Working Paper in Economics

202310

Portfolio Selection Under Non-Gaussianity And Systemic Risk: A Machine Learning Based Forecasting Approach.

Weidong Lin[*] and Abderrahim Taamouti[†]

[*] Department of Finance, NEOMA Business School (Rouen). Address: 1 Rue du Maréchal Juin, Mont-Saint-Aignan, 76130, France. E-mail: weidong.lin@neoma-bs.fr.
[†] Corresponding author. Department of Economics, University of Liverpool Management School. Address: Chatham St, Liverpool L69 7ZH. E-mail: Abderrahim.Taamouti@liverpool.ac.uk.

Portfolio Selection Under Non-Gaussianity And Systemic Risk: A Machine Learning Based Forecasting Approach*

Weidong Lin† Abderrahim Taamouti‡

August 11, 2023

ABSTRACT

The Sharpe-ratio-maximizing portfolio becomes questionable under non-Gaussian returns, and it rules out, by construction, systemic risk, which can negatively affect its out-of-sample performance. In the present work, we develop a new performance ratio that simultaneously addresses these two problems when building optimal portfolios. To robustify the portfolio optimization and better represent extreme market scenarios, we simulate a large number of returns via a Monte Carlo method. This is done by first obtaining probabilistic return forecasts through a distributional machine learning approach in a big data setting, and then combining them with a fitted copula to generate return scenarios. Based on a large-scale comparative analysis conducted on the US market, the backtesting results demonstrate the superiority of our proposed portfolio selection approach against several popular benchmark strategies in terms of both profitability and minimizing systemic risk. This outperformance is robust to the inclusion of transaction costs.

Keywords: Portfolio optimization; probability forecasting; quantile regression neural network; extreme scenarios; big data.

---

*The authors would like to thank very much the Editor Professor Dick van Dijk and two anonymous reviewers for their very useful comments.

†Department of Finance, NEOMA Business School (Rouen). Address: 1 Rue du Maréchal Juin, Mont-Saint-Aignan, 76130, France. E-mail: weidong.lin@neoma-bs.fr

‡ *Corresponding author.* Department of Economics, University of Liverpool Management School. Address: Chatham St, Liverpool L69 7ZH. E-mail: Abderrahim.Taamouti@liverpool.ac.uk.

Portfolio Selection Under Non-Gaussianity And Systemic Risk: A Machine Learning Based Forecasting Approach

Abstract

The Sharpe-ratio-maximizing portfolio becomes questionable under non-Gaussian returns, and it rules out, by construction, systemic risk, which can negatively affect its out-of-sample performance. In the present work, we develop a new performance ratio that simultaneously addresses these two problems when building optimal portfolios. To robustify the portfolio optimization and better represent extreme market scenarios, we simulate a large number of returns via a Monte Carlo method. This is done by first obtaining probabilistic return forecasts through a distributional machine learning approach in a big data setting, and then combining them with a fitted copula to generate return scenarios. Based on a large-scale comparative analysis conducted on the US market, the backtesting results demonstrate the superiority of our proposed portfolio selection approach against several popular benchmark strategies in terms of both profitability and minimizing systemic risk. This outperformance is robust to the inclusion of transaction costs.

Keywords: portfolio optimization; probability forecasting; quantile regression neural network; extreme scenarios; big data

1 Introduction

1.1 Motivation of the new performance measure

Deciding the best performance measure to use for constructing optimal portfolios is an evergreen question in asset allocation. Following the work of Roy (1952), Sharpe (1966) established the popular Sharpe ratio, initially termed as a reward-to-variability ratio, measuring the tradeoff between mean return and risk. However, this ratio suffers from several drawbacks as it inherently depends on the normality assumption of the return distribution. Such drawbacks include ignoring higher order moments of returns, but importantly using an inadequate measure of risk, namely standard deviation.

Although the Sharpe ratio has always been seen as a reward-to-risk performance measure, it is essentially a dispersion-type of ratio since its risk measure (i.e. standard deviation) only quantifies uncertainty. As argued by Rachev et al. (2008), risk is an asymmetric concept that needs to consider downside and upside outcomes of an investment differently. Thus, the Sharpe ratio becomes unsuitable for assessing risk-adjusted performance once the normality assumption is relaxed. To overcome this, alternative ratios under non-Gaussian (asymmetric) distributions have been developed; see Sortino and Satchell (2001) and Ortobelli et al. (2005). For example, to better measure downside risk in a non-Gaussian setting, the standard deviation can be replaced by either Value-at-Risk (VaR), Expected Shortfall (ES), or partial moments of different orders; see Biglova et al. (2004). Among the existing reward-to-risk ratios, the Rachev ratio of Biglova et al. (2004) is an advanced alternative since it is fully compatible with non-Gaussian (asymmetric) return distributions.

Recently, other challenges have been pressing investors and portfolio managers to prevent their investments against extreme market events. For instance, the portfolio performance is not only affected by the individual risks of portfolio assets, but also by the systemic risk of the entire financial market. Hence, relevant performance ratios cannot only consider the realistic aspects of return distributions (asymmetry and heavy-tailedness, etc.), but also incorporate the potential impacts of market distress. Unfortunately, none of the above-surveyed measures including the Rachev ratio addresses this concern. In the present paper, we address this issue by extending the unconditional Rachev ratio to account for non-Gaussian returns and allow for the occurrence of systemic events.

Systemic risk can be defined as the possibility of breakdown of the whole financial system, which is opposed to the risk relevant to individual entities within the system. The 2007-2008 financial turmoil and the subsequent crises (e.g. the euro crisis and the COVID-19 pandemic) are examples that illustrate the consequence of ignoring this type of risk. While the macroprudential literature has made substantial progress in developing monitoring tools for assessing the underlying systemic risk within the financial system, investors and asset managers still lack explicit guidance for controlling their portfolios' systemic risk. There are only a few studies that have examined the implications of systemic risk for investment decisions.

Biglova et al. (2014) studied the portfolio selection problem under systemic risk by proposing a conditional Rachev ratio ($\text{CoRR}^{\text{Biglova}}$), where systemic risk takes place when all portfolio assets are distressed. However, $\text{CoRR}^{\text{Biglova}}$ does not connect systemic risk with market distress and is not an ex-ante measure. Instead, it evaluates portfolio performance conditional on the occurrence of idiosyncratic (individual) risk events. Moreover, $\text{CoRR}^{\text{Biglova}}$ takes the expected portfolio's active return as a reward measure conditional on all asset prices co-moving in the right tail. And this assumption is hard to be satisfied in practice and might lead to an empty set if the number of portfolio assets is sufficiently large. Another effort was recently made by Lin et al. (2023), where the authors studied the tradeoff between reward and risk under systemic risk by introducing a conditional Sharpe ratio (CoSR). However, CoSR cannot account for non-Gaussian (asymmetric) returns. In this work, we extend the unconditional Rachev ratio by explicitly incorporating the occurrence of systemic events to account for both individual risk and systemic risk under non-Gaussian (asymmetric) return distributions.

Last but not least, the out-of-sample performance of optimal portfolios also depends on the quality of inputs of portfolio optimization. In general, portfolio selection models require estimating reward and risk measures using either historical or simulated return samples. The former approach has been often criticized under the mean-variance framework since the sample-based estimators are subject to substantial estimation errors that can lead to extreme portfolio weights. This is sometimes referred to as the *error maximization* (Michaud 1989). Nevertheless, reducing estimation error is of great importance not only to the Gaussian-based mean-variance model where the estimates of the first two moments of returns are required but also to other reward-to-risk models that work under more general distributional assumptions. In this paper, we adopt the latter approach by employing a distributional machine learning (ML) method for return prediction, where the resulting probabilistic return forecasts can help mitigate the estimation error of inputs to portfolio optimizers as discussed below.

1.2 Motivation of using ML techniques for return prediction

To obtain more robust estimators for portfolio optimization, ML models seem to be promising tools in obtaining more robust estimators for the input parameters of portfolio optimizers, see for example Kaczmarek and Perez (2021). In the past decades, the rapid development of computer technology combined with the availability of big data enables us to train more complicated models via ML algorithms, see Messmer (2017). Gu et al. (2020) define ML as a set of high-dimensional predictive statistical models, associated with regularization approaches for mitigating overfitting problems and efficient algorithms for hyperparameter tuning, respectively. With such advantages and an ever-increasing number of predictors, the ML techniques have become the favourite approach for improving stock return predictability in a big data setting; see Abe and Nakayama (2018), Feng et al. (2018), Chen et al. (2019), Jan and Ayub (2019), Gu et al. (2020, 2021) and Feng et al. (2021) among others.

Since the ML techniques have shown to be superior to the traditional statistical methods in terms of stock return prediction, many researchers have applied them to portfolio optimization and generated satisfying results; see Zhang et al. (2020), Babiak and Baruník (2020) and Huang et al. (2021) among others. However, to our knowledge, there is no existing work that explores the potential economic gains of utilizing ML-based probabilistic return forecasts in portfolio selection. The existing applications in FinTech literature focus mostly on obtaining point forecasts of stock returns without accounting for any predictive distributional information. Moreover, so far the efficiency of ML-based portfolios has been tested mainly for characteristic-sorted portfolios (e.g. long-short decile portfolios) without involving any portfolio optimization strategy. All these motivate us further to investigate the potential benefit of using a distributional ML approach in portfolio optimization.

Specifically, we solve the portfolio selection problem via a three-stage supervised learning model. We start by predicting conditional quantiles of cross-sectional returns using a distributional ML model, i.e., smooth pinball neural network (SPNN), based on which we estimate the conditional return densities of portfolio assets and the market. Next, we use t-copula to model the dependence among portfolio assets and the market, and generate scenarios for future returns. Lastly, based on the simulated returns, we solve the portfolio optimization problem dynamically by maximizing an ex-ante conditional Rachev ratio (CoRR), which accounts for systemic risk and non-Gaussianity.

To show the superiority of our portfolio selection approach, we perform a large-scale comparative study using nearly 600 US equities with 37 years of history from January 1985 to December 2021. Our set of predictors includes 94 firm-specific characteristics, 14 macroeconomic variables, and 74 industry dummies. We use the SPNN model to forecast monthly return quantiles for portfolio assets and the market index. Thereafter, at the beginning of each out-of-sample month, we use generated return scenarios to solve the portfolio optimization problems with CoRR and other performance measures. Finally, we measure the out-of-sample performance of all portfolio candidates by various metrics in terms of both profitability and systemic risk.

1.3 Contribution and paper structure

Our paper contributes to the literature in multiple ways. Firstly, we shed new light on reward-risk portfolio optimization by introducing a new performance measure that accounts for both non-Gaussianity (asymmetry) and systemic risk. This is achieved by explicitly incorporating the occurrence of systemic events into the portfolio's Rachev ratio. This proposed ratio is able to quantify the tradeoff between conditional expected reward and loss, where the conditional information is the market distress. The optimal portfolio obtained by maximizing this new measure is expected to deliver a resilient performance during crisis periods. Secondly, we enrich the asset pricing literature by utilizing a distributional ML model for predicting cross-sectional returns. We demonstrate its superiority in generating significant economic gains through a comparative backtesting analysis. Contrary to the majority of FinTech applications that focus on predicting conditional mean return, this paper takes advantage of the predictive information implied by the whole conditional distribution that is obtained using probabilistic return forecasts via a distributional ML approach. Lastly, we build a bridge between the literature on performance strategy and systemic risk. More specifically, the risk measure in our proposed performance ratio can be interpreted as the portfolio-level Conditional Expected Shortfall (CoES), which can be viewed as an extension of Conditional Value-at-Risk (CoVaR) as argued by Adrian and Brunnermeier (2016). The portfolio's CoES relative to the whole financial system refers to the ES of the portfolio's active return conditional on extreme market scenarios. Interestingly, if we consider portfolio loss instead of return by putting a minus sign, the resulting CoES becomes a reward measure.

The remaining paper is structured as follows. Section 2 formulates the return quantile prediction using the SPNN model. Section 3 defines the portfolio selection problem using our proposed performance criterion. Section 4 conducts a large-scale comparative study based on a high-dimensional dataset on the US market, in which we assess the out-of-sample portfolio performance of all candidate strategies. Section 5 concludes. The simulation algorithm for generating return scenarios can be found in Appendix A, while Appendix B describes how we estimate our proposed measure based on simulated returns. Appendices C, D and E contain some supplementary information on SPNN modelling. Figures and tables are included in Appendix F.

2 Smooth pinball neural network

Before we specify our model, let us first set some notations. We denote by $\boldsymbol{R} = (R_1, ..., R_V)$ the $1 \times V$ vector of predictand (monthly realized return) of V training samples, and $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_V)$, with $\boldsymbol{X}_v = (x_{1,v}, ..., x_{P,v})^T$, for $v = 1, ..., V$, the corresponding $P \times V$ matrix of P one-month lagged predictors, including firm-level features, interactions of each feature with macroeconomic variables, and industry dummies. Note that in the above notations, we do not use any subscript to distinguish between different entities (e.g. individual firms and the market portfolio), but we will do so in Section 3.

2.1 Model specification

Recently, Hatalis et al. (2019) proposed an advanced variant of the traditional quantile regression neural network (QRNN) of Taylor (2000), namely the smooth pinball neural network (SPNN). Formally, the cost function of SPNN is defined as

$$L = \frac{1}{M}\frac{1}{V}\sum_{m=1}^{M}\sum_{v=1}^{V}\rho_{\tau}^{(A)}\Big[\big(R_v - \hat{Q}_{R_v}(\tau_m|\boldsymbol{X}_v)\big)\Big] + p + \lambda||\boldsymbol{\beta}||_1, \tag{1}$$

where the M prespecified quantiles are equally spaced as $\tau_m = \frac{m}{M+1}$, the conditional quantiles are represented by a QRNN model $f(\cdot)$ with a set of parameters $\boldsymbol{\beta} = \{\boldsymbol{\beta}(\tau_m)\}_{m=1,\ldots,M}$ such that $\hat{Q}_{R_v}(\tau_m|\boldsymbol{X}_v) = f(\boldsymbol{X}_v, \hat{\boldsymbol{\beta}}(\tau_m))$, $\rho_{\tau}^{(A)}$ is the smoothed pinball loss proposed by Zheng (2011). In particular, p is the penalty term added for satisfying the non-crossing constraint $\hat{Q}_{R_v}(\tau_1|\boldsymbol{X}_v) \leq \cdots \leq \hat{Q}_{R_v}(\tau_M|\boldsymbol{X}_v),\ \forall v$, which is defined as

$$p = c\frac{1}{MV}\sum_{m=1}^{M}\sum_{v=1}^{V}\Big[\max\Big(0, \epsilon - \big(\hat{Q}_{R_v}(\tau_m|\boldsymbol{X}_v) - \hat{Q}_{R_v}(\tau_{m-1}|\boldsymbol{X}_v)\big)\Big)\Big]^2, \tag{2}$$

where $\hat{Q}_{R_v}(\tau_0|\boldsymbol{X}_v)$ is initialized to zero, ϵ denotes the minimum magnitude between two adjacent quantiles, and c denotes the penalty parameter. If all constraints are satisfied, then $p = 0$. Otherwise, once $\hat{Q}_{R_v}(\tau_m|\boldsymbol{X}_v) < \hat{Q}_{R_v}(\tau_{m-1}|\boldsymbol{X}_v)$, the squared difference between them is incorporated as a penalty into the objective function. Finally, the LASSO penalty term $\lambda||\boldsymbol{\beta}||_1$ is added to mitigate the overfitting problem, where $||\cdot||_1$ refers to the L1-norm and λ denotes the regularization parameter.

2.2 Related literature

SPNN is a further extension of the composite QRNN (CQRNN) proposed by Xu et al. (2017), by which we can estimate multiple conditional quantiles simultaneously and efficiently. CQRNN inherits one of the same capabilities as linear composite quantile regression (CQR) of Zou and Yuan (2008), i.e., combining multiple quantile regressions to better capture complex nonlinear relationships between the predictors and the predictand. CQRNN is a flexible model not only because it allows uncovering complex nonlinear patterns among variables taking advantage of ANN, but also because it helps enhance the process of estimation and prediction thanks to the property of CQR (Xu et al. 2017).

Although CQRNN improves the model efficiency and prediction accuracy, it fails to prevent the quantile crossover problem. Quantile crossing violates the requirement that the cumulative distribution function (CDF) should be monotonically increasing. To mitigate this issue, Cannon (2018) developed a monotonic CQRNN (MCQRNN) model that imposes monotonicity constraints on a standard multi-layer perceptron and integrates the model architecture of CQRNN to achieve simultaneous estimation. However, the stacked matrix of covariates complicates the network by adding overmuch parameters, which makes the estimation computationally inefficient and induces the propensity of overfitting. Instead, SPNN can be seen as an efficient alternative to MCQRNN.

3 Portfolio selection under non-Gaussianity and systemic risk

In this section, we first propose a new performance measure that allows for non-Gaussianity and accounts for systemic risk. Next, we formulate the portfolio selection problem using our proposed ratio.

3.1 Conditional Rachev ratio

Unlike Biglova et al. (2014), where they define systemic event (SE) by idiosyncratic (individual) risk events, in our paper, SE occurs when the market return goes below a certain threshold C over the next month, i.e., SE $= \{R_m < C\}$.[1] This definition is in line with the systemic risk literature, see, for example, Adrian and Brunnermeier (2016), Brownlees and Engle (2016), and Acharya et al. (2017). We assume that there exists a benchmark

[1] In this subsection, we omit the subscript t for simplicity.

systemic risk index (e.g. S&P 500 Index), that reflects broad market conditions. And the investors aim to maximize an ex-ante Rachev ratio conditional on a SE. By implementing our investment strategy, one can find portfolios that deliver the best tradeoff between reward and risk under non-Gaussianity and systemic risk.

In order to construct our new performance measure, we first briefly review a well-known systemic risk measure namely CoVaR proposed by Adrian and Brunnermeier (2016). The CoVaR corresponds to the value-at-risk (VaR) of firm i's return obtained conditioning on some SE denoted by $C(R_m)$, say $\text{CoVaR}_\alpha^{i|C(R_m)}$, is implicitly defined as

$$Pr(R_i \leq -\text{CoVaR}_\alpha^{i|C(R_m)}) = \alpha, \quad \alpha \in (0,1). \tag{3}$$

Following the similar idea of Capponi and Rubtsov (2022), we replace R_i with the portfolio's active return $(R_p - R_b)$ and $C(R_m)$ with SE, and obtain the CoVaR of our portfolio denoted by $\text{CoVaR}_\alpha^{p|\text{SE}}$. Given the above, we now define the conditional measure of risk (hereafter CoETL) which is used to build our performance measure:

$$\text{CoETL}(R_p;\alpha) := -E(R_p - R_b | R_p - R_b \leq -\text{CoVaR}_\alpha^{p|\text{SE}}). \tag{4}$$

The CoETL quantifies the conditional expected tail loss of a portfolio relative to a benchmark strategy when the market is in distress. Thus, CoETL can be used to measure portfolio-level systemic risk. Notice that CoETL can be interpreted as the portfolio's CoES, where CoES was initially mentioned by Adrian and Brunnermeier (2016) and later extended to the context of portfolio choice by Capponi and Rubtsov (2022). Here, if we denote $X = (R_b - R_p)$ as benchmark underperformance, then $-X = (R_p - R_b)$ stands for the active portfolio return. Consequently, the conditional measure of reward (hereafter

CoETP) can be formulated as

$$\text{CoETP}(R_p;\alpha) := E(R_p - R_b | R_p - R_b \geq \text{CoVaR}_{1-\alpha}^{p|\text{SE}}), \tag{5}$$

which measures the mean gains that are greater than the $(1-\alpha)$-conditional percentile of $(R_p - R_b)$. Finally, based on the terms (4) and (5), the conditional Rachev ratio (CoRR) is defined as

$$\text{CoRR}(R_p;\alpha,\beta) := \frac{\text{CoETP}(R_p;\alpha)}{\text{CoETL}(R_p;\beta)}, \tag{6}$$

where the two performance levels α and β can be set to different values, and more discussions about the choice of these numbers will be provided in empirical analysis.

To indicate the severity of SE, different choices of C can be adopted. In our paper, we follow Adrian and Brunnermeier (2016) and Acharya et al. (2017) and set C as the negatively signed VaR of market return, i.e.,

$$\text{SE} = \big\{R_m < -\text{VaR}_\alpha(R_m)\big\}. \tag{7}$$

In the empirical analysis, we adopt two threshold values namely $\text{VaR}_{1\%}(R_m)$ (hereafter C1) and $\text{VaR}_{5\%}(R_m)$ (hereafter C2). In terms of the choice of the benchmark rate, we follow Lin et al. (2023) and consider $R_b = R_m$.[2]

3.2 Portfolio selection problem

Suppose that there are N risky assets in our economy. Hereafter, we formulate the asset allocation problem based on the maximization of some performance measures. Before we describe our portfolio problem, let us first define some notations that will be used later on.

[2] Maximizing the absolute performance of the portfolio (i.e. $R_b = 0$) using CoSR and CoRR measures tends to result in extreme portfolio compositions since the absolute portfolio return is hard to be positive under extreme market conditions. Therefore, we focus on the case where our investors aim to benchmark to the market index (i.e. $R_b = R_m$) with the proposed approach.

Let $\boldsymbol{R}_t = (R_{1,t}, ..., R_{N,t})^T$ be the vector of monthly returns over month t, $R_{m,t}$ be the market return over month t, and $\boldsymbol{W}_t = (\omega_{1,t}, ..., \omega_{N,t})^T$ be the vector of portfolio weights held over month $t+1$. The portfolio return over next month is denoted by $R_{p,t+1} = \boldsymbol{W}_t^T \boldsymbol{R}_{t+1}$. $\mathbf{0}$ and $\mathbf{1}$ denote the column vector of zeros and ones, respectively.

A generic portfolio optimization problem when an investor's objective function is given by a performance measure $\rho(\cdot)$ can be described as follows

$$\boldsymbol{W}_t^* = \arg\max_{\boldsymbol{W}_t} \rho(R_{p,t+1}), \quad \text{s.t. } \mathbf{1}^T \boldsymbol{W}_t = 1, \tag{8}$$

where the different candidates of $\rho(\cdot)$ result in different optimal portfolios. In particular, the portfolio selection problem under CoRR is given by $\rho(R_{p,t+1}) = \text{CoRR}(R_{p,t+1}; \alpha, \beta)$.

In practice, it is often the case for investors to place additional constraints on the optimization. For instance, we might want to restrict the portfolio weights such that none of them is greater than a certain amount of the overall wealth invested in the portfolio, or we might want to prohibit short selling by allowing only long positions. The latter scenario is realistic in settings characterized by systemic risk in which financial regulators ban short-selling to reduce short-term investment with speculative motives. Hence, we consider no short-sale constraint ($\boldsymbol{W} \geq \mathbf{0}$) in our later exercise.

We consider three different types of benchmark strategies. The first includes CoRR portfolios constructed based on CQR of Zou and Yuan (2008) (hereafter CQR-CoRR). The second contains two different optimization criteria using SPNN, one is the unconditional Sharpe ratio (hereafter SPNN-SR), another is the conditional Sharpe ratio (CoSR) proposed by Lin et al. (2023) (hereafter SPNN-CoSR). The last consists of the well-diversified equal-weighted portfolio (1/N), which does not rely on any model estimation.

4 Empirical analysis

4.1 Data

Our empirical analysis is conducted on a monthly cross-sectional US dataset that spans from January 1985 to December 2021. Following Gu et al. (2020), we adopt 94 monthly firm characteristics. In addition, we consider 14 macroeconomic variables. Among those 8 are adopted by Gu et al. (2020), including dividend-price ratio (macro_dp), earnings-price ratio (macro_ep), book-to-market ratio (macro_bm), net equity expansion (macro_ntis), Treasury-bill rate (macro_tbl), term spread (macro_tms), default spread (macro_dfy), and stock variance (macro_svar); 6 are uncertainty indices proposed by Ludvigson et al. (2021), which covers total real uncertainty index (macro_TRU), economic real uncertainty index (macro_ERU), total macro uncertainty index (macro_TMU), economic macro uncertainty index (macro_EMU), total financial uncertainty index (macro_TFU), and economic financial uncertainty index (macro_EFU). Furthermore, we also include 74 industry dummies following Gu et al. (2020). In summary, the dimension of our predictor set is $94 \times (14 + 1) + 74 = 1484$.

The sample period of Gu et al. (2020) spans from March 1957 to December 2016. However, their original data involves a large number of variables with missing values.[3] After deleting missing data, the remaining sample spans from January 1985 to December 2021. To alleviate the computational burden associated with network training, we further restrict our data to firms existing throughout the whole sample period. The resulting balanced data panel contains 256,632 monthly observations with 577 firms in total.

[3] All data before January 1985 contains at least one variable with a large portion of missing observations. Thus, filling in those missing variables with the monthly cross-sectional medians as implemented by Gu et al. (2020) is impractical. We thus decide to only focus on the sample period without missing observations.

4.2 Asset choice

As argued by Lin et al. (2023), big financial institutions are preferred in systemic risk-based portfolio analysis since they are more exposed to market distress than non-financial counterparts. Their pre-analysis results have shown that the SE-based objective function is more relevant when the universe of portfolio assets covers large financial institutions that are potentially systemic, though not necessarily classified as Systemically Important Financial Institutions (SIFIs). Note that our aim is not to only minimize the systemic risk of a portfolio but also to maximize its profit under stressed market conditions. Those systemic firms might also exhibit positive active returns, so it may be profitable to invest in them as well. Therefore, we consider large financial firms in our portfolio analysis. Following the same filter criterion of Lin et al. (2023), we obtain a list of 38 portfolio assets including 17 SIFIs and 21 non-SIFIs. These firms are listed in Table 1.

4.3 SPNN modelling

We forecast return quantiles using a recursive window method. To achieve this, we first divide our original sample into two disjoint but consecutive subsamples. The first subsample - known as in-sample - is further decomposed into a training subsample $\mathcal{L}_1$ and a validation subsample $\mathcal{L}_2$ that we use to estimate and select the best SPNN model, respectively. The second subsample - known as out-of-sample - represents a testing subsample $\mathcal{L}_3$ on which we make final forecasts. The starting window covers 180 monthly observations, which spans from January 1985 to December 1999. The incremental size of estimation windows is a one-month period, resulting in an out-of-sample that includes 264 monthly observations spanning from January 2000 to December 2021.

It is well known that the ML models are prone to overfit the data, so it is critical to tune hyperparameters. Following Gu et al. (2020), we use the validation subsample

$\mathcal{L}_2$ to do the model selection. Specifically, for every iteration, we use as a validation subsample $\mathcal{L}_2$ the last one-year/12-month cross-sectional data of each in-sample for all 577 firms and the market. We estimate our SPNN model on $\mathcal{L}_1$ using different combinations of hyperparameters. The subsequent validation subsample $\mathcal{L}_2$ is exploited for determining optimal hyperparameters through evaluating the predicted conditional quantiles based on fitted models obtained on $\mathcal{L}_1$ with respect to each hyperparameter set. In particular, the hyperparameters are tuned by minimizing the quantile score (QS) over $\mathcal{L}_2$.

As for data preprocessing, we normalize covariates so each is scaled within the range $[0, 1]$. We first normalize the data on $\mathcal{L}_1$ when selecting optimal hyperparameters and then normalize all observations within the in-sample ($\mathcal{L}_1 + \mathcal{L}_2$) when making final forecasts. Due to the computational intensity of ML-based approaches, instead of recursively estimating the model for each month, we do it on an annual basis (i.e. every 12 months) and keep the estimates to make predictions for the following year.

4.4 Portfolio formation

After fitting SPNN models, we obtain quantile forecasts of monthly returns, based on which we estimate the conditional marginal return distributions following the method discussed in Appendix A.1. Combining the distributional forecasts with the fitted t-copula model, we generate 30,000 return scenarios at the beginning of each out-of-sample month. The portfolio optimization problem defined in (8) is solved on a monthly basis by maximizing the ex-ante CoRR measure based on generated return scenarios. To obtain a robust estimator of our CoRR measure, we follow Biglova et al. (2014) and set $\alpha = \beta = 10\%$.

We perform three steps to compute the final wealth and cumulative return at the k-th rebalancing, for $k \in \{0, ..., 263\}$. We first generate return scenarios based on the algorithms described in Appendix A, and obtain the optimal weights $\boldsymbol{W}^*_{k+1}$ for each of the performance measures under consideration. Then, we compute the final wealth as

$$\mathrm{FW}_{k+1} = \mathrm{FW}_k(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1}), \tag{9}$$

where $\boldsymbol{R}_{k+1}$ is the vector of realized returns over period $k+1$ and $\mathrm{FW}_0 = 1$. Lastly, the cumulative return is computed as

$$\mathrm{CR}_{k+1} = \mathrm{CR}_k + \ln(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1}), \tag{10}$$

where $\mathrm{CR}_0 = 0$. Note that the latter equation reports the cumulative performance of the portfolio net of wealth. That is, expression (9) implies that $\mathrm{FW}_{K+1} = \mathrm{FW}_0 \prod_{k=0}^{K}(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1})$. Taking logs of both sides of the latter equation, we obtain $(\ln \mathrm{FW}_{K+1} - \ln \mathrm{FW}_0) = \sum_{k=0}^{K} \ln(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1})$. Therefore, the growth in wealth due to the cumulative return on the portfolio is given by expression (10). By repeatedly computing FW_{k+1} and CR_{k+1} for different strategies, we obtain the ex-post paths of final wealth and cumulative return over the evaluation period.

4.5 Results

In this section, we first evaluate return quantile forecasts using standard diagnostic tests. Then we examine the predictive power of predictors using two variable importance measures namely mean squared sensitivity (MSS) and quantile causality measure (QC). Thereafter, we display backtesting results with and without accounting for transaction costs. Lastly, we calculate the portfolio's long-run marginal expected shortfall (LRMES) and CoES to compare the systemic risk of candidate strategies.

4.5.1 Evaluation of quantile forecasts

To present some insights on return quantile forecasts obtained from SPNN models, in Figure 1 we display the realized returns and the prediction intervals obtained using SPNN1. To conserve space, we only show relevant results for the market portfolio and three portfolio assets (CMA, WFC and JPM).[4] From Figure 1, we see that the SPNN1-based return quantile forecasts are able to capture most of the variation of realized returns, especially during crisis episodes.

To further assess the quality of quantile estimates, we backtest predicted quantile series using three kinds of tests, namely the Conditional Coverage (CC) test of Christoffersen (1998) (hereafter $\mathrm{LR_{CC}}$), the Dynamic Quantile test of Engle and Manganelli (2004) (hereafter DQ), and the Dynamic Binary test of Dumitrescu et al. (2012) (hereafter DB). Tables 4 to 8 report the p values for all candidate tests obtained from SPNN1-based quantile estimates, considering quantile levels 0.05, 0.25, 0.5, 0.75 and 0.95.[5] Specifically, the DB1-DB7 are specifications proposed in Dumitrescu et al. (2012), while the DQ1-DQ3 and DQVaR1-DQVaR3 specifications refer to the DQ tests with only lagged hits and with both lagged hits and the contemporaneous VaRs as defined by Engle and Manganelli (2004), respectively.

Turning our attention to the results displayed in Tables 4 to 8, the first notable results is that only a few number of p values are below 1% significance level (which suggests a rejection of the CC hypothesis). This confirms the validity of our SPNN1-based conditional quantile forecasts in most cases. Take $\tau = 0.05$ for example, even for the BK asset with some negative results, the majority of tests (8 out of 14) still favor our prediction model. In addition, due to the dichotomic nature of the dependent variable, the DQ test might not be an appropriate choice for the inference on the parameters and consequently on the hypothesis of validity of the quantiles under linear regression models. Therefore, the

[4] The results for the remaining portfolio assets are available upon request.

[5] For saving space, here we only report the test results for a subset of portfolio assets. In practice, we have also tested for other assets and overall this subset is representative of the whole sample.

positive results of the nonlinear regression-based DB tests for BK are still supportive of our model. The similar conclusion is also held for other quantile levels under consideration.

4.5.2 Variable importance

Next, we measure the variable importance within both training and testing subsamples. Gu et al. (2020) highlighted the importance of analyzing the contributions of individual predictors for better interpreting ML-based models. Unlike Gu et al. (2020) who computed the change in out-of-sample R^2 to measure the variable importance in the context of mean regression, hereafter we adopt two measures that are directly related to measuring the performance of quantile forecasts. As a first measure, we consider the Mean Squared Sensitivity (MSS) that measures the sensitivity of m-th output neuron with respect to p-th input variable (Zurada et al. 1994; Yeh and Cheng 2010):

$$\mathrm{MSS}_{p,m} = \sqrt{\frac{\sum_{t\in(\mathcal{L}_1+\mathcal{L}_2)} \left(s_{p,m}|_{\boldsymbol{X}_t}\right)^2}{|\mathcal{L}_1| + |\mathcal{L}_2|}}, \tag{11}$$

with

$$s_{p,m}\big|_{\boldsymbol{X}_t} = \frac{\partial \hat{Q}_{R_{t+1}}(\tau_m|\boldsymbol{X}_t)}{\partial x_{p,t}}(\boldsymbol{X}_t), \tag{12}$$

where $\boldsymbol{X}_t = (x_{1,t}, ..., x_{P,t})^T$ refers to the t-th observation of P predictors within the in-sample $(\mathcal{L}_1 + \mathcal{L}_2)$, $s_{p,m}\big|_{\boldsymbol{X}_t}$ denotes the sensitivity of m-th output neuron (which in our case is the τ_m-th conditional quantile) with respect to p-th input neuron evaluated at $\boldsymbol{X}_t$, and $|\mathcal{L}_i|$ denote the number of observations in set $\mathcal{L}_i$, for $i = \{1, 2\}$. The sensitivity term (12) is calculated using the chain rule, see Pizarroso et al. (2020) for more computational details. By computing MSS, we can measure the sensitivity of model estimation/prediction to the changes in a candidate predictor. In practice, for each predictor x_p, we compute the following average MSS

$$\widetilde{\text{MSS}}_p = \frac{1}{M}\sum_{m=1}^{M} \text{MMS}_{p,m}. \tag{13}$$

It is worth noting that MSS defined above is able to identify and rank predictors of QRNN models across all quantiles of interest.

Next, we consider the QRNN causality measure developed by Lin and Taamouti (2023), which is an extension of the Quantile Causality (QC) measure proposed by Song and Taamouti (2021). Specifically, for $\tau \in (0, 1)$, the QC of the p-th input variable in QRNN model is defined as

$$\text{QC}_p(\tau) = \ln\left[\frac{E\left[\rho_\tau\left(R_{t+1} - Q_{R_{t+1}}(\tau|\overline{\boldsymbol{X}}_t)\right)\right]}{E\left[\rho_\tau\left(R_{t+1} - Q_{R_{t+1}}(\tau|\boldsymbol{X}_t)\right)\right]}\right], \tag{14}$$

where $\overline{\boldsymbol{X}}_t$ denotes the information set of predictors available by month t, except for the p-th predictor. $\text{QC}_p(\tau)$ measures the degree of Granger causality from a certain predictor p to the τ-th quantile of the predictand given the past of the latter. QC quantifies the predictive information provided by the historical observations of p-th predictor regarding the prediction of τ-th conditional return quantile. Similar to the average measure $\widetilde{\text{MSS}}_p$, in our empirical analysis we compute the average QC for each predictor x_p as

$$\widetilde{\text{QC}}_p = \ln\left[\frac{\frac{1}{M|\mathcal{L}_3|}\sum_{m=1}^{M}\sum_{t\in\mathcal{L}_3}\rho_{\tau_m}\left(R_{t+1} - \hat{Q}_{R_{t+1}}(\tau_m|\overline{\boldsymbol{X}}_t)\right)}{\frac{1}{M|\mathcal{L}_3|}\sum_{m=1}^{M}\sum_{t\in\mathcal{L}_3}\rho_{\tau_m}\left(R_{t+1} - \hat{Q}_{R_{t+1}}(\tau_m|\boldsymbol{X}_t)\right)}\right], \tag{15}$$

where the marginal contribution of each predictor x_p is assessed using the out-of-sample $\mathcal{L}_3$ only, whose data does not overlap with those of training or tuning samples.

Based on the SPNN1 model, Figure 2 reports the variable importance measured by MSS for the 10 most influential firm-level predictors and all macroeconomic variables, while Figure 3 displays the corresponding results for QC measure.[6] Note that the variable

[6] To save space, hereafter we only report the variable importance results obtained by SPNN1 model. The corresponding results for other SPNN configurations are similar and are available upon request.

importance is normalized to sum up to one, which makes it easier to interpret the relative importance of the predictive power of each predictor compared to those of others. Variables with the highest (lowest) importance are displayed on the top (bottom).

The top 10 most influential firm-level features measured by MSS as shown in the top panel of Figure 2 can be grouped into five categories. The first group contains risk measures such as the total and idiosyncratic return volatility (retvol and idiovol); the second one considers liquidity variables including the dollar volume (dolvol), the bid-ask spread (baspread), the scaled average trading volume (turn), and the turnover-weighted number of zero trading days (zerotrade); the third group contains a single momentum predictor namely the short-term reversal (mom1m); the fourth group includes fundamental variables of the fundamental performance indicator (ms) and the price-sales ratio (sp); the last group consists of industry dummy (sic2). As for the macroeconomic variables, from the bottom panel of Figure 2, we see that all of them contribute significantly to the model training, but among those, the total financial uncertainty index (macro_TFU) is ranked as the most influential macro-level predictor.

Analogously, the rankings based on the QC measure as shown in Figure 3 draw similar conclusions. The results reveal a fairly small set of dominant firm-level predictors, which covers the risk measure retvol; the liquidity variable dolvol; the short-term reversal mom1m; the industry dummy sic2; the fundamental variables of ms and the fundamental health score (ps); the accounting variables of the number of years since first Compustat coverage (age), the SG&A ratio (operprof), the tax income (tb), and the sum of returns around earnings announcement (ear). For the macro variables, the results confirm again their predictive power and place the greatest emphasis on the macro_dp in this case.

To further illustrate the variable importance, Figures 4 to 6 display the time-varying rankings of the predictors in SPNN1 as measured by MSS and QC respectively. In particular, these figures rank the importance of individual predictors according to their average contribution in terms of predictive power over all quantiles of returns and across all in-sample and out-of-sample windows depending on the measure in use. Characteristics are sorted based on their average ranks over all windows, with the most (least) influential ones placed at the top (bottom). The results displayed in these figures again confirm the most influential firm- and macro-level predictors as identified before.

4.5.3 Backtesting results

In this section, we use the return quantile forecasts obtained from the fitted SPNN models to estimate conditional marginal return distributions, based on which we simulate returns using the copula method and solve the portfolio optimization problem thereafter. We perform a backtesting analysis to evaluate the economic gains of applying SPNN-based probabilistic return forecasts to portfolio selection under systemic risk. In particular, we compare the out-of-sample performance of SPNN-CoRR portfolios with those of several benchmark portfolios. The optimized portfolios were built recursively using different performance measures that are estimated from simulated returns obtained from different statistical models. Note that all portfolios are monthly rebalanced.

The backtesting results obtained based on SPNN1 model are displayed in Figure 7.[7] There are several noticeable features from these figures. Firstly, we observe that all portfolios perform less well during the 2007-2008 financial crisis and the recent COVID-19 pandemic. The SPNN1-SR and 1/N strategies lose all their values during the global financial crisis period, while the SPNN1-CoRR portfolios perform significantly better than others, even though they lost around half of their values since the last peak in 2007. In particular, the SPNN1-CoRR with C1 delivers the best out-of-sample performance. Secondly, the CQR-CoRR portfolios can be identified as strong competitors, where their performance

[7] We omit the backtesting results obtained by other SPNN configurations since the portfolio performance does not vary significantly. Our findings are in agreement with Gu et al. (2020), where the authors argued that "shallow" learning outperforms "deep" learning. Increasing the model complexity is not necessarily beneficial in terms of economic gains.

is better than that of SPNN1-CoSR portfolios but worse than that of SPNN1-CoRR portfolios. Thirdly, all portfolios that account for systemic risk in performance criteria show a strong upward trend in profitability throughout the evaluation period. This strong performance can be mainly attributed to their relatively stable performance during market distress. In short, our backtesting results confirm the benefits of combining SPNN-based return forecasts with the incorporation of systemic risk into the unconditional Rachev ratio when constructing optimal portfolios.

To further illustrate the outperformance of our approach against other benchmark strategies under investigation, we consider an additional exercise where we compare the performance of optimized portfolios using the same criterion based on different models and using the same model based on different criteria. Let us first focus on the first case, where we maximize the same CoRR criterion using different statistical models (CQR and SPNN1). The corresponding backtesting results are displayed in the top panel of Figure 8, from which we confirm the outperformance of the SPNN1 model against the CQR model, with the latter being considered as an advanced variant of quantile-based models. Next, we compare different performance criteria under the same model. Specifically, we consider three different performance measures under SPNN1, namely SR, CoSR and CoRR. As we can see from the bottom panel of Figure 8, the backtesting results demonstrate the superiority of our CoRR measure against other criteria under the same SPNN1 model. Thus, the results of these two cases consistently favor our approach.

Table 2 reports the values of several statistics that are used to measure ex-post portfolio performance. The results vary among different strategies depending on the performance criteria and statistical models used in portfolio optimization, with the exception being the 1/N portfolio which does not rely on any optimization or model estimation. Overall, the SPNN1-CoRR portfolios perform the best in terms of out-of-sample profitability. Moreover, the SPNN1-CoRR portfolio with C1 outperforms CQR-CoRR and SPNN1-CoSR portfolios by a wide margin, with the latters being considered as robust benchmarks. Specifically, using the SPNN1-CoRR portfolio with C1, investors would multiply their wealth by 41.9878, which is near twice that of the SPNN1-CoSR portfolio with C1 (17.4357). Unsurprisingly, the naive 1/N portfolio offers the lowest final wealth of 9.5409 and an annual return of 0.1080. The results for the Sharpe ratio, Sortino ratio and Calmar ratio again demonstrate the superiority of our proposed approach, where the SPNN1-CoRR portfolio with C2 delivers the highest values of Sharpe ratio (0.7690), while the SPNN1-CoRR portfolio with C1 presents the highest Sortino ratio (1.2714) and Calmar ratio (0.3740) among all competitors.

Besides the above-mentioned performance ratios, investors may consider alternative measures to gain deeper insights into their trading strategies. Therefore, we add maximum drawdown (MDD), average turnover rate (TO), and Farinelli-Tibiletti (FT) ratio as alternative metrics. Formally, the MDD is calculated as

$$\mathrm{MDD} = \max_{t_0 \leq t_1 \leq t_2 \leq T_0} \left\{ r_{p,t_0:t_1} - r_{p,t_0:t_2} \right\}, \tag{16}$$

where $r_{p,t_0:t_i}$, for $i \in \{1,2\}$ denotes the cumulative portfolio return from time t_0 to t_i, with t_0 and T_0 being the first and last month of evaluation period. The average TO is defined as

$$\mathrm{TO} = \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \left| \omega_{i,t+1} - \frac{\omega_{i,t}(1 + R_{i,t+1})}{1 + \sum_{j=1}^{N} \omega_{j,t} R_{j,t+1}} \right| \right), \tag{17}$$

where $\omega_{i,t}$ is the desired weight of portfolio asset i at time t. The FT ratio was proposed by Farinelli and Tibiletti (2008) to capture the asymmetric information of portfolio return distribution. Unlike the Sharpe ratio, which measures the tradeoff between reward and risk via two-sided type measures (by which the asymmetric deviations from the benchmark are equally weighted), the FT ratio is a one-sided type measure that describes the volatility above and below a benchmark. Formally, the FT ratio is given by

$$\mathrm{FT}(R_p; p, q) = \frac{\left(E(R_p - R_b)_+^p\right)^{1/p}}{\left(E(R_b - R_p)_+^q\right)^{1/q}}, \tag{18}$$

where $(X)_+ = \max(X, 0)$, and $p \geq 1$, $q \geq 1$ are the orders of the corresponding partial moments. The FT ratio is an alternative reward-risk measure that is compatible with skewed return distributions, see for example Bouaddi and Taamouti (2013). Note that the FT ratio implicitly embraces some well-known indices in the literature. For example, for $p = q = 1$, FT represents the Omega ratio of Keating and Shadwick (2002), while for $p = 1$ and $q = 2$, FT corresponds to the Upside Potential ratio of Sortino et al. (1999).

Table 2 reports the values of the above-mentioned alternative measures as well. Overall speaking, the SPNN1-CoRR portfolios possess the lowest MDD among all candidate strategies. In particular, the SPNN1-CoRR portfolio with C1 presents the lowest MDD of 0.4951, while the SPNN1-CoRR portfolio with C2 delivers the second lowest MDD of 0.5027. In terms of the FT ratios, the SPNN1-CoRR portfolio with C1 dominates other strategies in all cases. This indicates that our proposed approach achieves better performance under different asymmetric preferences depending on different choices of partial moment orders.

4.5.4 Effect of transaction costs

The calculation of transaction cost (TC) is based on TO as defined in (17). After accounting for a proportional TC of c, the portfolio return is now calculated as follows:

$$\tilde{R}_{p,t+1} = (1 + R_{p,t+1})\left(1 - c\sum_{i=1}^{N}\left|\omega_{i,t+1} - \frac{\omega_{i,t}(1 + R_{i,t+1})}{1 + \sum_{j=1}^{N}\omega_{j,t}R_{j,t+1}}\right|\right) - 1. \tag{19}$$

Given the major role that momentum predictors played in ML models, it is expected that SPNN-based trading strategies are characterized by relatively high TO, see Gu et al. (2020). Promisingly, as we can see from Table 2, the SPNN1-CoRR portfolio with C1 has a TO of 0.1194, which is lower than that of the CQR-CoRR portfolio with C1 (0.2117) and the SPNN1-CoSR portfolio with C1 (0.1772). The SPNN1-SR portfolio possesses the highest TO of 0.2779. Unsurprisingly, the 1/N portfolio delivers the lowest TO (0.0242) due to its well-diversified property.

Although the ML-based portfolios with relatively high TO are more flexible to adapt to the changes in market conditions than other benchmarks, their values are likely to decrease due to the higher rebalancing TC. To analyze the effect of TC, we set a moderate level of $c = 20$ basis points (bps) and recompute the ex-post paths of final wealth and cumulative return for all portfolios under consideration. Figure 9 illustrates the ex-post paths of final wealth and cumulative return after taking into account TC, whereas Table 3 reports the updated values of performance metrics. In short, we find that the inclusion of proportional TC does not alter our main conclusions. The SPNN1-CoRR portfolios still outperform all other competitors in terms of profitability and performance metrics. Remarkably, the final wealth of the SPNN1-CoRR portfolio with C1 (37.0063) is more than one and a half times that of the CQR-CoRR portfolio with C1 (21.7091) and is more than two and a half times that of the SPNN1-CoSR portfolio with C1 (14.4584).

4.5.5 Portfolio-level systemic risk

In this section, we define two portfolio-level systemic risk measures. The first one is the portfolio's LRMES (Lin et al. 2023):

$$\mathrm{LRMES}_p = \sum_{i=1}^{N} \omega_i \, \mathrm{LRMES}_i, \tag{20}$$

where LRMES_i indicates the expected loss of asset i over next month. The LRMES_p can be interpreted as the expected percentage drop in portfolio value under stressed market conditions, which we estimate using generated return scenarios. In the same spirit, we extend the CoES measure to a portfolio-level version as follows

$$\mathrm{CoES}_{\alpha}^{p|\mathrm{SE}} = \sum_{i=1}^{N} \omega_i \, \mathrm{CoES}_{\alpha}^{i|\mathrm{SE}}, \tag{21}$$

where $\mathrm{CoES}_{\alpha}^{i|\mathrm{SE}} = E(R_i | R_i \leq \mathrm{CoVaR}_{\alpha}^{i|\mathrm{SE}})$ refers to the expected tail loss of asset i conditional on market distress.[8] Compared to the portfolio's LRMES defined previously, the portfolio's CoES considers a more extreme scenario where both portfolio assets and the market can be in a low-return environment.

Figure 10 illustrates the time-varying portfolio's LRMES and CoES over the evaluation period. Overall speaking, the SPNN1-CoRR portfolios offer the best performance in terms of both systemic risk measures. The relatively low values of their LRMES and CoES indicate that they tend to suffer from less potential losses during crisis periods. Specifically, the SPNN1-CoRR portfolio with C1 provides the lowest LRMES over the first third of the evaluation period, while the SPNN1-CoRR portfolio with C2 becomes hard to beat over the rest of the period. The SPNN1-CoSR portfolio with C2 is a serious competitor that presents slightly higher LRMES in the middle of the evaluation period. Similarly, the SPNN1-CoRR portfolio with C1 delivers the lowest CoES among all candidate competitors throughout the out-of-sample period.

5 Conclusions

In this paper, we propose a novel performance ratio that simultaneously takes into account systemic risk and non-Gaussianity when building optimal portfolios. The proposed measure extends the unconditional Rachev ratio by explicitly incorporating the occurrence of extreme events. To robustify the portfolio optimization and better represent the extreme

[8] It is worth noting that CoES is subadditive and is able to account for distributional aspects within the conditional tail.

market events, we generate a large number of return scenarios via a Monte Carlo method. This is done by first obtaining probabilistic return forecasts via a quantile regression neural network (regarded as a distributional machine learning approach), and then simulating returns via a fitted t-copula model. Thereafter, a large-scale comparative analysis using US data is conducted to compare the out-of-sample performance of the proposed portfolio selection approach against benchmark strategies. The backtesting results demonstrate the superiority of our approach in terms of profitability, with its outperformance staying robust after the inclusion of moderate transaction costs. Furthermore, we compare the portfolio-level systemic risk among all candidates using LRMES and CoES measures. Our SPNN-CoRR portfolio is not only characterized by the highest profitability, but it also delivers the lowest systemic risk throughout the evaluation period.

References

Abe, M. and H. Nakayama (2018). Deep learning for forecasting stock returns in the cross-section. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 273–284. Springer.

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2017). Measuring systemic risk. *The Review of Financial Studies 30*(1), 2–47.

Adrian, T. and M. K. Brunnermeier (2016). Covar. *American Economic Review 106*(7), 1705–41.

Babiak, M. and J. Baruník (2020). Deep learning, predictability, and optimal portfolio returns. *arXiv preprint arXiv:2009.03394*.

Biglova, A., S. Ortobelli, and F. J. Fabozzi (2014). Portfolio selection in the presence of systemic risk. *Journal of Asset Management 15*(5), 285–299.

Biglova, A., S. Ortobelli, S. T. Rachev, and S. Stoyanov (2004). Different approaches to risk estimation in portfolio theory. *The Journal of Portfolio Management 31*(1), 103–112.

Bouaddi, M. and A. Taamouti (2013). Portfolio selection in a data-rich environment. *Journal of Economic Dynamics and Control 37*(12), 2943–2962.

Brownlees, C. and R. F. Engle (2016). SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies 30*(1), 48–79.

Cannon, A. J. (2011). Quantile regression neural networks: implementation in R and application to precipitation downscaling. *Computers & Geosciences 37*, 1277–1284. doi:10.1016/j.cageo.2010.07.005.

Cannon, A. J. (2018). Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. *Stochastic Environmental Research and Risk Assessment 32*(11), 3207–3225.

Capponi, A. and A. Rubtsov (2022). Systemic risk-driven portfolio selection. *Operations Research 70*(3), 1598–1612.

Chen, L., M. Pelger, and J. Zhu (2019). Deep learning in asset pricing. *arXiv preprint arXiv:1904.00745*.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841–862.

Dumitrescu, E.-I., C. Hurlin, and V. Pham (2012). Backtesting value-at-risk: from dynamic quantile to dynamic binary tests. *Finance 33*(1), 79–112.

Engle, R. F. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *22*, 367–381.

Farinelli, S. and L. Tibiletti (2008). Sharpe thinking in asset ranking with one-sided measures. *European Journal of Operational Research 185*(3), 1542–1547.

Feng, G., J. He, and N. G. Polson (2018). Deep learning for predicting asset returns. *arXiv preprint arXiv:1804.09314*.

Feng, G., N. Polson, and J. Xu (2021). Deep learning in characteristics-sorted factor models. *Available at SSRN 3243683*.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1), 429–450.

Hatalis, K., A. J. Lamadrid, K. Scheinberg, and S. Kishore (2019). A novel smoothed loss and penalty function for noncrossing composite quantile estimation via deep neural networks. *arXiv preprint arXiv:1909.12122*.

Huang, X., M. Guidolin, E. Platanakis, and D. Newton (2021). Dynamic portfolio management with machine learning. *Available at SSRN 3770688*.

Hüttel, F. B., I. Peled, F. Rodrigues, and F. C. Pereira (2022). Modeling censored mobility demand through censored quantile regression neural networks. *IEEE Transactions on Intelligent Transportation Systems 23*(11), 21753–21765.

Jan, M. N. and U. Ayub (2019). Do the FAMA and FRENCH Five-Factor model forecast well using ANN? *Journal of Business Economics and Management 20*(1), 168–191.

Kaczmarek, T. and K. Perez (2021). Building portfolios based on machine learning predictions. *Economic Research-Ekonomska Istraživanja*, 1–19.

Keating, C. and W. F. Shadwick (2002). An introduction to omega. *AIMA Newsletter*.

Lin, W., J. Olmo, and A. Taamouti (2023). Portfolio selection under systemic risk. *Journal of Money, Credit and Banking*, to appear.

Lin, W. and A. Taamouti (2023). Measuring Granger causality in quantile regression neural network. Technical report, Working paper, Durham University.

Ludvigson, S. C., S. Ma, and S. Ng (2021). Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics 13*(4), 369–410.

Masters, T. (1993). *Practical neural network recipes in C++*. Morgan Kaufmann.

McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

Messmer, M. (2017). Deep learning and the cross-section of expected returns. *Available at SSRN 3081555*.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized'optimal? *Financial analysts journal 45*(1), 31–42.

Ortobelli, S., S. T. Rachev, S. Stoyanov, F. J. Fabozzi, and A. Biglova (2005). The proper use of risk measures in portfolio theory. *International Journal of Theoretical and Applied Finance 8*(08), 1107–1133.

Ovadia, Y., E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek (2019). Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *Advances in neural information processing systems 32*.

Pizarroso, J., J. Portela, and A. Muñoz (2020). NeuralSens: sensitivity analysis of neural networks. *arXiv preprint arXiv:2002.11423*.

Quinonero-Candela, J., C. E. Rasmussen, F. Sinz, O. Bousquet, and B. Schölkopf (2005). Evaluating predictive uncertainty challenge. In *Machine Learning Challenges Workshop*, pp. 1–27. Springer.

Rachev, S., S. Ortobelli, S. Stoyanov, F. J. Fabozzi, and A. Biglova (2008). Desirable properties of an ideal risk measure in portfolio theory. *International Journal of Theoretical and Applied Finance 11*(01), 19–54.

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

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business 39*(1), 119–138.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris 8*, 229–231.

Song, X. and A. Taamouti (2021). Measuring Granger causality in quantiles. *Journal of Business & Economic Statistics 39*(4), 937–952.

Sortino, F. A. and S. Satchell (2001). *Managing downside risk in financial markets*. Elsevier.

Sortino, F. A., R. Van Der Meer, and A. Plantinga (1999). The dutch triangle. *The Journal of Portfolio Management 26*(1), 50–57.

Taylor, J. W. (2000). A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting 19*(4), 299–311.

Xu, Q., K. Deng, C. Jiang, F. Sun, and X. Huang (2017). Composite quantile regression neural network with applications. *Expert Systems with Applications 76*, 129–139.

Yeh, I.-C. and W.-L. Cheng (2010). First and second order sensitivity analysis of MLP. *Neurocomputing 73*(10-12), 2225–2233.

Zhang, Z., S. Zohren, and S. Roberts (2020). Deep learning for portfolio optimization. *The Journal of Financial Data Science 2*(4), 8–20.

Zheng, S. (2011). Gradient descent algorithms for quantile regression with smooth approximation. *International Journal of Machine Learning and Cybernetics 2*(3), 191–207.

Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics 36*(3), 1108–1126.

Zurada, J. M., A. Malinowski, and I. Cloete (1994). Sensitivity analysis for minimization of input data dimension for feedforward neural network. In *Proceedings of IEEE International Symposium on Circuits and Systems-ISCAS'94*, Volume 6, pp. 447–450. IEEE.

Online Appendix

A - Simulation algorithm

Although CoRR has no closed-form expression when the non-short-selling constraint is imposed, we can still apply a Monte-Carlo simulation-based procedure to solve the portfolio optimization problem. In practice, CoRR can be estimated using its empirical analogue that we can calculate from simulated returns over the subset of SE scenarios.

In this section, we discuss how we estimate the conditional marginal distributions (densities) of monthly returns. In particular, we consider a nonparametric estimation approach for predictive densities using conditional quantiles obtained from SPNN models. After fitting the marginal densities, we apply t-copula to model the dependence between assets and market returns. Lastly, we describe an algorithm for simulating return scenarios.

A.1 - Estimation of predictive densities

Let $\boldsymbol{X}_{j,t} = \{x_{j,p,t}\}_{p=1,...,P;\,t=1,...,T}$ for $j \in \{i, m\}$ with $i = 1, ..., N$ be the P-dimensional predictor set for monthly return of firm i or market index available at month t. Hereafter, we show how the conditional quantiles of returns obtained from SPNN, i.e. $\hat{q}_{j,t+1}(\tau_m) = \hat{Q}_{R_{j,t+1}}(\tau_m|\boldsymbol{X}_{j,t})$, can be utilized to approximate the conditional density $p_{j,t} = p(R_{j,t+1}|\boldsymbol{X}_{j,t})$. Formally, to recover the predictive probability density $\hat{p}_{j,t}(\cdot)$ based on conditional quantiles, we distinguish between the following three cases:

- If $\hat{q}_{j,t+1}(\tau_1) \leq R_{j,t+1} < \hat{q}_{j,t+1}(\tau_M)$ and τ_m and τ_{m+1} are such that $\hat{q}_{j,t+1}(\tau_m) \leq R_{j,t+1} < \hat{q}_{j,t+1}(\tau_{m+1})$, then

$$\hat{p}_{j,t} = \frac{\tau_{m+1} - \tau_m}{\hat{q}_{j,t+1}(\tau_{m+1}) - \hat{q}_{j,t+1}(\tau_m)}. \tag{22}$$

- If $R_{j,t+1} < \hat{q}_{j,t+1}(\tau_1)$, we assume a lower exponential tail

$$\hat{p}_{j,t} = z_1 \exp\Big(-\frac{|R_{j,t+1} - \hat{q}_{j,t+1}(\tau_1)|}{e_1}\Big), \tag{23}$$

where $z_1 = (\tau_2 - \tau_1)/(\hat{q}_{j,t+1}(\tau_2) - \hat{q}_{j,t+1}(\tau_1))$ and $e_1 = \tau_1/z_1$.

- If $R_{j,t+1} \geq \hat{q}_{j,t+1}(\tau_M)$, we assume an upper exponential tail

$$\hat{p}_{j,t} = z_M \exp\Big(-\frac{|R_{j,t+1} - \hat{q}_{j,t+1}(\tau_M)|}{e_M}\Big), \tag{24}$$

where $z_M = (\tau_M - \tau_{M-1})/(\hat{q}_{j,t+1}(\tau_M) - \hat{q}_{j,t+1}(\tau_{M-1}))$ and $e_M = \tau_M/z_M$.

The specifications (22) to (24) that can be viewed as a sort of semiparametric approach for estimating densities were proposed by Quinonero-Candela et al. (2005) and later exploited by other papers on distributional prediction and uncertainty analysis, see Cannon (2011), Ovadia et al. (2019), and Hüttel et al. (2022) among others. For the interior points of the support, this approach estimates the predictive density by interpolating the neighboring quantiles. While for the extreme points of the support (lower and upper tails), due to the lack of observations at the extremes, this approach uses some parametric functional forms (e.g. exponential function) to better estimate the tails of the predictive density of returns. Notice that the usage of exponential tails helps ensure that the estimated density function integrates to one.

In practice, the resulting estimated predictive densities can also be used to estimate CDF and its inverse (i.e. quantile function), see the documentation of R package **qrnn** (Cannon 2011).

A.2 - Dependence modelling and scenario generation

Once the predictive margins of portfolio assets and the market are obtained, we next model the joint return distribution via copula. An $(N+1)$-dimensional copula C is a multivariate distribution function on $[0,1]^{N+1}$, with standard uniform margins. Following Sklar's theorem (Sklar 1959), any multivariate distribution, which in our case is the multivariate distribution function of individual firm and market monthly returns, can be resolved into univariate margins and a certain copula function

$$F_{R_1,...,R_{N+1}}(u_1,...,u_{N+1}) = C\left(F_{R_1}(u_1),...,F_{R_{N+1}}(u_{N+1})\right), \tag{25}$$

where $u_j \sim U(0,1)$ for $j = 1,...,N+1$, $R_{N+1} = R_m$, and F_{R_j} denotes the marginal CDF of monthly return on an individual asset or market index.

In our empirical analysis, we adopt t-copula to model the dependence among monthly returns. The t-copula function is given by

$$C_{\nu,\boldsymbol{\mathcal{P}}}(u_1,...,u_{N+1}) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_{N+1})} \frac{\Gamma(\frac{\nu+N+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^{N+1}|\boldsymbol{\mathcal{P}}|}} \left(1 + \frac{x'\boldsymbol{\mathcal{P}}^{-1}x}{\nu}\right)^{-\frac{\nu+N+1}{2}} dx, \tag{26}$$

where Γ denotes the Gamma function, $\boldsymbol{\mathcal{P}}$ represents the correlation matrix, and ν refers to the degrees of freedom. We now generate return scenarios according to the following steps:

- Given historical monthly returns on firms and market, i.e, $\{R_{j,t}\}_{j=1,...,N+1;\, t=1,...,T}$, we estimate the empirical CDF, say $\hat{F}_{\nu_{j,t}}$, of return series $\{R_{j,t}\}$, i.e. $R_{j,t} \sim \hat{F}_{\nu_{j,t}}$.
- Convert historical monthly returns over each estimation window into standard uniforms using probability transformation: $u_{j,t} = \hat{F}_{\nu_{j,t}}(R_{j,t})$, where $u_{j,t} \sim U(0,1)$.
- Given $\{u_{j,t}\}_{j=1,...,N+1}$, we use moment method to estimate the degrees of freedom ν and the correlation matrix $\boldsymbol{\mathcal{P}}$ of the t-copula, see McNeil et al. (2015).

- Simulate dependent standard uniform vectors $\boldsymbol{u}_{t+1}^{(s)} = \left(u_{1,t+1}^{(s)}, \cdots u_{N+1,t+1}^{(s)}\right)$ for $s = 1, ..., S$, where S is the simulation sample size.

- Convert $\boldsymbol{u}_{t+1}^{(s)}$ to return scenarios via quantile transformation: $R_{j,t+1}^{(s)} = \hat{F}_{R_{j,t+1}}^{-1}(u_{j,t+1}^{(s)})$, where $\hat{F}_{R_{j,t+1}}^{-1}$ is the inverse CDF of the fitted j-th marginal empirical distribution deduced from $\hat{p}_{j,t}$ for $j \in \{i, m\}$. From this, we obtain S simulated return samples over month $t+1$ that possess the same dependence structure as the in-sample dataset.

B - CoRR estimation

Suppose that we have generated S return scenarios for each portfolio asset and market index. Let $\boldsymbol{R}_{i,t+1}^{sim} = (R_{i,t+1}^1, ..., R_{i,t+1}^S)^T$, $i \in \{1, ..., N\}$ and $\boldsymbol{R}_{m,t+1}^{sim} = (R_{m,t+1}^1, ..., R_{m,t+1}^S)^T$ denote the $S \times 1$ column vectors of simulated returns for asset i and market portfolio, respectively. Thereafter, $\boldsymbol{R}_{t+1}^{sim} = [\boldsymbol{R}_{1,t+1}^{sim}\ \boldsymbol{R}_{2,t+1}^{sim} \cdots \boldsymbol{R}_{N,t+1}^{sim}]$ denotes the $S \times N$ matrix storing simulated returns for all portfolio assets. Furthermore, $\#\mathrm{SE} = \sum_{s=1}^{S} I\{R_{m,t+1}^s < -\widehat{\mathrm{VaR}}_q(R_{m,t+1})\}$ is the number of SE scenarios based on the estimated market VaR.

To estimate the CoRR based on simulated returns, we first estimate the VaR of the market return. The one-month ahead VaR at coverage rate q is estimated using the empirical qth-quantile of the simulated market returns, say $\widehat{\mathrm{VaR}}_q(R_{m,t+1})$, for $q = 1\%$, 5%.[9] Analogously, the CoVaR of the portfolio return can be implicitly estimated by the α-th empirical quantile of the conditional probability distribution of portfolio active return:

$$Pr(\tilde{\boldsymbol{R}}_{p,t+1|SE}^{sim} \leq -\widehat{\mathrm{CoVaR}}_{\alpha}^{p|\mathrm{SE}}) := Pr(\boldsymbol{R}_{t+1|\mathrm{SE}}^{sim}\boldsymbol{W}_t - \boldsymbol{R}_{m,t+1|\mathrm{SE}}^{sim} \leq -\widehat{\mathrm{CoVaR}}_{\alpha}^{p|\mathrm{SE}}) = \alpha, \quad (27)$$

where $\boldsymbol{R}_{t+1|\mathrm{SE}}^{sim}$ and $\boldsymbol{R}_{m|\mathrm{SE}}^{sim}$ denote $\#\mathrm{SE} \times N$ matrix and $\#\mathrm{SE} \times 1$ column vector of the simulated returns for portfolio assets and market portfolio that satisfy SE condition (hereafter

[9] Specifically, if the generated S market return scenarios are sorted in ascendant order, then the $\widehat{\mathrm{VaR}}_q(R_{m,t+1})$ is calculated as the $[(1-q)S-1]$-th observation, which is just the empirical quantile of the simulated market return distribution.

we use the word "filtered" to refer to SE-truncated scenarios), respectively.

Let $\tilde{\boldsymbol{R}}^{sim}_{p,t+1|SE} = (\tilde{R}^1_{p,t+1|SE}, ..., \tilde{R}^{\#\mathrm{SE}}_{p,t+1|SE})^T$ refer to the $\#\mathrm{SE} \times 1$ vector of filtered return scenarios of portfolio active return, and $\#\mathrm{TLE} = \sum_{s=1}^{\#\mathrm{SE}} I\{\tilde{R}^s_{p,t+1|SE} \leq -\widehat{\mathrm{CoVaR}}^{p|\mathrm{SE}}_{\alpha}\}$ is the number of scenarios out of #SE that represents the conditional tail loss event (TLE). Using the above, the CoETL in (4) can be estimated as

$$\widehat{\mathrm{CoETL}}_t(R_{p,t+1}; \alpha) = -\frac{\sum_{s=1}^{\#\mathrm{SE}} \tilde{R}^s_{p,t+1|SE} I\{\tilde{R}^s_{p,t+1|SE} \leq \widehat{\mathrm{CoVaR}}^{p|\mathrm{SE}}_{\alpha}\}}{\#\mathrm{TLE}}. \tag{28}$$

Similarly, let $\#\mathrm{TPE} = \sum_{s=1}^{\#\mathrm{SE}} I\{\tilde{R}^s_{p,t+1|SE} \geq \widehat{\mathrm{CoVaR}}^{p|\mathrm{SE}}_{1-\alpha}\}$ be the number of scenarios that indicate conditional tail profit event (TPE). The CoETP can then be estimated as

$$\widehat{\mathrm{CoETP}}_t(R_{p,t+1}; \alpha) = \frac{\sum_{s=1}^{\#\mathrm{SE}} \tilde{R}^s_{p,t+1|SE} I\{\tilde{R}^s_{p,t+1|SE} \geq \widehat{\mathrm{CoVaR}}^{p|\mathrm{SE}}_{1-\alpha}\}}{\#\mathrm{TPE}}. \tag{29}$$

Combining the above estimators, we obtain the following estimator of CoRR at each month t:

$$\widehat{\mathrm{CoRR}}_t(R_{p,t+1}; \alpha, \beta) = \frac{\widehat{\mathrm{CoETP}}_t(R_{p,t+1}; \alpha)}{\widehat{\mathrm{CoETL}}_t(R_{p,t+1}; \beta)}. \tag{30}$$

C - SPNN configuration

We follow the same choice of neural network architectures as in Gu et al. (2020). The number of neurons within each layer is set in accordance with the geometric pyramid rule (Masters 1993). Specifically, we consider the following model configurations: (1) SPNN with a single hidden layer (32) (hereafter SPNN1); (2) SPNN with two hidden layers (32, 16) (hereafter SPNN2); (3) SPNN with three hidden layers (32, 16, 8) (hereafter SPNN3); (4) SPNN with four hidden layers (32, 16, 8, 4) (hereafter SPNN4); and (5) SPNN with five hidden layers (32, 16, 8, 4, 2) (hereafter SPNN5).

D - Training and regularization methods

The training of neural networks is very time-consuming due to the high degree of computational complexity involved in tuning a big number of parameters and processing a large amount of data. To improve the generalization power of fitted SPNN models and reduce the training cost, in addition to the LASSO penalization, we consider additional techniques including batch training, batch normalization, early stopping, and forecast averaging.[10]

E - Hyperparameters

We use a two-dimensional grid search to tune hyperparameters by minimizing the QS among all possible SPNN configurations over the validation set $\mathcal{L}_2$. The tuning parameters are the L_1 penalty parameter λ_1 and the learning rate of Adam optimizer lr. For the grid of values we keep following Gu et al. (2020) and set $\lambda_1 \in [10^{-5}, 10^{-3}]$ and $lr \in [10^{-3}, 10^{-2}]$.[11]

Our goal of model selection is modest in the sense of fixing a variety of hyperparameters ex-ante to reduce the computational cost, though tuning on a more extensive set of hyperparameters might help in terms of accuracy.[12] Unlike Gu et al. (2020) who set the batch size as 10,000, we apply a relatively small batch size of 32. Although a large batch size tends to give more precise estimates of the gradients, a small batch size ensures that each training iteration is fast and reduces memory usage as well. For the remaining hyperparameters, we follow the same choice of Gu et al. (2020). Specifically, the number of epochs is set to 100, the patience in early stopping is set to 5, and the number of ensemble models is set to 10.

[10] As argued by Gu et al. (2020), L_2-penalty provides similar regularization effect as early stopping. Therefore, we only apply L_1-penalty to the loss function as defined in (1).

[11] For the CQR benchmark model, we tune on $\lambda_1 \in [10^{-5}, 10^{-3}]$ only.

[12] We also tested for different combinations of $L1$-penalty, learning rate, dropout rate, and patience in early stopping, and the current setting is found to be most effective.

F - Figures and tables

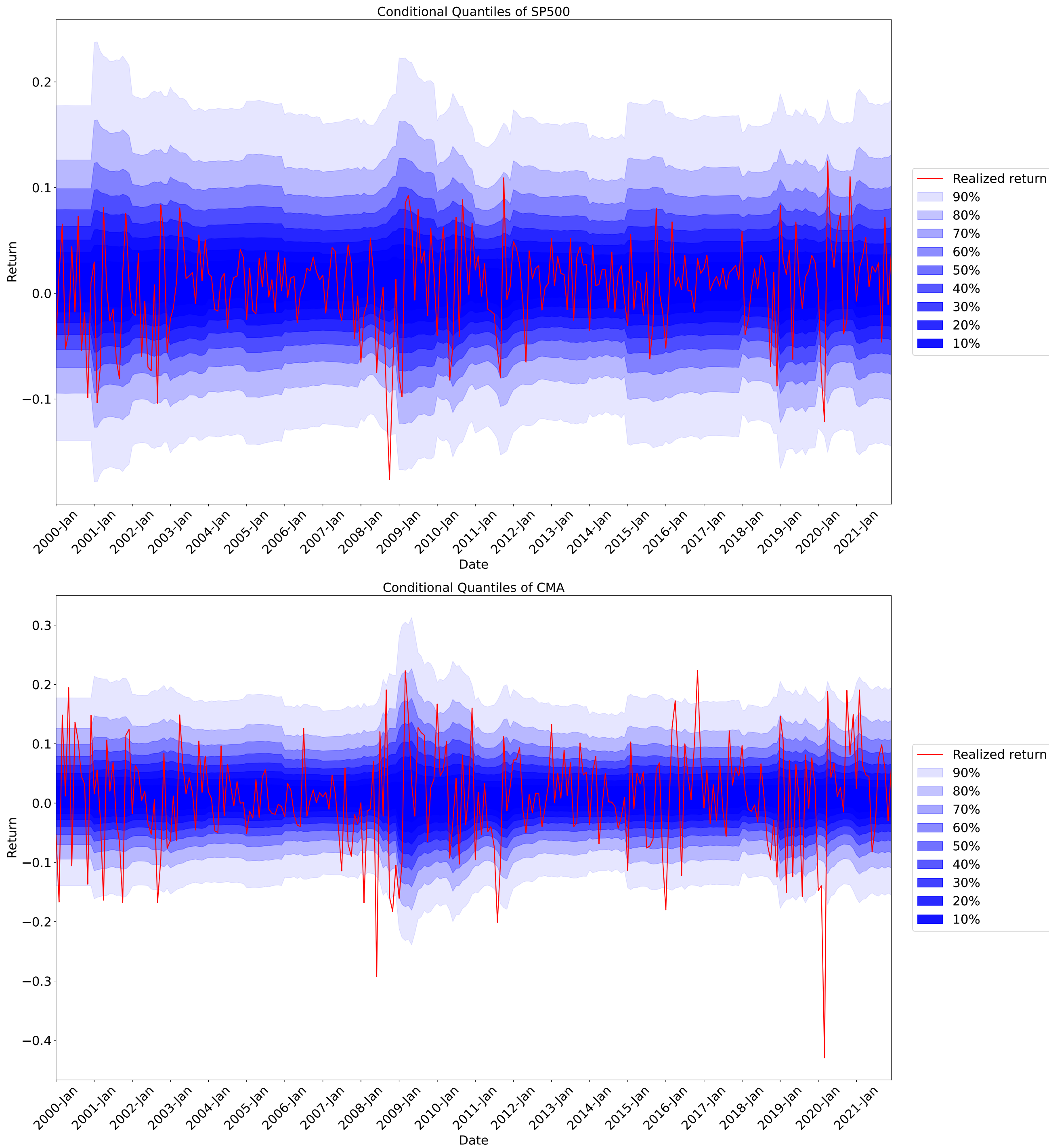
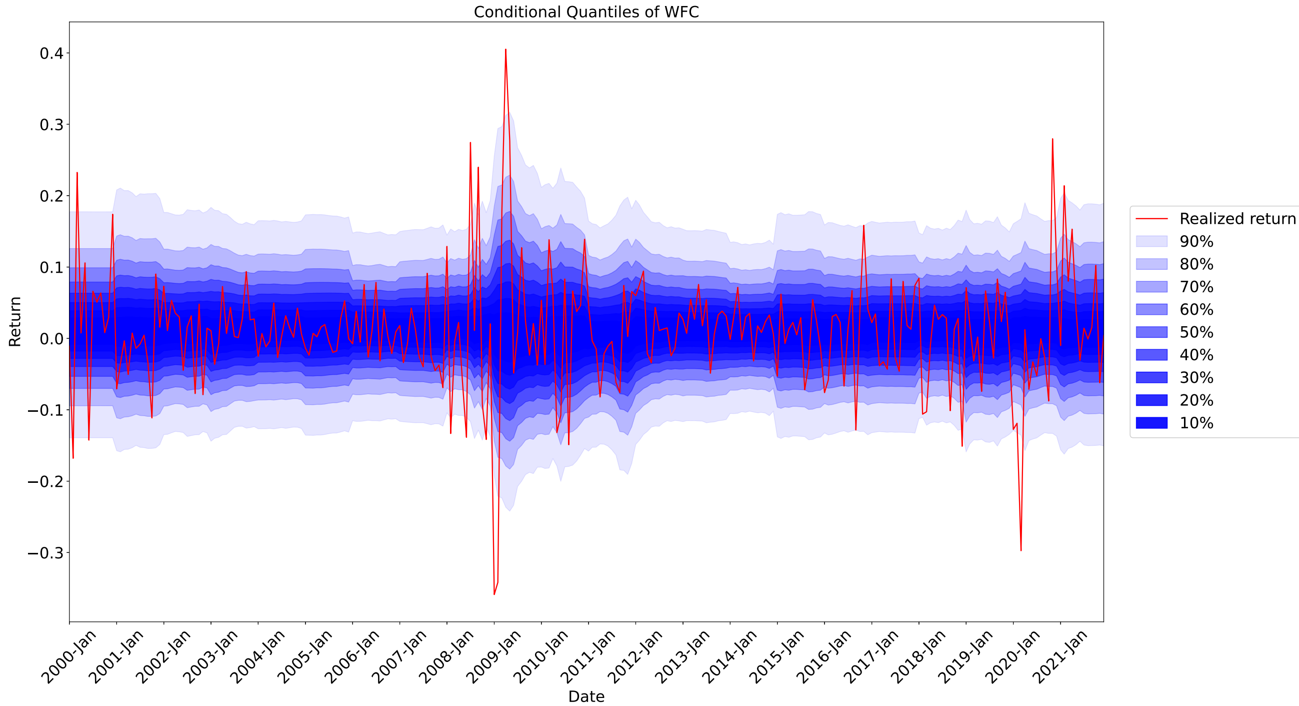


Figure 1: Predicted conditional return quantiles of S&P 500 Index (market portfolio) and three portfolio assets (CMA, WFC and JPM) obtained from SPNN1 throughout the out-of-sample period.



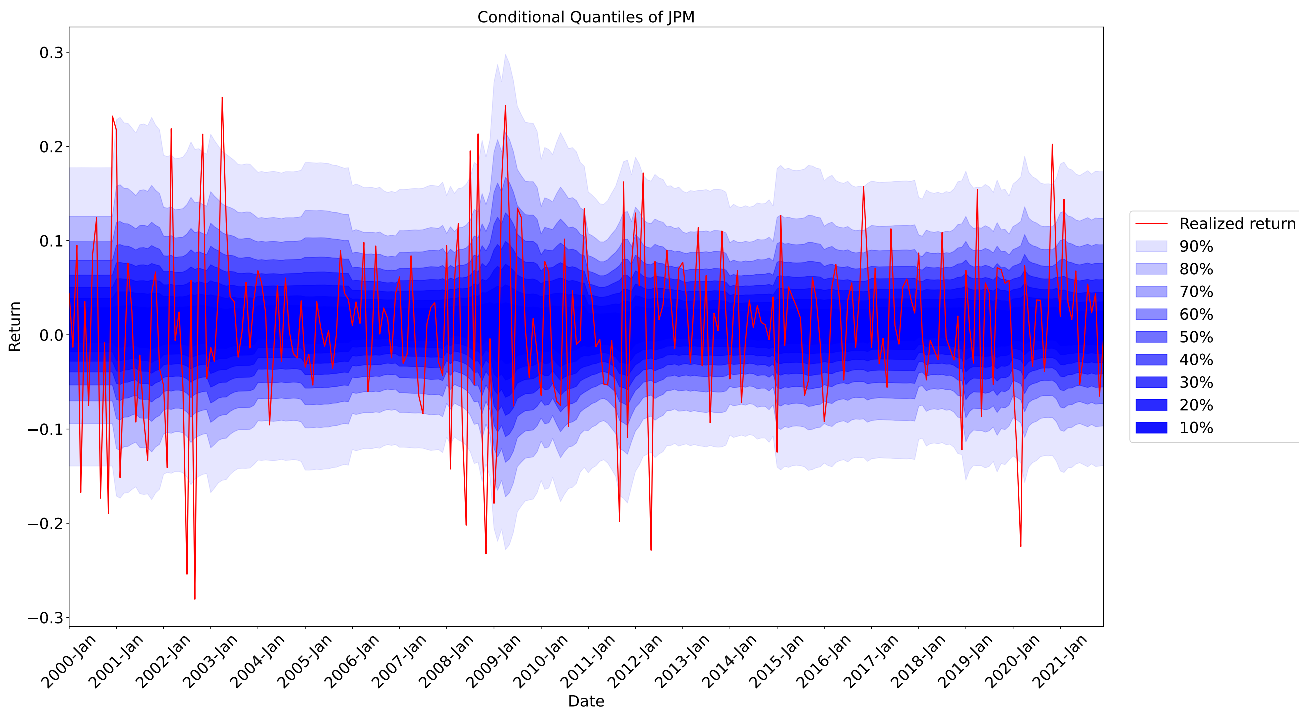
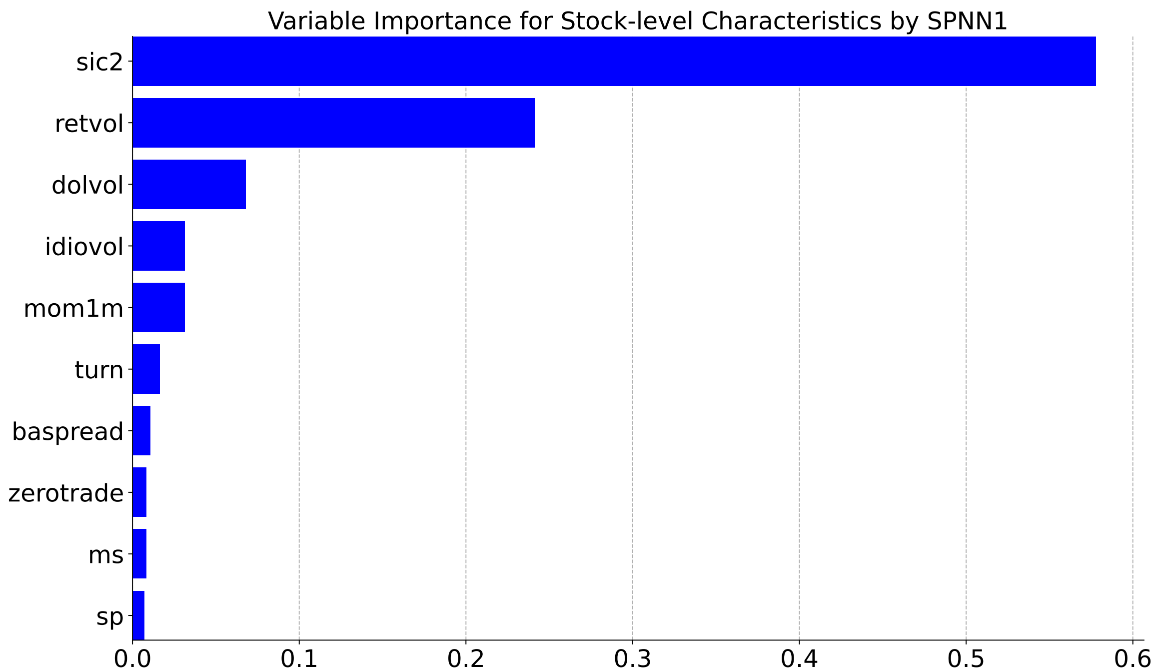


Figure 1: (continued)



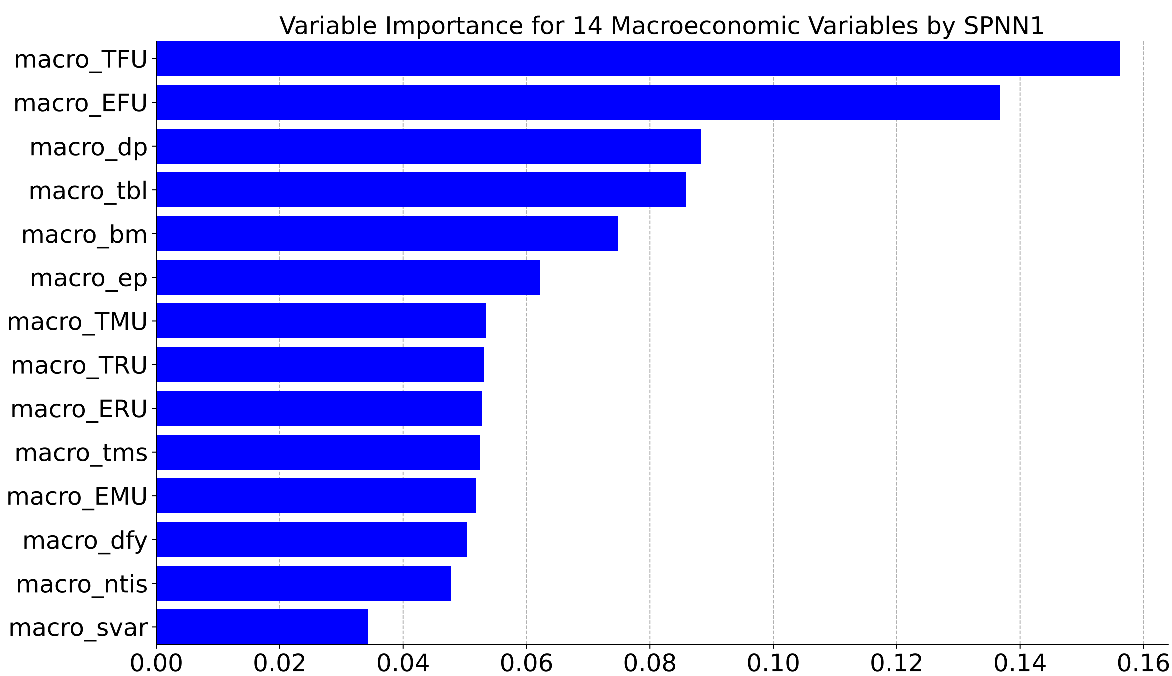


Figure 2: Top and bottom panels display the variable importance of top-10 most influential firm-level predictors and all macroeconomic variables measured by MSS based on SPNN1, respectively. Variable importance is an average across all quantiles and over all training samples. Variable importance is normalized to sum to one.

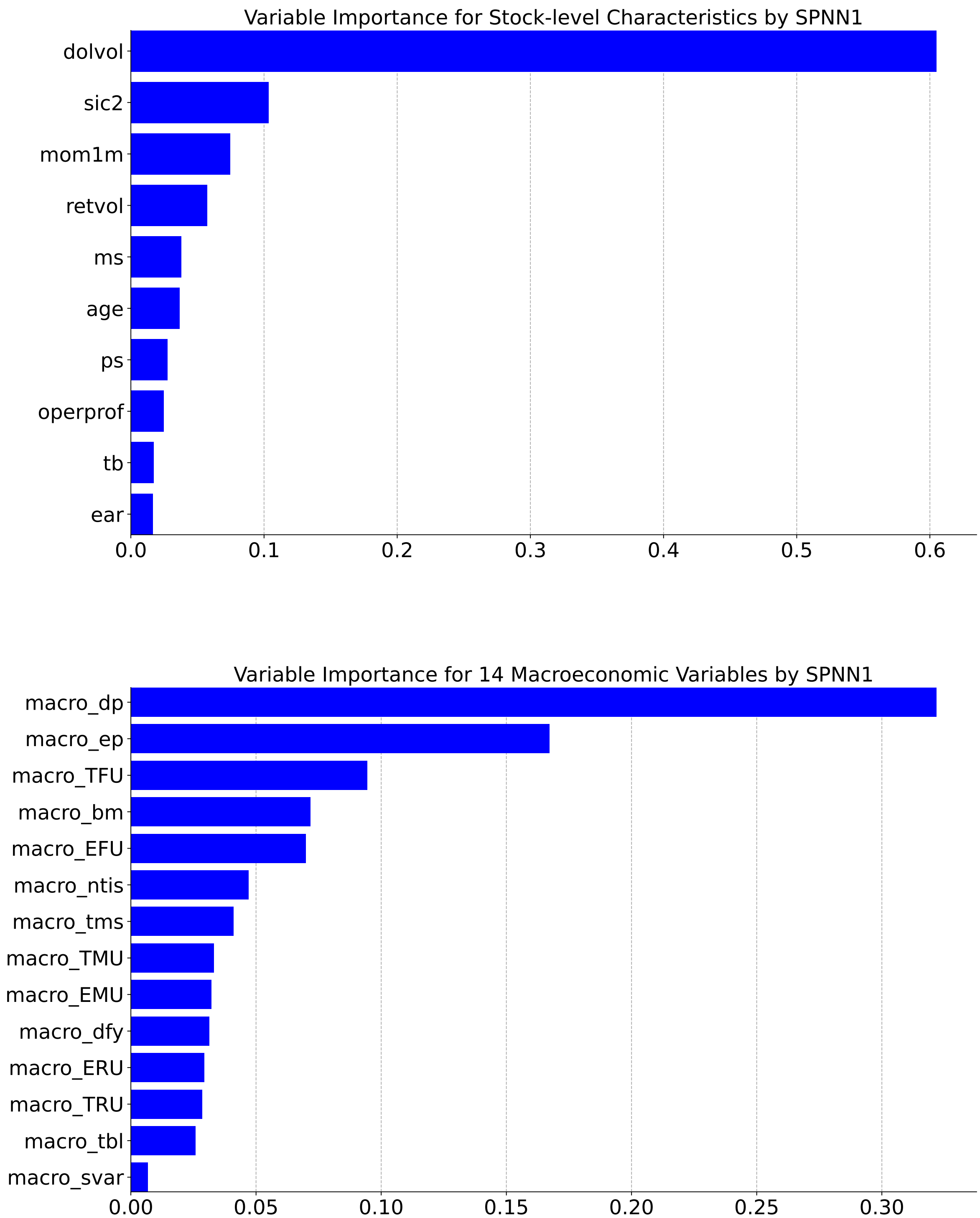


Figure 3: Top and bottom panels display the variable importance of top-10 most influential firm-level predictors and all macroeconomic variables measured by QC based on SPNN1, respectively. Variable importance is an average across all quantiles and over all testing samples. Variable importance is normalized to sum to one.

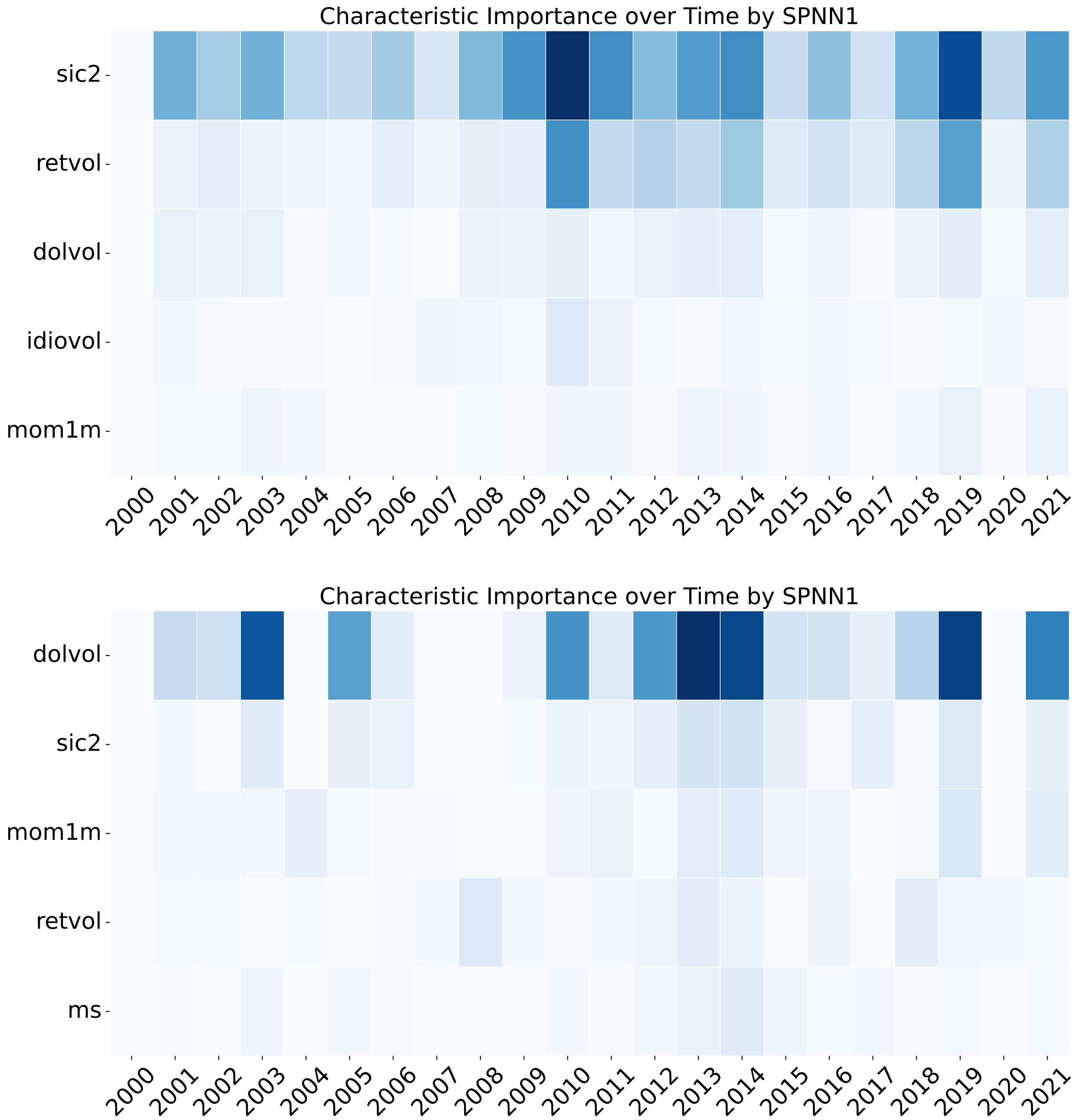


Figure 4: Time-varying variable importance of the top-5 most influential firm-level predictors measured by MSS (top panel) and QC (bottom panel) based on SPNN1, respectively. Predictors are ordered based on the average MSS value over recursive training, with the most influential features at the top and the least influential at the bottom. Columns correspond to the year-end of each in-sample window, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

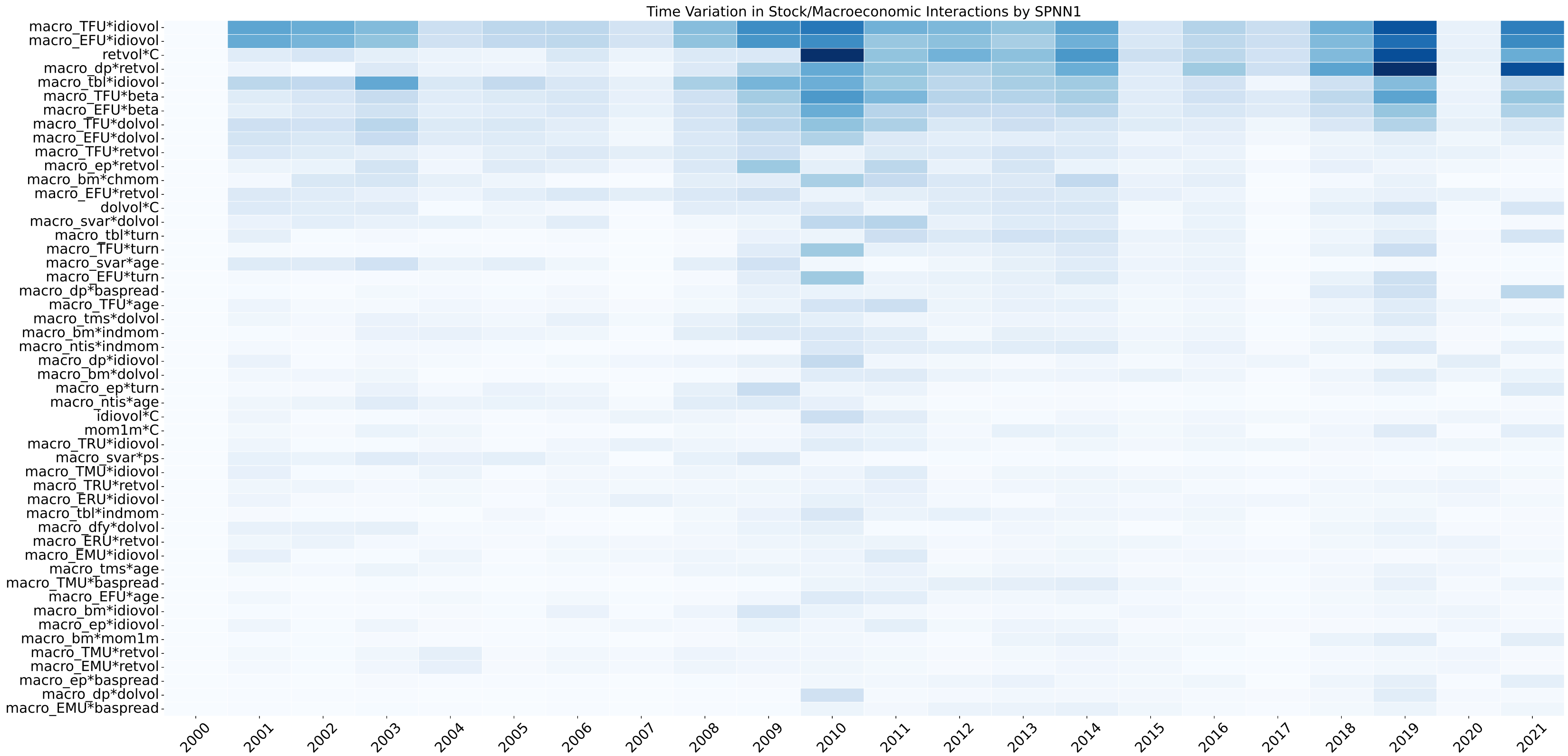


Figure 5: Time-varying variable importance of the top-50 most influential predictors of interactions between each firm characteristic with macroeconomic variables measured by MSS based on SPNN1. Columns correspond to the year-end of each in-sample window, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

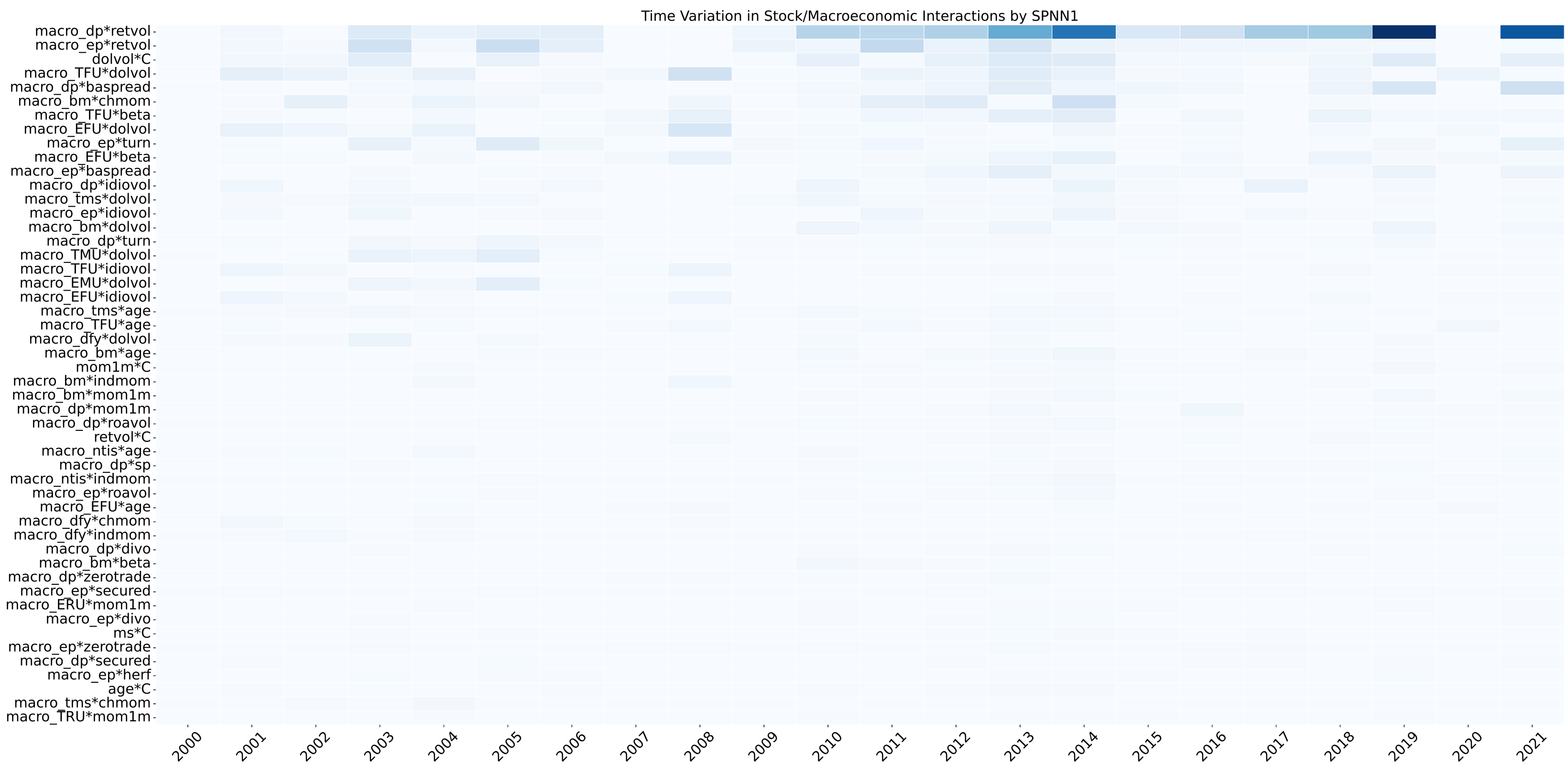


Figure 6: Time-varying variable importance of the top-50 most influential predictors of interactions between each firm characteristic with macroeconomic variables measured by QC based on SPNN1. Columns correspond to the year-end of each in-sample window, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

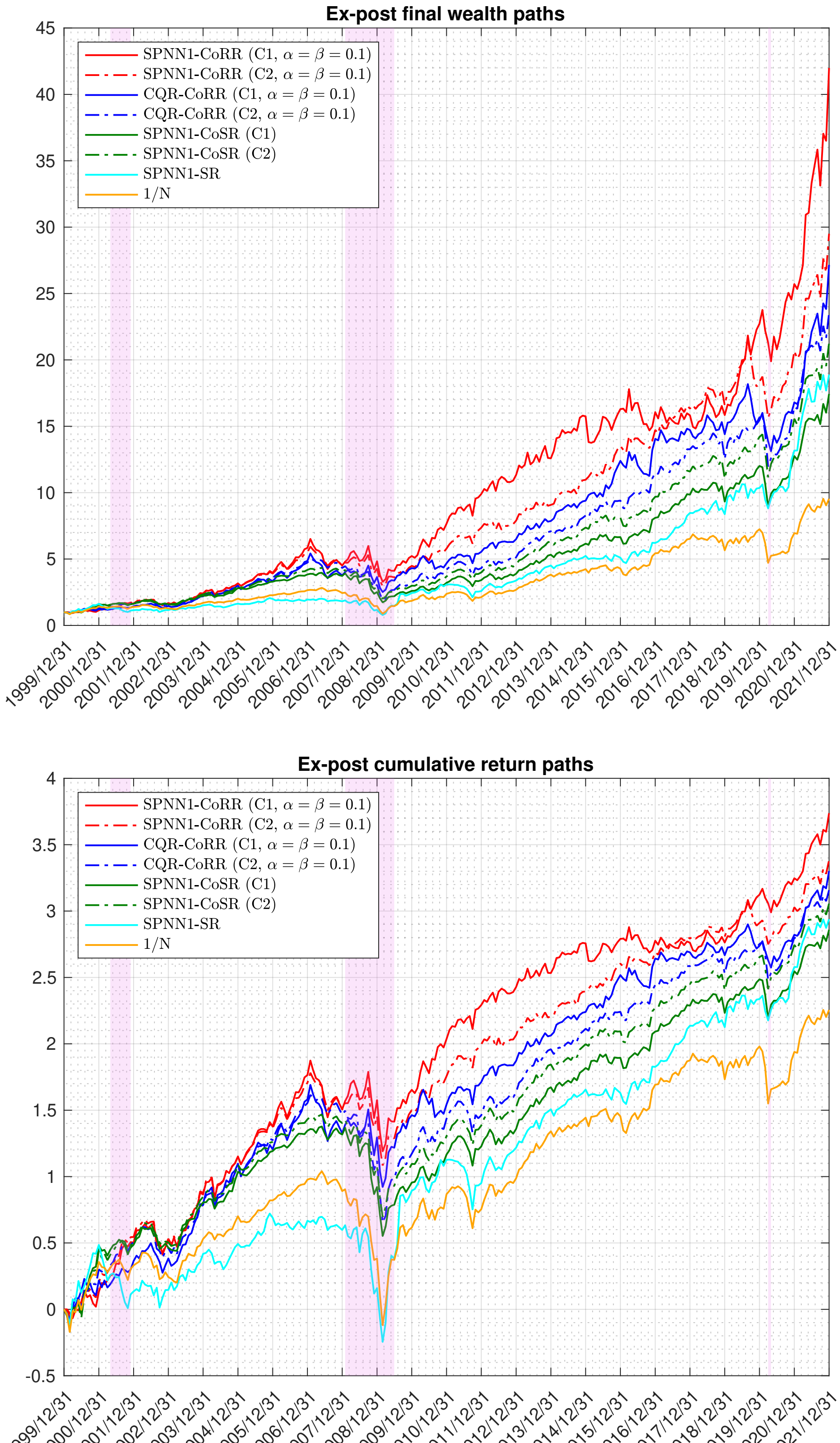


Figure 7: Ex-post paths of final wealth (top panel) and cumulative return (bottom panel) obtained using different strategies. The shaded areas indicate the NBER recession periods.

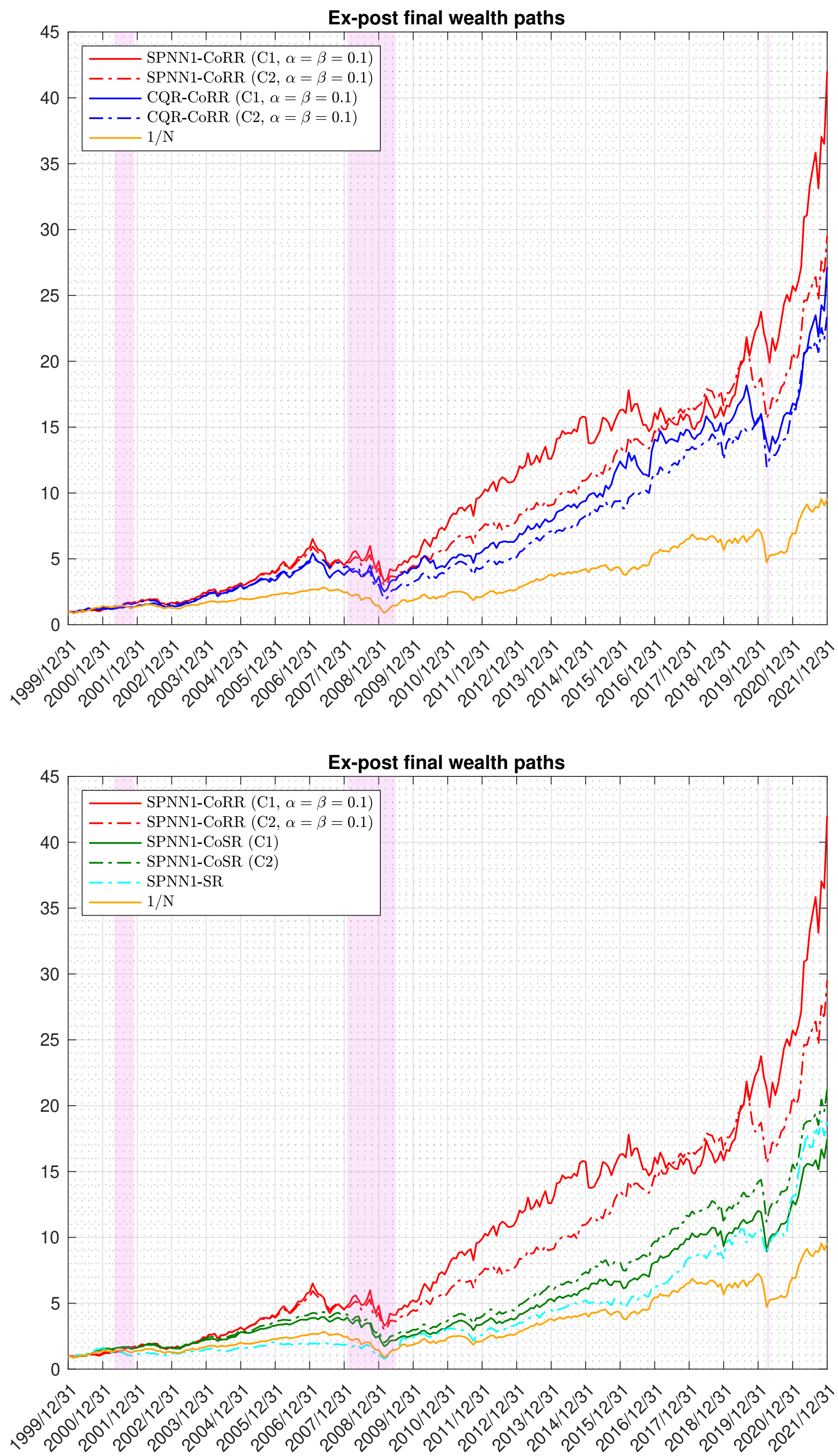


Figure 8: The top panel displays ex-post final wealth paths obtained using different models under the same CoRR measure, while the bottom panel displays ex-post final wealth paths obtained using different criteria under the same SPNN1 model.

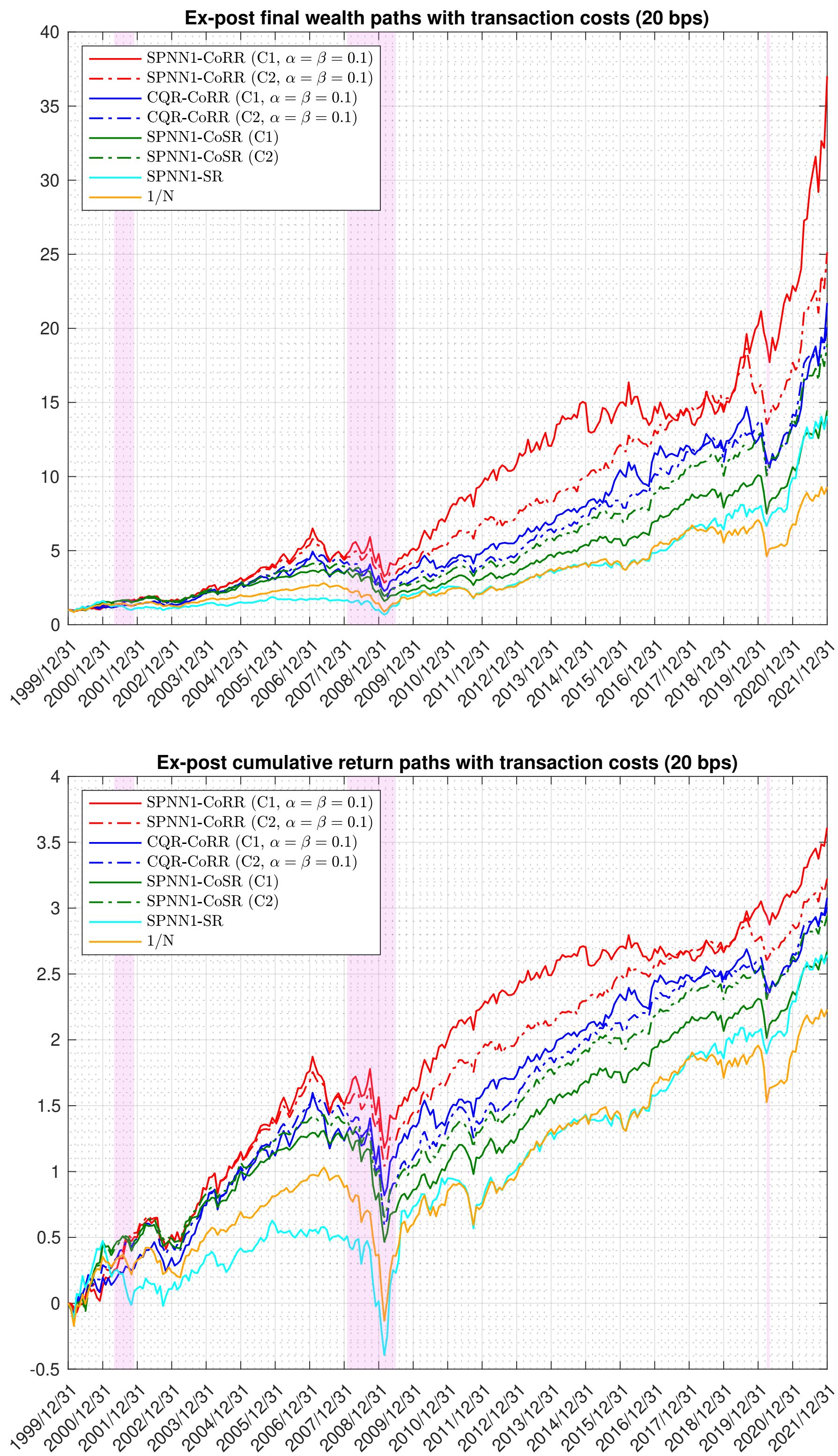


Figure 9: Ex-post paths of final wealth (top panel) and cumulative return (bottom panel) obtained using different strategies with 20 bps proportional TC. The shaded areas indicate the NBER recession periods.

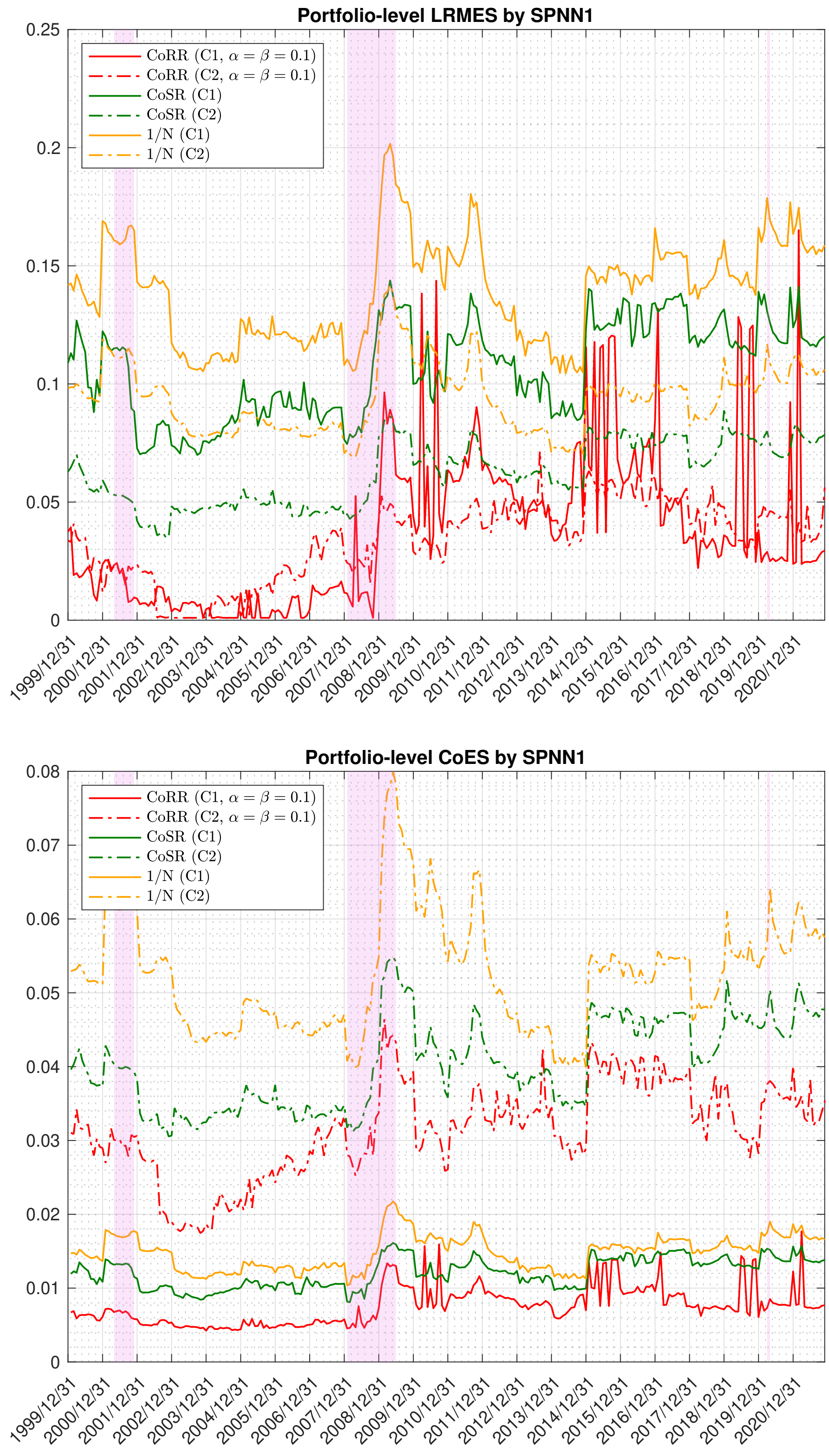


Figure 10: Portfolio-level LRMES (top panel) and CoES (bottom panel) based on SPNN1. The shaded areas indicate the NBER recession periods.

Table 1: List of portfolio assets

Ticker	Firm name	Ticker	Firm name
SNV	Synovus Financial Corp.	AFL	Aflac Incorporated
JEF	Jefferies Financial Group Inc.	NTRS	Northern Trust Corporation
CINF	Cincinnati Financial Corporation	AXP	American Express Company
CMA	Comerica Incorporated	BAC	Bank of America Corporation
L	Loews Corporation	PNC	The PNC Financial Services Group, Inc.
VNO	Vornado Realty Trust	AON	Aon plc
FITB	Fifth Third Bancorp	GL	Globe Life Inc.
RF	Regions Financial Corporation	CI	Cigna Corporation
MTB	M&T Bank Corporation	PGR	The Progressive Corporation
BEN	Franklin Resources, Inc.	PSA	Public Storage
WFC	Wells Fargo & Company	KEY	KeyBank
HBAN	Huntington Bancshares Incorporated	USB	U.S. Bancorp
MMC	Marsh & McLennan Companies, Inc.	SLM	SLM Corporation
HST	Host Hotels & Resorts, Inc.	AIG	American International Group, Inc.
CNA	CNA Financial Corporation	SEIC	SEI Investments Company
JPM	JPMorgan Chase & Co.	TFC	Truist Financial Corporation
HUM	Humana Inc.	STT	State Street Corporation
LNC	Lincoln National Corporation	ZION	Zions Bancorporation
BK	The Bank of New York Mellon Corporation	UNH	UnitedHealth Group Incorporated

Table 2: Backtesting results

Metric	SPNN1-CoRR(C1)	SPNN1-CoRR(C2)	CQR-CoRR(C1)	CQR-CoRR(C2)	SPNN1-CoSR(C1)	SPNN1-CoSR(C2)	SPNN1-SR	1/N
Final wealth	41.9878	29.5068	27.1558	23.5262	17.4357	21.2523	18.8947	9.5409
Annual return	0.1852	0.1663	0.1619	0.1544	0.1388	0.1490	0.1429	0.1080
MDD	0.4951	0.5027	0.5363	0.6173	0.5617	0.5376	0.6197	0.6863
TO	0.1194	0.1522	0.2117	0.1430	0.1772	0.1094	0.2779	0.0242
Sharpe ratio	0.7545	0.7690	0.6800	0.6952	0.6357	0.7194	0.4953	0.4083
Sortino ratio	1.2714	1.2614	1.1491	1.0958	1.0277	1.1684	0.9074	0.6667
Calmar ratio	0.3740	0.3309	0.3019	0.2501	0.2470	0.2772	0.2306	0.1573
FT ratio(p=1,q=1)	1.0833	0.9672	1.0196	0.9370	0.9782	0.8638	0.9779	0.8861
FT ratio(p=1,q=2)	0.7682	0.6639	0.7387	0.6131	0.6197	0.5897	0.6959	0.5888
FT ratio(p=1,q=3)	0.5998	0.5015	0.5741	0.4573	0.4598	0.4477	0.5466	0.4425
FT ratio(p=1,q=4)	0.5005	0.4092	0.4726	0.3747	0.3747	0.3673	0.4588	0.3631

Table 3: Backtesting results with 20 bps proportional TC

Metric	SPNN1-CoRR(C1)	SPNN1-CoRR(C2)	CQR-CoRR(C1)	CQR-CoRR(C2)	SPNN1-CoSR(C1)	SPNN1-CoSR(C2)	SPNN1-SR	1/N
Final wealth	37.0063	25.1234	21.7091	20.2266	14.4584	18.9336	14.0858	9.3000
Annual return	0.1784	0.1578	0.1502	0.1465	0.1291	0.1430	0.1278	0.1067
MDD	0.4998	0.5129	0.5412	0.6251	0.5700	0.5421	0.6395	0.6874
Sharpe ratio	0.7264	0.7274	0.6287	0.6575	0.5886	0.6885	0.4393	0.4027
Sortino ratio	1.2214	1.1911	1.0600	1.0358	0.9528	1.1178	0.8059	0.6581
Calmar ratio	0.3569	0.3077	0.2774	0.2343	0.2265	0.2638	0.1998	0.1552
FT ratio(p=1,q=1)	1.0891	0.9758	0.9810	0.9212	0.9598	0.8591	1.0054	0.8872
FT ratio(p=1,q=2)	0.7693	0.6677	0.7188	0.6073	0.6143	0.5866	0.7060	0.5901
FT ratio(p=1,q=3)	0.5999	0.5035	0.5616	0.4535	0.4565	0.4450	0.5516	0.4436
FT ratio(p=1,q=4)	0.5001	0.4104	0.4633	0.3717	0.3722	0.3649	0.4615	0.3640

Table 4: p values of CC tests for SPNN1 forecasts ($\tau = 0.05$)

Ticker	DB1	DB2	DB3	DB4	DB5	DB6	DB7	LR_{CC}	DQ1	DQ2	DQ3	DQVaR1	DQVaR2	DQVaR3
JEF	0.9500	0.3982	0.3415	0.3160	0.9674	0.5172	0.6621	0.3482	0.2104	0.1316	0.1319	0.3452	0.2554	0.3542
CINF	0.4612	0.4297	0.4065	0.3896	0.2895	0.3559	0.4947	0.2844	0.2901	0.4234	0.5646	0.3425	0.3493	0.5084
L	0.8011	0.5643	0.5712	0.6780	0.9159	0.7275	0.8433	0.5186	0.6810	0.7581	0.8136	0.8409	0.9411	0.9718
BK	0.6758	0.0279	0.0580	0.0865	0.7680	0.0576	0.0996	0.0272	0.0012*	0.0029*	0.0035*	0.0036*	0.0020*	0.0048*
TFC	0.6758	0.6094	0.7848	0.8843	0.7923	0.7552	0.8581	0.7113	0.7038	0.5896	0.7505	0.7967	0.8460	0.8406
ZION	0.4612	0.2921	0.6240	0.2985	0.6554	0.0060*	0.0071*	0.2844	0.2901	0.4234	0.3856	0.4793	0.7244	0.4041

NOTE: "∗" denotes rejection from the coverage test at 1% significance level.

Table 5: p values of CC tests for SPNN1 forecasts ($\tau = 0.25$)

Ticker	DB1	DB2	DB3	DB4	DB5	DB6	DB7	LR_{CC}	DQ1	DQ2	DQ3	DQVaR1	DQVaR2	DQVaR3
JEF	0.2059	0.1853	0.2948	0.1271	0.2388	0.2669	0.2387	0.1617	0.1668	0.2792	0.2770	0.2563	0.5004	0.5215
CINF	0.6350	0.2084	0.3335	0.4427	0.6609	0.2787	0.3777	0.2415	0.2304	0.3605	0.5035	0.3374	0.5798	0.7389
L	0.5777	0.2342	0.3659	0.3338	0.6676	0.3191	0.4384	0.1648	0.1578	0.2942	0.4046	0.2574	0.4788	0.5862
BK	0.7424	0.2224	0.3549	0.1870	0.4372	0.1654	0.2587	0.3880	0.3640	0.3980	0.4571	0.3733	0.4448	0.5280
TFC	0.6832	0.5864	0.6536	0.6773	0.8212	0.7141	0.8240	0.5075	0.4907	0.6999	0.7613	0.6635	0.8897	0.9517
ZION	0.6845	0.8556	0.6998	0.4849	0.4068	0.6825	0.0722	0.9105	0.9217	0.7278	0.7645	0.7253	0.5681	0.6916

Table 6: p values of CC tests for SPNN1 forecasts ($\tau = 0.5$)

Ticker	DB1	DB2	DB3	DB4	DB5	DB6	DB7	LR_{CC}	DQ1	DQ2	DQ3	DQVaR1	DQVaR2	DQVaR3
JEF	0.3429	0.2922	0.3631	0.4968	0.4795	0.4118	0.4873	0.4529	0.4520	0.3854	0.5432	0.5853	0.6357	0.7464
CINF	0.3429	0.3365	0.4380	0.5691	0.2502	0.3111	0.3061	0.4858	0.5397	0.5338	0.7002	0.5487	0.3545	0.4697
L	0.4666	0.3384	0.4689	0.6142	0.0017*	0.0045*	0.0062*	0.8284	0.7845	0.5517	0.7125	0.0044*	0.0164	0.0483
BK	0.3055	0.4728	0.6404	0.7541	0.4961	0.6167	0.0368	0.5823	0.5565	0.7436	0.8472	0.7474	0.8477	0.8400
TFC	0.2315	0.2349	0.3396	0.1172	0.1488	0.1668	0.2562	0.2385	0.2093	0.3704	0.2048	0.1541	0.3450	0.2973
ZION	0.5000	0.2189	0.1332	0.1706	0.0319	0.0115	0.0237	0.3938	0.3622	0.1253	0.1882	0.0165	0.0175	0.0475

NOTE: "∗" denotes rejection from the coverage test at 1% significance level.

Table 7: p values of CC tests for SPNN1 forecasts ($\tau = 0.75$)

Ticker	DB1	DB2	DB3	DB4	DB5	DB6	DB7	LR_{CC}	DQ1	DQ2	DQ3	DQVaR1	DQVaR2	DQVaR3
JEF	0.3178	0.3772	0.5224	0.2773	0.4334	0.4970	0.4893	0.3997	0.4026	0.5717	0.4417	0.5247	0.7863	0.6709
CINF	0.5777	0.5437	0.1439	0.1529	0.2405	0.6571	0.7924	0.6225	0.5975	0.1730	0.1090	0.7462	0.3775	0.0936
L	0.5143	0.7110	0.7939	0.6846	0.2019	0.3627	0.5035	0.6700	0.6636	0.8394	0.6788	0.2711	0.5370	0.5368
BK	0.3178	0.3826	0.4059	0.5108	0.2452	0.2518	0.3534	0.3923	0.3811	0.3603	0.5171	0.2859	0.3624	0.5568
TFC	0.7423	0.0101	0.1242	0.0965	0.0499	0.0004*	0.0001*	0.2266	0.2428	0.2473	0.1386	0.0955	0.0444	0.0472
ZION	0.3178	0.0698	0.2826	0.3519	0.1204	0.2816	0.1575	0.4237	0.4159	0.2726	0.4115	0.2095	0.1998	0.3787

NOTE: "$*$" denotes rejection from the coverage test at 1% significance level.

Table 8: p values of CC tests for SPNN1 forecasts ($\tau = 0.95$)

Ticker	DB1	DB2	DB3	DB4	DB5	DB6	DB7	LR_{CC}	DQ1	DQ2	DQ3	DQVaR1	DQVaR2	DQVaR3
JEF	0.6403	0.1014	0.1030	0.1632	0.1100	0.0807	0.0939	0.0935	0.0393	0.0591	0.0640	0.0360	0.0849	0.1137
CINF	0.2960	0.4872	0.4455	0.0980	0.2951	0.3642	0.5033	0.2417	0.3295	0.4947	0.0312	0.4212	0.5681	0.1208
L	0.0029*	0.0067*	0.0159	0.0320	0.0017*	0.0158	0.0268	0.0029*	0.0169	0.0427	0.0854	0.0403	0.1400	0.3039
BK	0.4612	0.0202	0.0433	0.0417	0.0693	0.0169	0.0280	0.0446	0.0257	0.0421	0.0445	0.0163	0.0054*	0.0147
TFC	0.0812	0.0071*	0.0000*	0.0000*	0.0214	0.0158	0.0048*	0.0742	0.1294	0.1003	0.0700	0.0675	0.1294	0.1356
ZION	0.0812	0.0273	0.0484	0.0879	0.0518	0.0396	0.0734	0.0742	0.1294	0.2433	0.3716	0.2052	0.4388	0.4845

NOTE: "$*$" denotes rejection from the coverage test at 1% significance level.