$ achieves a Sharpe ratio of 1.74 for stocks without earnings announcements, compared to only 1.07 for stocks with earnings announcements. Similar trends are evident for $Skew$, Vol^{-}, and VaR in Panels B, C, and D, respectively. Figure 7 shows that $R^1_{os}(\tau)$ and R^2_{os} values are higher in stocks without earnings announcements. These patterns contrast with the one documented by Engelberg, McLean, and Pontiff (2018), suggesting that biased expectations are unlikely to be a source of return distribution predictability.

5.5 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 $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$, 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 sample, and then averaging the $R^1_{os}(\tau)$ reductions across 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.

Figure 8 presents the resulting variable importances of the top ten firm-level characteristics for each model.[14] Panel A reports that for the 0.05 quantile, the models are in close agreement regarding the most influential predictors, which are primarily risk measures such as the total and idiosyncratic return volatility (retvol, idiovol), market beta (beta), and beta squared (betasq), followed by liquidity variables, including the bid-ask spread (baspread), dollar volume volatility (std_dolvol), firm size (mvel1), number of zero trading days (zerotrade), Amihud illiquidity (ill), share turnover and turnover volatility (turn and std_turn). Recent price trends are the third influential group, including stock momentum (mom6m, mom12m), and the maximum return (maxret). These set of predictors are consistent with Chen, Hong, and Stein (2001), who show that risk, liquidity and price trend variables are strong predictors of crashes. Furthermore, the significant importance of volatility, firm size, and illiquidity provides additional support for information frictions as a source of return distribution predictability, as these variables are common proxies for limits to arbitrage.

The last group of influential predictors includes valuation ratios and fundamental signals, such as book-to-market and industry-adjusted book-to-market (bm and bm_ia), dividend initiation and omission (divi and divo), return on assets (roaq), and earnings volatility (roavol). Multi-task neural networks extract predictive information from a more diverse set of characteristics compared to other models, which heavily lean towards total and idiosyncratic return volatility, and the bid-ask spread.

Panels B, D, and E of Figure 8 reveal that the significant predictors for the 0.25, 0.75, and 0.95 quantiles closely align with those for the 0.05 quantile, with the exception of the short-term reversal (mom1m) playing a more prominent role in these other quantiles. These findings largely agree with the literature on forecasting volatility (Andersen, Bollerslev, Christof-

[14]Refer to Table A.6 of the Internet Appendix in GKX for detailed definitions of these characteristics.

fersen, and Diebold (2005)) and Value-at-Risk (Liu and Stentoft (2021)), where models generally leverage predictors such as historical volatility and short-term reversal. However, the machine learning models gain accuracy from incorporating liquidity and momentum variables, which are not typically incorporated into Value-at-Risk models, highlighting their capability of drawing predictive information from a much broader set of variables.

Panel C of Figure 8 illustrates that for the conditional median, momentum (mom6m, mom12m) and illiquidity have a more significant influence. The models also leverage information from additional valuation ratios and fundamental signals such as earnings-to-price (ep), operating profitability (operprof), and the difference between percentage changes in sales and inventory (pchsaleinv). The MT models, which attain the highest $R^1_{os}(\tau)$ values for the median, feature mostly valuation ratios and fundamental signals in its most influential variables. It's expected that momentum, valuation ratios, and fundamental signals are influential for the conditional median, as it can be viewed as a measure of a stock's consistency. It's also noteworthy that, akin to Chen, Dolado, and Gonzalo (2021), we discover that the variables impacting the median differ from those influencing the other quantiles.

Our set of dominant predictor variables closely aligns with those identified by GKX, who pinpoint price trends, liquidity variables, and risk measures as crucial characteristics for predicting the conditional expectation. Moreover, our quantile-based method complements their approach, offering a more detailed insight into the variables that shape the conditional distribution of returns. For instance, we find momentum, liquidity, valuation ratios, and fundamental signals to be particularly influential for the distribution center. However, we further reveal that volatility and liquidity variables are the most potent predictors for other quantiles.

Figure 9 presents the variable importance of each macroeconomic variable (normalized to sum to one within each model). Panels A, B, and C reveal a general consensus among models that the dividend-price (DP) ratio is the most influential predictor for the 0.05, 0.25, and 0.5 quantiles. Conversely, Panels D and E illustrate that the earnings-price (EP) ratio holds the

most influence for the 0.75 and 0.95 quantiles. These findings highlight that the set of influential macroeconomic predictors varies between downside (0.05 and 0.25) and upside (0.95 and 0.75) quantiles, and that models largely agree with each other on the set of dominant macroeconomic variables for each quantile. The book-to-market ratio demonstrates considerable influence across all quantiles, a finding that aligns with the results reported in GKX, which identified the book-to-market ratio as the most influential macroeconomic variable in predicting the conditional expectation. As observed earlier, multi-task neural networks tend to extract predictive information from a broader set of macroeconomic variables.

6 Conclusion

A rapidly growing literature applies machine learning techniques to model the *conditional expectation* of asset returns. However, the research on applying machine learning to study the *conditional distribution* of asset returns is relatively scarce and not well understood. This paper closes this gap by introducing quantile machine learning techniques to improve our understanding of asset return quantiles, and their implications for asset pricing. Furthermore, investors can benefit from using forecasts of the distribution of future returns to measure risk and better manage their portfolios.

This paper utilizes quantile machine learning techniques to predict individual stock return quantiles, leveraging a set of 94 firm-level predictors and 8 macroeconomic variables from 1957 to 2016. Empirically, we demonstrate several results that enhance our knowledge of the cross-section of stock returns. First, we demonstrate that multi-task neural networks generally outperform linear, tree-based, and feed-forward neural network models across all quantiles. Second, our method of integrating over quantiles to form a robust risk premium yields superior out-of-sample predictive power and more profitable portfolios compared to the established machine learning methods in GKX. Third, we extend our methodology to predict volatility, skewness, and VaR, and demonstrate strong asymmetries in the relationship

between risks and returns. Fourth, we show that strategies incorporating machine learning Sharpe ratio forecasts significantly improve portfolio performance. Fifth, we present compelling evidence that return distribution predictability is influenced by investor biases, including underreaction to news, investor inattention, and information frictions. Lastly, we document that the set of influential variables varies across quantiles, enriching our understanding of how different predictors can influence different parts of the future return distribution.

The quantile machine learning approaches studied in this paper 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 (Bali, Goyal, Huang, Jiang, and Wen (2020), Bianchi, Büchner, and Tamoni (2021)), options (Bali, Beckmeyer, Mörke, and Weigert (2023), Goyenko and Zhang (2020)), 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)), all 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

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

Aleti, S., and T. Bollerslev (2024): "News and Asset Pricing: A High-Frequency Anatomy of the SDF," *Review of Financial Studies, forthcoming*.

Aleti, S., T. Bollerslev, and M. Siggaard (2023): "Intraday Market Return Predictability Culled from the Factor Zoo," *Available at SSRN 4388560*.

Allena, R. (2021): "Confident risk premiums and investments using machine learning uncertainties," *Available at SSRN 3956311*.

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez (2015): "Does realized skewness predict the cross-section of equity returns?," *Journal of Financial Economics*, 118(1), 135–167.

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*.

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.

Bali, T. G., A. Goyal, D. Huang, F. Jiang, and Q. Wen (2020): "Different strokes: Return predictability across stocks and bonds with machine learning and big data," *Georgetown McDonough School of Business Research Paper*, pp. 20–110.

Bali, T. G., J. Hu, and S. Murray (2019): "Option implied volatility, skewness, and kurtosis and the cross-section of expected stock returns," *Georgetown McDonough School of Business Research Paper*.

Bali, T. G., L. Peng, Y. Shen, and Y. Tang (2014): "Liquidity shocks and stock market reactions," *The Review of Financial Studies*, 27(5), 1434–1485.

Barber, B. M., and T. Odean (2008): "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors," *The review of financial studies*, 21(2), 785–818.

Barberis, N. (2018): "Psychology-based models of asset prices and trading volume," in *Handbook of behavioral economics: applications and foundations 1*, vol. 1, pp. 79–175. Elsevier.

Barberis, N., and M. Huang (2008): "Stocks as lotteries: The implications of probability weighting for security prices," *American Economic Review*, 98(5), 2066–2100.

Barberis, N., A. Mukherjee, and B. Wang (2016): "Prospect theory and stock returns: An empirical test," *The Review of Financial Studies*, 29(11), 3068–3107.

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

Barro, R. J. (2006): "Rare disasters and asset markets in the twentieth century," *The Quarterly Journal of Economics*, 121(3), 823–866.

Bégin, J.-F., C. Dorion, and G. Gauthier (2020): "Idiosyncratic jump risk matters: Evidence from equity returns and options," *The Review of Financial Studies*, 33(1), 155–211.

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

Bergstra, J., and Y. Bengio (2012): "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, 13(2).

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.

Bollerslev, T., S. Z. Li, and B. Zhao (2020): "Good volatility, bad volatility, and the cross section of stock returns," *Journal of Financial and Quantitative Analysis*, 55(3), 751–781.

Bowley, A. L. (1920): *Elements of Statistics*. P.S. King & Son.

Boyer, B., T. Mitton, and K. Vorkink (2010): "Expected idiosyncratic skewness," *The Review of Financial Studies*, 23(1), 169–202.

Brunnermeier, M. K., C. Gollier, and J. A. Parker (2007): "Optimal beliefs, asset prices, and the preference for skewed returns," *American Economic Review*, 97(2), 159–165.

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.

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*.

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 (2023): "Alpha go everywhere: Machine learning and international stock returns," *Available at SSRN 3489679*.

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

Cohen, L., and A. Frazzini (2008): "Economic links and predictable returns," *The Journal of Finance*, 63(4), 1977–2011.

Conrad, J., R. F. Dittmar, and E. Ghysels (2013): "Ex ante skewness and expected stock returns," *The Journal of Finance*, 68(1), 85–124.

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

Da, Z., U. G. Gurun, and M. Warachka (2014): "Frog in the pan: Continuous information and momentum," *The review of financial studies*, 27(7), 2171–2218.

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.

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

——— (2009): "Investor inattention and Friday earnings announcements," *The journal of finance*, 64(2), 709–749.

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.

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.

Farago, A., and R. Tédongap (2018): "Downside risks and the cross-section of asset returns," *Journal of Financial Economics*, 129(1), 69–86.

Feng, G., J. He, N. Polson, and J. Xu (2023): "Deep Learning in Characteristics-Sorted Factor Models," *Journal of Financial and Quantitative Analysis, Forthcoming*.

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.

Gabaix, X. (2012): "Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance," *The Quarterly Journal of Economics*, 127(2), 645–700.

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. (2014): "Conditional skewness with quantile regression models: SoFiE presidential address and a tribute to Hal White," *Journal of Financial Econometrics*, 12(4), 620–644.

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.

Giglio, S., and K. Shue (2014): "No news is news: do markets underreact to nothing?," *The Review of Financial Studies*, 27(12), 3389–3440.

Goyenko, R., and C. Zhang (2020): "The joint cross section of option and stock returns predictability with big data and machine learning," *Available at SSRN 3747238*.

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*.

Gul, F. (1991): "A theory of disappointment aversion," *Econometrica*, pp. 667–686.

Han, B., D. Hirshleifer, and J. Walden (2023): "Visibility bias in the transmission of consumption beliefs and undersaving," *The Journal of Finance*, 78(3), 1647–1704.

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

Harvey, C. R., and A. Siddique (2000): "Conditional skewness in asset pricing tests," *The Journal of Finance*, 55(3), 1263–1295.

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

Hirshleifer, D., K. Hou, S. H. Teoh, and Y. Zhang (2004): "Do investors overvalue firms with bloated balance sheets?," *Journal of accounting and economics*, 38, 297–331.

Hirshleifer, D., P.-H. Hsu, and D. Li (2013): "Innovative efficiency and stock returns," *Journal of financial economics*, 107(3), 632–654.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009): "Driven to distraction: Extraneous events and underreaction to earnings news," *The journal of finance*, 64(5), 2289–2325.

Hirshleifer, D., and S. H. Teoh (2003): "Limited attention, information disclosure, and financial reporting," *Journal of accounting and economics*, 36(1-3), 337–386.

Hong, H., T. Lim, and J. C. Stein (2000): "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies," *The Journal of Finance*, 55(1), 265–295.

Hong, H., and J. C. Stein (1999): "A unified theory of underreaction, momentum trading, and overreaction in asset markets," *The Journal of Finance*, 54(6), 2143–2184.

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.

Hou, K., and T. J. Moskowitz (2005): "Market frictions, price delay, and the cross-section of expected returns," *The Review of Financial Studies*, 18(3), 981–1020.

Huang, W., Q. Liu, S. G. Rhee, and F. Wu (2012): "Extreme downside risk and expected stock returns," *Journal of Banking & Finance*, 36(5), 1492–1502.

Huberman, G., and T. Regev (2001): "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar," *The Journal of Finance*, 56(1), 387–396.

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

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*, (22-63).

Jiang, L., J. Liu, L. Peng, and B. Wang (2022): "Investor attention and asset pricing anomalies," *Review of Finance*, 26(3), 563–593.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263–291.

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

Kapadia, N., and M. Zekhnini (2019): "Do idiosyncratic jumps matter?," *Journal of Financial Economics*, 131(3), 666–692.

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

Kelly, B., and H. Jiang (2014): "Tail risk and asset prices," *The Review of Financial Studies*, 27(10), 2841–2871.

Kim, T.-H., and H. White (2004): "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, 1(1), 56–73.

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.

Kraus, A., and R. H. Litzenberger (1976): "Skewness preference and the valuation of risk assets," *The Journal of Finance*, 31(4), 1085–1100.

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

Liu, F. (2020): "Can the Premium for Idiosyncratic Tail Risk be Explained by Exposures to its Common Factor?," *Available at SSRN 3711215*.

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*.

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

Lou, D. (2014): "Attracting investor attention through advertising," *The Review of Financial Studies*, 27(6), 1797–1829.

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

Markowitz, H. (1952): "Portfolio selection," *Journal of Finance*, 7, 77–91.

——— (1959): *Portfolio Selection*. New Haven: Yale University Press.

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

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.

Mitton, T., and K. Vorkink (2007): "Equilibrium underdiversification and the preference for skewness," *The Review of Financial Studies*, 20(4), 1255–1288.

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

Peng, L. (2005): "Learning with information capacity constraints," *Journal of Financial and Quantitative Analysis*, 40(2), 307–329.

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.

Rehman, Z., and G. Vilkov (2012): "Risk-neutral skewness: Return predictability and its sources," *Available at SSRN 1301648*.

Roy, A. D. (1952): "Safety first and the holding of assets," *Econometrica: Journal of the Econometric Society*, pp. 431–449.

Sharpe, W. F. (1965): "Risk-aversion in the stock market: Some empirical evidence," *The Journal of Finance*, 20(3), 416–422.

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

Stambaugh, R. F., J. Yu, and Y. Yuan (2015): "Arbitrage asymmetry and the idiosyncratic volatility puzzle," *The Journal of Finance*, 70(5), 1903–1948.

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

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*.

Xing, Y., X. Zhang, and R. Zhao (2010): "What does the individual option volatility smirk tell us about future equity returns?," *Journal of Financial and Quantitative Analysis*, 45(3), 641–662.

Yang, X., Z. Zhu, D. Li, and K. Zhu (2023): "Asset pricing via the conditional quantile variational autoencoder," *forthcoming in Journal of Business & Economic Statistics*.

Yi, C., and J. Huang (2017): "Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression," *Journal of Computational and Graphical Statistics*, 26(3), 547–557.

Zeiler, M. D. (2012): "Adadelta: An adaptive learning rate method," *arXiv preprint arXiv:1212.5701*.

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.

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.

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

This table reports the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$ for the entire panel of individual stock excess returns. We compare twelve models, including linear quantile regression with all covariates (LR), LR-4 (which preselects size, book-to-market, momentum, and return volatility as the only covariates), 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), and quantile multi-task neural networks with 1-3 shared layers (MT1,...,MT3). 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

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

The full sample spans from 1957 to 2016, and is divided into 3 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 $\mathcal{T}_3$ (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 30 out-of-sample years from 1987 to 2016.

τ	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
0.05	11.18	11.08	12.87	12.39	13.69	16.20	16.16	16.07	15.75	20.37	20.24	19.89
0.25	3.83	4.42	4.65	4.60	4.65	5.86	5.74	5.77	5.82	7.16	7.18	6.92
0.50	-0.86	0.36	0.27	0.24	0.29	0.55	0.61	0.52	0.54	0.91	0.92	0.66
0.75	0.04	1.25	1.65	1.25	1.54	1.88	1.88	1.96	1.88	2.79	2.70	0.89
0.95	6.06	8.09	10.29	10.53	12.78	13.52	13.26	13.26	13.13	16.20	16.21	15.87

Table 2: 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 stock-level quantile prediction performance among the machine learning models for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$. We compare twelve models in total including linear quantile regression with all covariates (LR), LR-4 (which preselects size, book-to-market, momentum, and return volatility as the only covariates), 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), and quantile multi-task neural networks with 1-3 shared layers (MT1,...,MT3). A positive number indicates that the column model outperforms the row model, and bold font signifies that the difference is significant at the 5% level.

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

	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
LR	0.01	**2.87**	1.25	**2.22**	**4.51**	**4.38**	**4.26**	**3.97**	**5.79**	**5.73**	**5.69**
LR-4		**2.45**	**2.46**	**4.80**	**8.64**	**8.30**	**8.60**	**8.04**	**9.60**	**9.38**	**10.26**
EN			-0.32	1.17	**4.62**	**4.38**	**4.24**	**3.77**	**6.25**	**6.27**	**5.97**
RF				**2.22**	**5.29**	**5.76**	**5.68**	**5.02**	**9.53**	**9.18**	**8.91**
GBT					**2.14**	**2.22**	**2.21**	1.83	**5.61**	**5.61**	**5.77**
NN1						0.11	-0.50	-1.81	**7.03**	**6.50**	**6.85**
NN2							-0.68	**-3.09**	**7.53**	**7.09**	**7.33**
NN3								-1.95	**7.72**	**7.08**	**7.57**
NN4									**7.66**	**7.34**	**7.71**
MT1										-0.68	**-2.14**
MT2											-1.50

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

	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
LR	1.35	**2.85**	1.79	1.88	**4.58**	**4.29**	**4.52**	**4.49**	**5.35**	**5.18**	**5.14**
LR-4		0.75	0.87	1.08	**6.00**	**5.35**	**6.48**	**6.01**	**7.09**	**6.84**	**5.85**
EN			0.13	0.21	**5.00**	**4.36**	**5.00**	**4.69**	**5.40**	**5.30**	**4.88**
RF				0.08	**3.92**	**3.19**	**3.62**	**3.77**	**6.87**	**6.74**	**5.24**
GBT					**4.49**	**3.93**	**4.43**	**4.34**	**6.64**	**6.88**	**5.51**
NN1						-1.71	-0.63	-0.53	**4.65**	**4.50**	**3.24**
NN2							0.99	1.08	**4.80**	**4.66**	**3.74**
NN3								0.33	**4.44**	**4.31**	**3.39**
NN4									**4.82**	**4.57**	**3.54**
MT1										0.37	-1.31
MT2											-1.30

Panel C: Diebold-Mariano Tests for the $\tau = 0.5$											
	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
LR	**2.72**	**3.46**	**2.51**	**2.62**	**3.52**	**3.44**	**3.40**	**3.67**	**3.64**	**3.59**	**3.77**
LR-4		-0.69	-0.72	-0.36	1.70	**2.99**	1.51	1.52	**4.64**	**4.09**	1.16
EN			-0.03	0.39	**2.52**	**2.52**	**2.18**	**2.90**	**3.25**	**3.19**	1.88
RF				0.50	1.51	1.95	1.36	1.44	**3.83**	**3.48**	1.35
GBT					1.47	**2.20**	1.34	1.43	**4.34**	**4.33**	1.24
NN1						1.59	-0.68	0.17	**2.67**	**2.54**	0.40
NN2							**-2.31**	-1.06	**2.22**	**2.11**	-0.02
NN3								0.76	**2.72**	**2.67**	0.48
NN4									**2.26**	**2.25**	0.36
MT1										0.05	-1.44
MT2											-1.46

Panel D: Diebold-Mariano Tests for the $\tau = 0.75$											
	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
LR	1.33	**2.21**	1.32	1.62	**2.01**	**2.05**	**2.11**	**2.03**	**2.85**	**2.72**	0.49
LR-4		1.26	-0.16	1.30	**2.92**	**2.24**	**3.10**	**3.36**	**4.54**	**4.65**	-0.35
EN			-1.39	-0.34	1.06	1.13	1.50	1.09	**4.22**	**3.91**	-0.59
RF				1.49	**2.58**	**2.61**	**3.18**	**2.61**	**6.85**	**6.42**	-0.33
GBT					1.46	1.32	1.90	1.57	**4.54**	**4.63**	-0.48
NN1						0.07	1.43	0.10	**4.71**	**4.85**	-0.66
NN2							1.35	-0.01	**6.15**	**6.05**	-0.66
NN3								-1.09	**5.29**	**5.21**	-0.70
NN4									**4.28**	**4.30**	-0.66
MT1										-1.43	-1.08
MT2											-1.04

Panel E: Diebold-Mariano Tests for the $\tau = 0.95$											
	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
LR	0.69	1.30	1.32	1.92	**2.09**	**2.04**	**2.04**	**2.01**	**2.77**	**2.85**	**2.78**
LR-4		**6.61**	**4.55**	**8.70**	**10.36**	**11.08**	**12.33**	**10.21**	**9.49**	**9.41**	**9.49**
EN			0.68	**5.83**	**8.79**	**9.11**	**9.95**	**8.97**	**9.41**	**9.79**	**9.97**
RF				**6.53**	**9.01**	**8.59**	**8.46**	**8.33**	**11.45**	**10.96**	**10.92**
GBT					1.31	0.90	0.92	0.70	**5.10**	**4.87**	**4.44**
NN1						-1.57	-1.38	**-2.01**	**6.35**	**6.23**	**5.78**
NN2							-0.09	-0.85	**5.76**	**5.69**	**5.45**
NN3								-0.63	**5.31**	**5.32**	**5.08**
NN4									**6.60**	**6.47**	**6.12**
MT1										-0.09	-1.62
MT2											**-2.48**

Table 3: Monthly Out-of-Sample Stock-Level Excess Return 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 entire panel of individual stock excess returns. The robust risk premium is defined as $\bar{Q} = 0.15Q_{0.05}+0.225Q_{0.25}+0.25Q_{0.5}+0.225Q_{0.75}+0.15Q_{0.95}$. We compare twelve models in total including linear quantile regression with all covariates (LR), LR-4 (which preselects size, book-to-market, momentum, and return volatility as the only covariates), elastic net quantile regression (EN), quantile random forest (RF), quantile gradient boosted regression trees (GBT), quantile neural networks with 1-4 layers (NN1,...,NN4), and quantile multi-task neural networks with 1-3 shared layers (MT1,...,MT3). 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

$$R^2_{os} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \bar{Q}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{i,t+1}}. \tag{14}$$

Panel B reports pairwise Diebold-Mariano test statistics, comparing the out-of-sample firm-level stock return prediction performance among the twelve models. A positive number indicates that the column model outperforms the row model, and bold font signifies that the difference is significant at the 5% level.

Panel A: Monthly Out-of-sample Stock-level Prediction Performance (Percentage R^2_{OS})											
LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
-1.44	0.2	0.34	0.19	0.49	0.47	0.47	0.45	0.52	1.16	1.1	0.68
Panel B: Comparison of Monthly Out-of-Sample Prediction Using Diebold-Mariano Tests											
	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
LR	-1.83	0.37	-0.07	1.05	1.28	1.20	1.23	1.49	**3.35**	**2.96**	0.78
LR-4		**2.64**	1.87	**2.14**	**2.24**	**2.28**	**2.23**	**2.34**	**2.89**	**2.75**	**2.96**
EN			-0.41	0.59	0.60	0.62	0.49	0.89	**3.16**	**2.60**	0.73
RF				1.18	0.79	0.79	0.72	0.94	**3.31**	**2.64**	0.77
GBT					-0.17	-0.18	-0.29	0.03	**2.82**	**2.50**	0.18
NN1						-0.08	-0.78	0.96	**3.17**	**2.69**	0.30
NN2							-0.48	1.02	**3.14**	**2.61**	0.31
NN3								1.52	**3.27**	**2.78**	0.36
NN4									**3.01**	**2.54**	0.19
MT1										-0.49	-1.31
MT2											-1.30

Table 4: 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 (decile 1), "High" denotes the portfolio with the highest robust risk premium (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 $\bar{Q}$												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.01	0.21	-0.05	-0.17	-0.10	-0.43	-0.49	-0.47	-0.51	-1.42	-1.37	-1.37
2	0.48	0.38	0.51	0.34	0.41	0.29	0.10	0.11	0.23	-0.28	-0.27	-0.41
3	0.51	0.56	0.53	0.50	0.52	0.45	0.43	0.44	0.41	0.12	0.09	0.14
4	0.65	0.66	0.65	0.59	0.57	0.65	0.57	0.53	0.60	0.38	0.42	0.46
5	0.64	0.77	0.69	0.70	0.73	0.66	0.71	0.68	0.70	0.68	0.67	0.65
6	0.82	0.84	0.83	0.70	0.81	0.71	0.77	0.79	0.80	0.88	0.76	0.79
7	0.87	0.90	0.84	0.86	0.85	0.91	0.96	0.92	0.84	1.00	1.01	0.95
8	0.90	0.89	1.00	0.87	0.91	0.98	1.06	1.07	1.13	1.14	1.15	1.12
9	0.94	1.00	0.97	1.03	1.08	1.28	1.29	1.34	1.30	1.38	1.29	1.29
High (H)	1.14	1.16	1.19	1.23	1.49	1.68	1.89	1.93	1.97	1.94	2.12	2.05
H-L	1.13	0.95	1.25	1.40	1.59	2.12	2.39	2.41	2.48	3.36	3.50	3.42
t	(4.79)	(2.74)	(4.90)	(4.92)	(4.12)	(6.70)	(6.55)	(6.53)	(7.05)	(9.49)	(9.38)	(8.96)
SR	0.89	0.50	0.91	0.91	0.76	1.21	1.29	1.34	1.36	1.71	1.63	1.57
Panel B: Risk-Adjusted Performance of Portfolios Sorted on $\bar{Q}$												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	0.48	0.24	0.54	1.33	0.79	1.4	1.83	1.95	1.85	2.65	2.77	2.65
t(α)	(2.79)	(1.34)	(2.87)	(5.15)	(2.89)	(5.28)	(5.42)	(4.91)	(5.88)	(9.38)	(9.5)	(8.33)
R^2	38	83	38	21	48	38	34	27	34	34	36	38
IR	0.48	0.3	0.51	0.98	0.52	1.02	1.22	1.27	1.26	1.66	1.61	1.55

Table 5: Drawdowns, Turnover, and Break-even Transaction Cost of Long-Short Portfolios

Panel A provides the maximum drawdowns ("Max DD"), most extreme negative monthly return ("Max 1M Loss"), average monthly percentage change in holdings ("Turnover"), and break-even transaction cost ("Break-even TC") for value-weighted long-short portfolios, sorted on robust risk premium forecasts. Panel B presents the results for equal-weighted long-short portfolios, sorted on robust risk premium forecasts.

Panel A: Drawdowns, Turnover, and Break-even Transaction Cost of Value-Weighted Portfolios Sorted on $\bar{Q}$												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Max DD(%)	37.48	90.40	36.51	65.36	99.78	51.30	62.10	34.34	60.81	27.92	39.55	48.08
Max 1M loss(%)	14.18	39.34	17.53	23.53	33.13	26.44	22.37	17.92	19.89	17.22	22.03	25.43
Turnover(%)	136.34	70.83	144.32	174.74	149.47	137.03	136.68	137.48	138.01	145.06	147.29	148.29
Break-even TC(%)	0.35	0.34	0.37	0.76	0.53	1.02	1.34	1.42	1.34	1.83	1.88	1.79
Panel B: Drawdowns, Turnover, and Break-even Transaction Cost of Equal-Weighted Portfolios Sorted on $\bar{Q}$												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Max DD(%)	42.87	130.28	38.18	26.28	64.43	20.50	18.03	20.13	17.97	15.25	19.24	42.03
Max 1M loss(%)	24.14	59.31	21.38	13.80	26.14	18.37	16.19	13.24	15.14	14.14	15.64	28.17
Turnover(%)	129.75	59.49	134.71	146.82	125.62	123.18	121.77	122.16	122.60	122.60	123.95	126.17
Break-even TC(%)	1.40	1.03	1.45	1.94	2.13	2.83	2.85	2.82	2.80	3.76	3.68	3.46

Table 6: Performance of Value-Weighted Portfolios Sorted on Conditional Volatility

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on robust conditional volatility forecasts. At the end of each month, we predict the one-month-ahead Vol, 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 Vol (decile 1), "High" denotes the portfolio with the highest Vol (decile 10), "H-L" denotes the long-short portfolio that buys the highest Vol 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 Value-Weighted Portfolios Sorted on Vol

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.69	0.69	0.66	0.72	0.64	0.59	0.64	0.59	0.63	0.56	0.56	0.56
2	0.71	0.76	0.70	0.78	0.71	0.72	0.68	0.74	0.71	0.73	0.70	0.70
3	0.77	0.70	0.76	0.76	0.77	0.78	0.79	0.79	0.77	0.75	0.77	0.77
4	0.80	0.81	0.77	0.74	0.75	0.84	0.74	0.81	0.85	0.75	0.79	0.79
5	0.83	0.85	0.90	0.84	0.78	0.81	0.89	0.86	0.83	0.87	0.82	0.82
6	0.73	0.62	0.70	0.72	0.82	0.88	0.82	0.82	0.76	0.88	0.87	0.87
7	0.61	0.53	0.65	0.72	0.83	0.60	0.64	0.62	0.65	0.70	0.78	0.65
8	0.57	0.09	0.50	0.68	0.32	0.34	0.37	0.37	0.44	0.70	0.49	0.58
9	0.26	-0.16	0.30	0.18	0.24	0.06	-0.02	0.02	-0.07	-0.02	0.21	-0.02
High (H)	-0.25	-0.65	-0.20	-0.15	-0.43	-0.70	-0.59	-0.47	-0.55	-0.39	-0.45	-0.42
H-L	-0.94	-1.35	-0.86	-0.87	-1.07	-1.29	-1.23	-1.06	-1.18	-0.95	-1.01	-0.98
t	(-1.42)	(-2.42)	(-1.31)	(-1.35)	(-1.64)	(-1.79)	(-1.69)	(-1.54)	(-1.69)	(-1.35)	(-1.42)	(-1.36)
SR	-0.29	-0.47	-0.27	-0.28	-0.33	-0.38	-0.37	-0.31	-0.35	-0.27	-0.29	-0.28

Panel B: Risk-Adjusted Performance of Value-Weighted Portfolios Sorted on Vol

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	-0.29	-0.49	-0.22	-0.09	-0.31	-0.56	-0.55	-0.31	-0.46	-0.36	-0.33	-0.28
t(α)	(-0.81)	(-1.7)	(-0.62)	(-0.33)	(-0.9)	(-1.57)	(-1.66)	(-1.08)	(-1.57)	(-1.05)	(-1.07)	(-0.84)
R^2	81	77	80	81	80	78	81	81	80	78	78	78
IR	-0.2	-0.35	-0.15	-0.07	-0.21	-0.35	-0.37	-0.21	-0.31	-0.22	-0.21	-0.17

Table 7: Performance of Value-Weighted Portfolios Sorted on Upside Volatility

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on upside volatility forecasts. At the end of each month, we predict the one-month-ahead Vol^+, 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 Vol^+ (decile 1), "High" denotes the portfolio with the highest Vol^+ (decile 10), "H-L" denotes the long-short portfolio that buys the highest Vol^+ 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 Value-Weighted Portfolios Sorted on Vol^+

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.47	0.67	0.45	0.54	0.62	0.46	0.43	0.50	0.51	0.18	0.26	0.47
2	0.59	0.67	0.60	0.64	0.68	0.67	0.62	0.62	0.60	0.44	0.51	0.58
3	0.71	0.77	0.70	0.70	0.65	0.73	0.72	0.70	0.74	0.74	0.74	0.70
4	0.85	0.73	0.89	0.71	0.81	0.71	0.68	0.77	0.69	0.84	0.84	0.80
5	0.79	0.74	0.78	0.86	0.67	0.70	0.71	0.76	0.80	0.83	0.77	0.80
6	0.77	0.75	0.82	0.83	0.67	0.79	0.78	0.89	0.87	0.73	0.78	0.78
7	0.89	0.56	0.85	0.77	0.63	0.88	0.94	0.85	0.84	0.88	0.78	0.77
8	1.04	0.60	0.90	0.90	0.89	0.86	1.01	0.90	0.90	0.79	0.90	0.83
9	0.80	0.52	0.83	0.87	0.64	0.89	0.88	1.04	0.96	1.17	1.14	1.19
High (H)	0.79	0.26	0.65	0.59	0.73	1.18	1.45	1.29	1.32	1.50	1.61	1.48
H-L	0.33	-0.41	0.20	0.05	0.11	0.71	1.02	0.79	0.82	1.31	1.34	1.01
t	(0.66)	(-0.84)	(0.39)	(0.11)	(0.23)	(1.15)	(1.75)	(1.37)	(1.47)	(2.11)	(2.03)	(1.66)
SR	0.14	-0.17	0.08	0.02	0.05	0.24	0.38	0.30	0.31	0.45	0.45	0.35

Panel B: Risk-Adjusted Performance of Value-Weighted Portfolios Sorted on Vol^+

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	0.49	-0.18	0.33	0.56	0.31	1.14	1.32	1.05	0.94	1.42	1.43	1.11
t(α)	(1.67)	(-0.69)	(1.12)	(2.01)	(1.12)	(3.55)	(3.84)	(3.36)	(2.79)	(3.42)	(3.78)	(3.11)
R^2	75	75	74	67	71	76	70	74	74	69	68	69
IR	0.41	-0.15	0.26	0.41	0.25	0.78	0.89	0.78	0.69	0.88	0.86	0.69

Table 8: Performance of Value-Weighted Portfolios Sorted on Downside Volatility

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on downside volatility forecasts. At the end of each month, we predict the one-month-ahead Vol^-, 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 Vol^- (decile 1), "High" denotes the portfolio with the highest Vol^- (decile 10), "H-L" denotes the long-short portfolio that buys the highest Vol^- 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 Value-Weighted Portfolios Sorted on Vol^-

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.81	0.73	0.82	0.82	0.67	0.71	0.78	0.75	0.75	0.72	0.68	0.67
2	0.81	0.75	0.79	0.79	0.79	0.78	0.73	0.78	0.78	0.83	0.81	0.84
3	0.76	0.71	0.75	0.74	0.74	0.76	0.74	0.76	0.75	0.75	0.75	0.74
4	0.74	0.87	0.75	0.70	0.74	0.73	0.74	0.76	0.79	0.71	0.75	0.72
5	0.75	0.70	0.68	0.64	0.73	0.80	0.69	0.78	0.75	0.72	0.74	0.73
6	0.54	0.68	0.57	0.59	0.80	0.78	0.78	0.70	0.73	0.83	0.79	0.83
7	0.43	0.30	0.47	0.67	0.70	0.53	0.54	0.52	0.47	0.60	0.58	0.58
8	0.40	-0.02	0.37	0.41	0.33	0.36	0.31	0.26	0.46	0.24	0.30	0.21
9	0.21	-0.40	0.13	0.17	0.20	-0.14	-0.07	-0.17	-0.13	-0.26	-0.20	-0.33
High (H)	-0.69	-1.04	-0.85	-0.68	-0.54	-1.18	-1.34	-1.13	-1.23	-1.43	-1.38	-1.30
H-L	-1.50	-1.77	-1.67	-1.49	-1.21	-1.89	-2.12	-1.88	-1.98	-2.15	-2.06	-1.97
t	(-2.49)	(-3.03)	(-2.72)	(-2.29)	(-1.86)	(-3.00)	(-3.14)	(-2.94)	(-3.12)	(-3.32)	(-3.03)	(-2.91)
SR	-0.53	-0.60	-0.56	-0.49	-0.38	-0.62	-0.68	-0.61	-0.65	-0.68	-0.63	-0.60

Panel B: Risk-Adjusted Performance of Value-Weighted Portfolios Sorted on Vol^-

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	-0.67	-0.83	-0.92	-0.73	-0.34	-1.18	-1.38	-1.18	-1.22	-1.49	-1.33	-1.22
t(α)	(-2.35)	(-2.72)	(-3.12)	(-2.26)	(-0.89)	(-4.45)	(-4.47)	(-4.38)	(-4.43)	(-4.51)	(-4.3)	(-3.82)
R^2	81	78	81	80	78	82	81	82	82	80	80	79
IR	-0.54	-0.61	-0.71	-0.54	-0.22	-0.9	-1.02	-0.91	-0.95	-1.04	-0.91	-0.81

Table 9: Performance of Value-Weighted Portfolios Sorted on Conditional Skewness

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on robust conditional skewness forecasts. At the end of each month, we predict the one-month-ahead Skew, 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 Skew (decile 1), "High" denotes the portfolio with the highest Skew (decile 10), "H-L" denotes the long-short portfolio that buys the highest Skew 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 Value-Weighted Portfolios Sorted on Skew												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.61	-0.01	0.23	0.55	-0.46	0.37	0.17	0.23	0.12	-0.08	0.03	-0.28
2	0.61	0.34	0.30	0.61	-0.02	0.45	0.39	0.45	0.43	0.31	0.17	0.06
3	0.63	0.37	0.54	0.60	0.33	0.62	0.47	0.55	0.60	0.51	0.34	0.47
4	0.64	0.45	0.53	0.56	0.72	0.69	0.57	0.58	0.60	0.63	0.47	0.80
5	0.53	0.61	0.64	0.64	0.63	0.72	0.60	0.53	0.59	0.73	0.66	0.95
6	0.77	0.76	0.63	0.75	0.69	0.68	0.69	0.69	0.76	0.78	0.75	0.99
7	0.76	0.78	0.77	0.72	0.69	0.63	0.71	0.76	0.77	0.86	0.82	1.03
8	0.84	0.73	0.87	0.74	0.70	0.75	0.79	0.82	0.75	0.79	0.90	1.02
9	0.97	0.81	0.99	0.73	0.84	0.79	0.85	0.89	0.78	0.92	1.00	1.13
High (H)	0.94	0.88	0.92	0.79	0.87	0.73	1.01	1.07	0.91	1.09	1.25	1.24
H-L	0.33	0.88	0.70	0.24	1.33	0.36	0.83	0.83	0.79	1.18	1.22	1.51
t	(1.33)	(1.90)	(2.67)	(2.41)	(3.03)	(1.68)	(4.40)	(4.86)	(4.26)	(4.02)	(3.52)	(4.32)
SR	0.28	0.37	0.53	0.42	0.59	0.34	0.83	0.91	0.81	0.77	0.67	0.83
Panel B: Risk-Adjusted Performance of Value-Weighted Portfolios Sorted on Skew												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	0.02	0.03	0.14	0.16	0.55	0.29	0.77	0.73	0.49	0.96	0.97	0.83
t(α)	(0.14)	(0.15)	(0.55)	(1.57)	(1.72)	(1.23)	(3.34)	(3.02)	(2.49)	(3.62)	(2.83)	(2.34)
R^2	47	85	35	12	51	10	18	10	14	23	20	14
IR	0.03	0.04	0.13	0.3	0.35	0.29	0.85	0.84	0.54	0.72	0.6	0.49

Table 10: Performance of Value-Weighted Portfolios Sorted on Value-at-Risk

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on Value-at-Risk forecasts. At the end of each month, we predict the one-month-ahead VaR, 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 VaR (decile 1), "High" denotes the portfolio with the highest VaR (decile 10), "H-L" denotes the long-short portfolio that buys the highest VaR 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 Value-Weighted Portfolios Sorted on VaR

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.76	0.72	0.74	0.74	0.70	0.72	0.72	0.69	0.71	0.67	0.66	0.64
2	0.75	0.73	0.77	0.72	0.76	0.74	0.75	0.76	0.75	0.78	0.78	0.78
3	0.80	0.81	0.79	0.77	0.75	0.80	0.78	0.80	0.79	0.72	0.74	0.76
4	0.75	0.73	0.74	0.72	0.77	0.85	0.83	0.81	0.76	0.71	0.77	0.74
5	0.79	0.68	0.77	0.75	0.66	0.78	0.76	0.76	0.80	0.83	0.84	0.82
6	0.56	0.42	0.62	0.76	0.78	0.82	0.70	0.68	0.70	0.91	0.83	0.85
7	0.43	0.15	0.38	0.55	0.68	0.53	0.63	0.58	0.53	0.68	0.64	0.58
8	0.44	-0.10	0.43	0.49	0.44	0.23	0.27	0.23	0.23	0.25	0.28	0.26
9	-0.08	-0.49	0.01	0.06	-0.03	-0.32	-0.20	-0.23	-0.22	-0.26	-0.20	-0.35
High (H)	-0.82	-0.57	-1.03	-0.65	-0.90	-1.26	-1.50	-1.29	-1.35	-1.16	-1.20	-1.02
H-L	-1.58	-1.29	-1.77	-1.39	-1.60	-1.98	-2.22	-1.98	-2.07	-1.83	-1.86	-1.67
t	(-2.50)	(-2.27)	(-2.76)	(-2.26)	(-2.27)	(-2.81)	(-3.30)	(-2.86)	(-3.03)	(-2.73)	(-2.67)	(-2.43)
SR	-0.52	-0.45	-0.58	-0.46	-0.47	-0.60	-0.71	-0.62	-0.66	-0.56	-0.55	-0.49

Panel B: Risk-Adjusted Performance of Value-Weighted Portfolios Sorted on VaR

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	-0.76	-0.33	-1	-0.55	-0.63	-1.18	-1.49	-1.22	-1.37	-1.18	-1.07	-0.87
t(α)	(-2.47)	(-0.87)	(-3.4)	(-1.69)	(-1.87)	(-3.72)	(-5.11)	(-3.87)	(-4.34)	(-3.54)	(-3.19)	(-2.51)
R^2	79	72	80	78	78	80	80	79	77	77	78	77
IR	-0.55	-0.22	-0.73	-0.39	-0.39	-0.8	-1.07	-0.83	-0.91	-0.75	-0.67	-0.53

Table 11: Performance of Sharpe-Managed Value-Weighted Portfolios

Panel A reports the realized Sharpe ratios of the Sharpe-Managed value-weighted decile portfolios. At the end of each month, we sort stocks into deciles based on their one-month-ahead robust risk premium forecast. The portfolios with the lowest and highest forecasts are labeled as "Low" and "High", respectively. We then form Sharpe-Managed portfolios by subdividing each portfolio into deciles based on their SR_Q forecasts. Within each decile, D1 (D10) denotes the portfolio with the lowest (highest) SR_Q forecasts. This results in deciles $\{Low_{D1}, ..., Low_{D10}\}$ and $\{High_{D1}, ..., High_{D10}\}$ within the Low and High portfolios, respectively. Here, Low_{D10} (Low_{D1}) and $High_{D10}$ ($High_{D1}$) denote subgroups of stocks with the highest (lowest) SR_Q forecasts from the Low and High portfolios, respectively. We reconstitute portfolios monthly using value weights. Panel B displays the monthly average returns of our Sharpe-Managed strategy, denoted as "H_{D10}-L_{D1}", which involves exclusively taking long positions on $High_{D10}$ and short positions on Low_{D1}. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the H_{D10}-L_{D1} long-short portfolios. Panel C 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 A: Realized Sharpe Ratios of Sharpe-Managed Value-Weighted Portfolios

	LR		LR-4		EN		RF		GBT		NN1		NN2		NN3		NN4		MT1		MT2		MT3	
	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10
Low (L)	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.3	0.0	-0.3	0.0	-0.5	-0.1	-0.4	-0.2	-0.7	-0.1	-0.5	-0.1	-1.1	-0.3	-1.2	-0.2	-1.2	-0.2
2	0.0	0.0	-0.1	0.3	0.3	0.1	0.1	0.1	-0.1	0.1	0.1	0.1	-0.1	0.1	0.1	0.0	0.1	0.0	-0.2	0.0	-0.2	0.1	-0.1	-0.1
3	0.1	0.4	-0.1	0.4	0.1	0.2	0.2	0.3	0.1	0.3	0.3	0.3	0.2	0.4	0.3	0.3	0.2	0.2	0.0	0.2	0.1	0.2	-0.1	0.1
4	0.1	0.3	0.1	0.5	0.2	0.3	0.1	0.1	0.0	0.2	0.2	0.5	0.2	0.4	0.2	0.4	0.0	0.4	0.4	0.2	0.3	0.3	0.1	0.3
5	0.1	0.4	0.0	0.5	0.1	0.4	0.2	0.3	0.1	0.3	0.3	0.8	0.2	0.6	0.0	0.7	0.4	0.7	0.3	0.4	0.2	0.3	0.3	0.4
6	0.0	0.5	0.1	0.7	0.1	0.5	0.2	0.4	0.3	0.4	0.2	0.7	0.3	0.8	0.3	0.7	0.4	0.8	0.5	0.6	0.4	0.6	0.4	0.7
7	0.2	0.5	0.0	0.8	0.2	0.6	0.3	0.4	0.2	0.4	0.1	0.8	0.3	0.8	0.3	0.8	0.1	0.9	0.6	0.8	0.6	0.8	0.5	0.9
8	0.2	0.7	0.3	0.9	0.3	0.5	0.2	0.5	0.2	0.5	0.4	1.0	0.4	0.9	0.4	1.0	0.4	1.1	0.6	1.0	0.4	1.1	0.4	1.0
9	0.3	0.8	0.2	0.9	0.3	0.9	0.0	0.8	0.4	0.5	0.2	1.1	0.3	0.9	0.2	1.1	0.4	1.2	0.6	1.2	0.5	1.0	0.5	1.3
High (H)	0.2	1.0	0.7	0.9	0.4	1.0	0.0	1.1	0.5	1.3	0.3	0.9	0.3	1.2	0.5	1.2	0.5	1.2	0.7	1.5	0.6	1.4	0.6	1.4

Panel B: Performance of Sharpe-Managed Value-Weighted Long-Short Portfolios

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
H_{D10}-L_{D1}	1.5	1.22	1.36	2.68	3.24	2.82	3.26	4.04	3.44	6.01	6.21	6.32
t	(4.57)	(2.19)	(3.96)	(4.75)	(5.38)	(6.11)	(6.52)	(7.92)	(6.6)	(10.01)	(12.24)	(11.29)
SR	0.85	0.42	0.79	0.95	1.04	1.15	1.31	1.65	1.42	2.25	2.51	2.41

Panel C: Risk-Adjusted Performance of Sharpe-Managed Value-Weighted Long-Short Portfolios

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	0.82	0.57	0.82	2.11	2.24	1.97	2.52	3.35	2.88	5.49	5.73	5.56
t(α)	(3.02)	(1.68)	(2.95)	(4.77)	(4.95)	(5.59)	(5.84)	(7.43)	(6.53)	(8.74)	(12.01)	(11.02)
R^2	40	70	43	36	46	42	35	30	32	23	25	27

Table 12: Value-Weighted Bivariate Portfolios Sorted on Changes and Levels of Machine Learning Forecasts

Panel A reports the monthly performance of value-weighted portfolios first sorted by the monthly change in the robust risk premium (Delta$\bar{Q}$), and then by the robust risk premium level ($\bar{Q}$) for the MT1 model. 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 $\bar{Q}$. Within each Delta$\bar{Q}$ tercile, "Low" denotes the portfolio with the lowest $\bar{Q}$ (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ (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. Panels B, C, and D report the monthly performance of value-weighted portfolios first sorted by DeltaSkew, DeltaVol^{-} and DeltaVaR, respectively, then by Skew, Vol^{-} and VaR.

	Panel A: Sorts based on Delta$\bar{Q}$ and $\bar{Q}$			Panel B: Sorts based on DeltaSkew and Skew			Panel C: Sorts based on DeltaVol^{-} and Vol^{-}			Panel D: Sorts based on DeltaVaR and VaR		
	Delta1	Delta2	Delta3	Delta1	Delta2	Delta3	Delta1	Delta2	Delta3	Delta1	Delta2	Delta3
Low (L)	-2.42	-0.52	-0.56	-0.49	0.01	0.09	0.91	0.69	0.43	0.83	0.51	0.34
2	-1.01	0.21	0.14	0.00	0.37	0.38	0.92	0.71	0.53	0.79	0.65	0.65
3	-0.24	0.33	0.71	0.13	0.64	0.67	1.00	0.78	0.43	0.68	0.74	0.69
4	-0.27	0.55	0.88	0.44	0.63	0.82	0.96	0.79	0.41	0.74	0.77	0.64
5	-0.11	0.75	0.98	0.59	0.68	0.83	0.90	0.63	0.52	0.98	0.69	0.50
6	0.47	0.87	1.30	0.55	0.79	1.03	0.72	0.63	0.26	0.68	0.83	0.48
7	0.51	0.79	1.35	0.71	0.80	1.16	0.73	0.70	0.00	0.46	0.86	0.13
8	0.85	1.11	1.62	0.76	0.64	0.96	0.28	0.78	-0.41	0.14	1.00	-0.44
9	0.89	1.17	1.76	0.77	0.96	1.22	0.08	0.57	-1.07	0.06	0.54	-0.77
High (H)	1.23	1.64	3.23	1.00	1.08	1.41	-0.72	-0.19	-2.51	-1.18	-0.45	-1.84
H-L	3.65	2.15	3.80	1.49	1.07	1.32	-1.62	-0.88	-2.94	-2.00	-0.96	-2.17
t	(9.37)	(5.91)	(7.05)	(4.51)	(3.29)	(4.19)	(-2.41)	(-1.56)	(-4.11)	(-2.98)	(-1.47)	(-3.02)
SR	1.78	1.18	1.55	0.90	0.62	0.78	-0.48	-0.31	-0.82	-0.60	-0.31	-0.59
α	2.89	1.39	3.49	1.17	0.85	1.25	-0.91	-0.62	-1.90	-1.23	-0.70	-1.07
t(α)	(8.09)	(4.61)	(6.02)	(4.03)	(2.86)	(3.40)	(-2.35)	(-2.34)	(-4.61)	(-3.02)	(-1.94)	(-2.57)

Table 13: Value-Weighted Bivariate Portfolios Sorted on Analyst Coverage and Machine Learning Forecasts

Panel A reports the monthly performance of value-weighted portfolios first sorted by analyst coverage (CVRG), and then by the robust risk premium for the MT1 model. 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 $\bar{Q}$. Within each CVRG tercile, "Low" denotes the portfolio with the lowest $\bar{Q}$ (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ (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. Panels B, C, and D report the monthly performance of value-weighted portfolios first sorted by CVRG, then by Skew, Vol^{-} and VaR, respectively.

	Panel A: Sorts based on Analyst Coverage and $\bar{Q}$			Panel B: Sorts based on Analyst Coverage and Skew			Panel C: Sorts based on Analyst Coverage and Vol^{-}			Panel D: Sorts based on Analyst Coverage and VaR		
	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3	CVRG1	CVRG2	CVRG3
Low (L)	-1.72	-1.71	-0.86	-0.27	-0.13	0.22	0.53	0.71	0.80	0.46	0.62	0.71
2	-0.23	-0.25	-0.19	0.24	0.12	0.43	0.63	0.67	0.80	0.60	0.72	0.75
3	0.29	-0.22	0.24	0.23	0.30	0.60	0.76	0.71	0.69	0.67	0.66	0.70
4	0.24	0.23	0.62	0.42	0.45	0.67	0.69	0.98	0.74	0.77	0.98	0.83
5	0.53	0.71	0.70	0.59	0.61	0.81	0.61	0.73	0.67	0.66	0.93	0.71
6	0.65	0.86	0.83	0.51	0.77	0.71	0.51	0.67	0.84	0.45	0.58	0.88
7	0.92	1.03	0.97	0.56	0.89	0.75	0.60	0.40	0.85	0.59	0.45	0.89
8	1.22	1.27	0.97	0.91	0.82	0.93	0.13	0.14	0.57	0.27	0.10	0.75
9	1.68	1.47	1.21	0.90	1.06	0.79	-0.07	-0.24	0.10	-0.40	-0.32	0.12
High (H)	2.59	2.29	1.39	0.99	1.31	0.90	-2.34	-1.77	-0.51	-2.05	-1.32	-0.39
H-L	4.31	3.99	2.25	1.26	1.44	0.68	-2.88	-2.48	-1.31	-2.51	-1.95	-1.10
t	(11.17)	(10.47)	(5.84)	(4.71)	(4.24)	(2.26)	(-4.25)	(-3.64)	(-2.06)	(-3.52)	(-2.82)	(-1.71)
SR	2.15	1.91	1.06	0.87	0.83	0.41	-0.87	-0.75	-0.41	-0.70	-0.58	-0.35
α	4.37	3.53	1.34	1.19	1.29	0.51	-2.49	-1.83	-0.53	-1.99	-1.20	-0.40
t(α)	(10.68)	(9.33)	(4.38)	(4.93)	(3.45)	(2.05)	(-6.35)	(-5.37)	(-1.66)	(-4.45)	(-3.34)	(-1.24)

Table 14: Value-Weighted Bivariate Portfolios Sorted on Arbitrage Index and Machine Learning Forecasts

Panel A reports the monthly performance of value-weighted portfolios first sorted by an index of arbitrage costs, and then by the robust risk premium for the MT1 model. 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 $\bar{Q}$. Within each AI quintile, "Low" denotes the portfolio with the lowest $\bar{Q}$ (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ (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. Panels B, C, and D report the monthly performance of value-weighted portfolios first sorted by AI, then by Skew, Vol^{-} and VaR, respectively.

	Panel A: Sorts based on Arbitrage Index and $\bar{Q}$					Panel B: Sorts based on Arbitrage Index and Skew					Panel C: Sorts based on Arbitrage Index and Vol^{-}					Panel D: Sorts based on Arbitrage Index and VaR				
	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5	AI1	AI2	AI3	AI4	AI5
Low (L)	-0.05	-0.32	-1.25	-2.29	-3.23	0.29	0.18	-0.13	-0.69	-1.50	0.82	1.04	1.12	1.00	1.01	0.69	0.86	0.89	0.91	0.90
2	0.49	0.00	-0.13	-0.41	-1.26	0.49	0.39	0.18	-0.45	-1.14	0.79	0.86	0.88	1.11	1.15	0.77	0.85	0.87	0.93	0.91
3	0.49	0.61	0.22	-0.06	-0.46	0.62	0.62	0.34	-0.03	-0.92	0.81	0.74	0.95	1.04	0.74	0.68	0.80	0.80	1.06	0.64
4	0.65	0.74	0.69	0.50	-0.19	0.65	0.93	0.44	0.02	-0.46	0.80	0.94	0.83	1.01	0.71	0.78	0.81	1.03	0.84	0.59
5	0.82	0.88	0.75	0.50	0.56	0.88	0.89	0.76	0.48	0.10	0.66	0.82	1.02	0.97	0.49	0.74	0.86	0.86	0.77	0.15
6	0.91	1.04	0.85	1.06	0.79	0.81	0.83	0.86	0.66	0.11	0.69	0.97	0.75	0.71	0.48	0.66	0.86	0.78	0.71	0.04
7	0.88	1.05	1.22	1.19	1.17	0.91	1.01	0.87	1.10	0.45	0.72	0.88	0.72	0.35	-0.24	0.81	0.92	0.80	0.37	-0.26
8	0.90	1.14	1.34	1.67	1.90	0.73	0.77	0.74	0.91	1.15	0.62	1.05	0.63	0.30	-0.41	0.64	1.05	0.60	0.24	-0.52
9	1.05	1.43	1.41	1.99	2.54	0.97	0.94	1.04	1.05	1.33	0.74	0.68	0.38	0.00	-1.06	0.80	0.81	0.33	-0.25	-0.86
High (H)	1.29	1.70	1.82	2.67	4.72	0.90	1.25	1.07	1.51	2.95	0.43	0.07	-0.78	-1.56	-3.06	0.54	0.23	-0.62	-1.01	-1.84
H-L	1.34	2.02	3.07	4.96	7.95	0.61	1.07	1.20	2.20	4.45	-0.40	-0.97	-1.90	-2.56	-4.07	-0.14	-0.63	-1.52	-1.92	-2.73
t	(5.66)	(5.29)	(8.50)	(12.31)	(13.58)	(2.51)	(3.33)	(3.23)	(6.07)	(10.00)	(-0.94)	(-1.62)	(-2.92)	(-3.70)	(-6.49)	(-0.32)	(-1.03)	(-2.30)	(-2.60)	(-3.94)
SR	0.98	1.00	1.57	2.68	3.10	0.46	0.58	0.61	1.14	2.07	-0.18	-0.33	-0.61	-0.77	-1.22	-0.06	-0.21	-0.48	-0.55	-0.72
α	0.82	1.20	2.31	4.37	8.21	0.44	0.86	1.07	1.85	4.28	-0.16	-0.53	-1.45	-1.88	-3.09	0.10	-0.18	-1.03	-1.15	-1.69
t(α)	(4.12)	(4.39)	(8.33)	(12.60)	(12.11)	(2.14)	(3.05)	(3.18)	(5.44)	(10.46)	(-0.79)	(-1.61)	(-5.20)	(-5.34)	(-6.79)	(0.46)	(-0.57)	(-3.41)	(-2.89)	(-2.94)

Table 15: Value-Weighted Bivariate Portfolios Sorted on Earnings Events and Machine Learning Forecasts

Panel A reports the monthly performance of value-weighted portfolios first sorted by earnings announcement status, then by the robust risk premium for the MT1 model. 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 $\bar{Q}$. Within each earnings announcement subsample, "Low" denotes the portfolio with the lowest $\bar{Q}$ (decile 1), "High" denotes the portfolio with the highest $\bar{Q}$ (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. Panels B, C, and D report the monthly performance of value-weighted portfolios first sorted by earnings announcement status, then by Skew, Vol^{-} and VaR, respectively.

	Panel A: Sorts based on Earnings Events and $\bar{Q}$		Panel B: Sorts based on Earnings Events and Skew		Panel C: Sorts based on Earnings Events and Vol^{-}		Panel D: Sorts based on Earnings Events and VaR	
	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)	-1.61	-0.57	-0.11	0.40	0.65	1.02	0.57	0.95
2	-0.44	0.08	0.15	0.53	0.72	1.06	0.67	0.89
3	-0.04	0.15	0.25	0.92	0.57	1.14	0.55	1.20
4	0.22	0.64	0.44	1.01	0.56	1.15	0.58	1.15
5	0.53	1.15	0.59	1.05	0.59	1.16	0.62	1.29
6	0.75	1.38	0.56	1.13	0.53	0.89	0.67	1.21
7	0.78	1.38	0.77	1.17	0.46	1.13	0.45	1.08
8	1.03	1.67	0.67	1.43	0.07	0.50	0.08	1.00
9	1.22	1.64	0.79	1.26	-0.47	0.00	-0.41	0.17
High (H)	1.82	2.26	0.98	1.75	-1.53	-0.47	-1.30	-0.38
H-L	3.43	2.82	1.09	1.35	-2.18	-1.49	-1.86	-1.33
t	(9.66)	(5.42)	(3.62)	(3.75)	(-3.28)	(-2.09)	(-2.67)	(-1.84)
SR	1.74	1.07	0.68	0.66	-0.66	-0.41	-0.55	-0.36
α	2.84	1.74	0.85	1.18	-1.59	-0.57	-1.25	-0.42
t(α)	(9.56)	(4.02)	(3.15)	(3.54)	(-5.50)	(-1.17)	(-4.01)	(-0.85)

Figure 1: Multi-task Neural Network Architecture

This figure illustrates an example of a multi-task neural network architecture for quantile forecasting. The "hard sharing" hidden layers within the MT interact and nonlinearly transform the input predictors into common latent features, imposing an economically motivated restriction of a common structure across quantiles. The MT 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.

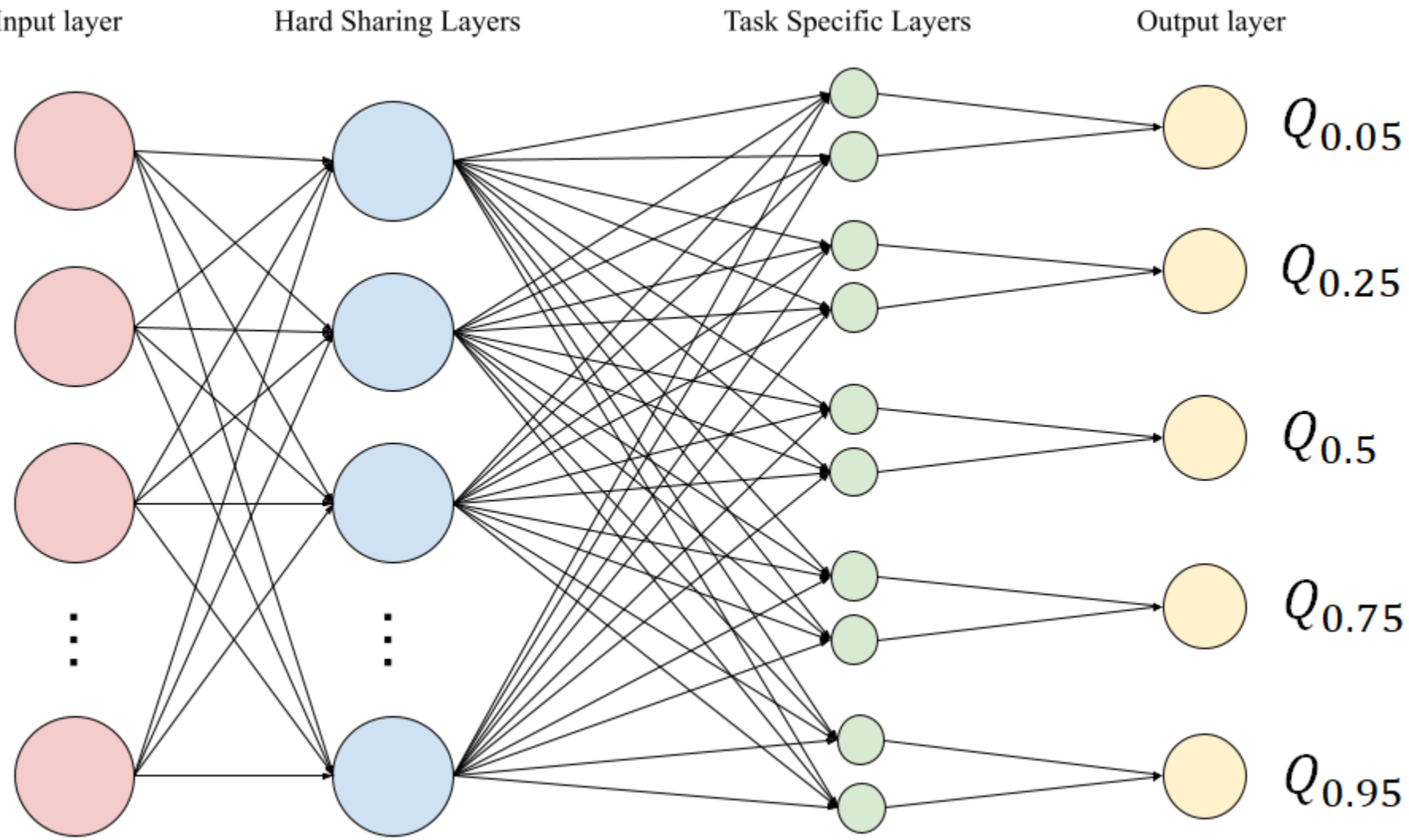


Figure 2: Robust Risk Premium Weights on Example Distribution

This figure illustrates how we derive the weights in Equation (3) 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 (4), 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.

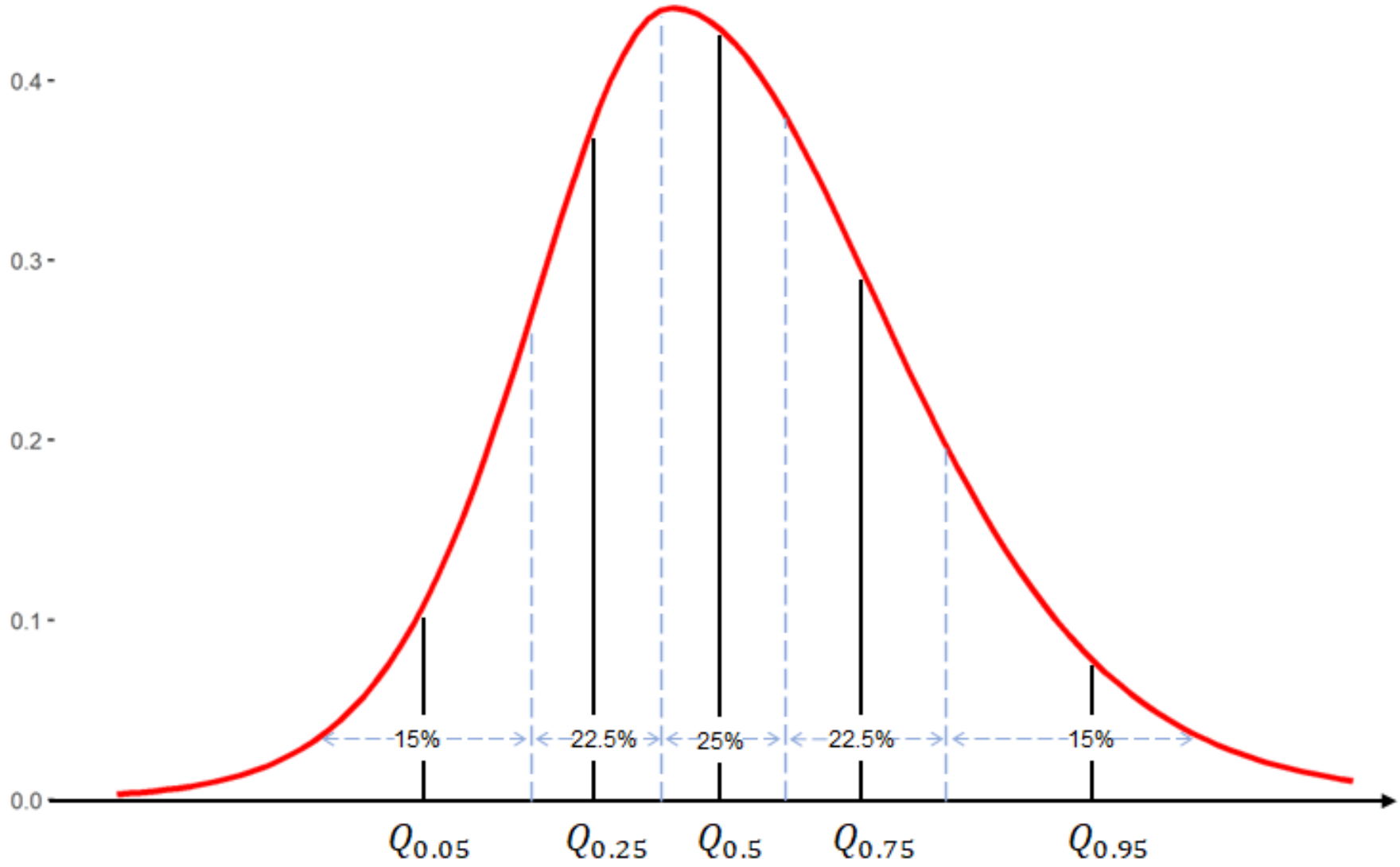


Figure 3: Cumulative Return of Long-Short Portfolios Sorted on Robust Risk Premium Measure $\bar{Q}$

This figure displays the cumulative log returns of long-short portfolios (value-weighted) sorted on robust risk premium measure $\bar{Q}$ for the twelve models. At the end of each month, we calculate the one-month-ahead forecasted conditional robust risk premium measure $\bar{Q} = 0.15Q_{0.05} + 0.225Q_{0.25} + 0.25Q_{0.5} + 0.225Q_{0.75} + 0.15Q_{0.95}$, then sort stocks into deciles based on $\bar{Q}$ into value-weighted portfolios. The High-Low portfolio corresponds to the long-short portfolio that buys stocks with the highest robust risk premium measure $\bar{Q}$ (decile 10) and sells stocks with the lowest (decile 1). The shaded periods show NBER recession dates.

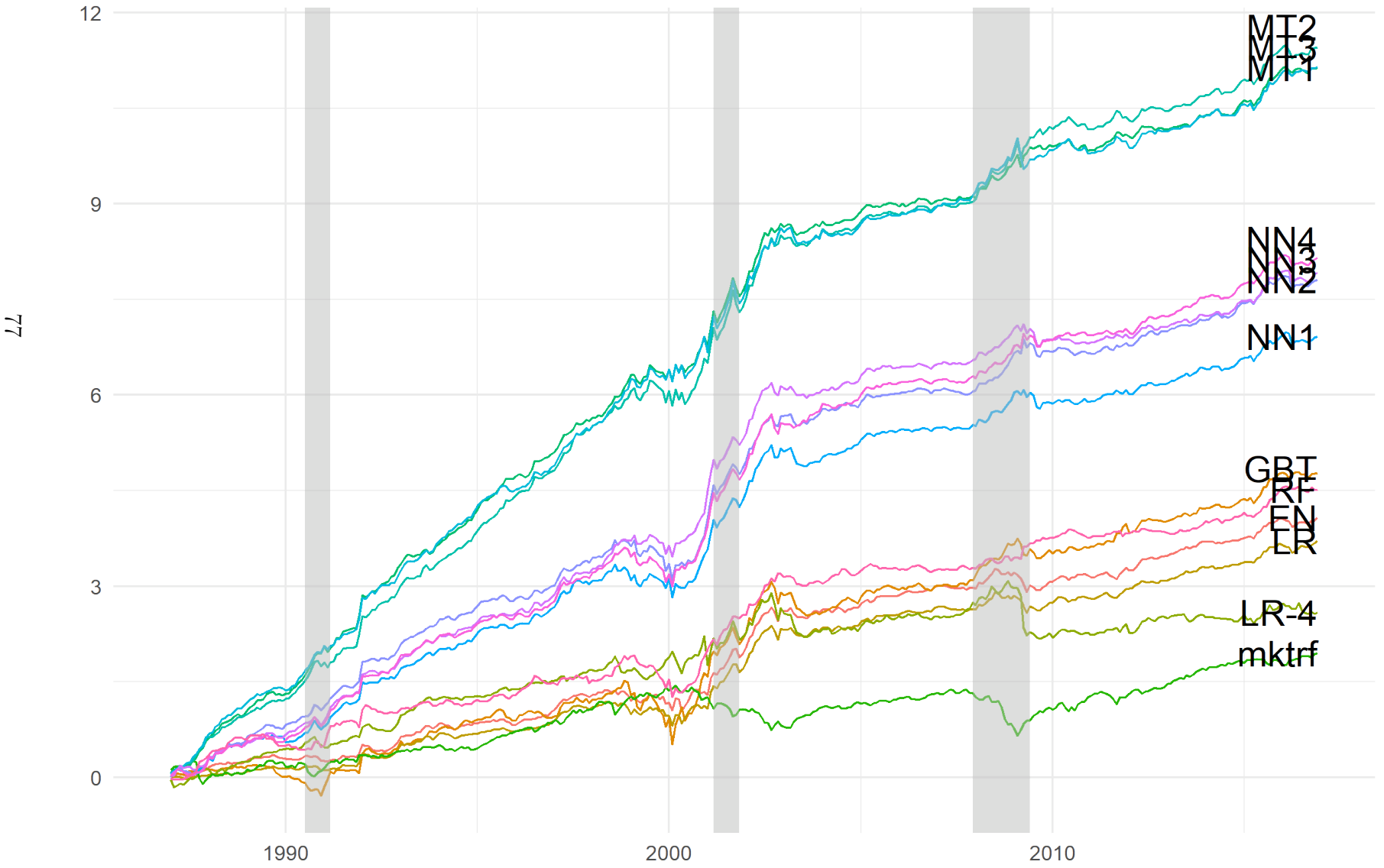


Figure 4: Quantile and Return Prediction Performance Conditional on Changes in Machine Learning Forecasts

This figure shows the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$ and R^2_{os} values (in percentages) of the MT1 model for stocks sorted by DeltaSkew. For each firm i, we define DeltaSkew as the Skew forecast at the end of month t minus the Skew forecast at the end of month $t-1$, then sort stocks into terciles based on their DeltaSkew. Stocks undergoing large positive or negative information shocks from month $t-1$ to t will exhibit correspondingly high positive or negative DeltaSkew values (Delta1 or Delta3), while stocks with minimal changes in forecasts will have only moderate DeltaSkew values (Delta2).

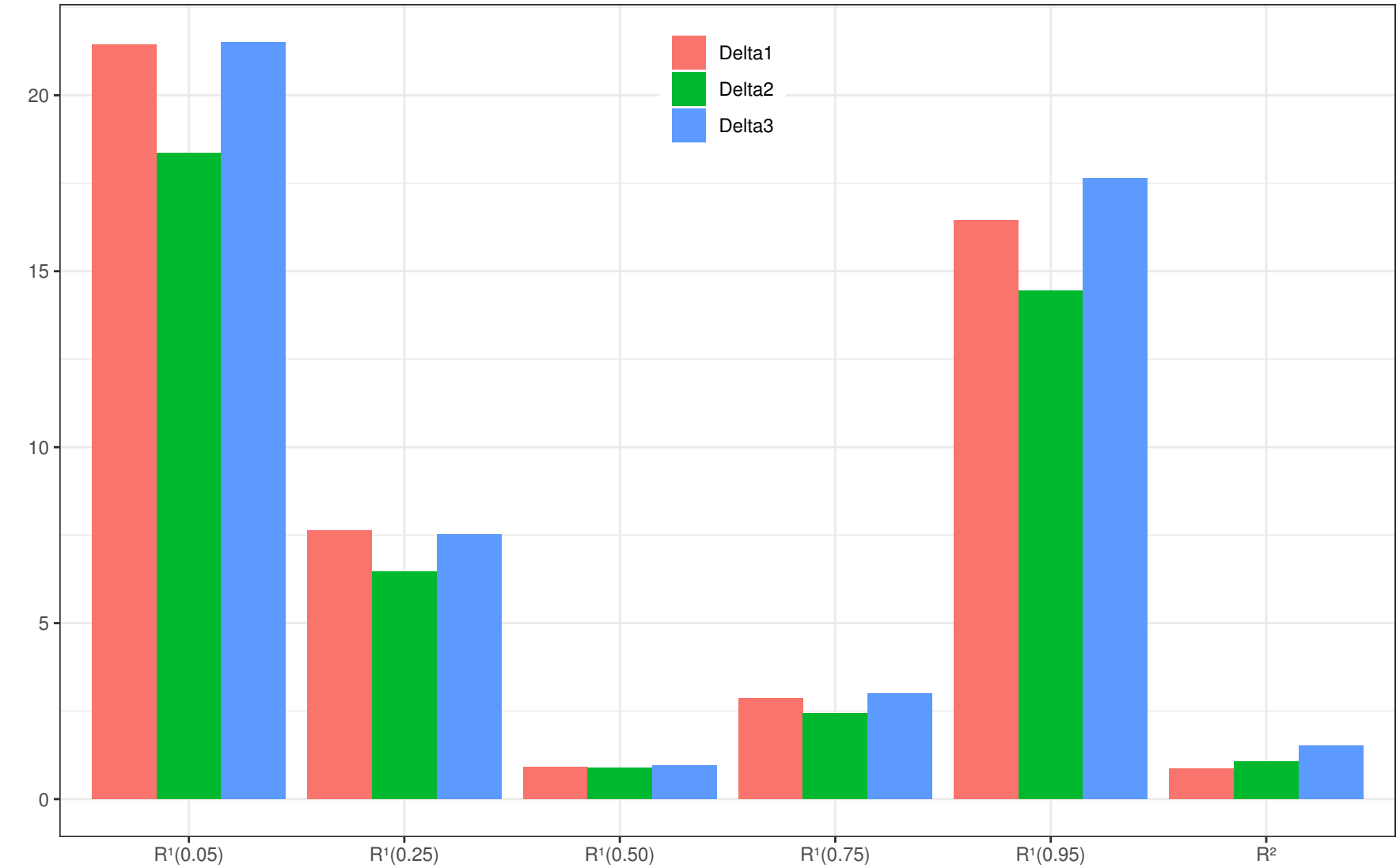


Figure 5: Quantile and Return Prediction Performance Conditional on Investor Inattention

This figure shows the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$ and R^2_{os} values (in percentages) of the MT1 model for stocks sorted by a proxy for investor attention. We sort stocks into terciles based on the number of analysts following the firm in the most recent month. Stocks with less analyst coverage (CVRG1) receive less investor attention, while stocks with more analyst coverage (CVRG3) receive more investor attention.

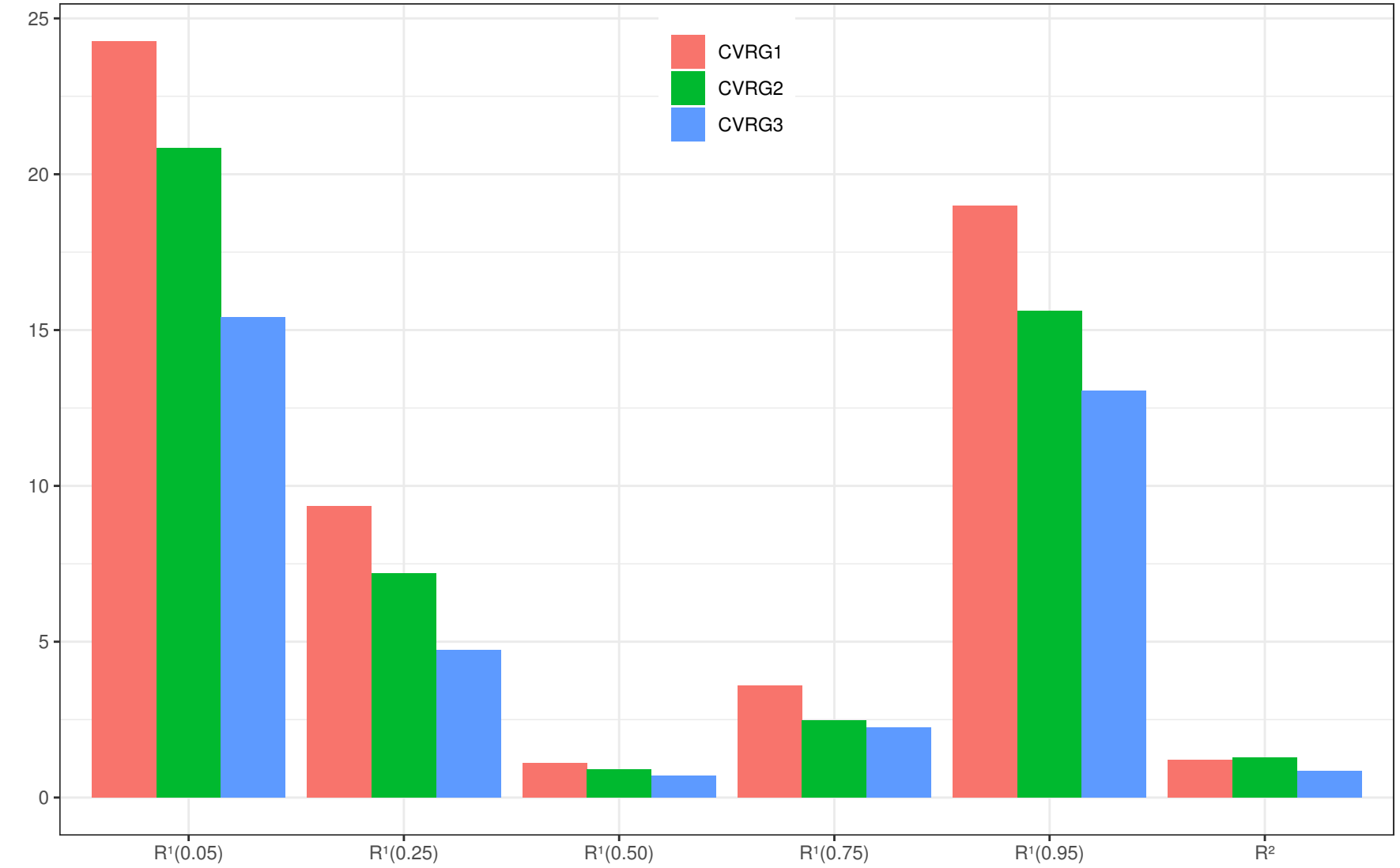


Figure 6: Quantile and Return Prediction Performance Conditional on Information Frictions

This figure shows the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$ and R^2_{os} values (in percentages) of the MT1 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.

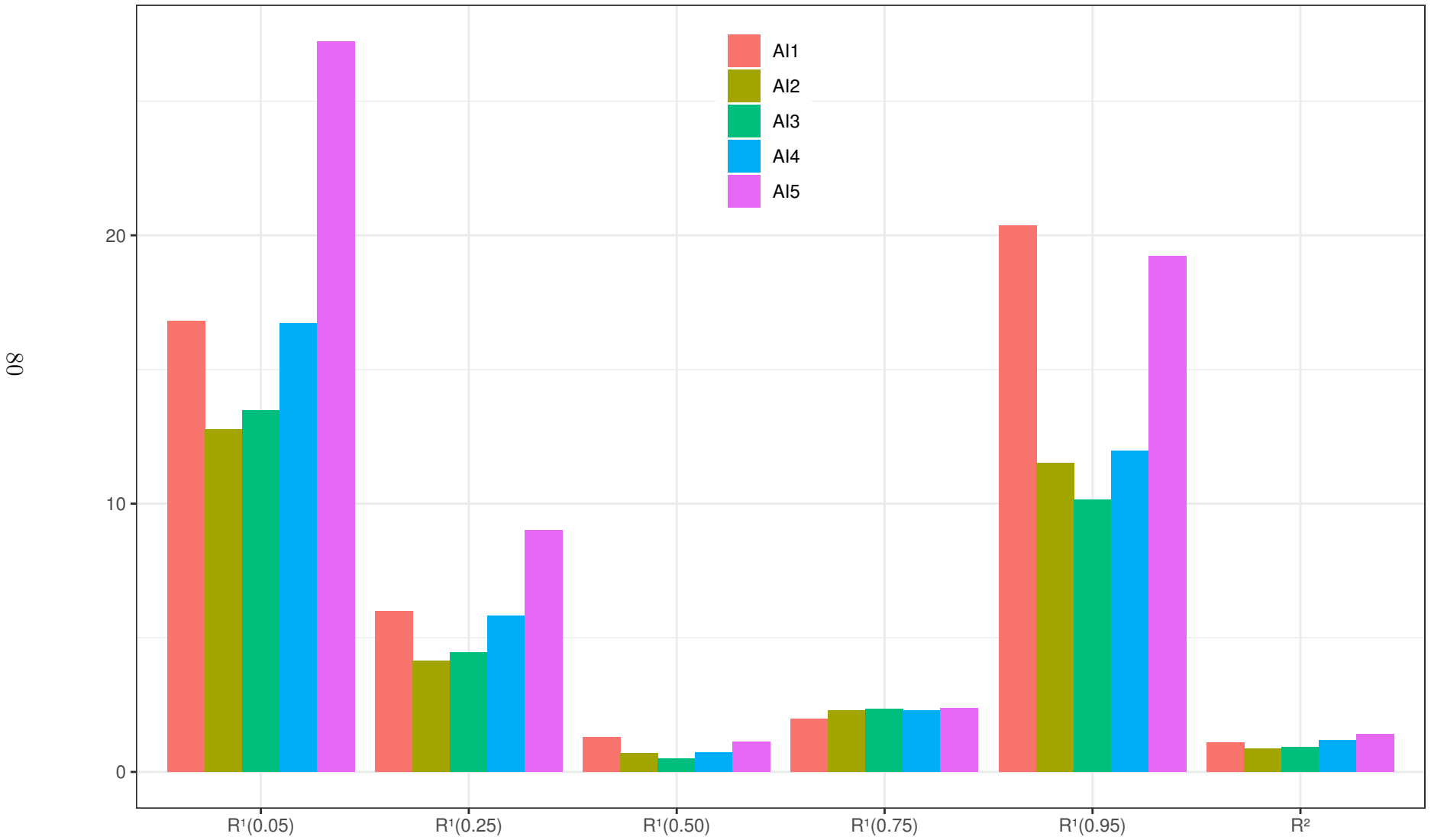


Figure 7: Quantile and Return Prediction Performance Conditional on Earnings Announcement

This figure shows the $R^1_{os}(\tau)$ values (in percentages) for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$ and R^2_{os} values (in percentages) of the MT1 model for stocks sorted by earnings announcement status. 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.

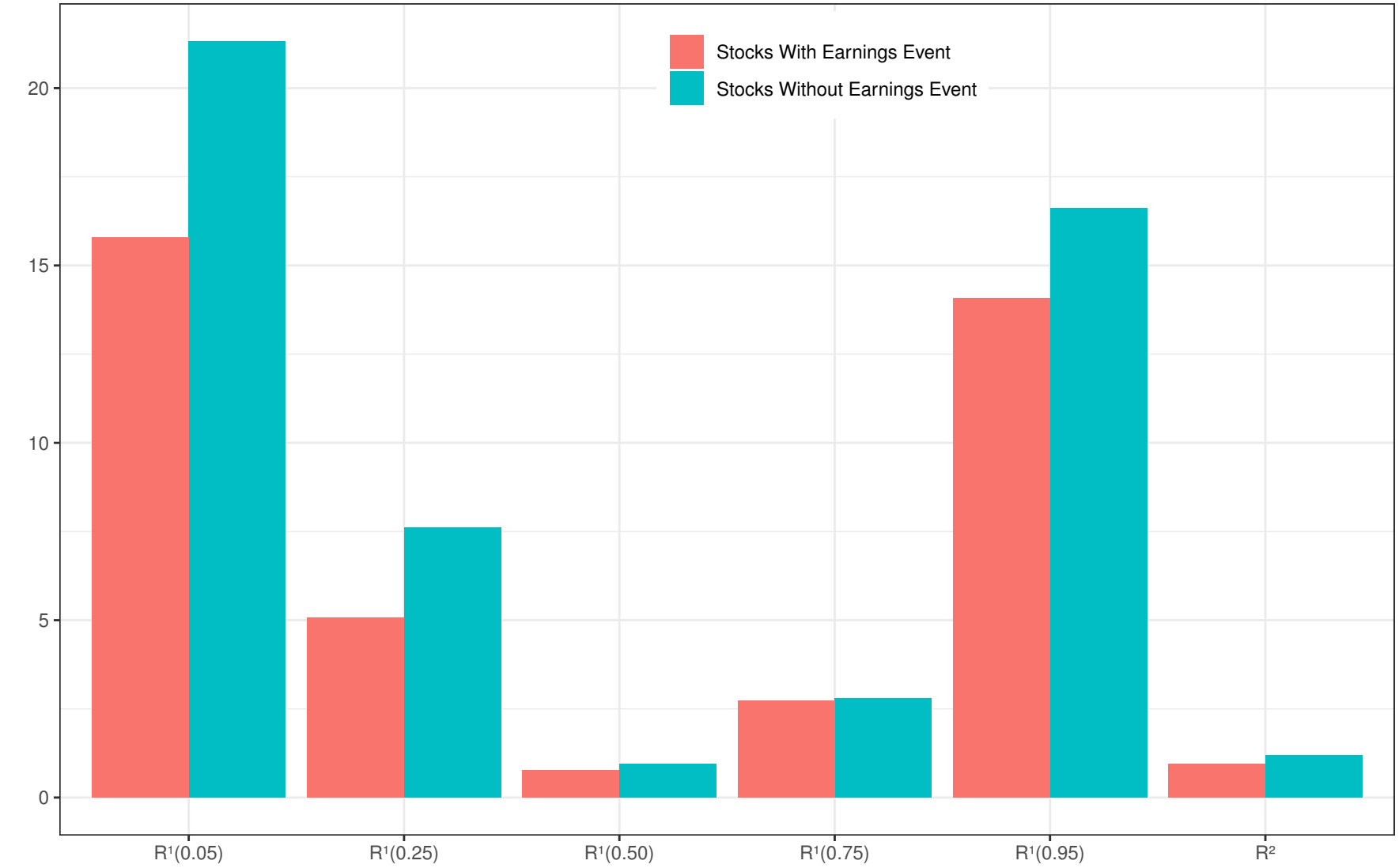
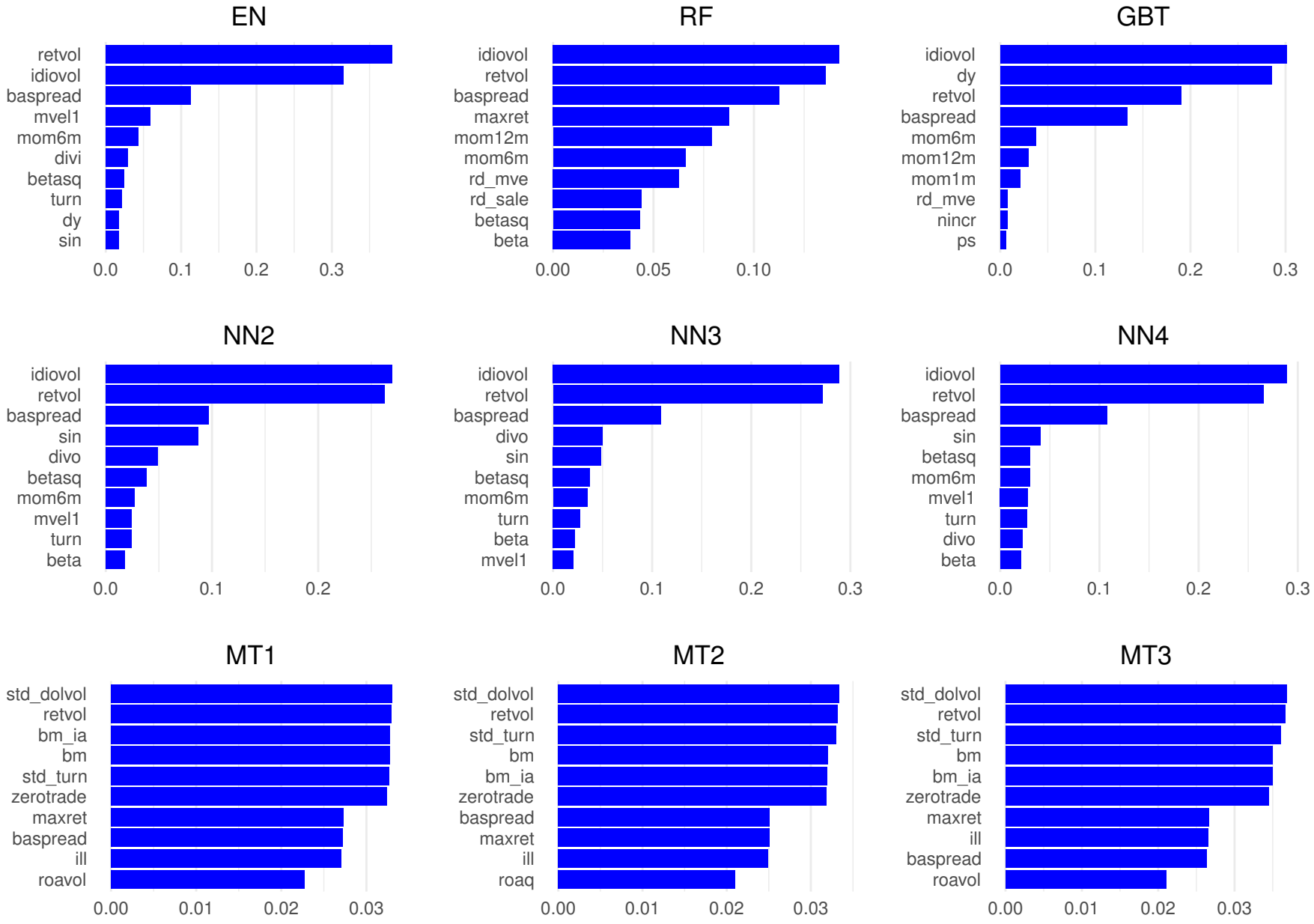
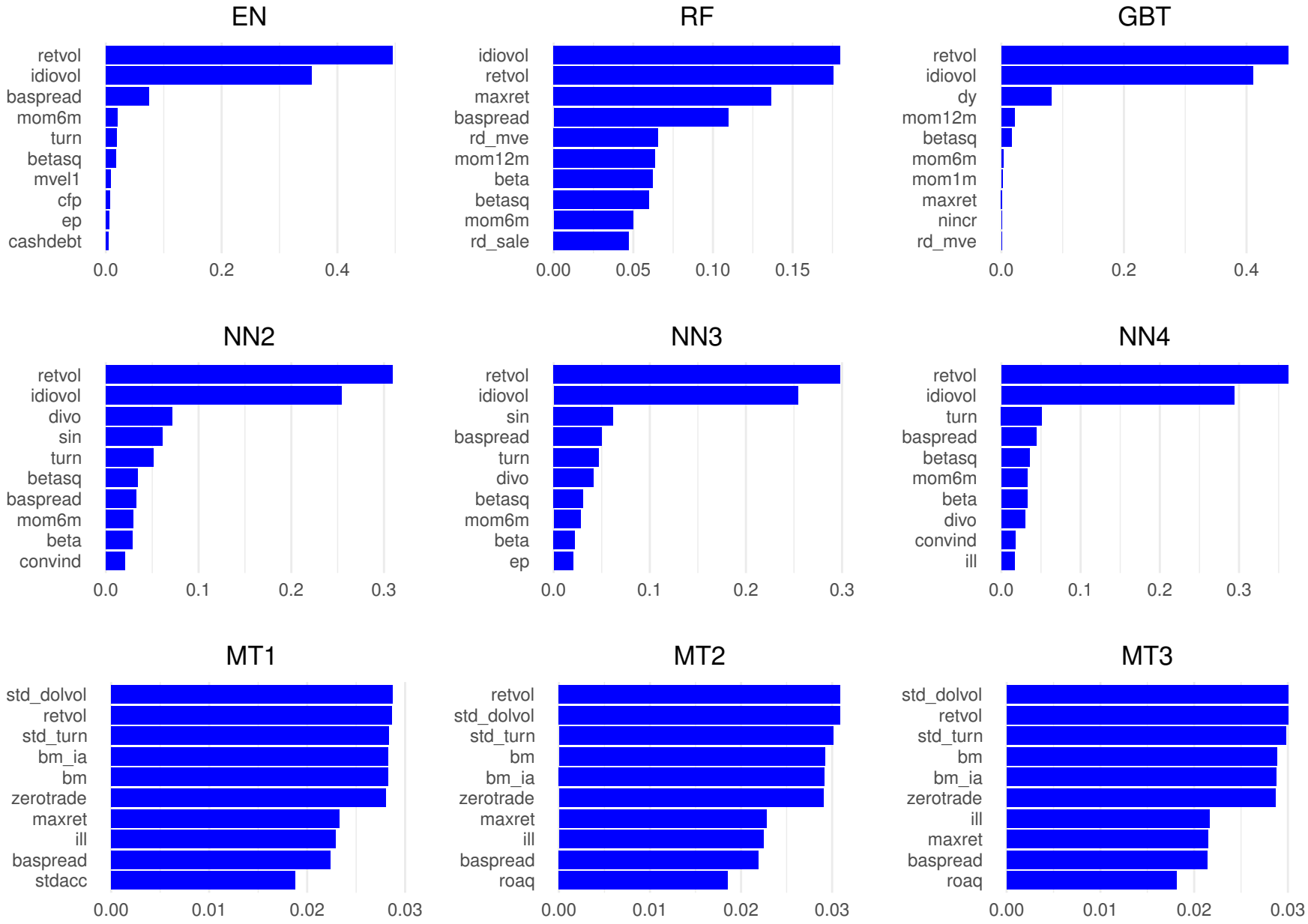


Figure 8: Characteristic Variable Importance by Model for Each Quantile

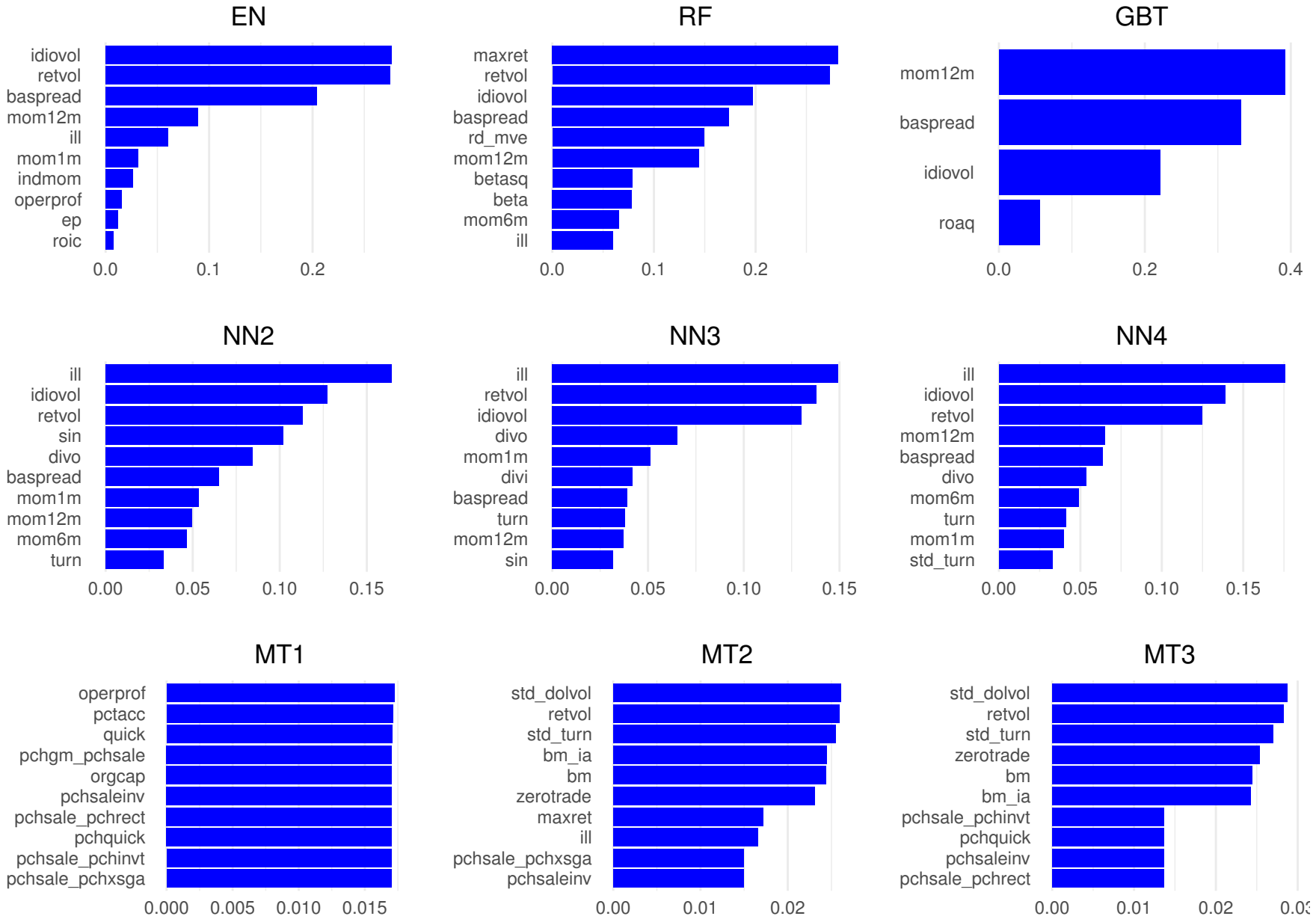
This figure reports the variable importance for the top 10 most influential characteristics in the models for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$. For each model and each quantile τ, the variable importance of the j^{th} predictor is measured by calculating the reduction in $R^1_{os}(\tau)$ from setting all values of the j^{th} predictor to zero within each sample, and then averaging the $R^1_{os}(\tau)$ reductions across samples as a single importance measure. For every τ, the variable importance within each model is normalized to sum to one.



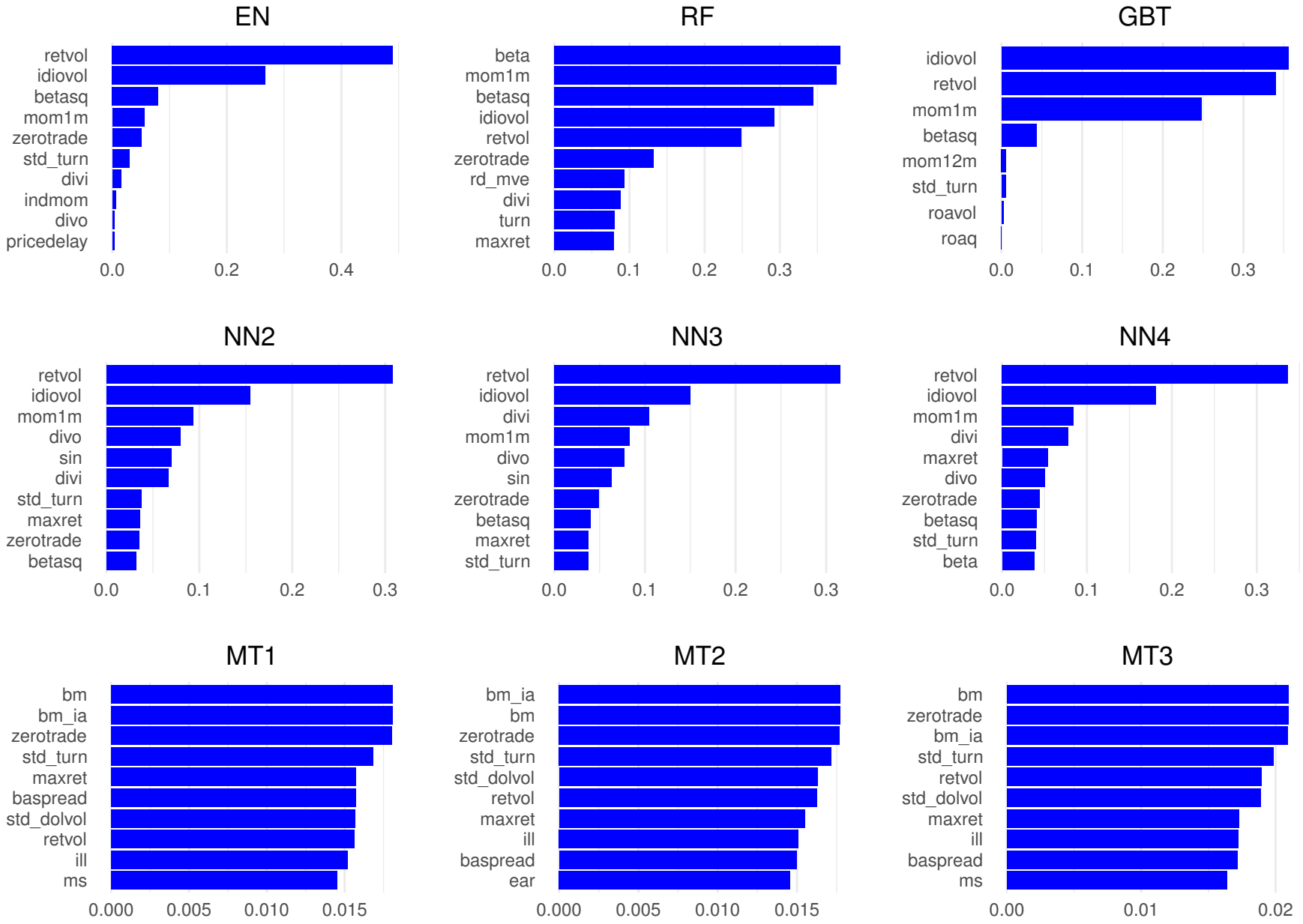
Panel A: Variable Importance Plot for $\tau = 0.05$



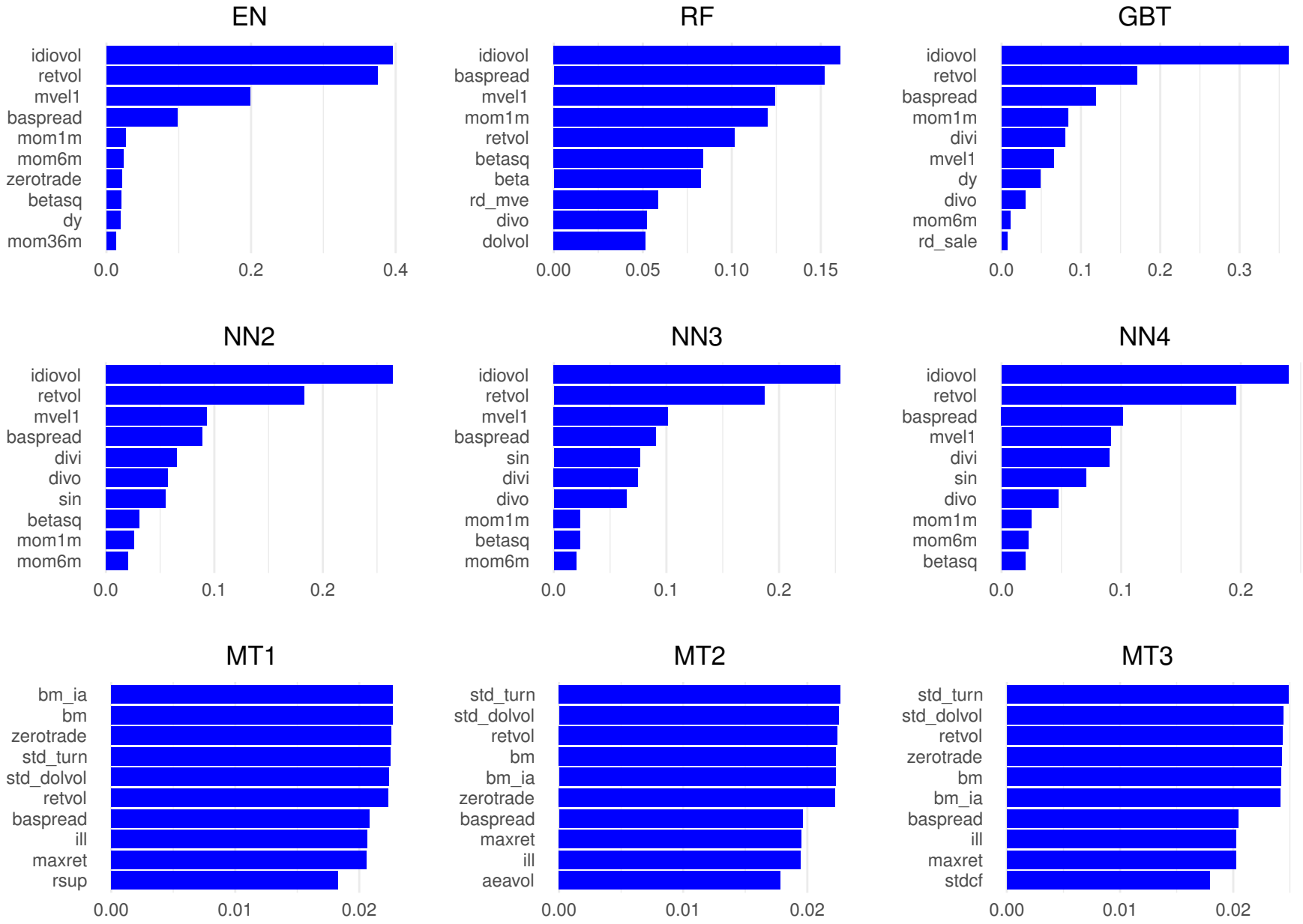
Panel B: Variable Importance Plot for $\tau = 0.25$



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



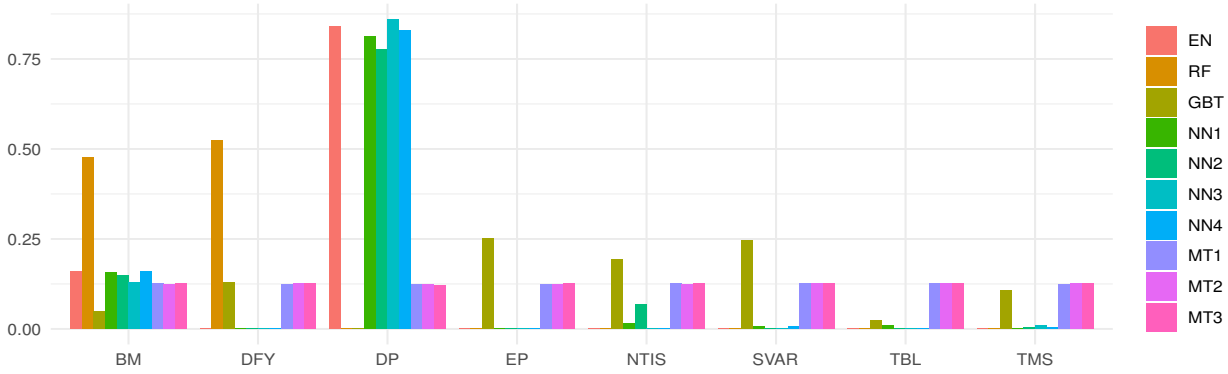
Panel D: Variable Importance Plot for $\tau = 0.75$



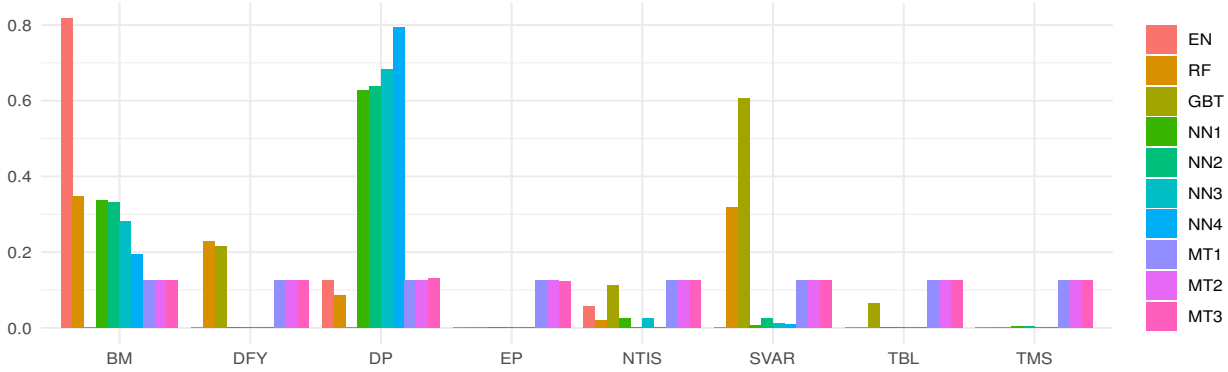
Panel E: Variable Importance Plot for $\tau = 0.95$

Figure 9: Macroeconomic Variable Importance by Model for each Quantile

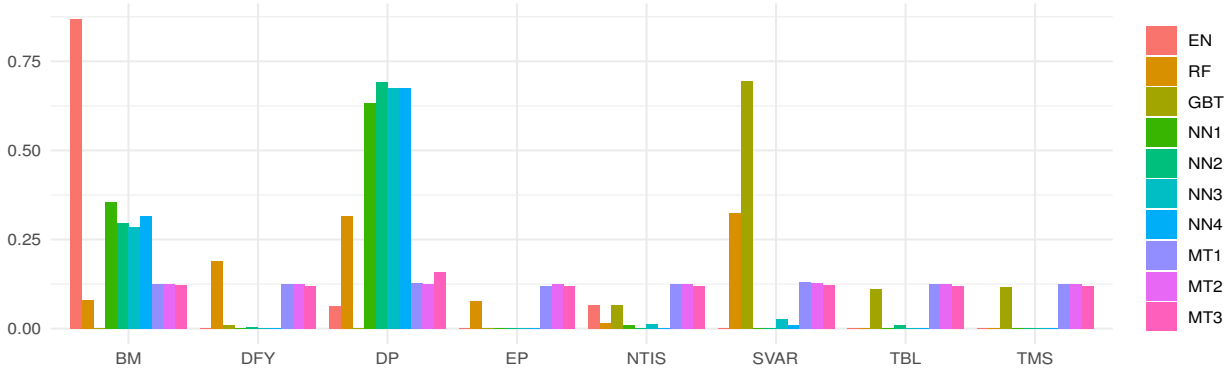
This figure reports the variable importance of the macroeconomic variables for all models in each quantile τ ($\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$). We use eight macroeconomic predictors including the book-to-market ratio (BM), default spread (DFY), dividend-price ratio (DP), earnings-price ratio (EP), net equity expansion (NTIS), stock variance (SVAR), Treasury-bill rate (TBL), and term spread (TMS). For each model and each quantile τ, 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 sample, and then averaging the $R^1_{os}(\tau)$ reductions across samples as a single importance measure. For every τ, the variable importance within each model is normalized to sum to one.



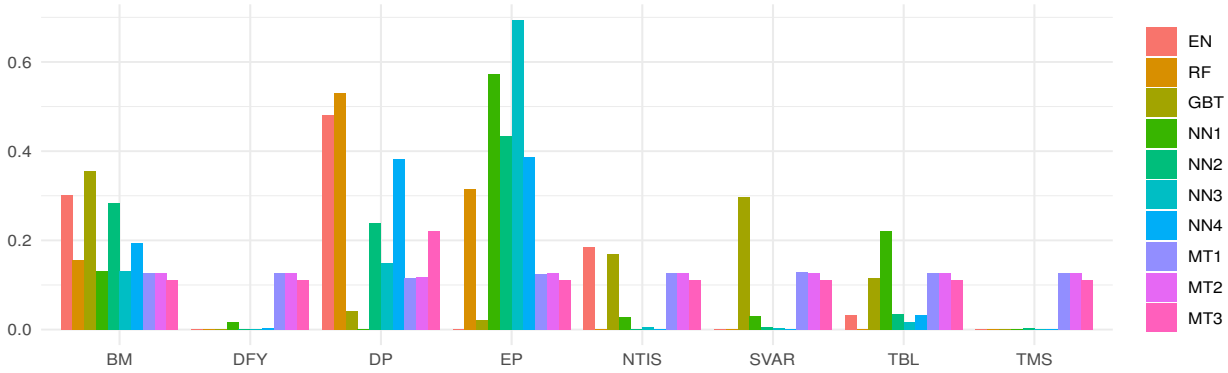
Panel A: Macroeconomic Variable Importance Plot for $\tau = 0.05$



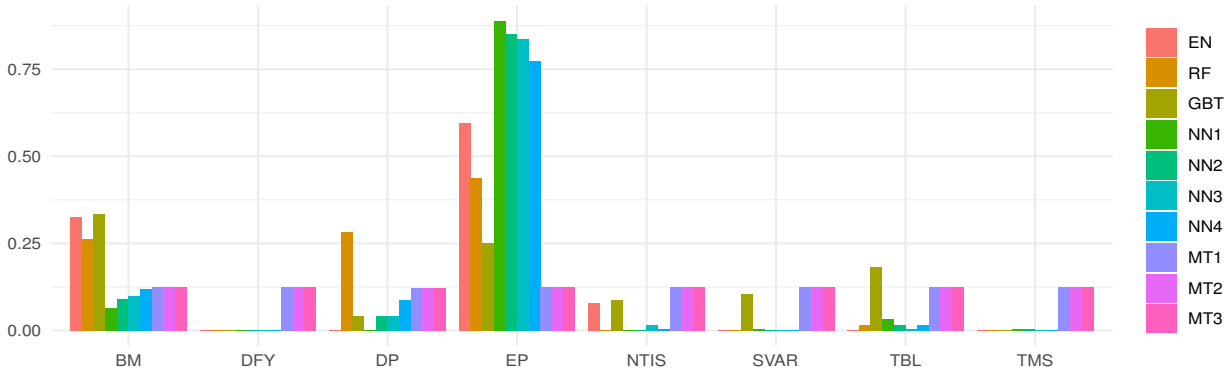
Panel B: Macroeconomic Variable Importance Plot for $\tau = 0.25$



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



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



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

Quantile Machine Learning and the Cross-Section of Stock Returns

Internet Appendix

Table of Contents:

- Section IA1 provides details on the **machine learning methods** 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**.
- Table IA1 presents the **set of hyperparameters** used in this paper.
- Table IA2 provides R^2_{os} values and value-weighted portfolios sorted on an **alternative risk premium** forecast based on Yang, Zhu, Li, and Zhu (2023).
- Table IA3 provides **equal-weighted** portfolios sorted on the robust risk premium.
- Table IA4 provides **equal-weighted** portfolios sorted on robust conditional volatility.
- Table IA5 provides **equal-weighted** portfolios sorted on robust conditional skewness.
- Table IA6 provides value-weighted portfolios sorted on **right-tail measure,** $Q_{0.95}$.
- Table IA7 provides Sharpe-Managed **equal-weighted** portfolios.
- Table IA8 provides a robustness check of the portfolios **with NYSE breakpoints** sorted on the robust risk premium.
- Table IA9 provides a robustness check of the portfolios **with NYSE breakpoints** sorted on robust conditional skewness.
- Table IA10 provides a robustness check of the portfolios sorted on the robust risk premium **excluding microcap stocks**.
- Table IA11 provides a robustness check of the portfolios sorted on robust conditional skewness **excluding microcap stocks**.

IA1 Machine Learning Models

This section introduces several methods for estimating the function g_τ in Equation (1). Since machine learning methods 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.

IA1.2.4 Multi-Task Neural Networks

With traditional feed-forward neural networks, a separate model is required for each quantile τ, which implicitly assumes that the functional forms of each quantile are independent. However, from an economic perspective, the conditional quantiles for a given firm-month are likely to be highly related, and the functional forms across quantiles should exhibit commonalities.

Multi-task learning enhances model generalization by concurrently training several related tasks. Multi-task neural network (MT) architectures are designed to jointly predict all targeted quantiles within a single model. An example of a multi-task neural network architecture is depicted in Figure 1. The "hard sharing" hidden layers within the MT interact and nonlinearly transform the input predictors into common latent features, imposing an economically motivated restriction of a common structure across quantiles. The MT 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.

Multi-task learning acts as a regularization technique 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 multi-task neural network to excel at jointly predicting multiple quantiles, the network is prevented from overfitting on quantile-specific noise, improving its generalizability. This simultaneous learning of related tasks, called inductive transfer, increases data efficiency by utilizing additional information from the training signals of related tasks (Caruana (1997)).

MTs are estimated using the sum of the quantile loss functions as the objective function. For $k = 1, .., K$ quantiles given τ_k, the MT objective function is

$$\mathcal{L}(\theta) = \sum_{k=1}^{K} \mathcal{L}_{\tau_k}(\theta), \tag{24}$$

where quantile objective function $\mathcal{L}_{\tau_k}(\theta)$ is from equation (2). When $g_\tau(\cdot)$ is linear, objective function (24) is referred to as the Hogg function (Koenker (2005)), which has been used to estimate linear multi-quantile models (see, e.g., Catania and Luati (2023), White, Kim, and Manganelli (2015)).

We evaluate MT architectures featuring up to three hard sharing layers and one task-specific layer for each quantile. For all MT architectures, we fix each task-specific layer to comprise 8 neurons for each quantile. The shallowest architecture, denoted as MT1, contains a single hard sharing layer with 32 neurons. The subsequent architecture, MT2, incorporates

two hard-sharing layers with 32 and 16 neurons, respectively. Lastly, MT3 includes three hard-sharing layers with 32, 16, and 8 neurons, respectively.

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

$$\begin{aligned}
&\bar{Q}_K = (\tau_1 + \frac{\tau_2 - \tau_1}{2})Q_{\tau_1} + (\frac{\tau_2 - \tau_1}{2} + \frac{\tau_3 - \tau_2}{2})Q_{\tau_2} + \cdots + \\
&(\frac{\tau_{K-1} - \tau_{K-2}}{2} + \frac{\tau_K - \tau_{K-1}}{2})Q_{\tau_{K-1}} + (\frac{\tau_K - \tau_{K-1}}{2} + (1 - \tau_K))Q_{\tau_K} = \\
&= (\frac{\tau_2 + \tau_1}{2})Q_{\tau_1} + (\frac{\tau_3 - \tau_1}{2})Q_{\tau_2} + \cdots + (\frac{\tau_K - \tau_{K-2}}{2})Q_{\tau_{K-1}} + (1 - \frac{\tau_K + \tau_{K-1}}{2})Q_{\tau_K}
\end{aligned} \tag{25}$$

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

$$E_t(r_{i,t+1}) = \int_{\mathbb{R}} x d\mu_{F_t}, \tag{26}$$

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

$$\lim_{\substack{\Delta_1 \to 0 \\ \Delta_2 \to 0}} \bar{Q}_K = E_t(r_{i,t+1}). \tag{27}$$

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

$$E_t(r_{i,t+1}) = \int_a^b x dF_t(x). \tag{28}$$

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

$$RS_t(K) = \sum_{k=1}^{K} c_k (F_t(x_k) - F_t(x_{k-1})), \tag{29}$$

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

$$\lim_{\delta_K \to 0} RS_t(K) = \int_a^b x dF_t(x) = E_t(r_{i,t+1}), \tag{30}$$

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

$$|RS_t(K) - E_t(r_{i,t+1})| < \epsilon. \tag{31}$$

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 $\Delta_2 = \max_{2 \leq k \leq K-1} \{Q_{\frac{\tau_{k+1}+\tau_k}{2}} - Q_{\frac{\tau_k+\tau_{k-1}}{2}}\} < \delta_K$, then

$$
\begin{aligned}
RS_t(K) &= \sum_{k=1}^{K} c_k(F_t(x_k) - F_t(x_{k-1})) \\
&= (F_t(Q_{(\frac{\tau_2+\tau_1}{2})}) - F_t(a))Q_{\tau_1} + (F_t(Q_{(\frac{\tau_3+\tau_2}{2})}) - F_t(Q_{(\frac{\tau_2+\tau_1}{2})}))Q_{\tau_2} + \cdots \\
&+ (F_t(Q_{(\frac{\tau_K+\tau_{K-1}}{2})}) - F_t(Q_{(\frac{\tau_{K-1}+\tau_{K-2}}{2})}))Q_{\tau_{K-1}} + (F_t(b) - F_t(Q_{(\frac{\tau_K+\tau_{K-1}}{2})})Q_{\tau_K} \\
&= (\frac{\tau_2+\tau_1}{2} - F_t(a))Q_{\tau_1} + (\frac{\tau_3+\tau_2}{2} - \frac{\tau_2+\tau_1}{2})Q_{\tau_2} + \cdots \\
&+ (\frac{\tau_K+\tau_{K-1}}{2} - \frac{\tau_{K-1}+\tau_{K-2}}{2})Q_{\tau_{K-1}} + (F_t(b) - \frac{\tau_K+\tau_{K-1}}{2})Q_{\tau_K} \\
&= (\frac{\tau_2+\tau_1}{2} - F_t(a))Q_{\tau_1} + (\frac{\tau_3-\tau_1}{2})Q_{\tau_2} + \cdots + (\frac{\tau_K-\tau_{K-2}}{2})Q_{\tau_{K-1}} + (F_t(b) - \frac{\tau_K+\tau_{K-1}}{2})Q_{\tau_K} \\
&= \bar{Q}_K - F_t(a)Q_{\tau_1} - (1 - F_t(b))Q_{\tau_K},
\end{aligned}
\tag{32}
$$

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

$$
|\bar{Q}_K - E_t(r_{i,t+1})| = |RS_t(K) - E_t(r_{i,t+1})| < \epsilon, \tag{33}
$$

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 $r_{i,t+1} \in (-\infty, \infty)$.[15]

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 $b > d > 0$,

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

$$
\begin{aligned}
&|\int_a^b xdF_t(x) - E_t(r_{i,t+1})| < \frac{\epsilon}{4}, \\
&|\int_{-\infty}^a xdF_t(x)| < \frac{\epsilon}{4}, \\
&|\int_b^{\infty} xdF_t(x)| < \frac{\epsilon}{4}.
\end{aligned} \tag{34}
$$

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

$$
|RS_t(K) - \int_a^b xdF_t(x)| < \frac{\epsilon}{4}, \tag{35}
$$

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

This section discusses in more detail the hyperparameters that are tuned for each of the models used in this paper using the validation set approach discussed in Section 3. For the linear EN quantile regression, 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. 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. For RF, the main hyperparameter is the depth of each tree.

Feed-forward neural networks require choosing many hyperparameters to find the best model flexibility. We guard against model overfitting by adding ℓ^1 and ℓ^2 regularization to the loss function with various degrees of shrinkage, and a max sum of squared incoming weights per unit as another form of shrinkage. We employ early stopping and optimize the number of training iterations, or epochs, as dual measures to mitigate the risk of overfitting. In addition, we implement the ensemble approach from Gu, Kelly, and Xiu (2020), by averaging predictions across a set of ten neural networks, thus mitigating prediction variance attributable to the inherent stochasticity of the optimization process. Models are estimated using stochastic gradient descent optimization with the ADADELTA (Zeiler (2012)) adaptive learning rate method for faster estimation without requiring manual tuning of the learning rate. Rho and epsilon are hyperparameters under the ADADELTA method that control the learning rate decay and rate time smoothing (to avoid dividing by zero), respectively.

For multi-task neural networks, we use the Relu activation in all hidden units and apply the ADAM optimizer to estimate the network weights. The regularization techniques for MT architectures resemble those for feed-forward neural networks, encompassing the addition of ℓ^1 regularization, implementation of early stopping, and the averaging of predictions across ten different MT configurations. Additionally, we employ batch normalization to control the

Table IA1: Hyperparameters

EN	$\lambda \in (10^{-4}, 10^{-1})$
	$\alpha = 0.5$
GBT	#Trees = 200∼800
	Tree depth = 1∼4
	Learning rate = {0.01,0.1}
	Sample rate = 0.5
	Patience = 5
RF	#Trees = 300
	#Features in each split = floor(#Predictors/3) = 34
	Tree depth = 20∼50
NN1-NN4	Activation Functions = Relu
	Epochs = {50, 100}
	Max w^2 (max squared sum of weights) = {10, 100, 1000, 3.4028235e+38}
	ℓ^1 = {0, 1e-5, 1e-4}
	ℓ^2 = {0, 1e-5, 1e-4}
	Rho (rate time decay) = {0.9, 0.95, 0.99, 0.999}
	Epsilon (rate time smoothing) = {1e-10, 1e-8, 1e-6, 1e-4}
	Patience = 5
MT1-MT3	Activation Functions = Relu
	Epochs = {100}
	ℓ^1 = {1e-5, 1e-4, 1e-3}
	Adam Learning Rate = {0.001, 00.1}
	Patience = 5

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

variance of predictors between hidden layers.

Table IA1 summarizes the set of hyperparameters that are tuned for each of the 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 3. The set of hyperparameters with the lowest quantile loss in the validation set is selected as the best model. For most nonlinear models, random search is used with 50 models. Random search has been shown to be more efficient than grid search (Bergstra and Bengio (2012)). For RF and MT1-MT3, we use grid search, since the number of hyperparameter configurations is limited.

LR and EN are estimated using a semismooth Newton coordinate descent algorithm implemented in the hqreg R package (Yi and Huang (2017)). RF is estimated using a fast

implementation of random forests for high-dimensional data implemented in the ranger R package (Wright and Ziegler (2015)). GBT and NN1-NN4 are implemented using the H2O.ai machine learning platform. MT1 - MT3 are implemented using the Keras Functional API and Tensorflow.

Table IA2: Performance of Alternative Quantile-Based Risk Premium Measure

We use the methodology in Yang, Zhu, Li, and Zhu (2023) and derive the quantile weights to form an alternative quantile-based risk premium measure: $\bar{Q}_{YZLZ} = 0.1Q_{0.05} + 0.3Q_{0.25} + 0.2Q_{0.5} + 0.3Q_{0.75} + 0.1Q_{0.95}$. Panel A reports the out-of-sample R^2_{os} values (in percentages) of $\bar{Q}_{YZLZ}$ forecasts for the entire panel of individual stock excess returns. The R^2_{os} values are generally lower than that of the robust risk premium measure, displayed in Panel A of Table 3. Panel B reports the monthly performance of value-weighted decile portfolios, sorted on $\bar{Q}_{YZLZ}$ forecasts. The Sharpe ratios are generally lower than that of the robust risk premium measure, displayed in Panel A of Table 4. These results demonstrate that the robust risk premium measure in Equation (3) is a better estimate of the risk premium than $\bar{Q}_{YZLZ} = 0.1Q_{0.05} + 0.3Q_{0.25} + 0.2Q_{0.5} + 0.3Q_{0.75} + 0.1Q_{0.95}$.

Panel A: Prediction Performance of $\bar{Q}_{YZLZ}$ (Percentage R^2_{OS})												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
	-1.3	0.14	0.28	0.38	0.39	0.39	0.39	0.39	0.43	1	0.96	0.07
Panel B: Performance of Value-Weighted Portfolios Sorted on $\bar{Q}_{YZLZ}$												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	-0.05	-0.28	-0.20	-0.30	-0.26	-0.73	-0.77	-0.67	-0.70	-1.48	-1.46	-1.51
2	0.34	0.32	0.36	0.28	0.31	0.20	0.03	0.06	0.14	-0.41	-0.43	-0.56
3	0.47	0.39	0.38	0.44	0.52	0.38	0.38	0.27	0.29	-0.03	0.05	0.08
4	0.62	0.55	0.67	0.57	0.51	0.55	0.47	0.46	0.52	0.30	0.32	0.35
5	0.61	0.61	0.63	0.69	0.64	0.64	0.58	0.58	0.65	0.58	0.55	0.54
6	0.75	0.75	0.71	0.68	0.66	0.70	0.75	0.73	0.73	0.76	0.71	0.80
7	0.83	0.83	0.86	0.86	0.81	0.80	0.85	0.83	0.80	0.93	0.87	0.89
8	0.85	0.84	0.89	0.86	0.84	0.91	1.00	0.98	0.97	1.04	1.10	1.04
9	0.90	0.94	0.95	0.98	0.99	1.10	1.10	1.20	1.15	1.23	1.17	1.17
High (H)	1.10	1.02	1.13	1.22	1.30	1.45	1.60	1.71	1.70	1.71	1.78	1.81
H-L	1.16	1.29	1.33	1.53	1.56	2.19	2.37	2.38	2.39	3.19	3.24	3.32
t	(4.49)	(2.91)	(4.78)	(4.69)	(3.21)	(5.92)	(6.13)	(6.05)	(6.51)	(8.11)	(8.29)	(8.06)
SR	0.82	0.58	0.88	0.90	0.64	1.16	1.21	1.24	1.25	1.46	1.46	1.47

Table IA3: Performance of Equal-Weighted Portfolios Sorted on Robust Risk Premium

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 (decile 1), "High" denotes the portfolio with the highest robust risk premium (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)), 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 A: Performance of Equal-Weighted Portfolios Sorted on $\bar{Q}$

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	-0.75	-0.03	-0.79	-0.59	-0.62	-1.12	-1.13	-1.06	-1.12	-1.79	-1.74	-1.71
2	0.26	0.27	0.22	0.18	0.26	0.10	0.07	0.03	-0.01	-0.28	-0.28	-0.24
3	0.45	0.56	0.45	0.44	0.39	0.45	0.39	0.40	0.36	0.20	0.20	0.24
4	0.67	0.59	0.67	0.54	0.47	0.52	0.54	0.52	0.55	0.45	0.47	0.52
5	0.79	0.75	0.73	0.66	0.61	0.63	0.69	0.63	0.68	0.67	0.68	0.66
6	0.88	0.84	0.91	0.76	0.75	0.76	0.77	0.77	0.80	0.89	0.82	0.84
7	0.96	0.83	0.97	0.91	0.84	0.95	0.91	0.90	0.94	1.07	1.05	1.04
8	1.11	0.94	1.12	0.97	0.93	1.14	1.09	1.15	1.08	1.27	1.30	1.28
9	1.28	1.11	1.31	1.23	1.24	1.37	1.40	1.42	1.44	1.65	1.64	1.58
High (H)	1.62	1.43	1.69	2.18	2.41	2.49	2.55	2.51	2.57	3.14	3.15	3.07
H-L	2.36	1.47	2.48	2.76	3.03	3.62	3.68	3.57	3.68	4.92	4.89	4.77
t	(8.40)	(3.52)	(8.71)	(9.17)	(8.36)	(11.60)	(12.02)	(11.70)	(12.94)	(14.46)	(14.14)	(13.30)
SR	1.77	0.71	1.81	1.85	1.75	2.45	2.53	2.46	2.64	3.19	3.05	2.80

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

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	1.82	0.61	1.96	2.85	2.68	3.48	3.47	3.45	3.43	4.61	4.56	4.36
t(α)	(9.58)	(2.72)	(9.47)	(9.36)	(9.82)	(8.48)	(9.79)	(9.14)	(11.35)	(13.73)	(13.22)	(13.03)
R^2	50	81	44	10	38	12	17	13	20	13	13	17

Table IA4: Performance of Equal-Weighted Portfolios Sorted on Conditional Volatility

Panel A reports the monthly performance of equal-weighted decile portfolios, sorted on robust conditional volatility forecasts. At the end of each month, we predict the one-month-ahead Vol, 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 Vol (decile 1), "High" denotes the portfolio with the highest Vol (decile 10), "H-L" denotes the long-short portfolio that buys the highest Vol 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)), 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 A: Performance of Equal-Weighted Portfolios Sorted on Vol												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.59	0.59	0.58	0.57	0.58	0.51	0.50	0.49	0.49	0.47	0.49	0.49
2	0.76	0.74	0.76	0.80	0.77	0.77	0.73	0.75	0.74	0.72	0.71	0.72
3	0.78	0.83	0.77	0.80	0.81	0.77	0.78	0.79	0.79	0.82	0.79	0.80
4	0.80	0.90	0.80	0.83	0.79	0.88	0.86	0.86	0.88	0.84	0.85	0.85
5	0.85	0.95	0.87	0.79	0.83	0.91	0.93	0.91	0.87	0.92	0.89	0.87
6	0.82	0.85	0.82	0.77	0.87	0.91	0.88	0.89	0.89	0.86	0.84	0.87
7	0.69	0.79	0.72	0.69	0.79	0.78	0.76	0.79	0.83	0.77	0.78	0.77
8	0.62	0.65	0.56	0.57	0.60	0.55	0.68	0.67	0.68	0.58	0.56	0.56
9	0.62	0.48	0.65	0.52	0.55	0.53	0.56	0.54	0.52	0.33	0.35	0.35
High (H)	0.76	0.52	0.75	0.94	0.67	0.67	0.60	0.60	0.58	0.98	1.03	1.02
H-L	0.16	-0.07	0.17	0.37	0.09	0.16	0.10	0.11	0.10	0.52	0.54	0.53
t	(0.25)	(-0.13)	(0.27)	(0.59)	(0.14)	(0.25)	(0.15)	(0.17)	(0.15)	(0.74)	(0.78)	(0.76)
SR	0.05	-0.03	0.06	0.12	0.03	0.05	0.03	0.04	0.03	0.15	0.16	0.16

Panel B: Performance of Equal-Weighted Portfolios Sorted on Vol												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	0.75	0.59	0.75	1.05	0.82	0.8	0.68	0.71	0.67	1.14	1.19	1.16
t(α)	(2.04)	(1.5)	(2.03)	(2.62)	(2.09)	(1.94)	(1.76)	(1.76)	(1.68)	(2.7)	(2.8)	(2.8)
R^2	72	62	71	67	69	69	70	70	70	70	69	69

Table IA5: Performance of Equal-Weighted Portfolios Sorted on Conditional Skewness

Panel A reports the monthly performance of equal-weighted decile portfolios, sorted on robust conditional skewness forecasts. At the end of each month, we predict the one-month-ahead Skew, 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 Skew (decile 1), "High" denotes the portfolio with the highest Skew (decile 10), "H-L" denotes the long-short portfolio that buys the highest Skew 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)), 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 A: Performance of Equal-Weighted Portfolios Sorted on Skew												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.43	0.06	0.05	0.27	-0.18	0.25	0.04	0.12	-0.11	-0.24	-0.10	-0.46
2	0.45	0.49	0.43	0.51	0.79	0.39	0.30	0.35	0.35	0.04	0.07	-0.07
3	0.45	0.60	0.43	0.57	0.74	0.59	0.42	0.60	0.55	0.28	0.36	0.32
4	0.56	0.51	0.58	0.61	0.76	0.77	0.63	0.57	0.69	0.53	0.54	0.59
5	0.57	0.74	0.79	0.65	0.73	0.75	0.66	0.64	0.75	0.70	0.70	0.76
6	0.74	0.91	0.85	0.74	0.84	0.77	0.79	0.80	0.86	0.83	0.81	0.93
7	0.79	1.00	0.92	0.74	0.81	0.82	0.89	0.84	0.93	0.99	0.89	1.09
8	0.95	1.02	1.05	0.87	0.88	0.91	1.02	1.00	0.97	1.09	1.10	1.11
9	1.07	0.97	1.08	0.97	0.93	0.96	1.18	1.05	1.01	1.19	1.23	1.26
High (H)	1.27	0.99	1.08	1.36	0.97	1.07	1.34	1.32	1.27	1.87	1.68	1.76
H-L	0.84	0.93	1.04	1.09	1.16	0.82	1.31	1.20	1.38	2.11	1.79	2.22
t	(3.02)	(1.80)	(3.85)	(6.31)	(2.63)	(2.93)	(6.08)	(5.60)	(6.40)	(6.81)	(4.88)	(6.24)
SR	0.64	0.36	0.74	1.29	0.53	0.64	1.20	1.09	1.46	1.37	1.06	1.32
Panel B: Performance of Equal-Weighted Portfolios Sorted on Skew												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	0.33	-0.1	0.32	0.98	0.33	0.64	1.25	1.3	1.07	2.05	1.59	1.47
t(α)	(1.51)	(-0.34)	(1.31)	(4.84)	(0.93)	(2.15)	(5.15)	(4.34)	(4.93)	(6.4)	(4.39)	(3.82)
R^2	33	76	45	11	54	10	9	10	18	20	15	16

Table IA6: Performance of Value-Weighted Decile Portfolios Sorted on $Q_{0.95}$ Forecasts

This table reports the monthly performance of value-weighted decile portfolios, sorted on $Q_{0.95}$ forecasts. At the end of each month, we predict the one-month-ahead $Q_{0.95}$, 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 $Q_{0.95}$ (decile 1), "High" denotes the portfolio with the highest $Q_{0.95}$ (decile 10), "H-L" denotes the long-short portfolio that buys the highest $Q_{0.95}$ 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 Value-Weighted Portfolios Sorted on $Q_{0.95}$

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.62	0.70	0.61	0.70	0.66	0.58	0.61	0.62	0.62	0.57	0.56	0.58
2	0.72	0.75	0.74	0.78	0.71	0.76	0.74	0.72	0.71	0.70	0.71	0.73
3	0.84	0.81	0.83	0.71	0.81	0.80	0.83	0.81	0.84	0.82	0.80	0.77
4	0.88	0.72	0.91	0.79	0.83	0.88	0.84	0.85	0.84	0.84	0.83	0.82
5	0.75	0.44	0.74	0.81	0.78	0.80	0.75	0.83	0.79	0.84	0.84	0.83
6	0.62	0.35	0.66	0.70	0.87	0.58	0.74	0.68	0.58	0.79	0.82	0.81
7	0.44	0.29	0.40	0.78	0.57	0.52	0.43	0.47	0.45	0.81	0.78	0.71
8	0.32	0.09	0.30	0.38	0.23	0.16	0.13	0.08	0.11	0.31	0.44	0.30
9	0.16	-0.08	0.14	0.18	-0.14	0.17	0.04	0.11	0.15	0.11	0.03	0.07
High (H)	0.49	-0.28	0.48	0.18	0.21	0.52	0.65	0.48	0.43	0.03	-0.01	0.06
H-L	-0.13	-0.98	-0.13	-0.52	-0.45	-0.06	0.04	-0.14	-0.19	-0.54	-0.57	-0.52
t	(-0.21)	(-1.85)	(-0.20)	(-0.83)	(-0.68)	(-0.09)	(0.05)	(-0.21)	(-0.28)	(-0.71)	(-0.75)	(-0.70)
SR	-0.04	-0.37	-0.04	-0.17	-0.13	-0.02	0.01	-0.04	-0.06	-0.15	-0.16	-0.14

Panel B: Risk-Adjusted Performance of Value-Weighted Portfolios Sorted on $Q_{0.95}$

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	0.7	-0.16	0.65	0.23	0.52	0.86	0.97	0.71	0.7	0.16	0.21	0.29
t(α)	(1.84)	(-0.41)	(1.75)	(0.7)	(1.3)	(2.07)	(2.24)	(1.84)	(1.75)	(0.4)	(0.56)	(0.86)
R^2	71	62	72	76	73	73	71	72	70	74	76	76
IR	0.42	-0.1	0.39	0.15	0.3	0.48	0.53	0.42	0.4	0.08	0.11	0.16

Table IA7: Performance of Sharpe-Managed Equal-Weighted Portfolios

Panel A reports the realized Sharpe ratios of the Sharpe-Managed equal-weighted decile portfolios. At the end of each month, we sort stocks into deciles based on their one-month-ahead robust risk premium forecast. The portfolios with the lowest and highest forecasts are labeled as "Low" and "High", respectively. We then form Sharpe-Managed portfolios by subdividing each portfolio into deciles based on their SR_Q forecasts. Within each decile, D1 (D10) denotes the portfolio with the lowest (highest) SR_Q forecasts. This results in deciles $\{Low_{D1}, ..., Low_{D10}\}$ and $\{High_{D1}, ..., High_{D10}\}$ within the Low and High portfolios, respectively. Here, Low_{D10} (Low_{D1}) and $High_{D10}$ ($High_{D1}$) represent subgroups of stocks with the highest (lowest) SR_Q forecasts from the Low and High portfolios, respectively. We reconstitute portfolios monthly using equal weights. Panel B displays the monthly average returns of our Sharpe-Managed strategy, denoted as "H_{D10}-L_{D1}", which involves exclusively taking long positions on $High_{D10}$ and short positions on Low_{D1}. "t" and "SR" respectively denote the Newey-West t-statistics and annualized Sharpe ratios for the H_{D10}-L_{D1} long-short portfolios. Panel C 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 A: Realized Sharpe Ratios of Sharpe-Managed Equal-Weighted Portfolios

	LR		LR-4		EN		RF		GBT		NN1		NN2		NN3		NN4		MT1		MT2		MT3	
	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10	D1	D10
Low (L)	-0.6	-0.2	-0.2	0.1	-0.5	-0.2	-0.6	-0.3	-0.7	-0.1	-1.0	-0.3	-0.9	-0.3	-1.0	-0.2	-0.9	-0.3	-1.3	-0.3	-1.4	-0.3	-1.3	-0.3
2	0.1	0.1	0.1	0.4	0.2	0.0	0.0	-0.1	-0.1	0.0	-0.1	-0.1	0.0	0.0	-0.1	-0.1	0.0	-0.1	-0.3	-0.1	-0.3	0.0	-0.3	0.0
3	0.1	0.2	0.2	0.4	0.2	0.0	0.1	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.1	0.0	0.1	0.1	0.1	-0.1	0.2
4	0.2	0.5	0.3	0.4	0.3	0.3	0.1	0.1	0.2	0.3	0.2	0.3	0.1	0.4	0.2	0.4	0.2	0.3	0.2	0.1	0.2	0.3	0.3	0.2
5	0.3	0.5	0.3	0.4	0.4	0.3	0.1	0.3	0.2	0.4	0.2	0.6	0.2	0.4	0.2	0.5	0.3	0.5	0.4	0.4	0.3	0.4	0.3	0.4
6	0.3	0.6	0.5	0.7	0.5	0.4	0.2	0.4	0.3	0.5	0.2	0.8	0.3	0.8	0.3	0.7	0.4	0.8	0.6	0.6	0.5	0.6	0.4	0.5
7	0.4	0.7	0.2	0.7	0.6	0.6	0.3	0.5	0.4	0.5	0.3	1.1	0.3	1.0	0.2	1.0	0.4	1.1	0.7	0.7	0.5	0.8	0.4	0.9
8	0.6	0.8	0.5	0.9	0.6	0.7	0.2	0.6	0.5	0.5	0.5	1.2	0.4	1.1	0.5	1.2	0.3	1.2	0.5	1.0	0.6	1.2	0.6	1.1
9	0.6	1.1	0.5	0.9	0.7	1.0	0.3	0.8	0.6	0.7	0.3	1.4	0.4	1.1	0.4	1.4	0.5	1.2	0.6	1.4	0.6	1.4	0.6	1.5
High (H)	0.7	1.4	0.8	1.1	0.8	1.2	0.3	1.5	0.8	1.7	0.6	1.4	0.6	1.5	0.6	1.5	0.7	1.6	0.7	1.8	0.7	1.8	0.7	1.8

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

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
H_{D10}-L_{D1}	2.59	1.65	2.2	5.1	6.25	5.15	4.96	5.25	5.22	8.36	8.66	8.87
t	(6.98)	(2.51)	(6.16)	(11.2)	(10.69)	(11.75)	(11.66)	(12.27)	(14.19)	(14.66)	(15.64)	(15.74)
SR	1.44	0.54	1.3	2.15	2.05	2.53	2.5	2.63	2.71	3.28	3.42	3.19

Panel C: Risk-Adjusted Performance of Sharpe-Managed Equal-Weighted Long-Short Portfolios

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
α	1.86	0.95	1.57	4.84	5.35	4.66	4.49	4.99	4.77	8.05	8.42	8.46
t(α)	(6.36)	(3.03)	(6.12)	(10.05)	(10.35)	(10.14)	(11.26)	(10.68)	(14.09)	(12.59)	(13.37)	(14.13)
R^2	55	74	56	18	33	27	32	25	34	6	5	8

Table IA8: Performance of Portfolios with NYSE Breakpoints Sorted on Robust Risk Premium

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 (decile 1), "High" denotes the portfolio with the highest robust risk premium (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, formed using NYSE breakpoints, which are sorted on robust risk premium forecasts.

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

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.17	0.26	0.13	0.08	0.17	0.05	-0.09	-0.04	-0.06	-0.46	-0.49	-0.47
2	0.55	0.51	0.52	0.42	0.52	0.39	0.40	0.34	0.42	0.20	0.28	0.10
3	0.58	0.64	0.62	0.53	0.57	0.57	0.58	0.51	0.53	0.46	0.30	0.50
4	0.71	0.70	0.70	0.64	0.65	0.68	0.59	0.60	0.69	0.56	0.54	0.59
5	0.78	0.78	0.79	0.62	0.68	0.69	0.73	0.66	0.66	0.74	0.72	0.69
6	0.81	0.86	0.82	0.74	0.66	0.69	0.79	0.82	0.82	0.80	0.84	0.84
7	0.86	0.84	0.85	0.77	0.83	0.81	0.83	0.79	0.82	0.95	0.93	0.92
8	0.83	0.86	0.92	0.83	0.85	0.92	0.92	0.94	0.89	1.00	0.96	0.94
9	0.95	0.91	0.96	0.85	0.92	1.09	1.10	1.09	1.08	1.17	1.12	1.13
High (H)	1.12	1.01	1.18	1.11	1.23	1.29	1.28	1.37	1.36	1.50	1.50	1.40
H-L	0.95	0.75	1.05	1.02	1.06	1.24	1.37	1.41	1.42	1.97	1.99	1.87
t	(4.53)	(2.49)	(4.79)	(5.01)	(3.59)	(4.96)	(5.75)	(5.94)	(6.10)	(7.95)	(7.61)	(6.59)
SR	0.84	0.48	0.89	0.92	0.66	0.92	1.04	1.06	1.09	1.34	1.25	1.18

Panel B: Performance of Equal-Weighted Portfolios Sorted on $\bar{Q}$ with NYSE Breakpoints

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	-0.52	0.03	-0.57	-0.31	-0.29	-0.56	-0.60	-0.58	-0.65	-0.94	-0.92	-0.85
2	0.40	0.35	0.33	0.35	0.36	0.40	0.41	0.38	0.36	0.27	0.34	0.38
3	0.55	0.58	0.53	0.48	0.41	0.49	0.52	0.52	0.46	0.53	0.49	0.52
4	0.75	0.61	0.69	0.61	0.55	0.58	0.60	0.59	0.61	0.59	0.59	0.60
5	0.85	0.72	0.79	0.65	0.63	0.67	0.70	0.67	0.64	0.69	0.69	0.69
6	0.89	0.82	0.88	0.77	0.74	0.76	0.80	0.72	0.80	0.88	0.79	0.82
7	0.97	0.85	0.94	0.84	0.78	0.85	0.84	0.82	0.89	1.00	0.98	0.95
8	1.04	0.89	1.05	0.91	0.81	1.02	0.96	0.98	0.93	1.06	1.13	1.08
9	1.25	1.03	1.23	1.01	1.03	1.13	1.10	1.14	1.09	1.38	1.33	1.32
High (H)	1.61	1.32	1.65	1.71	1.81	1.97	1.98	1.99	2.01	2.44	2.47	2.35
H-L	2.13	1.29	2.22	2.02	2.10	2.53	2.58	2.57	2.66	3.38	3.39	3.20
t	(8.36)	(3.52)	(8.61)	(9.32)	(7.08)	(11.04)	(11.66)	(11.57)	(12.39)	(14.40)	(14.43)	(12.85)
SR	1.80	0.74	1.80	1.84	1.52	2.31	2.44	2.45	2.55	3.10	2.95	2.59

Table IA9: Performance of Portfolios with NYSE Breakpoints Sorted on Conditional Skewness

Panel A reports the monthly performance of value-weighted decile portfolios, formed using NYSE breakpoints, sorted on robust conditional skewness forecasts. At the end of each month, we predict the one-month-ahead Skew, 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 Skew (decile 1), "High" denotes the portfolio with the highest Skew (decile 10), "H-L" denotes the long-short portfolio that buys the highest Skew 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, formed using NYSE breakpoints, which are sorted on robust conditional skewness forecasts.

Panel A: Performance of Value-Weighted Portfolios Sorted on Skew with NYSE Breakpoints

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.57	0.19	0.37	0.55	0.20	0.45	0.29	0.28	0.31	0.14	0.25	-0.06
2	0.54	0.51	0.46	0.62	0.67	0.52	0.43	0.48	0.54	0.48	0.52	0.34
3	0.69	0.53	0.56	0.59	0.56	0.65	0.60	0.65	0.56	0.49	0.66	0.51
4	0.63	0.73	0.74	0.58	0.68	0.69	0.61	0.54	0.64	0.68	0.70	0.74
5	0.66	0.80	0.63	0.69	0.71	0.79	0.62	0.67	0.66	0.83	0.74	0.87
6	0.74	0.75	0.80	0.74	0.64	0.65	0.67	0.64	0.71	0.85	0.74	0.86
7	0.80	0.77	0.80	0.70	0.82	0.65	0.76	0.70	0.78	0.85	0.69	0.91
8	0.77	0.76	0.93	0.67	0.72	0.69	0.75	0.81	0.80	0.84	0.90	0.99
9	1.03	0.82	1.01	0.76	0.86	0.71	0.86	0.89	0.78	0.89	0.84	1.07
High (H)	0.96	0.87	0.90	0.76	0.82	0.75	0.95	1.07	0.89	0.96	1.11	1.17
H-L	0.39	0.68	0.53	0.21	0.62	0.30	0.66	0.78	0.58	0.82	0.86	1.23
t	(1.66)	(1.68)	(2.31)	(2.19)	(2.05)	(1.41)	(3.47)	(4.55)	(3.26)	(3.14)	(2.83)	(4.10)
SR	0.33	0.34	0.46	0.41	0.39	0.29	0.67	0.84	0.62	0.59	0.54	0.77

Panel B: Performance of Equal-Weighted Portfolios Sorted on Skew with NYSE Breakpoints

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.46	0.17	0.23	0.30	0.45	0.31	0.16	0.23	0.08	-0.16	-0.03	-0.34
2	0.48	0.62	0.53	0.49	0.80	0.51	0.35	0.45	0.49	0.27	0.38	0.17
3	0.57	0.67	0.62	0.57	0.80	0.65	0.59	0.58	0.66	0.44	0.59	0.37
4	0.56	0.84	0.70	0.65	0.81	0.72	0.69	0.66	0.75	0.58	0.75	0.63
5	0.66	0.96	0.79	0.66	0.77	0.78	0.81	0.70	0.81	0.73	0.76	0.73
6	0.66	1.00	0.91	0.67	0.84	0.80	0.82	0.74	0.89	0.88	0.86	0.85
7	0.75	1.02	0.91	0.77	0.94	0.81	0.90	0.90	0.94	0.88	0.87	0.86
8	0.86	0.92	1.01	0.78	0.94	0.88	0.99	0.96	0.97	1.01	0.96	0.95
9	1.09	0.95	1.04	0.92	0.96	0.88	1.11	1.04	1.05	1.09	1.07	1.09
High (H)	1.21	0.97	1.09	1.27	0.96	1.06	1.36	1.32	1.27	1.62	1.46	1.44
H-L	0.75	0.80	0.86	0.97	0.51	0.76	1.20	1.09	1.18	1.79	1.49	1.78
t	(2.72)	(1.71)	(3.30)	(6.18)	(1.40)	(2.76)	(5.74)	(5.23)	(5.93)	(6.39)	(4.63)	(5.76)
SR	0.57	0.37	0.63	1.27	0.29	0.59	1.10	1.00	1.31	1.30	0.99	1.17

Table IA10: Performance of Portfolios Sorted on Robust Risk Premium 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 (decile 1), "High" denotes the portfolio with the highest robust risk premium (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 robust risk premium forecasts, for all stocks above the 20th percentile of NYSE cap weights.

Performance of Value-Weighted Portfolios Sorted on $\bar{Q}$ Excluding Microcaps												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.07	0.23	-0.02	0.09	0.11	-0.04	-0.22	-0.15	-0.10	-0.52	-0.56	-0.63
2	0.50	0.49	0.48	0.45	0.51	0.28	0.25	0.19	0.23	-0.01	0.02	0.11
3	0.53	0.59	0.63	0.59	0.53	0.45	0.47	0.50	0.45	0.48	0.38	0.34
4	0.73	0.58	0.63	0.65	0.55	0.70	0.63	0.60	0.63	0.55	0.63	0.68
5	0.63	0.77	0.65	0.66	0.72	0.65	0.72	0.68	0.71	0.75	0.68	0.70
6	0.73	0.86	0.77	0.74	0.66	0.73	0.72	0.79	0.73	0.81	0.88	0.84
7	0.91	0.81	0.88	0.70	0.79	0.74	0.86	0.85	0.84	0.97	0.92	0.94
8	0.88	0.83	0.89	0.82	0.90	0.90	0.90	0.89	0.87	0.93	0.91	0.91
9	0.92	0.92	0.98	0.90	0.85	1.04	0.98	0.98	1.02	1.17	1.08	1.09
High (H)	1.04	0.97	1.06	1.01	1.02	1.14	1.18	1.21	1.17	1.37	1.35	1.31
H-L	0.96	0.74	1.08	0.92	0.91	1.18	1.40	1.35	1.27	1.89	1.91	1.93
t	(4.14)	(2.32)	(4.46)	(3.76)	(2.40)	(3.82)	(4.34)	(4.41)	(3.97)	(5.78)	(5.66)	(5.43)
SR	0.75	0.41	0.80	0.73	0.46	0.70	0.83	0.83	0.74	1.03	0.99	1.01
Performance of Equal-Weighted Portfolios Sorted on $\bar{Q}$ Excluding Microcaps												
	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	-0.03	0.15	-0.11	0.07	0.14	-0.15	-0.23	-0.19	-0.16	-0.53	-0.49	-0.50
2	0.48	0.44	0.46	0.43	0.52	0.40	0.47	0.43	0.41	0.17	0.20	0.28
3	0.58	0.55	0.58	0.62	0.59	0.58	0.59	0.58	0.54	0.51	0.55	0.52
4	0.75	0.67	0.71	0.69	0.68	0.72	0.68	0.63	0.69	0.68	0.68	0.70
5	0.72	0.79	0.76	0.79	0.68	0.73	0.76	0.72	0.75	0.72	0.70	0.74
6	0.77	0.84	0.79	0.80	0.75	0.81	0.78	0.79	0.80	0.89	0.84	0.88
7	0.89	0.82	0.88	0.84	0.87	0.84	0.88	0.88	0.88	0.99	0.97	0.94
8	0.94	0.93	0.95	0.97	0.92	0.97	0.92	0.98	0.96	1.05	1.08	1.04
9	0.97	0.97	1.03	0.98	0.98	1.07	1.07	1.04	1.05	1.26	1.20	1.18
High (H)	1.19	1.09	1.20	1.05	1.11	1.28	1.33	1.38	1.32	1.51	1.53	1.46
H-L	1.22	0.94	1.31	0.98	0.97	1.43	1.56	1.57	1.48	2.04	2.01	1.96
t	(4.49)	(2.63)	(4.73)	(3.80)	(2.37)	(4.19)	(4.42)	(4.63)	(4.40)	(6.13)	(6.02)	(5.46)
SR	0.86	0.52	0.90	0.82	0.50	0.85	0.93	0.97	0.90	1.16	1.11	1.07

Table IA11: Performance of Portfolios Sorted on Conditional Skewness Excluding Microcaps

Panel A reports the monthly performance of value-weighted decile portfolios, sorted on robust conditional skewness 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 Skew, 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 Skew (decile 1), "High" denotes the portfolio with the highest Skew (decile 10), "H-L" denotes the long-short portfolio that buys the highest Skew 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 robust conditional skewness forecasts, for all stocks above the 20th percentile of NYSE cap weights.

Panel A: Performance of Value-Weighted Portfolios Sorted on Skew Excluding Microcaps

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.58	0.19	0.35	0.55	-0.02	0.45	0.38	0.35	0.36	0.25	0.24	-0.04
2	0.53	0.44	0.46	0.63	0.48	0.54	0.45	0.48	0.60	0.46	0.50	0.30
3	0.65	0.49	0.55	0.60	0.63	0.71	0.56	0.67	0.59	0.48	0.54	0.57
4	0.70	0.66	0.60	0.63	0.64	0.71	0.61	0.55	0.60	0.65	0.59	0.65
5	0.61	0.75	0.70	0.61	0.66	0.74	0.67	0.64	0.64	0.81	0.64	0.90
6	0.73	0.80	0.70	0.79	0.68	0.69	0.78	0.66	0.69	0.74	0.63	0.90
7	0.83	0.76	0.78	0.72	0.74	0.71	0.70	0.75	0.78	0.78	0.80	0.93
8	0.76	0.71	0.93	0.63	0.78	0.70	0.77	0.84	0.86	0.91	0.77	0.93
9	0.96	0.84	0.97	0.81	0.78	0.74	0.83	0.83	0.68	0.85	0.98	1.07
High (H)	0.96	0.86	0.93	0.72	0.86	0.78	0.95	1.05	0.93	0.96	0.92	1.06
H-L	0.38	0.67	0.57	0.16	0.88	0.33	0.57	0.70	0.57	0.71	0.68	1.11
t	(1.52)	(1.71)	(2.31)	(1.71)	(2.33)	(1.69)	(3.30)	(4.32)	(3.04)	(2.54)	(2.05)	(3.48)
SR	0.31	0.33	0.46	0.32	0.44	0.34	0.61	0.80	0.59	0.45	0.38	0.63

Panel B: Performance of Equal-Weighted Portfolios Sorted on Skew Excluding Microcaps

	LR	LR-4	EN	RF	GBT	NN1	NN2	NN3	NN4	MT1	MT2	MT3
Low (L)	0.52	0.17	0.30	0.61	0.13	0.40	0.26	0.28	0.31	0.26	0.17	-0.11
2	0.52	0.50	0.49	0.65	0.66	0.65	0.43	0.53	0.58	0.48	0.53	0.32
3	0.60	0.48	0.54	0.69	0.71	0.69	0.67	0.72	0.61	0.46	0.66	0.60
4	0.63	0.72	0.61	0.71	0.70	0.77	0.70	0.63	0.70	0.66	0.74	0.72
5	0.66	0.86	0.76	0.77	0.77	0.80	0.70	0.74	0.75	0.77	0.69	0.82
6	0.75	0.90	0.75	0.73	0.80	0.78	0.77	0.77	0.78	0.85	0.68	0.88
7	0.80	0.88	0.82	0.77	0.82	0.83	0.84	0.80	0.77	0.83	0.86	0.95
8	0.80	0.88	0.97	0.78	0.88	0.78	0.89	0.85	0.87	0.94	0.91	0.96
9	0.95	0.91	0.98	0.78	0.86	0.76	0.95	0.85	0.83	0.98	1.00	1.04
High (H)	1.02	0.93	1.02	0.75	0.92	0.77	1.03	1.09	1.04	1.02	1.02	1.07
H-L	0.51	0.76	0.72	0.14	0.78	0.37	0.77	0.81	0.73	0.76	0.84	1.18
t	(1.91)	(1.73)	(2.64)	(1.05)	(2.02)	(1.61)	(4.23)	(5.12)	(3.88)	(2.68)	(2.61)	(3.57)
SR	0.38	0.36	0.52	0.23	0.41	0.34	0.78	0.90	0.76	0.48	0.49	0.69