## *Systemic Risk Based Portfolio Selection*

LIN, WEIDONG

Durham University

Doctor of Philosophy

L1R001 Economics

# Systemic Risk Based Portfolio Selection

*Author:*

Weidong Lin

*Supervisor:*

Abderrahim Taamouti

March 2023

# Declaration

This thesis has been composed by myself and contains no material that has been accepted for the award of any other degree at any university. To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgement has been made.

# Abstract

This dissertation examines portfolio selection under systemic risk using performance measures. In the first chapter, we propose a novel performance measure to construct optimal portfolios that explicitly incorporate the occurrence of systemic event. Investors maximize an ex-ante modified Sharpe ratio that is conditional on some systemic event, with the latter interpreted as a low market return environment. We solve the portfolio optimization problem analytically under the absence of short-selling constraint and numerically when short-selling constraint is imposed. The approach is made operational by embedding it in a multivariate dynamic setting via dynamic conditional correlation and copula models. In the second chapter, we further enhance the portfolio selection approach proposed in the first chapter by using machine learning techniques. Specifically, the optimal portfolio is solved through a three-step supervised learning model. First, the smooth pinball neural network is employed to predict conditional marginal return distribution. Secondly, we use copula to model dependence between portfolio assets and the market, based on which we generate return scenarios. Lastly, we maximize the ex-ante conditional Sharpe ratio based on simulated returns. Unlike the previous chapter, where we use statistical models to forecast return distributions, in this chapter we take advantage of a distributional machine learning model along with a set of predictors that includes more than 1,000 predictive signals. In the last chapter, following the similar idea of conditional Sharpe ratio, we propose another systemic risk-based performance measure namely the conditional Rachev ratio. This measure inherits the advantage of unconditional Rachev ratio in the sense that it can account for asymmetric information of portfolio return distribution. Moreover, we build a link between our new measure and the well-know $CoVaR$ measure in the finance literature. In each chapter, we construct a comparative analysis using data on the US stock market. Overall speaking, all the backtesting results demonstrate the superiority of our proposed approaches against popular benchmark strategies in terms of profitability and systemic risk, where the outperformance is robust to the inclusion of transaction costs.

**Keywords:** Portfolio selection; systemic risk; machine learning; probabilistic forecasting; copula; scenario analysis; performance strategy.

To my parents, for always loving and supporting me.

# Acknowledgements

The completion of this doctoral degree is the culmination of massive hard work. Not only from myself, but also from everyone around me in inspiring and encouraging me to make it. Although only a few individuals are mentioned below, I am truly obliged to all the people who helped me along the way.

First and foremost, I would like to express my deep sense of thanks to my supervisor Professor Abderrahim Taamouti for his continuous support, patience and encouragement throughout my Ph.D. study. I am grateful to him for allowing me to explore the subjects that I am most interested in and letting me become an independent researcher with critical thinking. I could not have imagined having a better advisor and mentor other than him. Along with my supervisor, I extend my gratitude to Professor Jose Olmo for his insightful suggestions and expert guidance. I also want to give special thanks to all the anonymous reviewers and editors of my papers for their valuable comments and feedbacks.

I give the biggest thanks to my parents who have been providing me with unconditional support throughout my lifetime. I am so gracious for everything that they have done for me. Their understanding, encouragement and love have helped me get through all the hard times. I am and will be forever indebted to them. I hope today I make them proud.

Last but not least, I want to also thank my fellow doctoral friends in DUBS. Special thanks to Dr. Kaveh Salehzadeh Nobari, Mr. Andrew He, Dr. Christian Engels, Mr. Jingyuan Di, Mr. Tao Chen, Ms. Shan Jia, Mr. Ricky Xu, Ms. Emma Zhang, Ms. Emma Sun, Mr. David Woroniuk, Mr. Kaisheng Luo, Dr. William Hong, and Ms. Xiao Liu. Kaveh, your hardworking continues to inspire me, and I can't wait to see you in London. Andrew, I will never forget our amazing discussions and ideas on machine learning and systemic risk, and all the Dominos pizza we ate together in the PGR room. Jingyuan and Tao, I appreciate the days that we spent together.

# Contents

# List of Tables

# Chapter 1

# Portfolio selection under systemic risk

## 1.1 Introduction

Systemic risk is defined as the risk of collapse of an entire financial system, as opposed to risk associated with any single individual entity or component of the system. It also refers to the risk imposed by poorly understood interlinkages and interdependencies between assets and institutions in the financial market, where the failure of a single entity or cluster of entities can trigger the failure of more institutions, see Allen and Carletti (2013).

The global financial crisis of 2007-2008 and subsequent crises (e.g. COVID-19 crisis) provide ample evidence of the importance of containing this risk. More formally, Ben Bernanke, as previous Chairman of the US Federal Reserve, defined systemic risk as "*developments that threaten the stability of the financial system as a whole and consequently the broader economy, not just that of one or two institutions.*" For a brief discussion on the elements of a systemic risk monitor that help identify risks to financial stability, readers can consult Liang (2013). In this paper, we formally incorporate the occurrence of systemic event to the construction of optimal portfolios. This new approach is better suited to accommodate market turbulences and, as a result of this, it is able to outperform popular alternatives such as the classical mean-variance, global minimum variance and equally-weighted portfolios out-of-sample.

Prevalent financial regulations such as Basel capital requirements seek to control firms' individual risks without accounting for systemic event (Acharya et al. 2017). Empirical ev-

idence shows, however, that the interconnection among financial institutions has increased significantly in recent years, generating the risk of potential system-wide distress with major knock-on effects on the real economy. Financial institutions in the same sector have linkages and connections which can become channels for spreading poor performance from one to the others. A message that stems from this literature, see Board (2016), is that it is necessary for regulators to monitor systemically important financial institutions (SIFIs) whose failures may impose negative spillover effects on the wider financial system. Benoit et al. (2013, 2017) differentiate between two distinct approaches that measure the systemic risk contribution of financial institutions. The first method looks at different sources of systemic risk such as financial contagion, bank panics, liquidity problems, etc. It relies on the use of confidential data directly provided by financial institutions to regulators. Following this idea, various regulatory models are proposed to identify the transmission channels of systemic risk and supervise inter-bank behaviors with the aim of enhancing the stability of the financial system. Gourieroux et al. (2012), for example, propose a new regulation mechanism which requires periodic reporting by financial institutions of their structural information, which is used to quantify the bilateral exposures concerning equities, lendings, or derivatives. The second method depends on market trading data such as the prices of stocks, bonds, and CDSs.

Many financial economists have developed their own measures to quantify firms' contribution to the overall risk of the financial system (see, for example, Acharya et al. 2017, Acharya et al. 2012, Brownlees and Engle 2016 and Adrian and Brunnermeier 2016). While distinguished from traditional risk measures, the systemic risk measures proposed by these authors focus on the *interconnection* among financial firms. Prominent systemic risk measures are the CATFIN of Allen et al. (2012), the CoVaR of Adrian and Brunnermeier (2016) and its extension to a multivariate setting by Girardi and Ergün (2013), the SRISK of Brownlees et al. (2012); Brownlees and Engle (2016) and its extension to a multifactor model by Engle et al. (2014), the systemic expected shortfall (SES) of Acharya et al. (2017), and econometric measures of connectedness and systemic risk in finance and insurance sectors, such as Hong et al. (2009), Battiston et al. (2012), Billio et al. (2012), Helbing (2013), Ang and Longstaff (2013), Diebold and Yılmaz (2014), and Hautsch et al. (2014). Bisias et al. (2012) present a survey that covers over thirty systemic risk indices.

Although the existing systemic risk measures are helpful for financial regulators, portfolio managers are still looking for practical guidance under which they can account for

systemic event during their decision-making process. A general approach for constructing optimal portfolios is to maximize a reward-to-risk ratio. Modern portfolio theory pioneered by Markowitz (1952) stresses the idea that portfolio diversification leads to a risk reduction. Following this idea, Tobin (1958) developed further the concept of optimal portfolio allocation by arguing that agents would diversify their asset allocation. An alternative strategy to solve the optimal portfolio allocation exercise is to maximize the investors' expected utility, which was first proposed by Von Neuman and Morgenstern (1944). In this framework, the optimal portfolio decision is obtained as a result of the maximization of the expected utility derived from the portfolio return.

Unfortunately, none of these two paradigms is devised to properly take into account the occurrence of systemic event. Both approaches incorporate the possibility of joint dependence between the assets within the portfolio through the presence of cross-correlation between the returns on the portfolio constituents or through more sophisticated measures considering joint dependence in the tails. A seminal example is the literature on optimal portfolio allocation under tail quantile restrictions using value-at-risk (VaR) and expected shortfall (ES), see Duffie and Pan (1997) and Jorion (2007) for a comprehensive review of VaR models. More specifically, in an optimal asset allocation context, tail quantiles act as constraints in the asset allocation optimization exercise. These mean-risk models discussed in Fishburn (1977) can be considered as an extension of standard mean-variance formulations that interpret portfolio risk as the probability of tail events and that implicitly incorporate the occurrence of such events through VaR measures. The relevant literature includes Basak and Shapiro (2001), Campbell et al. (2001), Bassett et al. (2004), Engle and Manganelli (2004) and Ibragimov and Walden (2007), as seminal examples.

Whereas the macroprudential literature has made substantial progress in developing monitoring tools for assessing the underlying systemic risk in a financial system (see Tente et al. 2019, among others), the portfolio management literature has not evolved in parallel. This branch of the empirical finance literature has not explored systematically yet the implications of systemic event on the construction of investment portfolios. Our main contribution in this paper is to bring the attention of academics and financial practitioners to this important problem that has been overlooked until recently. To do this, we apply methods from the emerging macroprudential literature on systemic risk to the optimal portfolio allocation problem.

The marginal expected shortfall (MES) proposed by Brownlees et al. (2012) has received

much attention recently. This measure accounts for the comovements between individual firms and the market under stressed market conditions. It is defined as the expected percentage loss of a firm's equity value in times of a market decline. Motivating by this measure of systemic risk, we propose a modified mean-variance objective function to reflect investor's risk-return tradeoff. In particular, we propose a modified Sharpe ratio that is conditional on a systemic event, with the latter interpreted as a low market return environment. We solve the portfolio allocation problem analytically under the absence of short-selling restrictions and numerically when short-selling restrictions are imposed. This approach for obtaining an optimal portfolio allocation is made operational by embedding it in a multivariate dynamic setting. To do this, we consider two different processes for modelling multivariate financial returns and set up the portfolio allocation problem in an out-of-sample setting. The first model fits the return data to a GARCH type process and models the joint dependence between the return vector of portfolio constituents and the market portfolio using a dynamic conditional correlation (DCC) model introduced in Engle (2002). The second approach models the joint dependence between returns on portfolio assets and the market index using a Student's t-copula model. In contrast to standard approaches for portfolio selection, our proposed methodology is conditional on the occurrence of systemic event. To do this, we simulate the multivariate returns using a Monte Carlo scenario generation method.

We evaluate the portfolio performance on the US stock market. We choose a group of large financial institutions as portfolio assets, and the S&P 500 Index as benchmark rate. Our out-of-sample evaluation period spans from the beginning of 2007 to the end of 2020, hence covering two major financial crises with important systemic event (i.e. the bankruptcy of Lehman Brothers and the outbreak of COVID-19). We compare the ex-post wealth paths and portfolio-level systemic risk metric against three competitors. The first competitor is the unconditional Sharpe ratio that represents the classic mean-variance approach, the second portfolio is the naive equally-weighted portfolio that reflects full diversification and is shown to work well in financial applications (DeMiguel et al. 2009), and the third competitor is the global minimum variance portfolio (GMVP) which is often shown to outperform the mean-variance portfolio in many empirical studies (see, for example, Jagannathan and Ma 2003 and DeMiguel et al. 2009). The results of our empirical study show the outperformance of our portfolio against these three competitors in terms of profitability and systemic risk, especially during crisis periods.

The rationale for the excellent performance of our model is its positive exposure to assets that are more resilient in periods of market distress. Our portfolio clearly outperforms competitors under market distress and remains competitive in non-crisis periods. Interestingly, the proposed portfolio is less diversified than benchmark portfolios during crisis times since we only invest on a few stocks with low long-run MES level. In these periods, our strategy invests on those stocks that are expected to experience small losses under stressed market conditions. Underdiversification is the result of optimal strategies aiming to minimize exposure to systemic event. This is done by reducing the set of eligible assets to a small group of stocks with small systemic risk. This empirical finding provides an alternative interpretation to the presence of underdiversification observed in financial markets, see Mitton and Vorkink (2007) and references therein. Interestingly, our results can also be related to a recent literature on time series, see Farmer et al. (2019), which finds pockets of predictability. These pockets are short periods of time over which there is predictability of returns within longer periods with little or no evidence of predictability. In our setting, we interpret these *pockets* as periods of systemic risk that drive the overall performance of the proposed portfolio based on the maximization of a conditional Sharpe ratio objective function.

Our paper also contributes to a relatively scarce literature on systemic risk-based portfolio selection. There are a few studies on the implications of systemic risk in the investment decisions of financial institutions. Biglova et al. (2014) study portfolio selection under systemic risk using the Co-Rachev ratio as objective function. In their setting, systemic risk takes place when all assets in the investment portfolio are distressed, i.e., below their individual VaR thresholds. However, this definition can be ambiguous since the poor performance of individual assets in a portfolio does not necessarily imply a poor state of the whole financial system. Another exception is Capponi and Rubtsov (2022). These authors consider the problem of maximizing portfolio returns conditional on a systemic event given by the realization of an extremely adverse market outcome. These authors seek the portfolio that performs best in a low return environment and when the market is in distress. To solve the portfolio allocation problem, Capponi and Rubtsov (2022) impose the restrictive assumption that the distribution of the portfolio and market returns follows a bivariate Student's t distribution. More importantly, none of these papers explicitly focus on finding the best trade-off between return and risk under stressed market conditions. Our paper bridges this gap.

The rest of the paper is organized as follows. Section 2 introduces our novel objective function defined as a modified Sharpe ratio conditional on the occurrence of systemic event. Section 3 presents the investors' optimal portfolio allocation problem under systemic risk. This section derives analytically the solution without short-selling restrictions and proposes numerical methods to obtain the solution under the presence of short-selling restrictions. Section 4 introduces the simulation of return scenarios under a DCC model and a Student's t-copula for modelling the joint conditional distribution of asset and market portfolio returns. Section 5 discusses an application of our optimal asset allocation strategy to a portfolio of 23 assets and presents several robustness checks. Conclusions are in Section 6. Appendix A reviews several prominent systemic risk measures. A description of the simulation of return scenarios is provided in Appendix B. The last appendix collects the figures.

## 1.2 Our objective function under market distress

The mean-variance framework developed by Markowitz (1952) is one of the cornerstones for portfolio theory. Optimal portfolios are obtained by maximizing the expected return on an investment portfolio conditional on a given level of risk that is proxied by the variance of the portfolio return. Alternative formulations consider risk measures given by tail events such as VaR and ES, see Duffie and Pan (1997) and Jorion (2007) for a comprehensive review of VaR models. In these models the objective function is the expected portfolio return that is constrained by a tail quantile restriction on the asset allocation optimization exercise.

Based on these objective functions, the literature in financial economics has developed performance measures to evaluate investment strategies. A natural performance measure based on the seminal mean-variance framework is the Sharpe ratio (Sharpe 1966a), which is originally proposed for measuring the performance of mutual funds. This measure is defined as the ratio between the expected portfolio excess return (i.e. the expected portfolio return minus risk-free rate) and its standard deviation. Sharpe (1994) later revised this measure by referring the portfolio performance with respect to a certain benchmark rate $R_b$, which can change over time, such that the revised Sharpe ratio is defined as

$$SR(R_p) = \frac{E(R_p - R_b)}{std(R_p - R_b)}. \tag{1.1}$$

In the remainder of this paper, when referring to the Sharpe ratio, we will consider expression (1.1). It is typical to use the Sharpe Ratio to evaluate and compare the ex-post portfolio performance among different investment strategies.

Interestingly, Biglova et al. (2009) argue that the maximization of the Sharpe ratio allows one to obtain a market portfolio that is optimal in the sense that it is not dominated in stochastic dominance of second order by non-satiable risk-averse investors. This result suggests that using the Sharpe ratio and related performance measures as the investor's objective function in a portfolio allocation setting is a fruitful strategy (see Rachev et al. 2008 for a review of performance measures). The choice of a performance measure allows one to explicitly introduce the risk measure along with the corresponding reward measure in the portfolio choice optimization problem without having to specify a risk aversion coefficient.

Although the Sharpe Ratio works well in Gaussian settings, it is not a suitable performance measure in settings characterised by skewness and heavy tails of the return distributions. In order to capture higher moments of the return distributions on the performance of investment portfolios, many authors have developed their own ratios such as Gini ratio (Shalit and Yitzhaki 1984), Mean Absolute Deviation ratio (Konno and Yamazaki 1991), Mini-max ratio (Young 1998), Sortino-Satchell ratio (Sortino and Satchell 2001), Rachev ratio (Biglova et al. 2004) and others (see Farinelli et al. 2008 for a detailed survey). In this paper, we focus on tail risk measures capturing systemic risk. In particular, we propose a conditional performance measure that incorporates the occurrence of systemic risk without imposing any distributional assumptions.

Our objective function for optimal portfolio allocation is inspired by the conditional performance measure proposed by Biglova et al. (2014). These authors study the portfolio selection problem in the presence of systemic risk and propose a conditional version of Rachev ratio (CoRR), which is defined as:

$$CoRR(R_p;\alpha,\beta) = \frac{E(R_p - R_b | R_1 \geq -VaR_{1-\beta}(R_1), ..., R_n \geq -VaR_{1-\beta}(R_n))}{-E(R_p - R_b | R_1 \leq -VaR_{\alpha}(R_1), ..., R_n \leq -VaR_{\alpha}(R_n))}, \tag{1.2}$$

where $VaR_q(X) = -inf\{x|P(X \leq x) > q\}$ is the VaR of the random variable $X$ that is interpreted as a financial return on an investment portfolio. The interpretation of this measure is different from standard systemic risk formulations. CoRR does not link systemic risk to the occurrence of distress in the financial system, instead, it evaluates portfolio per-

formance conditional on the occurrence of idiosyncratic events in all assets in the portfolio (i.e. all asset returns are above (or below) their individual VaR levels). Moreover, CoRR takes the expected portfolio return as a reward measure conditional on all asset prices co-moving in the tail. This assumption may be difficult to be satisfied in practice and might lead to an empty set if the set of assets in the portfolio is sufficiently large.

Unlike Biglova et al. (2014), we define a systemic event when the return on the market index is below a certain threshold $C$ over a time horizon $h$. Following the related literature, we assume that there exists a benchmark systemic risk index, which is the S&P 500 Index in our case, that reflects broad market conditions. The goal of our investors is to maximize the Sharpe ratio conditional on the systemic risk index being below a threshold level $C$ between $t$ and $t+h$, and we set the horizon $h$ to one month (i.e. 22 trading days). Our investment strategy aims to find portfolios that perform best under stressed market conditions.

We start by introducing several assumptions and notations used throughout the paper. In our economy there is no risk-free asset and there are $N \geq 2$ risky assets (firms) with stochastic simple returns denoted by $R_t = (R_{1,t}, ..., R_{N,t})^T$. The return on the financial system is proxied by a market portfolio return $R_{m,t}$. The logarithmic returns of the firm $i$ and the market are denoted, respectively, as $r_{i,t} = log(1 + R_{i,t})$ and $r_{m,t} = log(1 + R_{m,t})$. The mean vector of returns is denoted by $\mu_t = E(R_t)$, while $\Sigma_t = E[(R_t - \mu_t)(R_t - \mu_t)^T]$ represents the covariance matrix of returns. The vector of portfolio weights is denoted by $W_t = (\omega_{1,t}, ..., \omega_{N,t})^T$ such that $\sum_{i=1}^{N} \omega_{i,t} = 1$. Let $R_{p,t} = W_t^T R_t$ be an investment portfolio with expected return given by $\mu_{p,t} = W_t^T \mu_t$. Similarly, $\mu_{m,t}$ and $\sigma_{m,t}$ denote the expected return and standard deviation of the market portfolio return reflecting the performance of the financial system. The column vector $\sigma_t = (\sigma_{1m,t}, ..., \sigma_{Nm,t})^T$ contains covariances of each asset with the market portfolio. Hereafter, we use $I\{x\}$ to denote the indicator function that equals 1 if condition $x$ is met and 0 otherwise. $\mathbf{1}$ and $\mathbf{0}$ are column vectors of ones and zeros, respectively, whose dimension are understood from the context.

In the next section we will be concerned with building portfolios under stressed market scenarios. Different definitions of SE can be adopted. For instance, Acharya et al. (2017) consider SE as extreme tail events that happen rarely on a daily basis. In particular, they focus on those "moderately bad days" defined as the worst 5% of daily market outcomes, $SE_t = \{R_{m,t} \leq -VaR_{5\%}(R_{m,t})\}$, while Biglova et al. (2014) define SE as all assets in the portfolio being below their individual VaR levels, $SE_t = \{R_{1,t} \leq -VaR_\alpha(R_{1,t}), ..., R_{N,t} \leq -VaR_\alpha(R_{N,t})\}$. We follow Brownlees et al. (2012); Brownlees and Engle (2016) and define

a systemic event as a severe drop of the market index below a threshold $C$ over a time horizon $h$, that is:

$$SE_{t:t+h} = \{R_{m,t:t+h} < C\}, \tag{1.3}$$

where $R_{m,t:t+h}$ is the multiperiod simple market return between $t$ and $t+h$. We also follow related literature and define the magnitude of the market decline ($C$) as a function of the length of the time horizon ($h$). Acharya et al. (2012) set $C$ equal to $-2\%$ and $h$ equal to one trading day to estimate the daily MES; Brownlees and Engle (2016) set $C$ equal to $-10\%$ and $h$ equal to one month for computing the monthly MES (i.e. LRMES); Engle et al. (2014) focus on long-run market stress and fix $C$ equal to $-40\%$ and $h$ equal to six months. In the empirical section, we use $C = 0$ and $-40\%$ as threshold values, which on a monthly basis correspond to $C = 0$ and $-6.7\%$ respectively.

We construct a new performance measure that will be used to build optimal portfolios under stressed market conditions. To do this, we incorporate systemic risk directly into the reward and risk measures. In order to account for the interconnection between individual assets and the financial market we propose to use the first and second moments of the excess portfolio return conditional on the occurrence of a systemic event. Our new performance measure is defined as:

$$CoSR_t(R_{p,t}) := \frac{CoER_t(R_{p,t})}{CoSD_t(R_{p,t})} = \frac{W_t^T \mu_{t|SE} - \mu_{m,t|SE}}{\sqrt{W_t^T \Sigma_{t|SE} W_t + \sigma^2_{m,t|SE} - 2W_t^T \sigma_{t|SE}}}. \tag{1.4}$$

Following the spirit of the Sharpe ratio and similar performance measures, the CoSR is defined as a ratio of a conditional reward measure over a conditional risk measure. The conditional reward measure CoER is defined as

$$\begin{aligned} CoER_t(R_{p,t}) :&= E_t(R_{p,t:t+h} - R_{m,t:t+h} | SE_{t:t+h}), \\ &= E_t(W_t^T R_{t:t+h} - R_{m,t:t+h} | SE_{t:t+h}), \\ &= W_t^T \mu_{t|SE} - \mu_{m,t|SE}, \end{aligned} \tag{1.5}$$

where $\mu_{t|SE} = E_t(R_{t:t+h}|SE_{t:t+h})$ denotes the column vector of conditional expected returns on individual assets, while $\mu_{m,t|SE} = E_t(R_{m,t:t+h}|SE_{t:t+h})$ represents the conditional expected market return. Inspired by the formulation of LRMES, we add the market index as a benchmark to enable us measure portfolio performance under stressed market scenarios. Analogously, we define the risk measure CoSD as the conditional second moment of

the portfolio excess return, that is:

$$\begin{aligned} CoSD_t(R_{p,t}) :&= [Var_t(R_{p,t:t+h} - R_{m,t:t+h}|SE_{t:t+h})]^{1/2} \\ &= \left[Var_t(W_t^T R_{t:t+h} - R_{m,t:t+h}|SE_{t:t+h})\right]^{1/2} \\ &= \left(W_t^T \Sigma_{t|SE} W_t + \sigma^2_{m,t|SE} - 2W_t^T \sigma_{t|SE}\right)^{1/2}, \end{aligned} \tag{1.6}$$

where $\Sigma_{t|SE} = Var_t(R_{t:t+h}|SE_{t:t+h})$ denotes the conditional covariance matrix of asset returns, $\sigma^2_{m,t|SE} = Var_t(R_{m,t:t+h}|SE_{t:t+h})$ denotes the conditional variance of market return, and $\sigma_{t|SE} = cov_t(R_{t:t+h}, R_{m,t:t+h}|SE_{t:t+h})$ is the column vector of conditional covariances between individual assets and the market portfolio.

## 1.3 Portfolio allocation under systemic risk

In this section, we present the portfolio allocation problem of an investor that is concerned with maximizing the modified Sharpe ratio conditional on the market being under distress. We describe first the generic portfolio optimization problem when the investor's objective function is given by a performance measure $\rho(\cdot)$. In this setting, the investor's optimal portfolio is obtained as

$$W^* = \arg\underset{W}{max}\, \rho(R_p), \quad \text{s.t. } \mathbf{1}^T W = 1. \tag{1.7}$$

Different performance measures $\rho(\cdot)$ will lead to different optimal portfolios. In the empirical application, we will consider the Sharpe ratio as the relevant objective function of interest under short-selling restrictions ($W \geq \mathbf{0}$).

In what follows, we present the optimization problem of an investor with objective function given by the CoSR measure defined above. To simplify the problem, we note that this measure can be expressed as a function of the portfolio weights as $CoSR = \frac{W^T \mu}{\sqrt{W^T \Sigma W}}$, with $\mu = E(R - R_m \cdot \mathbf{1}|SE)$ and $\Sigma = Var(R - R_m \cdot \mathbf{1}|SE)$ be the conditional mean vector and conditional covariance matrix of excess returns on individual assets respectively. The solution to the optimization problem is

$$W^{CoSR} = \arg\underset{W}{max}\, \{CoSR\}, \quad \text{s.t. } \mathbf{1}^T W = 1. \tag{1.8}$$

This portfolio optimization problem can be solved analytically under the absence of short-

selling constraints. To do this, we first solve for the conditional efficient frontier among all assets. That is, given a desired conditional expected excess return level $e$, we find the portfolio weights $W^*$ that minimize the risk measure.[1] The optimization problem becomes

$$W^* = \arg\min_W \ \frac{1}{2} CoSD, \quad \text{s.t. } \mu^T W = e, \text{ and } \mathbf{1}^T W = 1. \tag{1.9}$$

Expression (1.9) is a convex optimization problem since the objective function is convex and is subject to affine constraints. Furthermore, the Slater's condition is satisfied, hence the first order conditions are necessary and sufficient for an optimum. The Lagrangian of this problem is $\mathcal{L} = \frac{1}{2} W^T \Sigma W - \lambda_1(\mu^T W - e) - \lambda_2(\mathbf{1}^T W - 1)$, that yields the following first order condition with respect to $W$: $\frac{\partial \mathcal{L}}{\partial W} = \Sigma W - \lambda_1 \mu - \lambda_2 \mathbf{1} = 0$. Assuming that $\Sigma$ is full rank, we obtain $W = \lambda_1 \Sigma^{-1} \mu + \lambda_2 \Sigma^{-1} \mathbf{1}$. Now we need to solve for multipliers $\lambda_1$ and $\lambda_2$. Using the portfolio constraints $\mu^T W = e$ and $\mathbf{1}^T W = 1$, we have

$$\begin{cases} \lambda_1 \mu^T \Sigma^{-1} \mu + \lambda_2 \mu^T \Sigma^{-1} \mathbf{1} = e, \\ \lambda_1 \mathbf{1}^T \Sigma^{-1} \mu + \lambda_2 \mathbf{1}^T \Sigma^{-1} \mathbf{1} = 1. \end{cases} \tag{1.10}$$

Let $s_{\mu\mu} = \mu^T \Sigma^{-1} \mu$, $s_{1\mu} = \mu^T \Sigma^{-1} \mathbf{1}$ and $s_{11} = \mathbf{1}^T \Sigma^{-1} \mathbf{1}$, and $A = \begin{pmatrix} s_{\mu\mu} & s_{1\mu} \\ s_{1\mu} & s_{11} \end{pmatrix}$, with $A = \widetilde{\mu}^T \Sigma^{-1} \widetilde{\mu}$, and $\widetilde{\mu} = (\mu \ 1)^T$. The system of equations (1.10) can be rewritten in matrix form as $A\lambda = \widetilde{e}$, with $\lambda = (\lambda_1 \ \lambda_2)^T$ and $\widetilde{e} = (e \ 1)^T$. The matrix $A$ is positive definite and, hence, invertible such that $\lambda = A^{-1} \widetilde{e}$. Replacing the value of $W$ obtained above, we obtain the optimal portfolio weights $W^* = \Sigma^{-1} \widetilde{\mu} A^{-1} \widetilde{e}$. The portfolio $W^*$ is the minimum conditional variance portfolio for a given conditional mean $e$ and such that $\mathbf{1}^T W = 1$ is satisfied. The conditional variance frontier can be expressed as

$$CoSD^* = W^{*T} \Sigma W^* = \widetilde{e}^T A^{-1} \widetilde{e} = \frac{s_{11} e^2 - 2 s_{1\mu} e + s_{\mu\mu}}{s_{11} s_{\mu\mu} - s_{1\mu}^2}. \tag{1.11}$$

Now we can find the portfolio with maximum CoSR among all portfolios $W^*$ located on

[1] The conditional variance of the portfolio's excess return (i.e., CoSD) is divided by two in the optimization problem. This is merely for algebraic convenience and does not change the solution to the optimization problem.

the efficient frontier. Hence, the optimization problem (1.8) can be written as

$$W^{CoSR} = \arg\underset{W^*}{max}\, \frac{CoER}{CoSD^*} = \arg\underset{W^*}{max}\, \frac{e}{\sqrt{\frac{s_{11}e^2 - 2s_{1\mu}e + s_{\mu\mu}}{s_{11}s_{\mu\mu} - s_{1\mu}^2}}}. \tag{1.12}$$

The first order condition of this problem with respect to the objective expected reward $e$ is $\partial\left(\frac{e}{\sqrt{s_{11}e^2 - 2s_{1\mu}e + s_{\mu\mu}}}\right)/\partial e = 0$, which yields $e = \frac{s_{\mu\mu}}{s_{1\mu}}$. Therefore, the optimal portfolio weights defining the CoSR portfolio satisfy

$$W^{CoSR} = \Sigma^{-1}\widetilde{\mu}^T A^{-1} \begin{pmatrix} \frac{s_{\mu\mu}}{s_{1\mu}} \\ 1 \end{pmatrix} = \begin{pmatrix} \Sigma^{-1}\mu & \Sigma^{-1}\mathbf{1} \end{pmatrix} \begin{pmatrix} \frac{1}{s_{1\mu}} \\ 0 \end{pmatrix} = \frac{\Sigma^{-1}\mu}{\mu^T\Sigma^{-1}\mathbf{1}}. \tag{1.13}$$

It is often the case that we want to place additional constraints on the optimization - for instance we might want to restrict the portfolio weights so that none of the weights are greater than 25% of the overall wealth invested in the portfolio, or we might want to prohibit short selling allowing only long positions. This is a realistic scenario in settings characterised by systemic risk in which financial regulators ban short-selling to reduce short-term investment with speculative motives. Unfortunately, under short-selling restrictions ($W \geq \mathbf{0}$) the optimization problem (1.8) cannot be solved analytically and thus a numerical procedure must be employed. In our empirical application, we use the Solver function *fmincon* built in Matlab.

## 1.4 Simulation of return scenarios

Although CoSR has no closed-form expression in dynamic models when short-selling restrictions are imposed, we can still use a Monte Carlo simulation-based procedure to implement our systemic risk-based portfolio. The dynamic CoSR measure can be calculated using its empirical analog calculated from simulated returns over the subset of simulated crisis scenarios.

This section discusses two alternative multivariate settings to model dynamics of the returns of constituents of the investment portfolio and the market portfolio. First, we consider a semiparametric model in which the conditional mean and covariance matrix of the vector of returns is modelled parametrically. The return distribution is left unmodelled beyond these two moments and will be simulated using naive nonparametric bootstrap methods. As a robustness check, we also use a fully parametric model that allows for

heavy tails and joint tail dependence in return distributions. To do this, we consider a Student's t-copula model for modelling the multivariate conditional distribution of returns.

The following subsections describe both approaches to generate the vector of assets and market portfolio returns. A detailed algorithm describing the simulation scheme is presented in Appendix B.

### 1.4.1 GARCH-DCC Modelling

The DCC model proposed by Engle (2002) can be seen as an extension to the constant conditional correlation (CCC) model developed by Bollerslev (1990), which captures the time-varying correlation of multivariate data. In this subsection, we use the GARCH-DCC model to describe the volatility dynamics and conditional correlations between returns on portfolio assets and the market index.

Let $r_t$ be an $(N+1) \times 1$ vector of logarithmic returns. The last return, $r_{N+1,t}$ is the return on the market index, i.e. $r_{N+1,t} = r_{m,t}$. We propose an AR(1)-GJR-GARCH(1,1) model for the dynamics of returns such that

$$\begin{aligned} r_{i,t} &= \alpha_{i,0} + \alpha_{i,\mu}\, r_{i,t-1} + \xi_{i,t}, \\ \xi_{i,t} &= \sigma_{i,t}\ \varepsilon_{i,t}, \end{aligned} \tag{1.14}$$

where $\xi_{i,t}$ is the error term and $\varepsilon_{i,t}$ is an innovation process with $E_{t-1}(\varepsilon_{i,t}) = 0$ and $E_{t-1}(\varepsilon_{i,t}^2) = 1$; $\alpha_{i,0}$ and $\alpha_{i,\mu}$ are the parameters of the autoregressive process with $|\alpha_{i,\mu}| < 1$ to ensure stationarity of the process $r_{i,t}$ for $i = 1, \ldots, N+1$. The DCC model of Engle (2002) is estimated in two steps. In the first step, the univariate GARCH models for each time series of returns are fitted and estimates of their conditional variances are thus obtained. In the second step, the standardized residuals $\varepsilon_{i,t} = \xi_{i,t}/\sigma_{i,t}$ are used to estimate the time-varying correlation matrix. More formally, the conditional variance process is defined as $H_t = D_t P_t D_t$, with $P_t = [\rho_{ij,t}]$ the conditional correlation matrix and $D_t$ a diagonal matrix with time-varying standard deviations on the diagonal. Thus,

$$\begin{aligned} D_t &= diag(\sigma_{1,t}, ..., \sigma_{N+1,t}), \\ P_t &= diag(q_{11,t}^{-1/2}, ..., q_{N+1\,N+1,t}^{-1/2})\, Q_t\, diag(q_{11,t}^{-1/2}, ..., q_{N+1\,N+1,t}^{-1/2}). \end{aligned} \tag{1.15}$$

To capture potential leverage effects that may be empirically relevant in periods of financial distress, the idiosyncratic conditional variance terms $\sigma_{i,t}^2$ are modelled as univariate GJR-

GARCH models. For the GJR-GARCH(1,1) model the elements of $H_t$ can be expressed as:

$$\sigma_{i,t}^2 = \omega_i + (\alpha_i + \gamma_i I\{\xi_{i,t-1} < 0\})\xi_{i,t-1}^2 + \beta_i\, \sigma_{i,t-1}^2,\ \ i = 1, ..., N+1. \tag{1.16}$$

The quantity $Q_t = [q_{ij,t}]$ in (1.15) is a symmetric positive definite matrix which is specified as

$$Q_t = (1 - \theta_1 - \theta_2)\bar{Q} + \theta_1 \varepsilon_{t-1}\varepsilon_{t-1}^T + \theta_2 Q_{t-1}, \tag{1.17}$$

where $\bar{Q} = E(\varepsilon_t\, \varepsilon_t^T)$ is the unconditional covariance matrix of the standardized residuals $\varepsilon_t$ obtained from the first step estimation; $\theta_1$ and $\theta_2$ are non-negative scalars satisfying $0 < \theta_1 + \theta_2 < 1$. The correlation estimator is given by $\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}}$. Hereafter, we will refer to the above specified model as GARCH-DCC.

### 1.4.2 GARCH-Copula Modelling

An $(N+1)$-dimensional copula $C$ is a multivariate distribution function on $[0,1]^{N+1}$ with standard uniform marginal distributions. Following Sklar's theorem (Sklar 1959), any multivariate distribution, in our case the multivariate distribution function of the innovations of the above GARCH processes, can be decomposed into univariate margins and a certain copula, that is

$$F_{\varepsilon_1,...,\varepsilon_{N+1}}(u_1, ..., u_{N+1}) = C\left(F_{\varepsilon_1}(u_1), ..., F_{\varepsilon_{N+1}}(u_{N+1})\right), \tag{1.18}$$

where $u_i$ is uniformly distributed on $(0,1)$, $F_{\varepsilon_1,...,\varepsilon_{N+1}}$ denotes the joint cumulative distribution function and $F_{\varepsilon_i}$ the corresponding marginal distribution functions of the innovations $\varepsilon_i$, for $i = 1, \ldots, N+1$.

In this subsection, we use a t-copula function to model the mutual dependence among standardized residuals. This copula function is given by

$$C_{\nu,\boldsymbol{\rho}}^t(u_1, ..., u_{N+1}) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_{N+1})} \frac{\Gamma(\frac{\nu+N+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^{N+1}|\boldsymbol{\rho}|}} \left(1 + \frac{\boldsymbol{x}'\boldsymbol{\rho}^{-1}\boldsymbol{x}}{\nu}\right)^{-\frac{\nu+N+1}{2}} d\boldsymbol{x}, \tag{1.19}$$

where $\Gamma$ is the gamma function, $\boldsymbol{\rho}$ is a correlation matrix, $\nu$ represents the degree of freedom both in margins and copula function. Note that if the t-copula and univariate t margins share the same degree of freedom $\nu$, then we obtain a multivariate t distribution with $\nu$ degree of freedom as in (1.19). In our case, we assume that $F_{\varepsilon_1}, ..., F_{\varepsilon_{N+1}}$ are univariate

t distributions with different degree of freedom parameters $\nu_1, ..., \nu_{N+1}$, thus we obtain a multivariate distribution function $F_\nu$ which has been termed as meta-$t_\nu$ distribution function (see Fang et al. 2002 for more details). In the following, we will refer to this model as GARCH-Copula.

### 1.4.3 CoSR estimation

To obtain the estimator of CoSR, we first estimate individual elements contained in $\mu_t$ based on the Monte Carlo average of the simulated arithmetic $h$-period firm returns, that is

$$\widehat{\mu}_{i,t} = \frac{\sum_{s=1}^{S} R^s_{i,t:t+h} \, I\{R^s_{m,t:t+h} < C\}}{\#SE}, \tag{1.20}$$

where $S$ is the number of Monte Carlo simulations and $\#SE = \sum_{s=1}^{S} I\{R^s_{m,t:t+h} < C\}$ is the number of scenarios out of $S$ affected by market distress. For each asset in the portfolio the filtered mean vector (average h-period ahead return conditional on a market distress episode) is given by $\widehat{\mu}_t = (\widehat{\mu}_{1,t}, ..., \widehat{\mu}_{N,t})^T$. Similarly, $\mu_{m,t}$ can be estimated as

$$\widehat{\mu}_{m,t} = \frac{\sum_{s=1}^{S} R^s_{m,t:t+h} \, I\{R^s_{m,t:t+h} < C\}}{\#SE}. \tag{1.21}$$

Thus the estimator of CoER can be written as

$$\widehat{CoER}_t = W_t^T \widehat{\mu}_t - \widehat{\mu}_{m,t}, \tag{1.22}$$

where $W_t$ denotes the vector of portfolio weights that is known at time $t$. As for the CoSD, we first estimate the covariance matrix $\Sigma_{t|SE}$ using the Monte Carlo sample counterpart, with element $(i, j)$ defined as

$$\widehat{\Sigma}_{t(i,j)|SE} = \frac{\sum_{s=1}^{S} \left(R^s_{i,t:t+h} - \widehat{\mu}_{i,t}\right) \left(R^s_{j,t:t+h} - \widehat{\mu}_{j,t}\right) I\{R^s_{m,t:t+h} < C\}}{\#SE - 1} \tag{1.23}$$

for $i, j = 1, \ldots, N$. Then, we estimate $\sigma^2_{m,t|SE}$ as

$$\widehat{\sigma}^2_{m,t|SE} = \frac{\sum_{s=1}^{S} \left(R^s_{m,t:t+h} - \widehat{\mu}_{m,t}\right)^2 I\{R^s_{m,t:t+h} < C\}}{\#SE - 1}. \tag{1.24}$$

Analogously, we obtain the estimator of $\sigma_{im,t|SE}$ as

$$\widehat{\sigma}_{im,t|SE} = \frac{\sum_{s=1}^{S} \left(R^s_{i,t:t+h} - \widehat{\mu}_{i,t}\right)\left(R^s_{m,t:t+h} - \widehat{\mu}_{m,t}\right) I\{R^s_{m,t:t+h} < C\}}{\#SE - 1}, \tag{1.25}$$

and hence $\widehat{\sigma}_{t|SE} = (\widehat{\sigma}_{1,t}, ..., \widehat{\sigma}_{N,t})^T$. Combining the above estimators together, we obtain the estimator of CoSD, that is

$$\widehat{CoSD}_t = \left(W_t^T \widehat{\Sigma}_{t|SE} W_t + \widehat{\sigma}^2_{m,t|SE} - 2W_t^T \widehat{\sigma}_{t|SE}\right)^{1/2}. \tag{1.26}$$

The estimator of $CoSR_t$ is expressed as $\widehat{CoSR}_t = \frac{\widehat{CoER}_t}{\widehat{CoSD}_t}$.

## 1.5 Empirical analysis

This section illustrates the performance of our systemic risk-based optimal portfolios. We compare the ex-post final wealth and cumulative logarithmic returns of portfolios obtained by maximizing two performance measures: the traditional Sharpe ratio (SR) corresponding to the mean-variance strategy and our CoSR measure that incorporates systemic event. We also add the naive equally-weighted portfolio $\omega_i = 1/N$, for $i = 1, \ldots, N$, and the GMVP as benchmarks. Finally, we compute portfolio's LRMES as the relevant portfolio-level systemic risk measure, which is defined below as the weighted sum of LRMES across the portfolio constituents.

### 1.5.1 Dataset

We use stock price data from the US market. Our sample contains 23 big financial firms that are either SIFIs or non-SIFIs. The Financial Stability Board (FSB), in consultation with Basel Committee on Banking Supervision (BCBS) and national authorities, has just identified the latest list of global systemically important financial institutions (G-SIFIs) in November of 2020.[2] The overall number of G-SIFIs contained in the list is 30, specifically 20 of them are traded on the US market. Besides, the Board of Governors of the Federal Reserve System also maintains a list of domestic systemically important financial institutions (D-SIFIs). This list includes those financial institutions not being big enough for G-SIFIs status, but still possess high enough domestic systemically importance, making

[2] https://www.fsb.org/wp-content/uploads/P111120.pdf

them subject to the most stringent annual Stress Test (USA-ST) from the Federal Reserve. Despite the lack of any official D-SIFIs designation, the institutions being subject to the USA-ST can be considered to be D-SIFIs in the US.[3] According to the list released by Federal Reserve as of March 2014, 17 banks traded on US stock market were identified as D-SIFIs.[4]

The intensity of the computational simulation methods that we propose makes difficult to work with large sets of assets. In addition, the definition of the systemic risk measures also involves knowledge of financial information on firms beyond the stock price, which is not readily available for some firms. These two factors reduce the number of firms that we can consider in our empirical application. Thus we consider 16 firms within the group of SIFIs contained in the above two lists. All firms within the top three buckets (3.5%, 2.5% and 2.0%) of G-SIFIs list are included in our dataset.[5] A few remarks on computational complexity are given in the last section of Appendix B. In addition to the SIFIs, we also add 7 non-SIFIs into our dataset since we aim to find the best tradeoff between risk and return rather than only minimizing the underlying systemic risk of our portfolios. Our choice of non-SIFIs is motivated by Brownlees and Engle (2016), these authors also use these firms in their empirical study on systemic risk.

Historical return data on the stocks included in our dataset are retrieved from the Wharton Database website[6] over the period from January 3, 2000 to December 31, 2020 (5284 daily observations for each stock), and the panel is balanced since all firms have been trading continuously during the sample period. The price sequences are adjusted for splits based on split adjustment factors reported by both CRSP and Compustat. We proxy the market index with the S&P Composite Index, which will be later used as our benchmark when solving the portfolio optimization problem.

The full list of tickers and company names grouped by subindustry are Depositories: Bank of America (BAC), Citigroup (C), Synovus Financial (SNV), Truist Financial Corporation (TFC), HSBC Holdings (HSBC), JP Morgan Chase & Co (JPM), Barclays (BCS), Morgan Stanley (MS), State Street (STT), ING Groep (ING), Keycorp (KEY), Northern Trust (NTRS), PNC Financial Services (PNC) and Wells Fargo & Co (WFC); Insurance companies: Lincoln National (LNC), Progressive (PGR) and Global Life (GL); Broker-

[3] https://www.govinfo.gov/content/pkg/CHRG-113hhrg80873/pdf/CHRG-113hhrg80873.pdf
[4] https://www.federalreserve.gov/newsevents/press/bcreg/ccar_20140326.pdf
[5] The bucket approach is defined in Table 2 of the Basel Committee document (see https://www.bis.org/publ/bcbs255.pdf).
[6] https://wrds-www.wharton.upenn.edu/

Dealers companies: Goldman Sachs (GS) and Schwab Charles (SCHW); and other financial companies: American Express (AXP), Franklin Resources (BEN), Blackrock (BLK) and Capital One Financial (COF). The reason for only including large financial institutions in our dataset is that they are more exposed to systemic risk than non-financial firms, especially during crisis times.

For illustrative purposes, Figure 1.1 presents a descriptive analysis of two big financial institutions (Citigroup and Goldman Sachs) as well as two non-financial counterparts (Squibb and Boeing). The main aim of this exercise is to highlight the systemic risk of large financial institutions as opposed to non-financial firms of similar size. By doing so, we aim to motivate the importance of our portfolio strategy for portfolios of assets that exhibit large individual systemic risk.

The left panel of Figure 1.1 reports the relative price movements for these firms. The initial level of each price series has been normalized to unity to facilitate the comparison of relative performance, and no dividend adjustments are explicitly taken into account. The evolution of S&P 500 Index in the out-of-sample period (2007-2020) is reported in the top right panel of Figure 1.1. The S&P 500 Index has experienced four dramatic declines over the analyzed period. The first one happened during 2007-2009 due to the subprime crisis, the second one took place over 2010-2012 due to the European sovereign debt crisis, the third one occurred at the beginning of 2016 due to a decline in oil prices, and the latest one broke out at the beginning of 2020 due to the Covid-19 pandemic. The bottom panels of Figure 1.1 illustrate the dynamics of SRISK and LRMES (see Appendix A for definitions of both systemic risk measures) for these four firms over the evaluation period. During the subprime crisis, both financial firms suffered great losses with a drawdown of around 80%, while the non-financial firms performed much better, with relatively small drops in asset prices.

The comparison of the SRISK and LRMES measures between financial and non-financial firms during the different crisis episodes reveals that financial firms contribute more to the overall market disruption than non-financial firms. We also observe the buildup of the systemic risk measure at the start of the different crises for the two financial firms but not for the non-financial firms. In particular, the SRISK of non-financial firms delivers lower volatilities and is always below zero throughout the out-of-sample period. It is interesting to note, for example, that despite the increase in the SRISK of Boeing during the Covid-19 pandemic its value remains negative. Brownlees and Engle (2016) argue that a negative

SRISK indicates that the firm faces expected capital surpluses conditional on a market decline, i.e. the firm functions well and does not contribute to the overall systemic risk during times of crisis. Similar insights are obtained from the analysis of the dynamics of LRMES. This measure displays quite different patterns across firm groups over time. The LRMES of financial firms increases significantly before each crisis, which reflects the fact that the interconnections between financial institutions and the market become stronger during difficult times. However, the LRMES of non-financial firms does not exhibit violent fluctuations before or during crisis times. The lack of sensitivity of both systemic risk measures for both non-financial firms confirms the weak linkage between non-financial firms and the market.

These results show that our objective function is more relevant when the universe of assets includes large firms that are potentially systemic, although not necessarily classified as SIFIs. Therefore, in the remaining, we only focus on large financial firms when studying optimal portfolio allocation under market distress periods since these firms are more likely to affect and be affected by market declines during systemic risk episodes.

### 1.5.2 Empirical methodology

We demonstrate the superiority of the proposed portfolio selection procedure under stressed market conditions by comparing the results of the portfolios obtained from maximizing our CoSR measure against competitors used in the literature. We backtest our model over the period January 2007 to December 2020. The backtesting period has been chosen to include most of the recent financial crises. In particular, we use a rolling window of 1,500 daily historical returns to estimate the model parameters and then simulate 30,000 return scenarios from the above processes for each asset contained in the portfolio at the beginning of each month.

The portfolio optimization problem (1.8) with short-selling constraints is solved on a monthly basis by maximizing the proposed performance ratio CoSR based on generated return scenarios. To generate the return scenarios, we follow the two strategies discussed above. First, we apply a GARCH-DCC model for the dynamics of returns. After fitting the model, we use nonparametric bootstrap to resample the standardized residuals. These pseudo-samples are used as inputs of the GARCH and DCC filters respectively, to get the simulated monthly returns. The second approach is to use a GARCH-Copula model. After

fitting the model, we simulate 30,000 independent random trials of mutually dependent standardized residuals over a one-month horizon based on the fitted t-copula. Using the simulated standardized residuals as inputs to the GARCH filter, we obtain 30,000 simulated monthly cumulative returns. We can estimate the reward and risk measures using the generated return distributions, i.e. compute the first and second conditional moments by filtering realizations that satisfy the SE condition. In particular, following Capponi and Rubtsov (2022), we choose the following two specifications for the systemic event threshold $C$: i) $C = 0$, i.e., rebalancing occurs when the market index experiences negative returns, and ii) $C = -6.7\%$ for monthly rebalancing, which corresponds to a 40% decrease in the market index over a six-month period. Although the second specification better captures a SE (i.e. a significant drop in the market index), we still want to see the differences in portfolio allocation between milder and stronger definitions of systemic risk. Thus we also test our portfolios on less severe market declines, which are represented by the first specification.

For comparison purposes, we also evaluate the performance of our CoSR portfolio against three other performance criteria, namely the mean-variance (SR) portfolio obtained from maximizing the Sharpe ratio, the equally-weighted portfolio (1/N), and the GMVP. The first refers to the portfolio on the mean-variance efficient frontier that has the highest expected return per unit of risk, the second strategy represents a well-diversified portfolio of assets, and the last is the portfolio on the mean-variance efficient frontier with minimum variance. Moreover, the portfolio strategy maximizing CoSR is related to the SR portfolio since it is obtained by adjusting the latter to account for systemic risk events (see equations (1.5) and (1.6)). To avoid the construction of portfolios with large negative allocations to all assets under stressed market conditions, we assume that short-selling is not allowed in our model. Furthermore, we assume that our investors have an initial wealth of $FW_0 = 1$ and an initial cumulative logarithmic return $CR_0 = 0$ at the beginning of the backtesting period.

Three main steps are performed to calculate the ex-post final wealth and cumulative return at the $k$-th recalibration ($k = 0, 1, 2, ..., 168$). Firstly, we choose a performance ratio. Second, we generate return scenarios based on the algorithms described above and obtain the solution $W^*_{k+1}$ to the optimization problem (1.7). This step is performed in Matlab using the *fmincon* function. Following Kresta et al. (2015), we randomly choose 20 starting points in order to find the global instead of local minimum when solving (1.7).

Secondly, the ex-post final wealth is given by

$$FW_{k+1} = FW_k(1 + W_k^{*T} R_{k+1}), \tag{1.27}$$

where $R_{k+1}$ is the ex-post vector of simple returns between $k$ and $k+1$. Thirdly, the ex-post cumulative logarithmic return is given by

$$CR_{k+1} = CR_k + ln(1 + W_k^{*T} R_{k+1}). \tag{1.28}$$

Note that the latter measure reports the cumulative performance of the portfolio net of wealth. That is, expression (1.27) implies that $FW_{K+1} = FW_0 \prod_{k=0}^{K}(1 + W_k^{*T} R_{k+1})$. Then, taking logs, we obtain $\ln FW_{K+1} - \ln FW_0 = \sum_{k=0}^{K} \ln(1 + W_k^{*T} R_{k+1})$. Therefore, the growth in wealth due to the cumulative return on the portfolio is given by expression (1.28), with $CR_0 = 0$.

By repeatedly computing $FW_{k+1}$ and $CR_{k+1}$ for different performance ratios we obtain the wealth and cumulative return path evolutions over the evaluation period and the final wealth and total return accumulated at the end of the period. For simplicity, we neglect transaction costs for now. The influence of transaction costs will be further studied later.

### 1.5.3 Empirical results

In this section, we present the backtesting results. First, we show the results of the portfolio optimization exercise using the GARCH-DCC and GARCH-Copula models, respectively. Second, we study the influence of adding transaction costs to the results. We also compute confidence intervals to our estimates of final wealth paths to account for the uncertainty arising from model estimation.

The empirical results of the portfolio optimization backtesting using the GARCH-DCC model are depicted in Figure 1.2. There are several noticeable features from these figures. Firstly, all portfolios perform badly during the 2007-2008 financial crisis, no matter which model is chosen. In general, the CoSR portfolio with $C = -6.7\%$ outperforms the other competitors throughout the evaluation period. Final wealth is maximized when investors use the CoSR as objective function, the second strategy is the SR portfolio and the worst performance with regards to final wealth is the GMVP. In contrast, when the systemic event is defined by a milder threshold (i.e. $C = 0$) the results vary. In this scenario the

CoSR portfolio does not outperform the competing portfolios consistently but it is still more resilient to crises than the other three portfolios. Losses are significantly smaller during these periods. This observation also reflects the importance of choosing a proper systemic event threshold for portfolio selection. Conditioning on a mild threshold may jeopardize return at the expense of a more conservative portfolio allocation.

Table 1.1 confirms that the CoSR portfolio with $C = -6.7\%$ provides the best performance. An investor would multiply their wealth by 2.280 using the SR strategy, by 1.343 using the equally-weighted strategy, by 1.323 using the GMVP, while following the proposed CoSR strategy the final wealth would be around triple (2.794 for $C = 0$ and 3.021 for $C = -6.7\%$ respectively). Similarly, the annual return of the CoSR portfolio with $C = -6.7\%$ is 8.22%, which is about two percentage points above the SR portfolio given by 6.06%. The annual return for the equally-weighted portfolio and the GMVP are 2.13% and 2.02% respectively.

Another factor the investor would care about is the risk of the strategy. The CoSR strategy not only outperforms the other competing strategies in terms of profitability but also the maximum drawdown decreases, which is an important indicator of portfolio performance for portfolio managers. While SR, 1/N and GMVP strategies lost near 70% (74.22%, 71.74%, and 67.21%, respectively) of their values during the 2007-2008 financial crisis, the maximum drawdown of CoSR was around 60% for both thresholds. Similar findings are obtained for the other three major crisis episodes. In these cases there is also a drop in profitability of the strategy but this drop is smaller compared to the 2007-2009 period. To

Table 1.1: Final wealth and maximum drawdown of particular wealth paths based on GARCH-DCC model.

| Strategy | SR | CoSR(C=0) | CoSR(C=-6.7%) | 1/N | GMVP |
|---|---|---|---|---|---|
| Final Wealth | 2.280 | 2.794 | 3.021 | 1.343 | 1.323 |
| Annual Return | 6.06% | 7.62% | 8.22% | 2.13% | 2.02% |
| Maximum Drawdown | 74.22% | 61.55% | 58.75% | 71.74% | 67.21% |

add robustness to the results, we repeat the analysis for the copula model. The results are very similar to those obtained for the GARCH-DCC model. The empirical results of the portfolio optimization backtesting are depicted in Figure 1.3. There are several noticeable features from this figure. All portfolios perform badly during 2007-2008 financial crisis, no matter which model is chosen. The SR, 1/N and GMVP strategies lose almost all of their value during that period, while the CoSR portfolio performs much better but still lose

more than 50% of its value. The SR portfolio is a serious competitor and reports similar profitability figures to the CoSR during the first half of the evaluation period, however, from the second semester of 2016, the CoSR portfolio consistently beats the SR portfolio. Overall, the CoSR portfolio has a strong upward trend in profitability that results in superior performance over time. This strong performance is due to its relatively stable performance in times of market downturns. Table 1.2 summarizes earnings and maximum drawdown of different strategies. The SR portfolio provides the worst performance in terms of final wealth whereas the maximum drawdown is comparable to the maximum drawdown of the equally-weighted portfolio (73.86% for SR and 71.74% for 1/N). Both systemic event thresholds provide similar performance, where the CoSR portfolio with $C = 0$ provides the highest value of final wealth (annual return) and the lowest maximum drawdown.

Table 1.2: Final wealth and maximum drawdown of particular wealth paths based on GARCH-Copula model.

| Strategy | SR | CoSR(C=0) | CoSR(C=-6.7%) | 1/N | GMVP |
|---|---|---|---|---|---|
| Final Wealth | 1.299 | 2.423 | 2.134 | 1.343 | 1.423 |
| Annual Return | 1.88% | 6.53% | 5.56% | 2.13% | 2.55% |
| Maximum Drawdown | 73.86% | 59.98% | 61.07% | 71.74% | 67.43% |

**Portfolio diversification for portfolios of SIFI firms:**

As an additional robustness exercise, we repeat the portfolio allocation exercise for the subset of the firms in our study that are classified by the Financial Stability Board (FSB) and the Basel Committee on Banking Supervision (BCBS) as G-SIFIs and by the Board of Governors of the Federal Reserve System as D-SIFIs. In particular, we consider 16 firms. This exercise may be interesting to highlight the importance of portfolio diversification in a setting where all the assets in the portfolio are affected by systemic risk. Note that in the above exercises some firms were within the pool of SIFIs but others were not.

The results of this exercise are reported in Figure 1.4. The top panel of this figure shows the slight outperformance of the CoSR portfolio with $C = -6.7\%$ compared to other competitors over the first half of the evaluation period, however, as expected, there is a sizeable drop in profitability for all portfolios compared to the portfolios also considering non-SIFIs, see Figure 1.2. The analysis of the GARCH-Copula model shows similar results,

however, in this case, the equally-weighted portfolio is the top contender, followed by the CoSR with $C = 0$.

**Portfolio turnover and transaction cost:**

We use the definition of portfolio turnover in Kirby and Ostdiek (2012), which is consistent with the concept used in the mutual fund industry. This measure provides an indication of the variability of the portfolio weights over time. Table 1.3 reports the turnover rates for all the portfolios under investigation. This table shows that portfolio optimization strategies based on the maximization of CoSR are characterized by relatively high turnover rates. Unsurprisingly, the turnover rates are much smaller for the equally-weighted portfolio than for the remaining competitors. In contrast, both CoSR portfolios take larger values, which suggests that these portfolios are more flexible than the competitors to adapt to changes in market conditions.

Table 1.3: Comparison of turnover rates.

| Strategy | SR | CoSR(C=0) | CoSR(C=-6.7%) | 1/N | GMVP |
|---|---|---|---|---|---|
| GARCH-DCC | 0.250 | 0.356 | 0.298 | 0.025 | 0.236 |
| GARCH-Copula | 0.241 | 0.431 | 0.361 | 0.025 | 0.253 |

On the other hand, an increase in portfolio turnover entails an increase in transaction costs due to higher fees and other costs derived from modifying the portfolio allocation. We proceed to analyze the impact on portfolio performance of including these costs. To do this, we recompute the ex-post final wealth and the total return for all portfolios considering proportional transaction costs. In order to stress test the impact of transaction costs, we adopt 5 basis points as proportional transaction costs. Tables 1.4 and 1.5 report the results in this case. Figures 1.5 and 1.6 also illustrate the difference in portfolio performance for the DCC and copula models, respectively. The presence of transaction costs does not alter the results.[7]

[7] Unreported results show the effect of transaction costs of different magnitude on portfolio performance. More specifically, we obtain the results of the CoSR portfolio with $C = -6.7\%$ for the best performing strategy - GARCH-DCC approach - assuming transaction costs that range from 0 to 10 basis points. The results confirm the profitability of the CoSR strategy across different levels of the transaction costs. The CoSR strategy always outperforms the 1/N portfolio and GMVP.

Table 1.4: Recomputation results based on GARCH-DCC model with transaction costs.

| Strategy | SR | CoSR(C=0) | CoSR(C=-6.7%) | 1/N | GMVP |
|---|---|---|---|---|---|
| Final Wealth | 2.186 | 2.632 | 2.873 | 1.338 | 1.272 |
| Annual Return | 5.74% | 7.16% | 7.83% | 2.10% | 1.73% |
| Maximum Drawdown | 74.33% | 61.76% | 58.96% | 71.77% | 67.37% |

Table 1.5: Recomputation results based on GARCH-Copula model with transaction costs.

| Strategy | SR | CoSR(C=0) | CoSR(C=-6.7%) | 1/N | GMVP |
|---|---|---|---|---|---|
| Final Wealth | 1.247 | 2.254 | 2.008 | 1.338 | 1.364 |
| Annual Return | 1.59% | 5.98% | 5.11% | 2.10% | 2.24% |
| Maximum Drawdown | 73.99% | 60.23% | 61.31% | 71.77% | 67.60% |

**Portfolio systemic risk measure:**

Our portfolios are constructed to maximize the Sharpe ratio conditional on the market being under distress. This ratio can be viewed as a measure of risk-adjusted profitability under market distress, with the latter interpreted as a systemic event. In order to assess the underlying systemic risk of such portfolios we define portfolio's LRMES as

$$LRMES_{p,t} = \sum_{i=1}^{N} \omega_{i,t} LRMES_{i,t}. \tag{1.29}$$

This measure is a weighted combination of the LRMES of the individual firms at each point in time. Interestingly, the portfolio's LRMES can be interpreted as the expected percentage drop in portfolio value under stressed market conditions. Thus a lower value of $LRMES_p$ reflects a lower level of potential loss during crisis times. This quantity can be estimated based on the generated return scenarios obtained from the GARCH-DCC and GARCH-Copula models.

Figure 1.7 displays portfolios' LRMES paths obtained from different investment strategies over the out-of-sample evaluation period for the GARCH-DCC and GARCH-Copula models, respectively. The LRMES of the CoSR portfolio is relatively stable across the evaluation period and is always lower than for the other benchmark portfolios. This forward-looking measure can serve as an early warning indicator or monitoring tool for

both portfolio managers and financial regulators who aim to control the losses of their portfolios, especially during crisis times.

An important feature of the portfolio allocation exercise is to study the variation of the portfolio across assets and over time. The optimal weights are shown in Figure 1.8. Here, we set $C = -6.7\%$ for both GARCH-DCC and GARCH-Copula models when computing optimal weights and LRMES. Firms that receive greater allocations of wealth under the optimal CoSR portfolio strategy are more attractive from a systemic risk-return perspective. Interestingly, the empirical results in Figure 1.8 show that the optimal CoSR portfolio is less diversified than the SR portfolio during crisis times after accounting for systemic risk. For instance, the CoSR portfolio implies a relatively high investment proportion in PGR while the SR portfolio invests more in BEN across the evaluation period. An interpretation of this result is that investors anticipate a systemic risk event in advance. As a result, investors prefer to sacrifice diversification benefits and gain from the reduced exposure of their portfolios to stressed market conditions (see also Capponi and Rubtsov 2022). These insights of the model provide an alternative interpretation to the presence of underdiversification compared to standard mean-variance efficient allocations, see Mitton and Vorkink (2007) and references therein. In our model, underdiversification takes place because the CoSR portfolio is less likely to suffer great losses during a market slide. Figure 1.9 shows that the LRMES of PGR is always lower than the LRMES of BEN. This difference becomes even larger during distress episodes.

### 1.5.4 Estimation effects on optimal portfolio allocation

Throughout the study, we have considered two different specifications (GARCH-DCC and GARCH-Copula) to model the joint dynamics of financial returns. This exercise has provided robustness to our results against the presence of model uncertainty. Another related exercise is to study the impact of parameter uncertainty. In this case the objective is to assess the impact of parameter estimation on the outcome of the model. In our setting, the outcomes of the model are estimates of the final wealth and portfolio return. This exercise is particularly important in our setting as our model is heavily parametrized as it is custom in multivariate time series models. Alternative nonparametric solutions suffer instead from the curse of dimensionality as the number of variables in the model grows beyond a few dimensions.

In this section we assess the impact of estimation error. The uncertainty arises because of the parameter estimation error but also because of the randomness in choosing starting values in the portfolio optimization. As mentioned before, we randomly choose 20 starting points when solving portfolio optimization problems in order to find global instead of local optima. In the backtesting exercise we follow a rolling window approach starting initially from the beginning of 2007 with a window size of 1,500 observations. After fitting the different models within each window, the estimated parameters for predicting one-month ahead returns are obtained before re-estimating the same model with additional observations. The prediction of returns obtained from each model and the corresponding portfolio optimization are done on a monthly basis by updating the in-sample dataset. Motivated by the need of gauging the underlying estimation uncertainty, the whole procedure is repeated multiple times with the same methodology. By doing so, we obtain multiple portfolio path realizations throughout the out-of-sample period.

Figures 1.10 and 1.11 show the ex-post final wealth paths for different strategies after accounting for estimation uncertainty. For instance, the curve "CoSR_Average" reflects the average of the 200 portfolio paths, which is embedded into the corresponding 90% confidence bounds centered around the average. The grey shadow area reflects the uncertainty arising from the model estimation, return prediction and portfolio optimization underlying the 200 simulation exercises. The corresponding results for other competitors are also displayed therein. The results of both approaches displayed in Figures 1.10 and 1.11 confirm the statistical significance of the previous evidence on the superiority of the CoSR portfolios over the competing benchmark portfolios in all cases.

### 1.5.5 An alternative objective function for portfolio allocation

An alternative strategy to incorporate systemic risk in the portfolio allocation problem is to replace the denominator in (1.1) by the LRMES of portfolio's excess return. By doing this, we develop a new performance measure that we call mean-MES ratio (MMR):

$$\begin{aligned} MMR_t(R_{p,t}) :&= \frac{E_t(R_{p,t:t+h} - R_{m,t:t+h})}{-E_t(R_{p,t:t+h} - R_{m,t:t+h}|SE_{t:t+h})}, \\ &= \frac{W_t^T \mu_t - \mu_{m,t}}{\mu_{m,t|SE} - W_t^T \mu_{t|SE}}. \end{aligned} \tag{1.30}$$

If we set $C = VaR_\alpha(R_{m,t:t+h})$, then the above expression can be rewritten as

$$MMR_t(R_{p,t}) = \frac{W_t^T \mu_t - \mu_{m,t}}{LRMES_{p,t} - ES_{m\alpha,t}}, \tag{1.31}$$

where $ES_{m\alpha,t} = ES_\alpha(R_{m,t:t+h})$, and $ES_\alpha$ is defined as $ES_\alpha(X) = -E(X|X \leq VaR_\alpha(X))$ if we assume a continuous distribution for the probability law of $X$. The risk measure in the denominator can be decomposed into the difference between portfolio's LRMES and the ES of market return. MMR is able, by construction, to measure the tradeoff between portfolio's mean return and systemic risk, which formulates a new mean-ES model that accounts for systemic risk.

In what follows, we present the backtesting results for the portfolios obtained under the MMR objective function. We first show the results of the portfolio optimization exercise using GARCH-DCC and GARCH-Copula models, respectively. Then we study the systemic risk of MMR portfolio and compare against the CoSR portfolio proposed as our main objective function above. We also compute the confidence intervals of the ex-post final wealth paths to account for the uncertainty arising from the model estimation procedure.

**Backtesting results**

The backtesting results of GARCH-DCC and GARCH-Copula model including the MMR optimal portfolios are illustrated in Figure 1.12 and 1.13, respectively. These portfolios provide the best out-of-sample performance in terms of cumulative return over the evaluation period. The second competitors are the CoSR portfolios studied earlier whereas the remaining competitors perform clearly below these two investment portfolios that are focused on minimizing the effect of systemic event. Table 1.6 extends Table 1.1 by replacing the CoSR statistics by the MMR values. An investor will multiply his/her wealth by 2.280 using SR strategy, by 1.343 using 1/N strategy, by 1.323 using GMVP, while following the proposed MMR portfolio the final wealth would be more than sextuple (6.627) for $C = 0$ and triple (3.734) for $C = -6.7\%$. Similarly, the MMR portfolio with $C = 0$ gives an annual return of 14.46%, which is more than double the annual return of the SR portfolio (6.06%). The MMR portfolio with $C = -6.7\%$ performs slightly worse but still beats the other competitors with an annual return of 9.87%. The annual return for the naive and GMVP are 2.13% and 2.02%, respectively.

To add robustness to the results, we repeat the analysis using the GARCH-Copula

model. The backtesting results are illustrated in Figure 1.13. The SR, 1/N and GMVP lost almost all of their value during that period, while the MMR portfolios perform much better but still lost more than half of their value. The MMR portfolio with $C = 0$ provides the best performance, which is the same as we conclude from the GARCH-DCC model. However, the level of profitability is much lower compared to the previous counterparts. The CoSR presents strong performance in the second part of the evaluation period clearly beating the other portfolios but not the MMR in terms of profitability.

Table 1.7 summarizes the earnings and maximum drawdown of the different portfolios. Results for the CoSR portfolios are found in Table 1.2 and not reported here again. The MMR portfolio with mild systemic event threshold provides the best performance in terms of final wealth (3.222), while the MMR portfolio with $C = -6.7\%$ gives the lowest maximum drawdown (63.30%) among the competitors.

Table 1.6: Backtesting results based on GARCH-DCC model.

| Strategy | SR | MMR(C=0) | MMR(C=-6.7%) | 1/N | GMVP |
|---|---|---|---|---|---|
| Final Wealth | 2.280 | 6.627 | 3.734 | 1.343 | 1.323 |
| Annual Return | 6.06% | 14.46% | 9.87% | 2.13% | 2.02% |
| Maximum Drawdown | 74.22% | 34.96% | 41.45% | 71.74% | 67.21% |

Table 1.7: Backtesting results based on GARCH-Copula model.

| Strategy | SR | MMR(C=0) | MMR(C=-6.7%) | 1/N | GMVP |
|---|---|---|---|---|---|
| Final Wealth | 1.299 | 3.222 | 2.040 | 1.343 | 1.423 |
| Annual Return | 1.88% | 8.72% | 5.22% | 2.13% | 2.55% |
| Maximum Drawdown | 73.86% | 63.57% | 63.30% | 71.74% | 67.43% |

The MMR portfolio is clearly a strong portfolio candidate under market distress in terms of cumulative return, however, its exposure to systemic risk is significantly larger than for the CoSR portfolio. Figure 1.14 presents the dynamics of the LRMES of the different portfolios over the evaluation period. For both GARCH-DCC and GARCH-Copula methodologies and different values of $C$, the CoSR portfolio exhibits values of the LRMES statistic well below the other portfolios. This observation provides strong support to the CoSR against the MMR portfolio once we jointly consider the profitability measures given by the ex-post final wealth and cumulative return and the systemic risk measure given by portfolio's LRMES.

Another advantage of CoSR strategies compared to MMR portfolios is the excess variability in final wealth and cumulative return of the latter class of investment strategies. The results of the robustness exercise for both GARCH-DCC and GARCH-Copula accounting for estimation uncertainty obtained from 200 trials are illustrated in Figure 1.15 and 1.16, respectively. The solid lines reflect the average of 200 portfolio paths, while the shaded areas represent the corresponding 90% confidence bounds centered around the average. The simulations suggest that MMR portfolios tend to suffer bigger losses than CoSR portfolios under market distress after accounting for estimation uncertainty. It is also worth noting that MMR portfolios are more sensitive to estimation error than the CoSR strategies, which makes their performance more volatile (the variance of the final wealth paths is much bigger than other competitors).

## 1.6 Conclusion

Although the existing systemic risk measures are helpful for financial regulators, portfolio managers are still looking for practical guidance under which they can account for systemic event during their decision-making process. A general approach for constructing optimal portfolios is to maximize a reward-to-risk ratio. In this paper we propose a systemic Sharpe ratio as the investor's objective function that conditions on the market return being under the threshold of a systemic event. By doing so, we propose a methodology for portfolio construction that explicitly incorporates the sensitivity of portfolio performance to systemic risk events. Using this objective function, we solve the portfolio allocation problem analytically under the absence of short-selling restrictions and numerically when short-selling restrictions are imposed. This approach for obtaining an optimal portfolio allocation is made operational by embedding it in a dynamic setting and simulating the returns on the portfolio assets using Monte Carlo return scenario analysis.

We have applied the above model to a basket of 23 assets of big financial firms trading in the US stock market over an out-of-sample evaluation period spanning 2007 to 2020. The results of the empirical study confirm the outperformance of our systemic risk portfolio against the standard mean-variance formulation, the naive equally-weighted portfolio, and the global minimum variance portfolio. The systemic risk portfolio is, by construction, more resilient in periods of market distress and remains competitive in non-crisis periods. This portfolio is less diversified than benchmark portfolios during crisis times. In these periods,

the systemic risk strategy invests on those stocks that are expected to experience a small loss under stressed market conditions. In contrast to an emerging literature that suggests that the presence of underdiversification in financial markets is a rational response to a preference for positive skewness, we find that investors take conservative positions on a few stocks that are resilient against systemic risk to shield against potential large drawdowns in portfolio value.

# Appendix A - A brief review of systemic risk measures

This section reviews the different definitions of systemic risk measures proposed in the macro-finance literature. In particular, we review four prominent market-based measures that constitute the building blocks of an emerging literature on systemic risk. First, we review the MES and SES measures introduced in Acharya et al. (2017). The first measure is defined as the expected decrease of an institution's net equity return conditional on a market decline:

$$MES_{i,t|t-1} = -E_{t-1}(R_{i,t}|R_{m,t} < C), \tag{1.32}$$

where $C$ denotes a given threshold defining the magnitude of a systemic event and $E_{t-1}(\cdot)$ denotes the expectation operator conditional on the information available to the individual up to time $t-1$. This measure gauges how a specific institution's risk exposure contributes to the system-wide risk. Financial institutions with higher MES contribute more to the overall risk of the financial market, thus these institutions can be seen as systemically dangerous. The SES extends the MES and measures the amount an institution's equity would drop below its target level (defined as the prudential capital fraction $k$ of assets) in case of a future crisis when aggregate capital is less than $k$ times aggregate assets:

$$\frac{SES_{i,t|t-1}}{W_{i,t}} = kL_{i,t} - 1 - E_{t-1}\left(R_{i,t}\Big|\sum_{i=1}^{N} W_{i,t} < k\sum_{i=1}^{N} A_{i,t}\right), \tag{1.33}$$

where $L_{i,t} = (A_{i,t}/W_{i,t})$ is the quasi leverage ratio, $A_{i,t} = (D_{i,t} + W_{i,t})$ is the total assets, $D_{i,t}$ is the book value of debt, $W_{i,t}$ is the market value of equity, and $N$ denotes the number of financial firms within the system. Acharya et al. (2017) also show that the conditional expectation term can be expressed as a linear function of MES:

$$SES_{i,t|t-1} = W_{i,t}(kL_{i,t} - 1 + \theta MES_{i,t|t-1} + \Delta_i), \tag{1.34}$$

where $\theta$ and $\Delta_i$ are constant terms (see also Benoit et al. 2017). The higher the SES is, the higher the contribution of the financial institution to the system's overall risk. Acharya et al. (2017) provide detailed theoretical justification for the positive correlation between SES and a firm's MES and leverage.

The second measure that we review is SRISK. It was initially proposed by Acharya et al. (2010), and later extended to a conditional version by Brownlees et al. (2012); Brownlees

and Engle (2016). SRISK measures the expected capital shortfall conditional on a systemic event and can be expressed as a function of the institution's size, leverage, and long-run MES (LRMES):

$$SRISK_{i,t+h|t} = max\left[0; W_{i,t}\big(kL_{i,t} - 1 + (1-k)LRMES_{i,t+h|t}\big)\right], \tag{1.35}$$

where $LRMES_{i,t+h|t}$ denotes the expected equity loss of firm $i$ conditional on a systemic event over a time horizon $h$ (usually one month or six months):

$$LRMES_{i,t+h|t}(C) = -E_t(R_{i,t:t+h}|R_{m,t:t+h} < C). \tag{1.36}$$

Here we denote the multiperiod simple firm (market) return between time $t$ and $t+h$ as $R_{i,t:t+h}$ ($R_{m,t:t+h}$) and the systemic event (SE) as $\{R_{m,t:t+h} < C\}$. SRISK extends the MES by taking account of both size and debt of financial institutions. The institution with highest SRISK is seen as the main contributor to a crisis and is taken as the most systemically important. It is worth noting that the mathematical expressions for SRISK and SES are almost the same, since they are both comprised of three components: firm size, leverage, and marginal risk. At the aggregate level, SRISK can be thought of as a stress test on the whole financial system, where the adverse case scenario is defined as a 10% (40%) decrease of the market index over a one-month (six-month) time horizon.

Another important systemic risk measure is $\Delta$CoVaR. The marginal risk measure given by VaR is used for measuring the tail risk of a portfolio or an individual firm, however, it fails to take into account spillover effects from other institutions and is highly pro-cyclical. To overcome these shortcomings, Adrian and Brunnermeier (2016) modify the VaR measure and present $\Delta$CoVaR. This measure is defined as the difference between the VaR of the financial system conditional on a particular institution being in distress and the VaR of the financial system conditional on that institution being in its median state:

$$\Delta CoVaR_{i,t}(\alpha) = CoVaR_t^{m|C(r_{i,t})} - CoVaR_t^{m|C_{Median}(r_{i,t})}, \tag{1.37}$$

where the CoVaR corresponds to the VaR of the market return conditional on a certain event $C(r_{i,t})$ observed for firm $i$:

$$P\big(R_{m,t} \leq CoVaR_t^{m|C(r_{i,t})} \big| C(r_{i,t})\big) = \alpha. \tag{1.38}$$

Various definitions of $C(r_{i,t})$ can be adopted to define a systemic event for firm $i$, for example, Adrian and Brunnermeier (2016) define it as the institution's loss equals to its VaR, while Girardi and Ergün (2013) consider the case where the institution's loss exceeds its VaR level. A higher $\Delta$CoVaR indicates higher systemic risk of the financial institution for a given level of idiosyncratic risk.

# Appendix B - Simulation of stock returns

## Algorithm for GARCH-DCC Model

This section describes the simulation algorithm for constructing return predictions based on nonparametric bootstrap approach (see Brownlees and Engle 2016). Specifically, we are interested in computing the portfolio's expected excess return over next month conditional on a systemic event during that period. In the following, we assume that all parameters are known, while in practice we estimate the model parameters using all available information up to the current time. Let $h$ denote the length of the forecasting horizon on which the returns will be simulated, which in our case is 22 trading days. The details of the bootstrap procedure are discussed as follows:

- Construct the GJR-DCC standardized innovations: $\epsilon_t = (L_t)^{-1}\varepsilon_t, \quad t = 1, ..., T$, where $L_t$ denotes the lower-triangular matrix in the Cholesky decomposition of the correlation matrix $R_t$.

- Sample with replacement $S \times h$ vectors of standardized innovations $\epsilon_t$. This provides us with $S$ pseudo-samples, $\epsilon^s_{T+t}, \quad t = 1, ..., h, \ s = 1, ..., S$, of length $h$ (i.e. 22 days) of the GJR-DCC innovations.

- Use the pseudo-samples obtained from the previous step as inputs of the GJR-DCC filters, and set initial values as the last values of returns $r_{i,T}$, error terms $\xi_{i,T}$, pseudo correlation matrix $R_T$ and variances $\sigma^2_{i,T}$. This yields $S$ simulated paths of logarithmic returns between period $T+1$ and $T+h$ for all firms and the market index conditional on the realized process up to $T$, that is, we obtain $r^s_{i,T+t} | \, \mathcal{F}_T, \quad t = 1, ..., h\,, \quad s = 1, ..., S$, with $\mathcal{F}_T$ denoting the information set available to the individual at time $T$.

- Calculate the arithmetic multi-period return for each simulated path as

$$R^s_{i,T:T+h} = exp\left(\sum_{t=1}^{h} r^s_{i,T+t}\right) - 1. \tag{1.39}$$

The simulated monthly returns are utilized to solve the portfolio optimization problems specified in (1.7) and (1.9).

## Algorithm for GARCH-Copula Model

In the GARCH-Copula setting, we generate future return scenarios according to the following procedure:

- Given the standardized residuals $\varepsilon_t = (\varepsilon_{1,t}, ..., \varepsilon_{N+1,t})$ $(t = 1, ..., T)$ from the fitted GARCH models, we estimate the cumulative distribution function (CDF) $F_{\nu_i}$ of each series with a univariate t location-scale distribution, $\varepsilon_{i,t} \sim F_{\nu_i} \quad t = 1, ..., T$.

- Transform the marginal distribution functions to uniforms with the empirical CDFs: $u_{i,t} = F_{\nu_i}(\varepsilon_{i,t}) \quad t = 1, ..., T$, where $u_{i,t} \sim U(0, 1)$.

- Given the transformed uniform margins, now we estimate the scalar degrees of freedom parameter $\nu$ and the linear correlation matrix $\boldsymbol{\rho}$ of the t-copula using the Matlab function *copulafit* (see The MathWorks 2019).

- Simulate mutually dependent returns by first simulating the corresponding dependent standardized residuals $\epsilon^s_{T+t}$ $(t = 1, ..., h,\ s = 1, ..., S)$. To do so, we first simulate dependent uniform variates $u^s_{T+t}$ $(t = 1, ..., h,\ s = 1, ..., S)$ using the Matlab function *copularnd* (see The MathWorks 2019).

- Transform those simulated uniform variates $u^s_{T+t}$ into standardized residuals via the inverse marginal CDF of each series: $\epsilon^s_{i,T+t} = F^{-1}_{\nu_i}(u^s_{i,T+t}) \quad t = 1, ..., h,\ s = 1, ..., S$, where $F^{-1}_{\nu_i}$ is the inverse CDF of the fitted $i$th marginal distribution. This step delivers simulated standardized residuals (pseudo samples) consistent with those obtained from the GARCH filter above. Note that these residuals are independent in time but dependent at any point in time.

- Using simulated standardized residuals $\epsilon^s_{i,T+t}$ as the i.i.d. input noise process, we reintroduce the autocorrelation as well as heteroskedasticity observed in historical

return data via the Matlab function *filter* (see The MathWorks 2019). We set initial values as the last values of returns $r_{i,T}$, standardized residuals $\varepsilon_{i,T}$ and variances $\sigma^2_{i,T}$. This yields $S$ simulated paths of logarithmic returns between period $T+1$ and $T+h$ for all firms and the market index conditional on the realized process up to $T$, that is

$$r^s_{i,T+t} | \ \mathcal{F}_T, \quad t = 1, ..., h, \quad s = 1, ..., S. \tag{1.40}$$

- Calculate the arithmetic multi-period return for each simulated path as

$$R^s_{i,T:T+h} = exp\left(\sum_{t=1}^{h} r^s_{i,T+t}\right) - 1. \tag{1.41}$$

Given the simulated return distributions, we can easily compute the conditional (unconditional) moments of portfolio excess returns and solve the portfolio optimization problems specified in (1.7) and (1.9).

## Computational Remarks

Despite the computational benefits of scenario analysis, return simulation and solving the nonlinear programming problem can still be a computationally exhaustive task - ultimately depending on the sample size of Monte Carlo simulation and the sample size of robustness check. Getting access to rich computing resources is crucial. This application is feasible only through access to high performance clusters (HPC) or cloud computing resources.[8] Specifically, RAM could be a limiting factor in our case since we have to cache all simulated return scenarios in order to dynamically solve the subsequent portfolio optimization problems. Due to the computational intensity, we set the sample size of return simulations as 30,000 and the number of iterations as 200. The single run of model estimation and return simulation procedure took around 5.7 hours on a MacBook Pro 2018 with 2.6 GHz 6-Core i7 processor and 16 GB RAM, while the multiple runs for purpose of robustness exercise took around 2.5 days on HPC. Solving the portfolio optimization problems took around 7 minutes and 25.3 hours for single run and multiple runs, respectively. In order to accelerate computation speed, we also employ the Parallel Computing Toolbox build in Matlab.[9]

[8] The robustness exercise made use of the facilities of the Hamilton HPC Service of Durham University (https://www.dur.ac.uk/arc/platforms/).

[9] https://www.mathworks.com/products/parallel-computing.html

# Appendix B - Figures

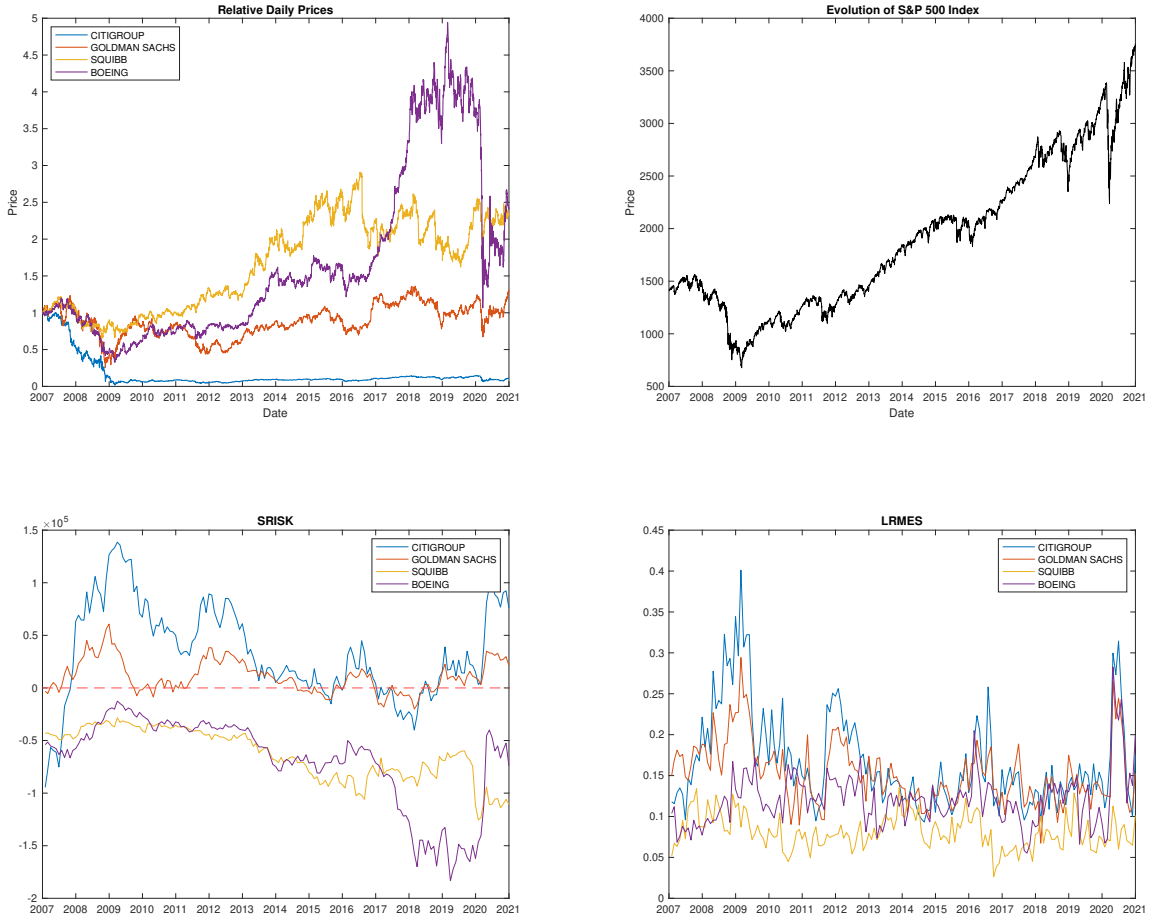



Figure 1.1: Left top panel reports the relative price developments between Jan 3, 2007 and Dec 31, 2020. Right top panel reports the S&P 500 Index. Dynamics of SRISK index on the left bottom panel and LRMES on the right bottom panel. In term of the simulation approach we follow Brownlees and Engle (2016).

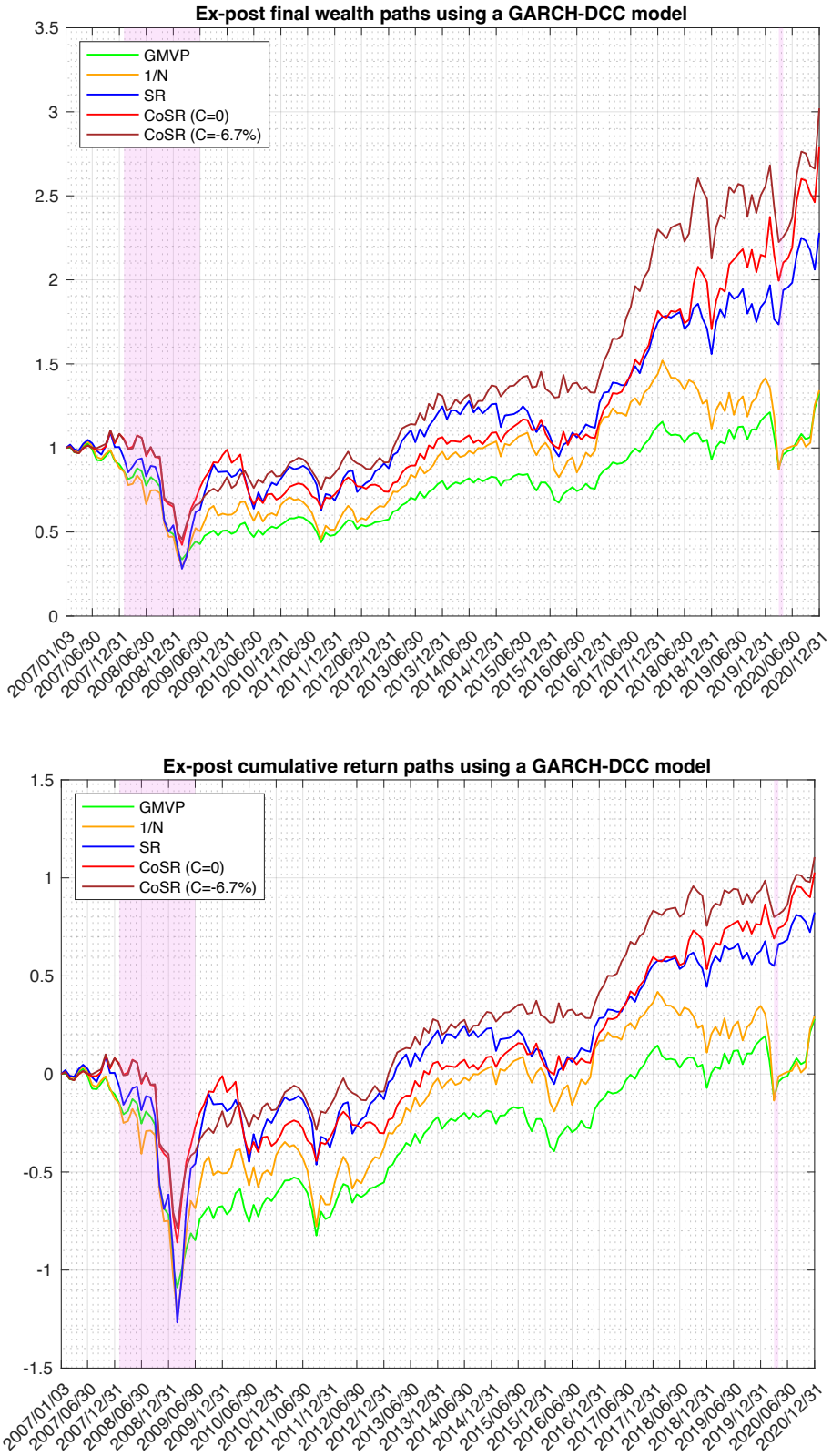



Figure 1.2: Top panel compares *ex post* final wealth paths and bottom panel compares the *ex post* cumulative return obtained using different strategies based on GARCH-DCC model. The shaded areas correspond to NBER recession periods.

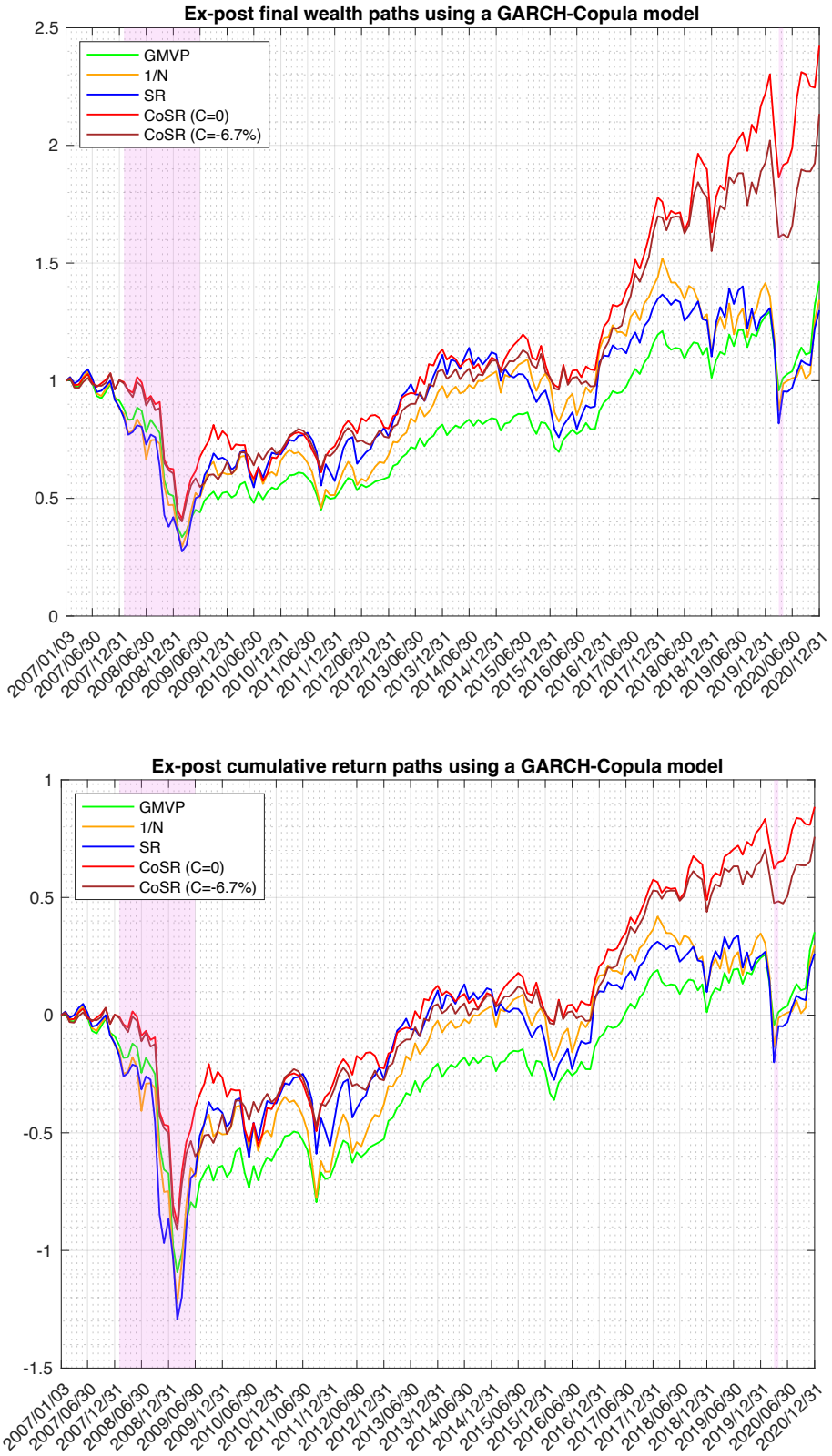



Figure 1.3: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies based on GARCH-Copula model. The shaded areas correspond to NBER recession periods.

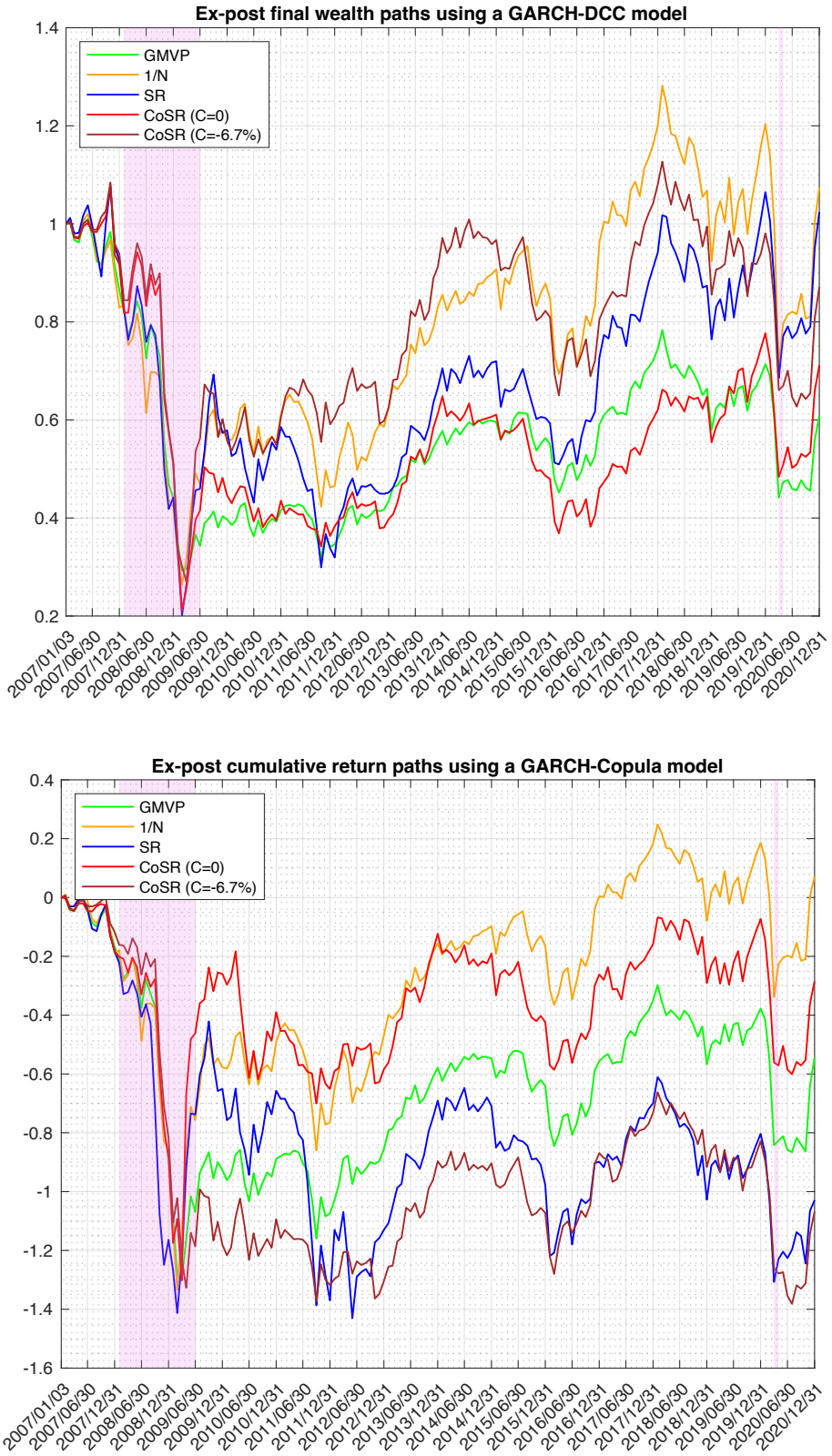



Figure 1.4: Ex-post final wealth paths obtained using **only** SIFIs as portfolio assets based on GARCH-DCC model (top panel) and GARCH-Copula model (bottom panel) respectively. The shaded areas correspond to NBER recession periods.

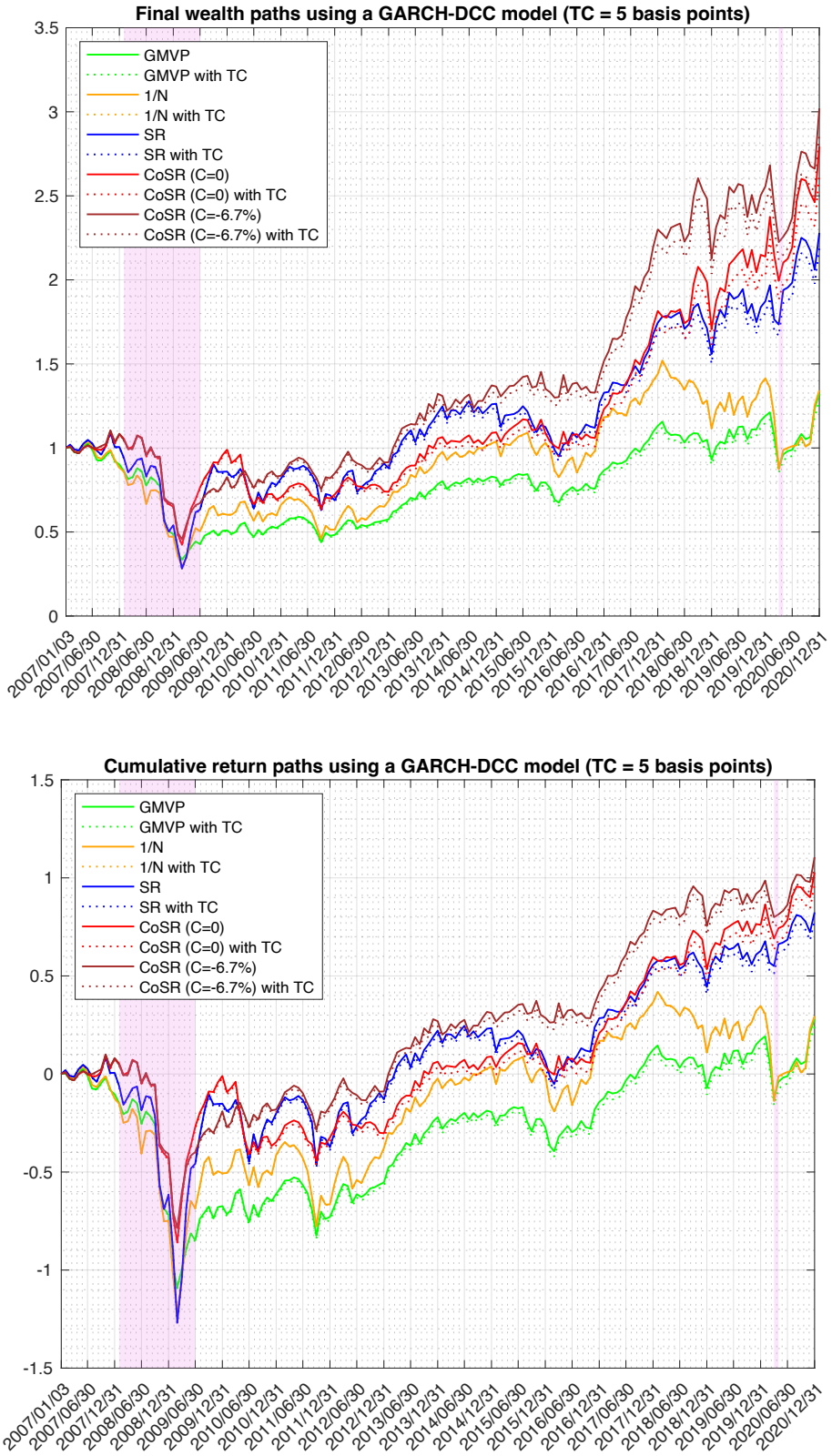



Figure 1.5: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies based on GARCH-DCC model with proportional transaction costs. The shaded areas correspond to NBER recession periods.

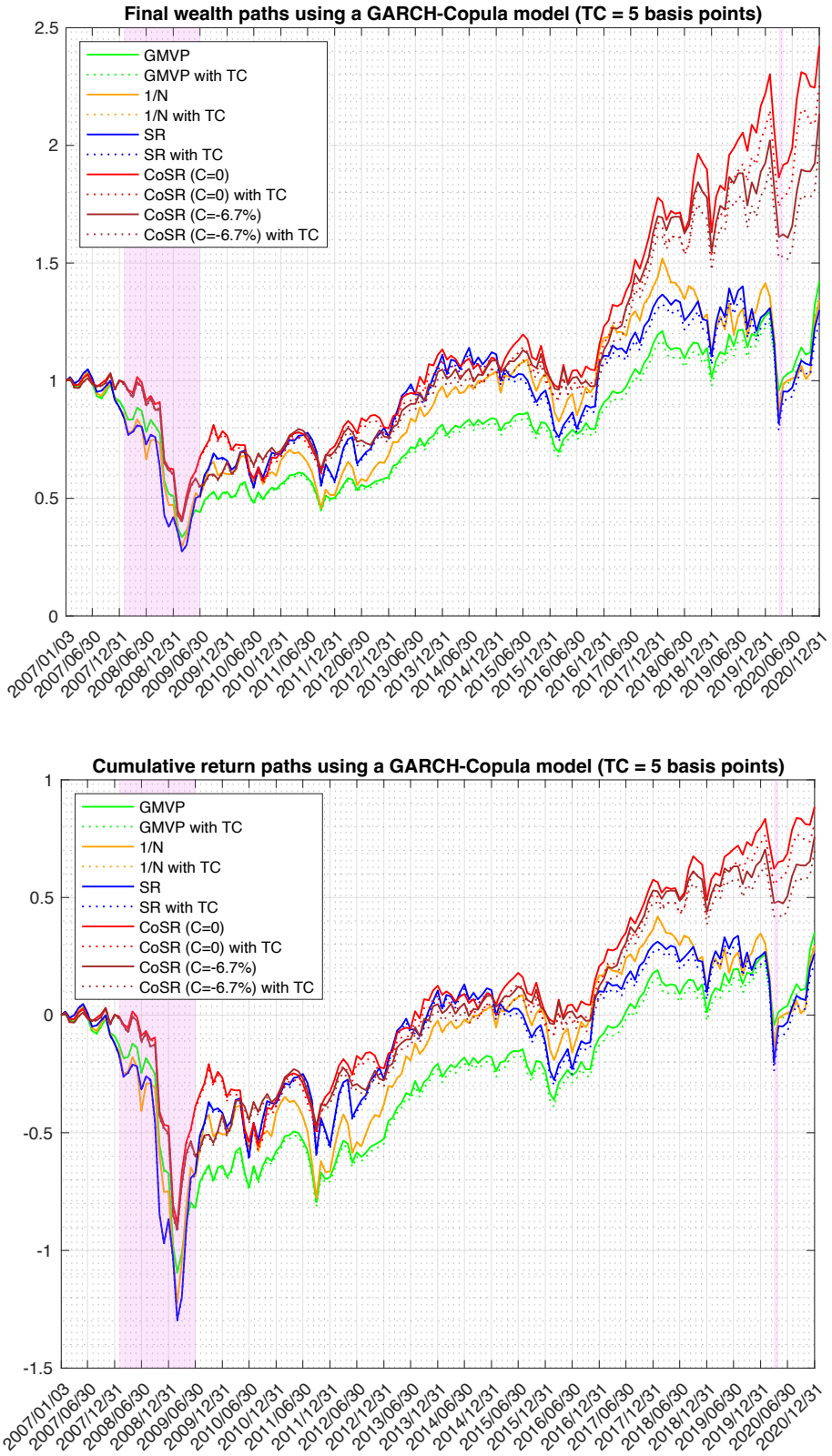



Figure 1.6: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies based on GARCH-Copula model with proportional transaction costs. The shaded areas correspond to NBER recession periods.

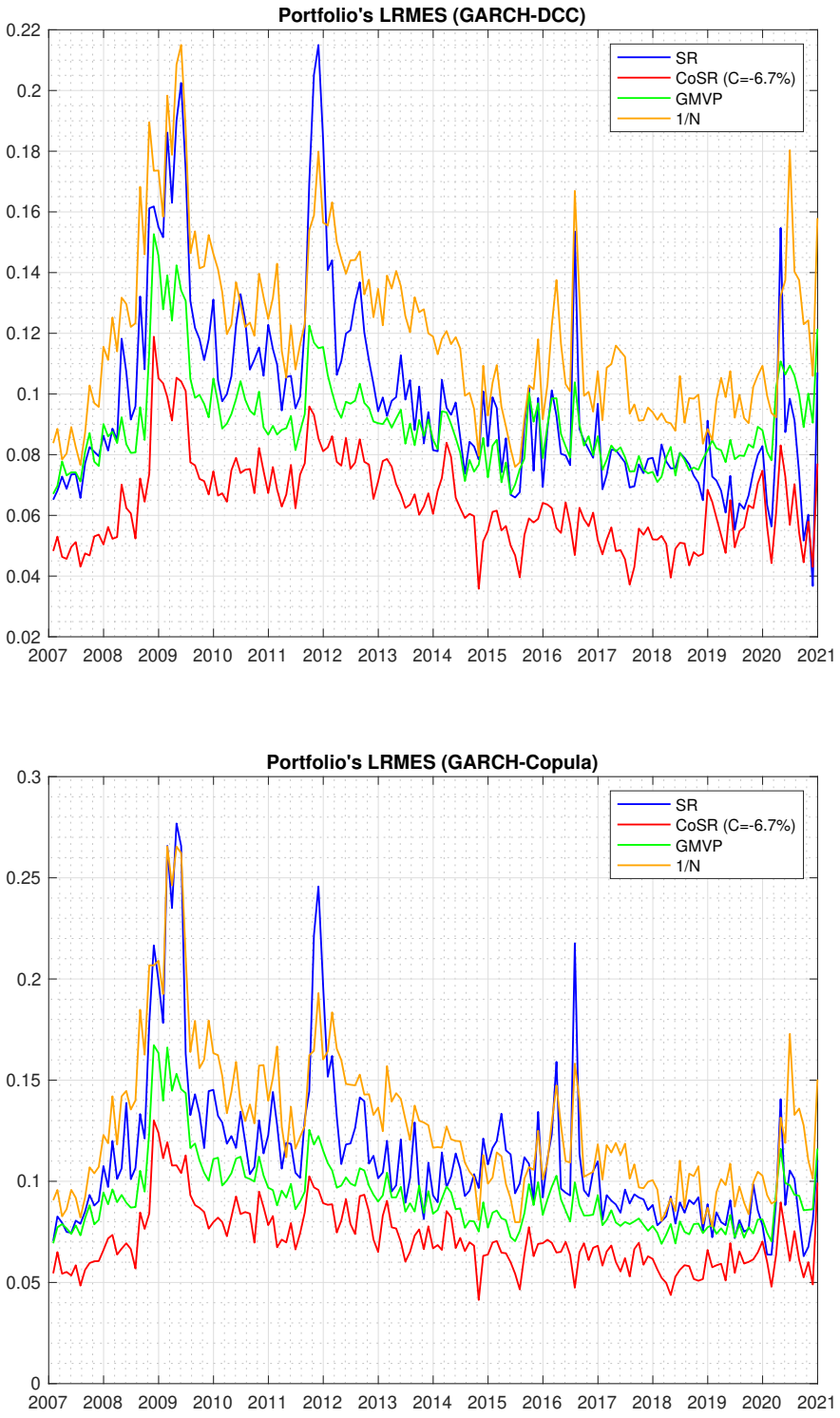



Figure 1.7: Portfolio's LRMES paths based on GARCH-DCC (top panel) and GARCH-Copula model (bottom panel).

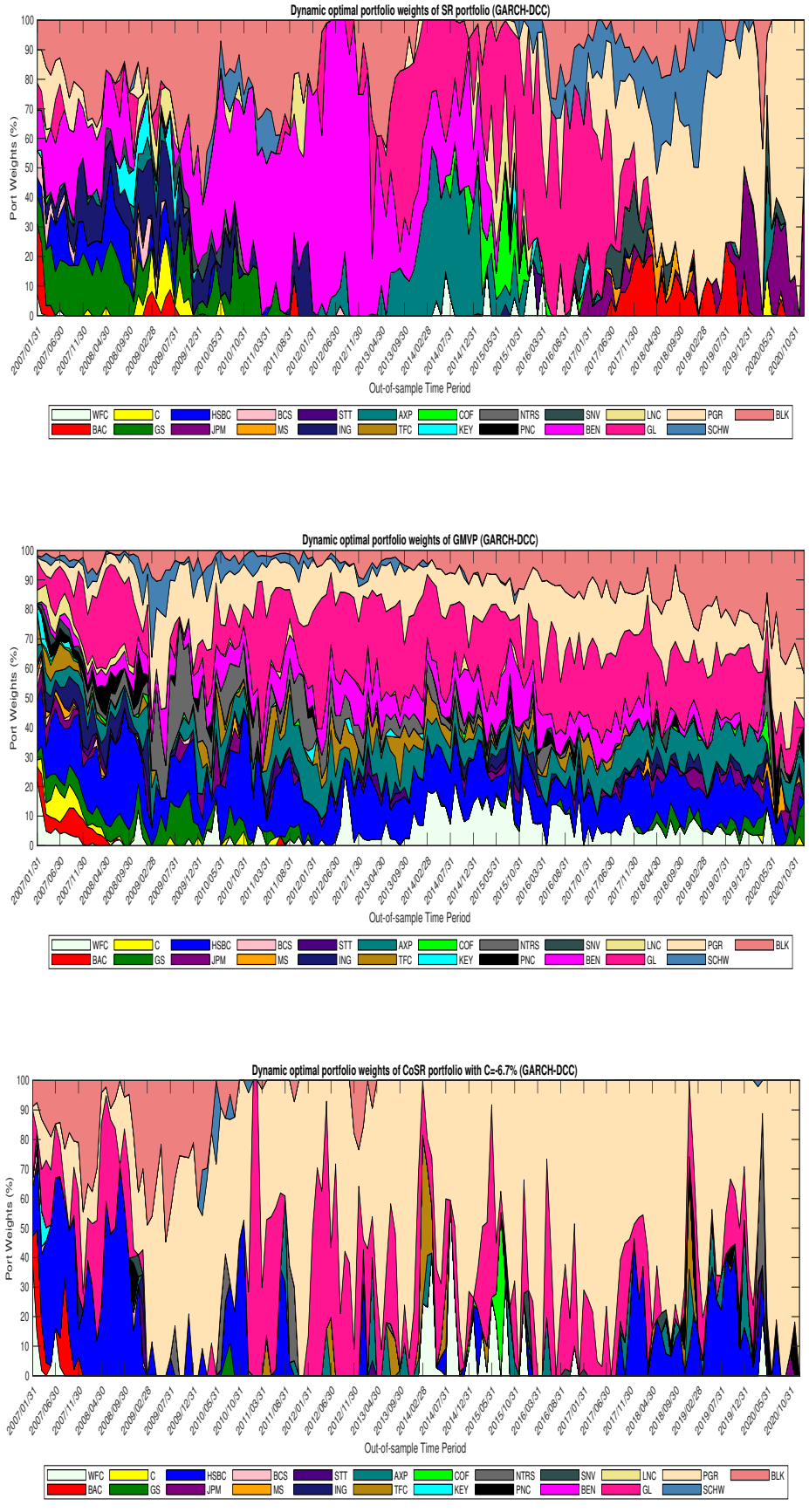



Figure 1.8: Time-varying portfolios' composition based on GARCH-DCC model. Top panel reports the portfolio weights under the SR strategy, middle panel under the GMVP strategy, and bottom panel under the CoSR strategy, respectively.

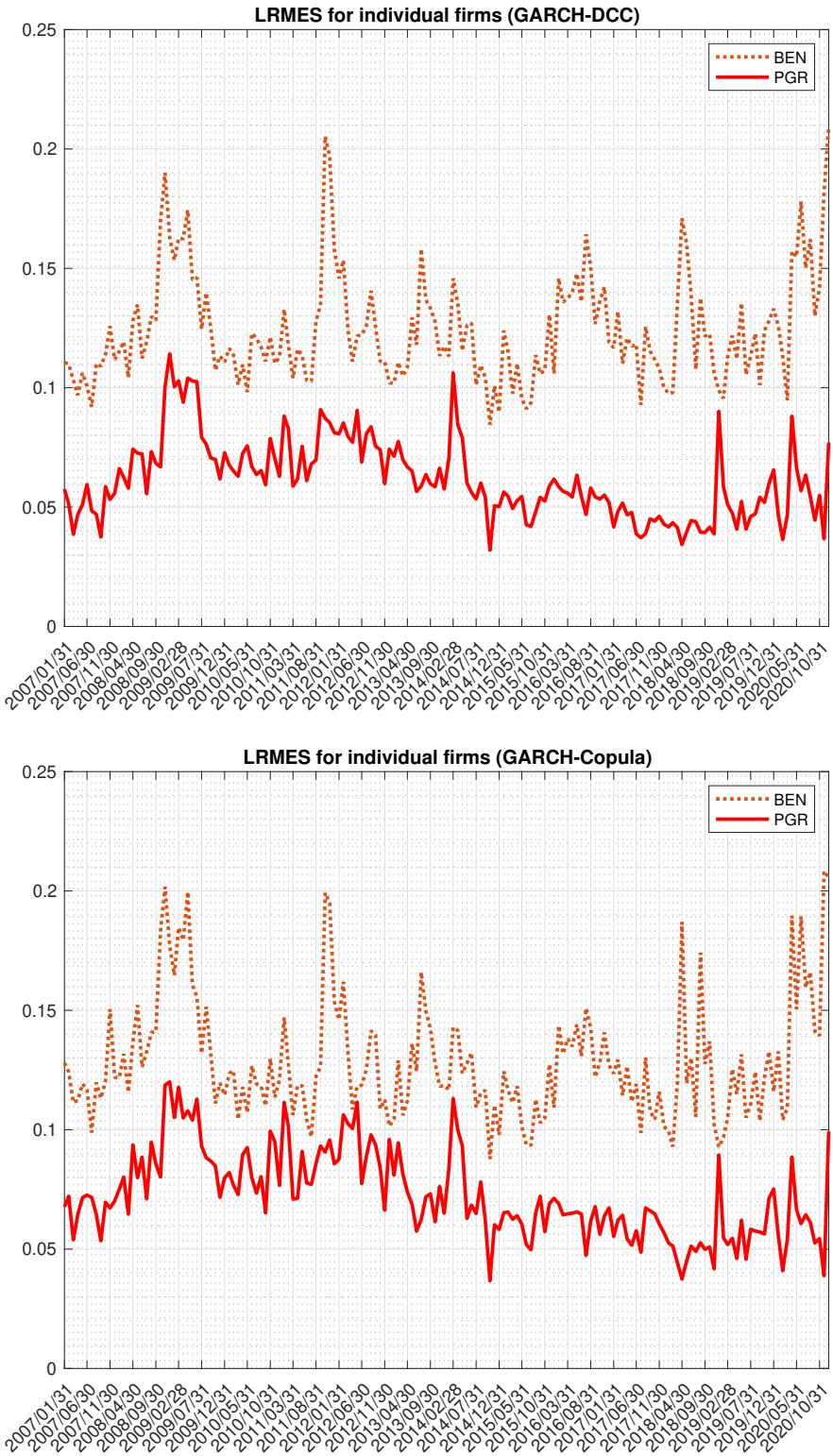



Figure 1.9: Comparison of individual firm's LRMES based on GARCH-DCC model (top panel) and GARCH-Copula model (bottom panel).

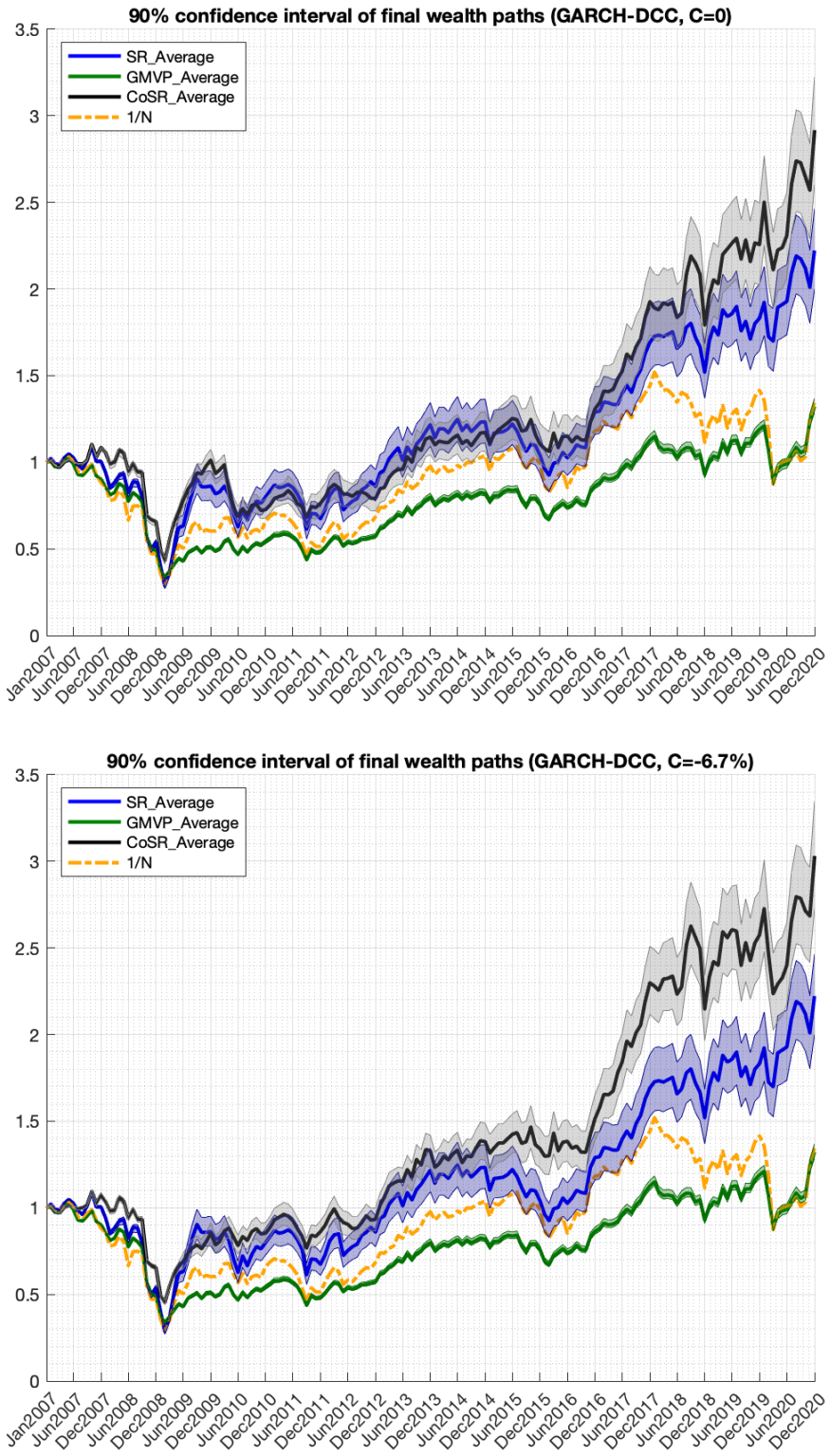



Figure 1.10: Comparison between different strategies accounting for estimation uncertainty using the GARCH-DCC model. Top panel considers a systemic event given by $C = 0$ and bottom panel considers a systemic event given by $C = -6.7\%$.

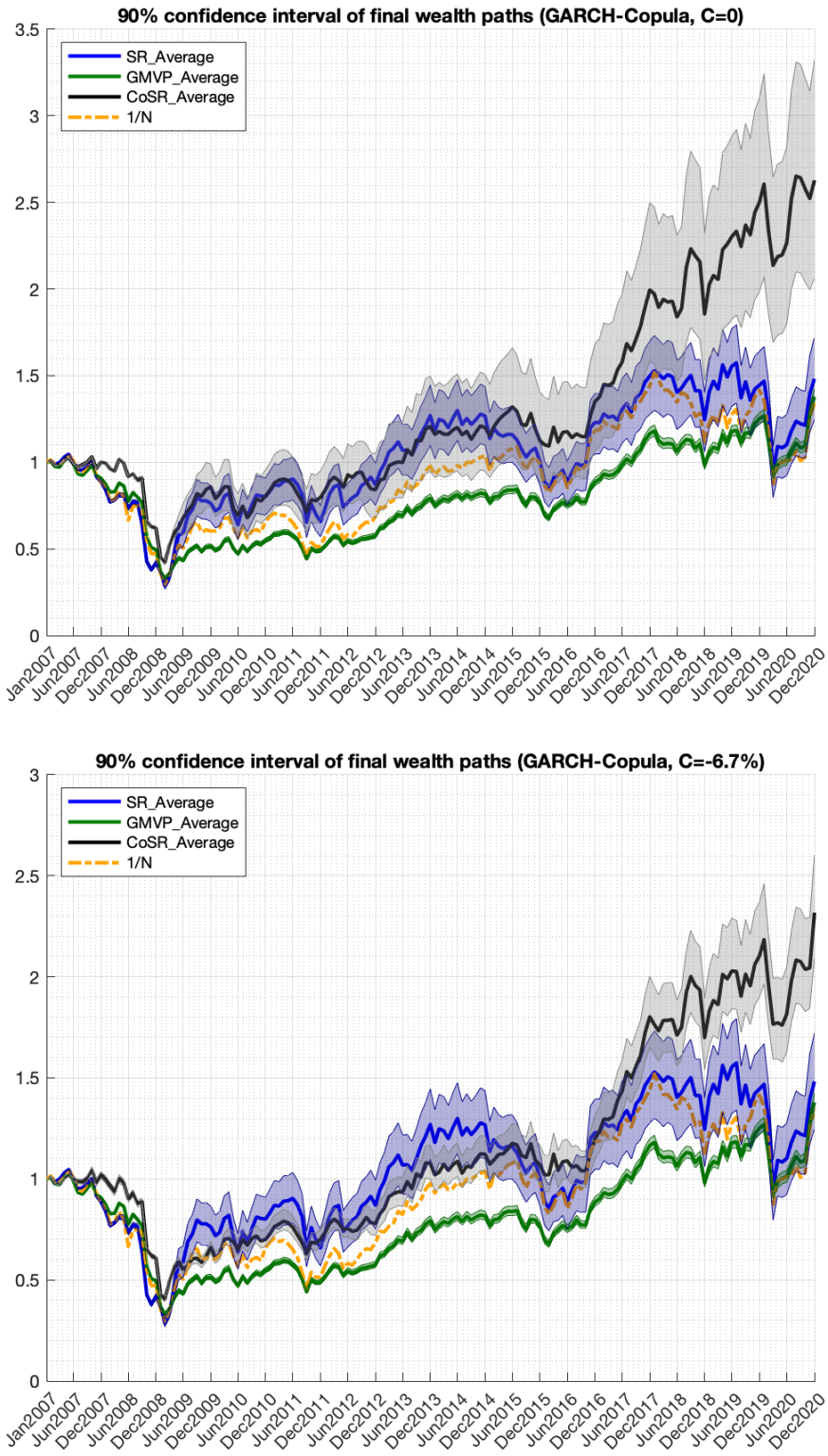



Figure 1.11: Comparison between different strategies accounting for estimation uncertainty using the GARCH-Copula model. Top panel considers a systemic event given by $C = 0$ and bottom panel considers a systemic event given by $C = -6.7\%$.

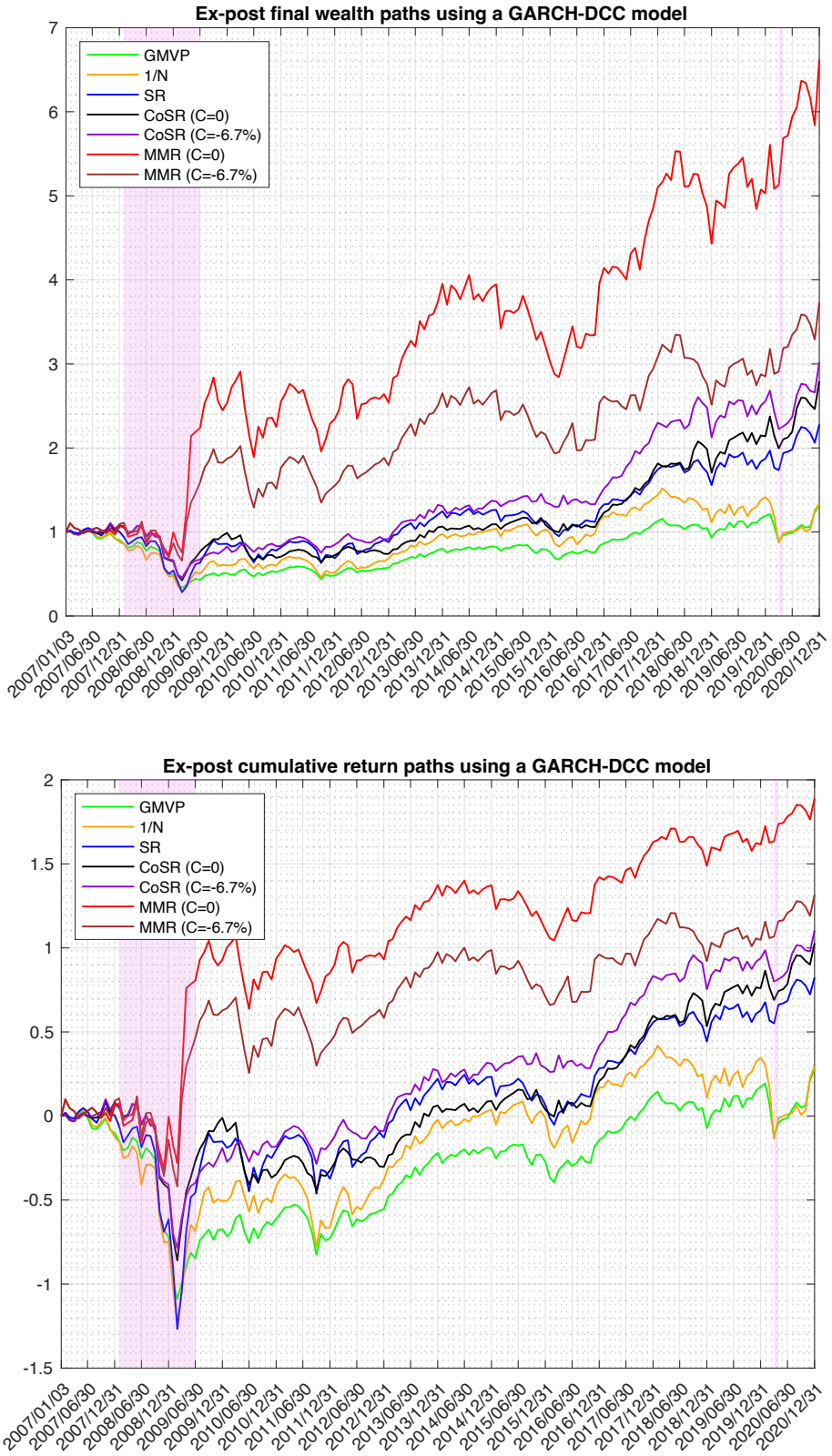



Figure 1.12: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies based on GARCH-DCC model ($S = 30,000$). The shaded areas correspond to NBER recession periods.

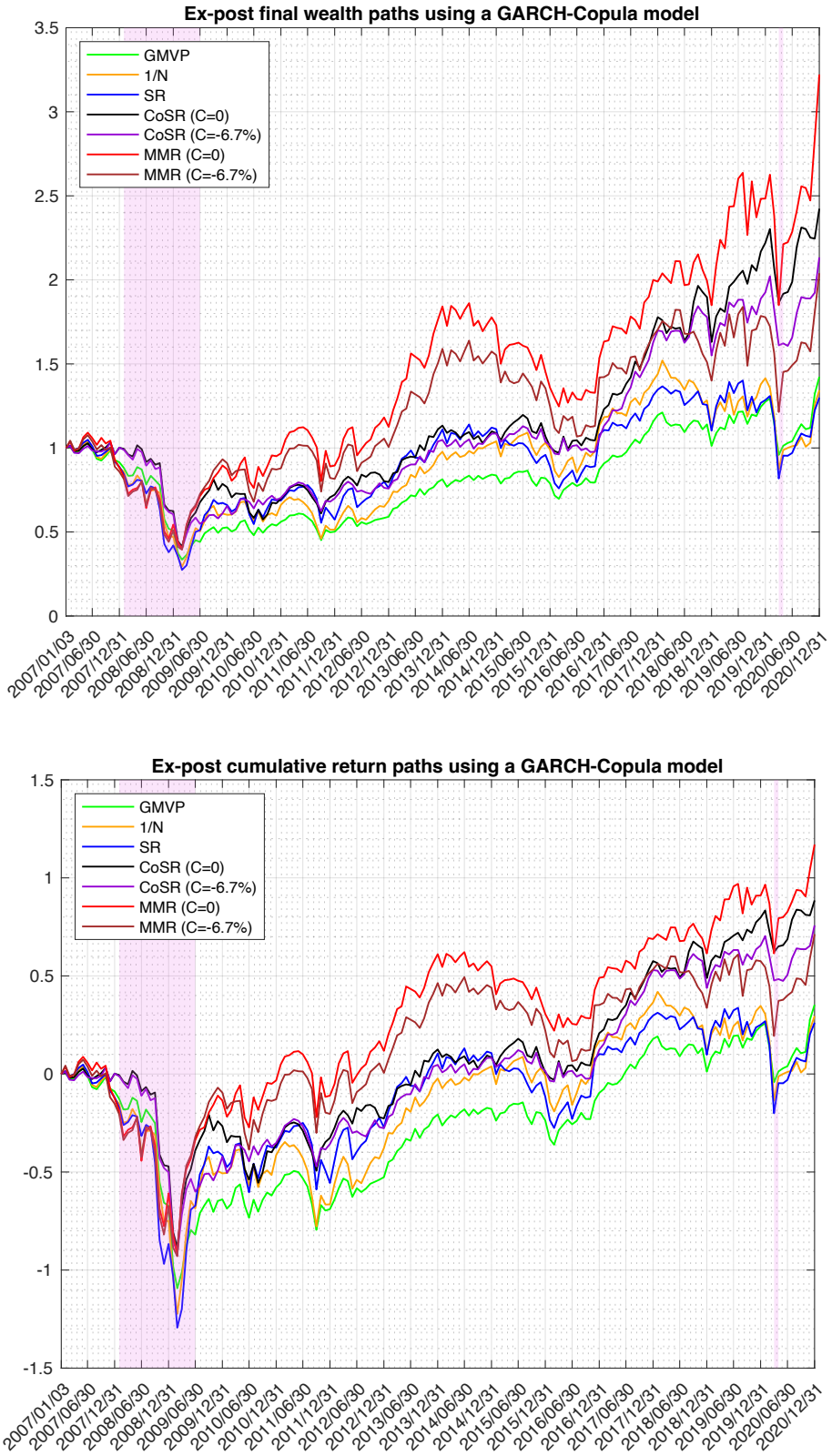



Figure 1.13: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies based on GARCH-Copula model ($S = 30,000$). The shaded areas correspond to NBER recession periods.

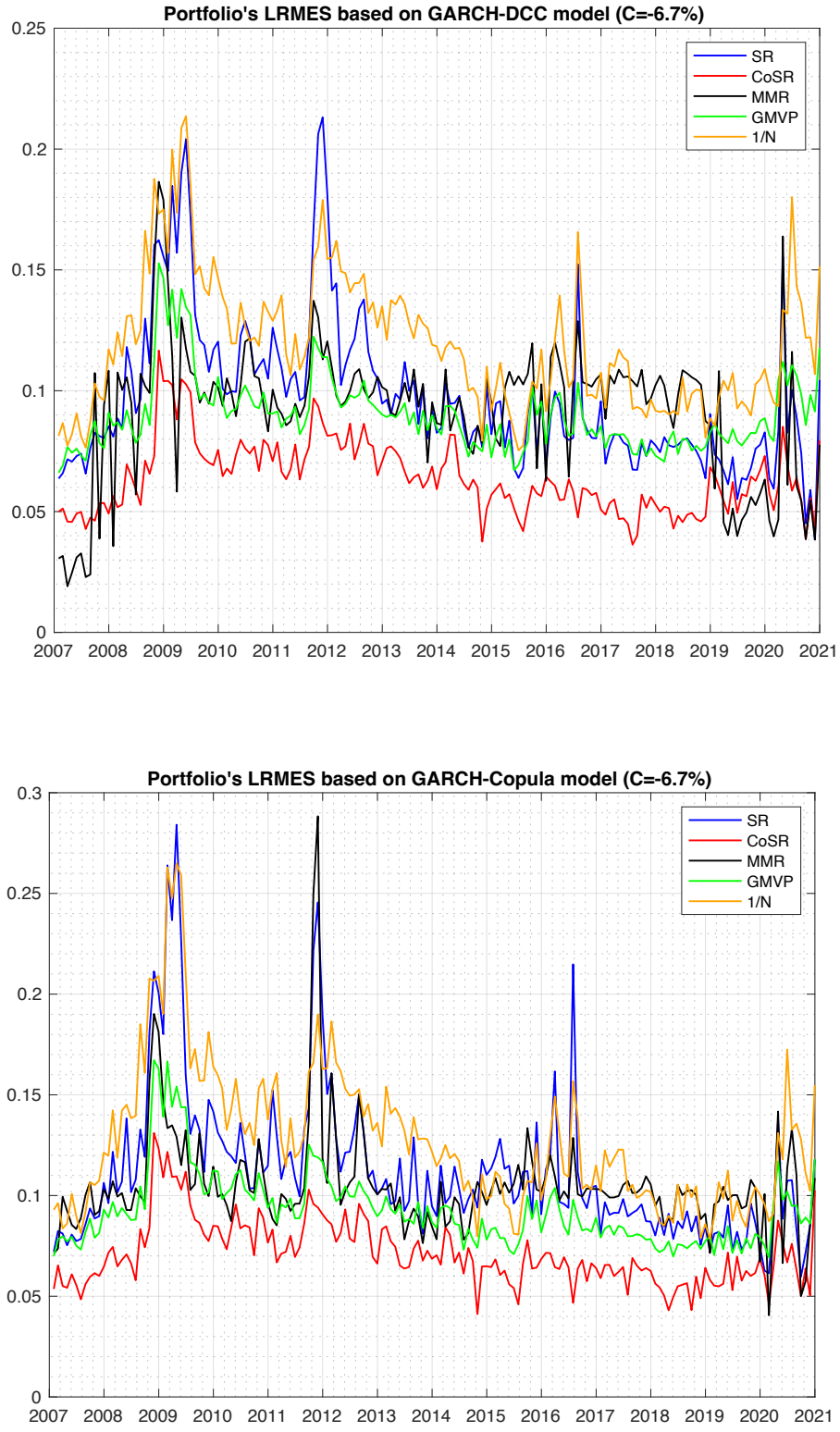



Figure 1.14: Portfolio's LRMES paths based on GARCH-DCC (top panel) and GARCH-Copula model (bottom panel).

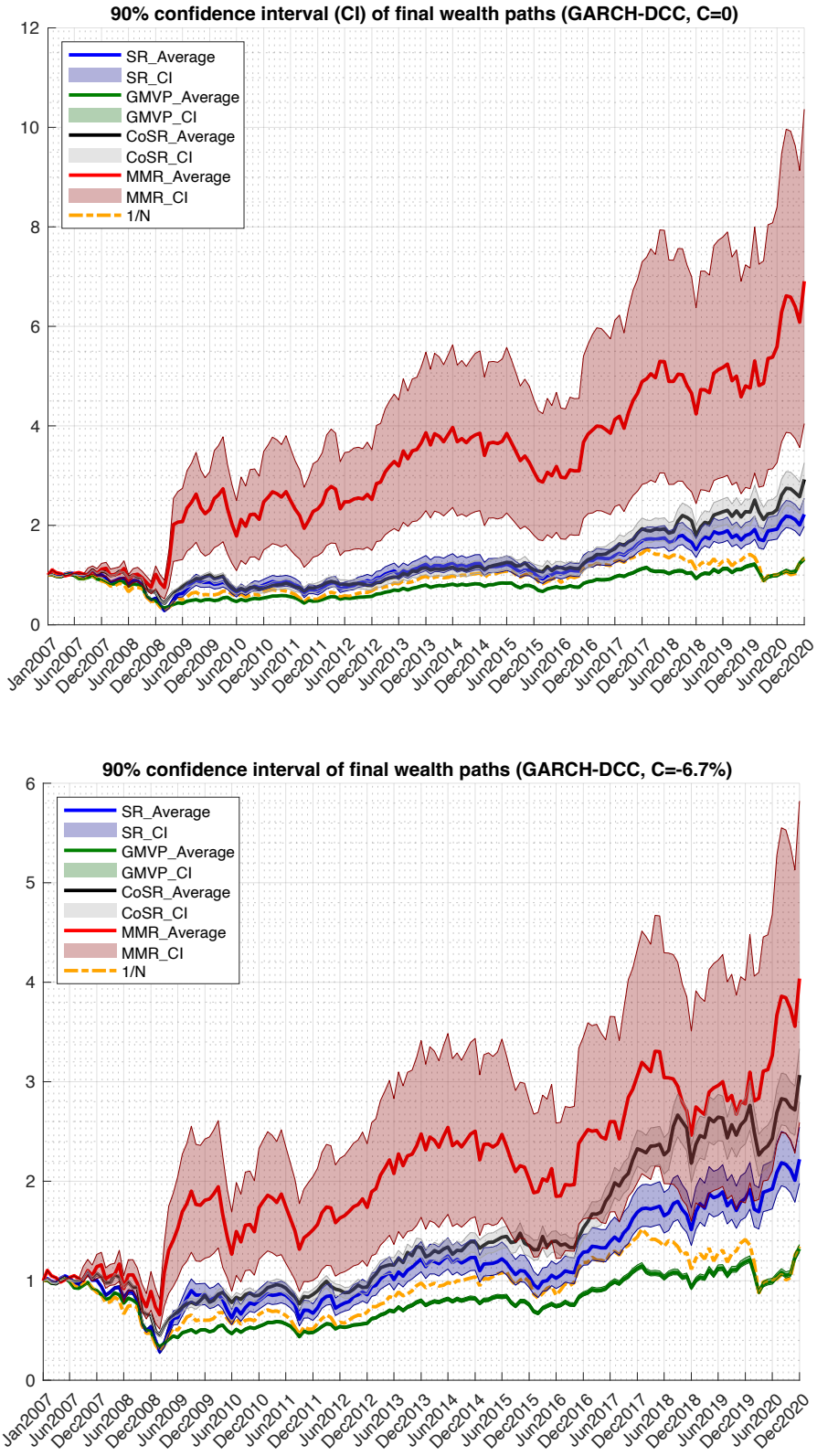



Figure 1.15: Comparison between different strategies accounting for estimation uncertainty using the GARCH-DCC model. Top panel considers a systemic event given by $C = 0$ and bottom panel considers a systemic event given by $C = -6.7\%$.

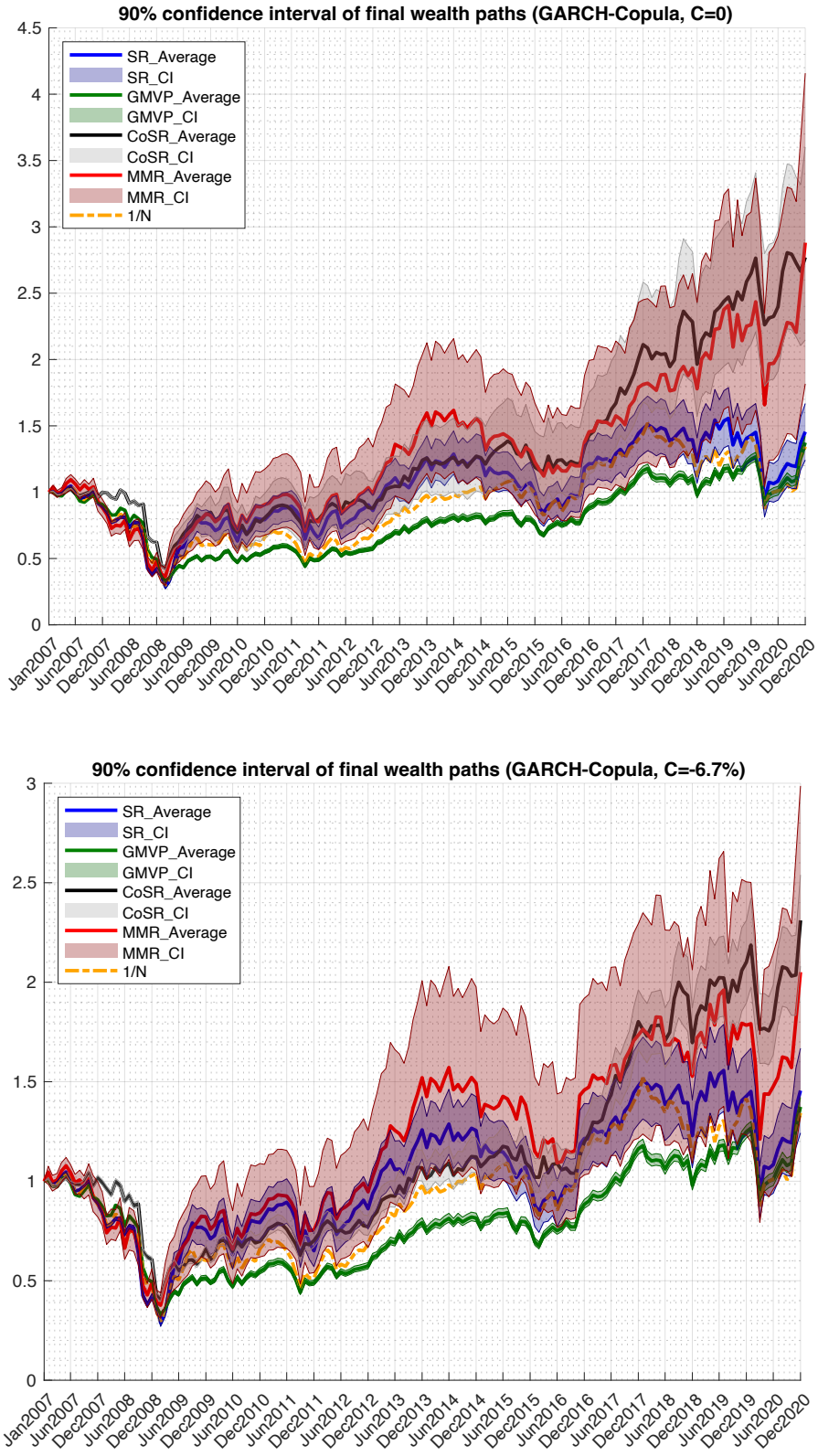



Figure 1.16: Comparison between different strategies accounting for estimation uncertainty using the GARCH-Copula model. Top panel considers a systemic event given by $C = 0$ and bottom panel considers a systemic event given by $C = -6.7\%$.

# Chapter 2

# Machine learning based portfolio selection under systemic risk

## 2.1 Introduction

In the first chapter, we proposed a conditional Sharpe ratio (CoSR) that takes into account systemic risk. The approach is made operational by embedding it in a multivariate dynamic setting using GARCH-DCC and GARCH-Copula models, and the backtesting results were shown to be promising for the proposed CoSR portfolio when compared to popular benchmark strategies. However, these advanced statistical methods fail to take advantage of the rich predictive information contained in firm- and macro-level predictors when generating return scenarios. This can be achieved by using machine learning (ML) techniques. Furthermore, ML can better explain the nonlinear relationships between the explicative and response variables when compared to traditional methods. Therefore, in this chapter, we further enhance the portfolio selection approach proposed in the first chapter by employing a distributional ML model that allows us to obtain more accurate probabilistic return forecasts in a high-dimensional setting, and thus a more robust estimator of CoSR measure.

Machine learning (ML) is a tool that can explicitly describe complex relationships and uncover patterns within high-dimensional datasets that might help improve forecast accuracy. In the empirical finance literature, Gu et al. (2020) and others narrowed the definition of ML down to a set of high-dimensional statistical prediction models, combined with optimization algorithms for parameter searching and regularization methods for overfitting mitigation. In this paper, we combine ML and a novel methodology to build optimal port-

folios that deal with systemic risk. The new approach enables us to incorporate information from high-dimensional datasets and accommodate market turbulences and, as a result of this, it is able to outperform many popular alternative portfolios.

Stock return predictability is of great importance to investors as it is a key ingredient for asset allocation, risk management, asset pricing, etc. Many studies try to explain cross-sectional stock returns using various predictors such as size, book-to-market, and momentum factor; see Harvey et al. (2016) and references therein. The increasing number of available factors might provide richer predictive information by incorporating big data into stock return modelling. However, as argued by Gu et al. (2020) and others, traditional prediction methods (e.g. linear models) are unable to fit complex patterns and tend to break down when the number of covariates is close to the number of observations or when the predictors are highly correlated. Thus, thanks to its ability to handle high-dimensional datasets and complex nonlinear relationships, ML is the tool we need to confront the challenge of improving prediction accuracy and consequently portfolio performance.

Since the ML techniques have shown promising superiority against traditional statistical methods in stock return prediction, many researchers have applied these models to portfolio optimization and generated satisfying results; see Zhang et al. (2020); Babiak and Baruník (2020); and Huang et al. (2021); among others. However, as far as we know, the existing literature has not yet explored the potential economic gains of combining ML-based probabilistic return forecasts with portfolio optimization. The applications in FinTech focus mostly on point forecasts of stock returns without accounting for any uncertainty. Moreover, so far the efficiency of portfolios constructed using ML techniques has been tested mainly for characteristic-sorted portfolios (e.g. long-short decile portfolios), which further motivates us to investigate whether the ML approach to probabilistic forecasting will help our investors when forming optimal portfolios.

Starting from the mean-variance paradigm of Markowitz (1952), the tradeoff between return and risk has become the focus of research on portfolio optimization. A general approach for building optimal portfolios consists of maximizing an ex-ante reward-risk performance measure to obtain the so-called *market portfolio*, which is based on diverse perceptions of reward and risk measures such as the well-known Sharpe ratio (Sharpe 1994); see Rachev et al. (2008) for a thorough review of various reward-risk ratios that have been proposed for portfolio optimization. Although the existing performance ratios have led to the development of many major theories and practices on optimal asset allocations,

by construction, they are unable to take into account systemic risk that might affect the portfolio's risk beyond the effect of individual assets' risks. In other words, these ratios only focus on measuring the performance of portfolio assets without accounting for systemic events happening in the market. Hence, investors need new portfolio strategies that can help them overcome systemic risk during their decision-making process. Systemic risk is defined as the risk of collapse of the whole financial system, as opposed to the risk associated with any individual entity of the system. It also refers to the risk imposed by poorly understood interlinkages and interdependencies among individual assets, where the failure of a single entity or cluster of entities can trigger the failure of more institutions within the same market, see Allen and Carletti (2013). The global financial crisis of 2007-2008 and subsequent crises (e.g. euro crisis and COVID-19 pandemic) provide ample evidence of the importance of containing systemic risk.

While the macroprudential literature has made substantial progress in terms of developing monitoring tools for assessing the underlying systemic risk in a financial system, portfolio selection literature has not evolved in parallel. Only a few studies examined the implications of systemic risk for investment decisions; see Capponi and Rubtsov (2022) and references therein. Recently, Lin et al. (2022) were interested in solving the tradeoff between reward and risk under stressed market conditions by introducing a conditional Sharpe ratio (CoSR), in which they incorporate the occurrence of systemic events into the performance measure. However, none of the above-mentioned papers utilizes ML techniques for predicting returns when building optimal portfolios. The present paper bridges this gap by merging the literature on portfolio selection under systemic risk with the one on cross-sectional return prediction using ML.

Theoretically, solving the portfolio optimization problem by maximizing a specific performance ratio requires knowing the true distribution of future portfolio returns. In practice, however, we might just need to estimate the reward and risk measures to solve the portfolio problem. The estimation of these measures can be done in two different ways that involve either using historical observations of return or simulated return scenarios. It has been argued that the optimal portfolios obtained based on the former approach are unlikely to beat the naive portfolio, which can be mainly attributed to their extreme weights over the out-of-sample period; see DeMiguel et al. (2009) among others. This might be caused by the estimation errors that are known to affect sample-based estimators and make the latter less effective when they are used as inputs in an optimization problem. Therefore, it

is important to find ways to robustify these portfolio optimization inputs.

Although extensive effort has already been devoted to alleviating the aforementioned problem by developing estimators that take into account estimation errors, in order to enhance the out-of-sample performance of the mean-variance portfolio relative to popular benchmarks; see for example Branger et al. (2019), few papers in the literature resort to promising ML tools. In this paper, we fill this gap by using a distributional ML approach to generate return scenarios that we use to estimate the reward and risk measures. Specifically, we formulate the portfolio selection problem as a three-stage supervised learning process that considers systemic risk when building optimal portfolios. We start by predicting quantiles of cross-sectional stock returns using the smooth pinball neural network (SPNN) model, based on which we estimate the conditional marginal distributions of returns on portfolio assets and the market. Thereafter, we apply t-copula to model the dependence between individual assets and the market, and generate scenarios for future returns. Lastly, we solve the portfolio optimization problem dynamically by maximizing CoSR based on the simulated return scenarios.

Furthermore, we perform a large-scale empirical study using nearly 600 US stocks with 37 years of history from 1985 to 2021. Our set of predictors includes 94 firm characteristics, 14 macroeconomic variables, and 74 industry dummies. For the returns of individual assets and the market, we calculate their monthly quantile forecasts using SPNN. Thereafter, on each month within our out-of-sample period, we solve the portfolio optimization problem dynamically using different objective functions. Specifically, we feed the CoSR optimizers with input parameters that we estimate using the return scenarios generated from a hybrid model that combines SPNN and copula. In addition, for comparison, we calculate sample-based tangency portfolio (SR), sample-based minimum variance portfolio (MVP), and equally weighted portfolio (1/N) as benchmark strategies. Finally, we calculate and report the out-of-sample portfolio performance of different strategies via a backtesting analysis. We also test the significance of the difference in Sharpe ratios between our approach and that of each benchmark portfolio.

Our paper contributes to the literature in two ways. First, we shed new light on the performance ratio-based portfolio selection using a distributional ML approach. This is done by incorporating SPNN-based probabilistic forecasts of stock returns into a conditional Sharpe ratio. The ability of ML methods to capture complex and nonlinear patterns that characterize big datasets results in more accurate forecasts of future return distributions,

and hence in more robust estimates that are then fed into the portfolio optimizer, which in turn enhances the portfolio performance relative to popular benchmark portfolios. Second, we further improve the portfolio selection process by explicitly incorporating the occurrence of systemic events into the construction of optimal portfolios, which leads to portfolios that are less likely to suffer great losses during market distress. After accounting for the conditional tail risk of portfolios, our backtesting results show that our proposed approach not only performs well on stressed scenarios but also performs steadily when the market is doing well.

The rest of the paper is organized as follows. Section 2 introduces the stock return quantile prediction via SPNN. Section 3 formulates the portfolio selection problem under systemic risk. In this same section, we discuss the method of probabilistic forecasting of returns, and illustrate how we model dependence through copula and generate return scenarios. Section 4 uses a high-dimensional dataset from the US market to conduct a large-scale empirical analysis in which we compare the out-of-sample portfolio performance of our proposed approach with those of several popular benchmark strategies. Section 5 concludes. Figures and tables that displayed in Appendix A and B, respectively.

## 2.2 Quantile regression neural network

We start by reviewing the traditional quantile regression (QR), which is a building block of Quantile Regression Neural Network (QRNN). We then introduce the mathematical formulation of QRNN and its advanced variant Smooth Pinball Neural Network (SPNN). Before we describe our quantile models, let us first set some notations. Using the terminology of the literature on neural networks, we denote by $\boldsymbol{R} = (R_1, ..., R_V)$ the $1 \times V$ vector of monthly returns for $V$ 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 $P \times V$ matrix of $P$ covariates across $V$ training samples, 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), but we will do so in Section 2.3.

### 2.2.1 Model specification

Initially proposed by Koenker and Bassett (1978), the quantile regression (QR) model estimates the relationship between predictors and a conditional quantile of the response

variable. Formally, the $\tau$-th conditional quantile of the predictand $R_v$ is given by

$$Q_{R_v}(\tau|\boldsymbol{X}_v) = \boldsymbol{X}_v^T\boldsymbol{\beta}(\tau), \quad \forall v \in \{1, ..., V\}, \quad \tau \in (0, 1), \tag{2.1}$$

where $\boldsymbol{\beta}(\tau) = [\beta_0(\tau), ..., \beta_P(\tau)]^T$ is the vector of the regression coefficients and can be estimated by solving the following optimization problem

$$\hat{\boldsymbol{\beta}}(\tau) = Arg \min_{\boldsymbol{\beta}(\tau)} \frac{1}{V} \sum_{v=1}^{V} \rho_\tau \left[R_v - \boldsymbol{X}_v^T\boldsymbol{\beta}(\tau)\right], \tag{2.2}$$

where the asymmetric loss function $\rho_\tau$ (known as pinball loss function) is defined as

$$\rho_\tau(u) = \begin{cases} \tau u & u \geq 0 \\ (\tau - 1)u & u < 0 \end{cases}. \tag{2.3}$$

The fitted conditional quantile is expressed as

$$\hat{Q}_{R_v}(\tau|\boldsymbol{X}_v) = \boldsymbol{X}_v^T\hat{\boldsymbol{\beta}}(\tau). \tag{2.4}$$

QR provides a more complete picture of the conditional distribution of $\boldsymbol{R}$ than conditional mean regression and does not make assumptions on the distribution of the target variable. Moreover, QR is robust to outliers and can thus be estimated more accurately than conventional moments regression. The QR model defined in (2.1) is, however, unable to capture possible nonlinear relationships between $\boldsymbol{R}$ and $\boldsymbol{X}$. To overcome this issue, Taylor (2000) originally introduced the quantile regression neural network (QRNN) that combines QR with ANN to depict the complex nonlinear relationships between predictors and the response variable without pre-specifying a functional form. Thus, instead of using a linear function, the conditional quantile is approximated by a neural network $f(\cdot)$ under the QRNN framework. Formally, the conditional $\tau$-th quantile of $R_v$ based on QRNN model with a single hidden layer can be formulated as

$$Q_{R_v}(\tau|\boldsymbol{X}_v) = f\left(\boldsymbol{X}_v, \boldsymbol{H}(\tau), \boldsymbol{O}(\tau)\right) = g_2\Big[\sum_{k=1}^{K} o_k(\tau) g_1\big(\sum_{j=1}^{P} h_{j,k}(\tau) x_j^v\big)\Big], \tag{2.5}$$

where $\boldsymbol{H}(\tau) = (h_{1,1}(\tau), ..., h_{P,K}(\tau))^T$ is the weight vector that links the input and hidden layer, $\boldsymbol{O}(\tau) = (o_1(\tau), ..., o_K(\tau))^T$ is the weight vector responsible for connecting the hid-

den and output layer, and $K$ is the number of hidden neurons. The activation functions $g_1(\cdot)$ and $g_2(\cdot)$ are generally specified as a sigmoid/rectifier function and a linear function, respectively. The set of parameters $\boldsymbol{\beta}(\tau) \equiv \{\boldsymbol{H}(\tau), \boldsymbol{O}(\tau)\}$ can be estimated by solving

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\boldsymbol{\beta}(\tau)}{\text{Argmin}}\, L(\tau) = \underset{\boldsymbol{\beta}(\tau)}{\text{Argmin}}\, \frac{1}{V}\sum_{v=1}^{V} \rho_\tau\Big[\big(R_v - f(\boldsymbol{X}_v, \boldsymbol{\beta}(\tau))\big)\Big], \tag{2.6}$$

and the fitted conditional quantiles are obtained as $\hat{\boldsymbol{Q}}_{\boldsymbol{R}}(\tau|\boldsymbol{X}) = f(\boldsymbol{X}, \hat{\boldsymbol{\beta}}(\tau))$. Figure 2.1 illustrates the architecture of a QRNN model with a single hidden layer.

### 2.2.2 Approximation of pinball loss function

The parameters of neural networks are typically determined through some gradient-based nonlinear optimization algorithms by which the gradients are calculated using the backpropagation algorithm, see Cannon (2011). The gradient of (2.6) can be computed analytically by updating backpropagation equations based on the least absolute error function, see Hanson and Burr (1988). However, the loss function $\rho_\tau$ is non-differentiable at the origin ($u = 0$), which thus requests a smooth approximation in order to apply gradient-based optimization methods. To smooth $\rho_\tau$, one can resort to the Huber norm introduced by Huber (2004), which is defined as:

$$h(u) = \begin{cases} \dfrac{1}{2}u^2 & |u| \le \varepsilon \\ \varepsilon(|u| - \dfrac{1}{2}\varepsilon) & \text{otherwise} \end{cases}, \tag{2.7}$$

where $\varepsilon$ is a given threshold magnitude, see Cannon (2018) and Xu et al. (2017) and references therein. The check function is approximated by

$$\rho_\tau^{(A)}(u) = |\tau - I_{\{u<0\}}|\, h(u), \tag{2.8}$$

where $I_{\{u<0\}}$ is an indicator function that values as one when $u < 0$ and zero otherwise. As $\varepsilon$ converges to zero, the approximate error function converges to the exact QR error function; see Xu et al. (2017). An alternative way to smooth the loss function was proposed by Zheng (2011), which smoothes $\rho_\tau$ using a logistic function, i.e., for $\tau \in (0, 1)$,

$$\rho_\tau^{(A)}(u) = \tau u + \alpha\, \ln\big(1 + \exp(-\frac{u}{\alpha})\big), \tag{2.9}$$

where $\alpha > 0$ is the smoothing parameter. As argued by Arends et al. (2020), the loss function in equation (2.9) combines Huber loss and pinball loss together. Zheng (2011) has shown that $\rho_\tau^{(A)}(u) = \rho_\tau(u)$ as $\alpha \to 0^+$ in the limit. For illustration, Figure 2.2 in Appendix B displays the pinball loss function (red curve) for $\tau = 0.5$, the Huber approximation (blue curve) for $\varepsilon = 10$, and the logistic approximation (black curve) for $\alpha = 0.2$, respectively.

Replacing the loss function in equation (2.6) by a smoothed $\rho_\tau^{(A)}$, we obtain the following updated objective function

$$L^{(A)}(\tau) = \frac{1}{V}\sum_{v=1}^{V} \rho_\tau^{(A)}\Big[\big(R_v - f(\boldsymbol{X}_v, \boldsymbol{\beta}(\tau))\big)\Big]. \tag{2.10}$$

We can minimize (2.10) using standard gradient-based optimization algorithms to obtain the estimate of $\boldsymbol{\beta}(\tau)$. Cannon (2011) implemented this optimization procedure in R using the quasi-Newton optimization algorithm for calculating the Huber loss, while Hatalis et al. (2019) applied the logistic loss in Python based on the TensorFlow Platform. We adopt the logistic loss (2.9) in our empirical analysis.[1]

### 2.2.3 Smooth pinball neural network

To further enhance the performance of estimating quantiles, Xu et al. (2017) extended the original QRNN model to composite quantile regression neural network (CQRNN), which can be used to estimate multiple conditional quantiles (for different values of $\tau$) simultaneously and efficiently. CQRNN shares the same goal as the one of linear composite quantile regression (CQR) developed by Zou and Yuan (2008), namely combining the strength across multiple quantile regressions to better capture the complex nonlinear relationships between the predictors and the predictand (Cannon 2018). CQRNN is similar to QRNN by structure, where the difference lies in the objective function, which is now summed over $M$ values of $\tau$:

$$L_C^{(A)} = \frac{1}{M}\sum_{m=1}^{M} L^{(A)}(\tau_m), \tag{2.11}$$

where $\tau$ is equally spaced as $\tau_m = \frac{m}{M+1}$ for $m \in \{1, \cdots, M\}$. The expression in (2.11) is a composite version of the objective function in equation (2.10) since it evaluates multiple regression quantiles synthetically. CQRNN is a flexible model not only because it allows

[1] We have also tried for the Huber loss and the backtesting results are similar to those of using logistic loss.

us to uncover complex nonlinear patterns among variables through the properties of ANN, but also because it helps improve the estimation efficiency and prediction accuracy 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 property that a cumulative distribution function (CDF) is a monotonically increasing function. As stated by Ouali et al. (2016), quantile crossing is a serious modelling problem that may result in an invalid predictive distribution of the predictand. Similarly, Bang et al. (2016) argued that this problem reduces the estimation accuracy of regression quantiles and can cause trouble to the subsequent analysis and interpretation of the model. In order to mitigate this issue, Cannon (2018) developed a monotonic CQRNN (MCQRNN) model that imposes partial monotonicity constraints on the neural network weights and stacks covariates into an input matrix. MCQRNN imposes monotonicity constraints on a standard Multi-Layer Perceptron (MLP) and then it 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.

Recently, Hatalis et al. (2019) proposed an efficient alternative to MCQRNN namely smooth pinball neural network (SPNN) that introduces a set of constraints into the CQRNN framework. To prevent quantile crossing, the constraint $Q_{R_v}(\tau_1|\boldsymbol{X}_v) \leq \cdots \leq Q_{R_v}(\tau_M|\boldsymbol{X}_v)$, $\forall v$, needs to be satisfied. However, it is hard to solve the optimization problem via gradient-based methods with such constraints. To fix this issue, Hatalis et al. (2019) suggested adding a penalty term to the objective function (2.11), where the penalty term $p$ is defined as

$$p = c\frac{1}{MV}\sum_{m=1}^{M}\sum_{v=1}^{V}\Big[\max\Big(0, \epsilon - \big(\hat{Q}_{R_v}(\tau_m|\boldsymbol{X}_v) - \hat{Q}_{R_v}(\tau_{m-1}|\boldsymbol{X}_v)\big)\Big)\Big]^2, \tag{2.12}$$

where $\hat{Q}_{R_v}(\tau_0|\boldsymbol{X}_v)$ is initialized to zero, $\epsilon$ denotes the minimum difference value between two neighbouring quantiles, and $c$ is the penalty parameter. The objective function of SPNN is now given by

$$L_S = L_C^{(A)} + p + \lambda||\boldsymbol{\beta}||_1, \tag{2.13}$$

where $\boldsymbol{\beta} \equiv \{\boldsymbol{H}, \boldsymbol{O}\} = \{\boldsymbol{H}(\tau_m), \boldsymbol{O}(\tau_m)\}_{m=1,\ldots,M}$ represents the composite parameters of neural network (i.e. parameters across all values of $\tau$). Note that the $l_1$ norm $||\cdot||_1$ is applied in (2.13) to mitigate the overfitting problem, where $\lambda$ denotes the regularization

parameter. The training of SPNN can be conducted using standard gradient-based optimization algorithms. In our paper, we adopt SPNN for completing prediction tasks due to its virtues of simultaneously estimating multiple quantiles and preventing quantile crossing.

## 2.3 Portfolio selection under systemic risk

In this section, we first review the CoSR-based portfolio selection problem. Then we discuss the simulation scheme for generating multivariate return scenarios, which is done by combining the predicted conditional marginal return densities from SPNN with a fitted t-copula. Lastly, we show how to calculate the CoSR estimator based on simulated returns.

### 2.3.1 Portfolio selection problem

We consider an economy with $N$ risky assets. Hereafter, we present the portfolio allocation problem of an investor that wishes to select the weights of $N$ assets by maximizing an ex-ante CoSR measure following Lin et al. (2022). Before we describe our portfolio problem, let us first set some notations. We denote by $\boldsymbol{R}_t = (R_{1,t}, ..., R_{N,t})^T$ the vector of monthly returns over month $t$, $R_{m,t}$ the market return over month $t$, and $\boldsymbol{W}_t = (\omega_{1,t}, ..., \omega_{N,t})^T$ the vector of portfolio weights held over month $t+1$. The portfolio return is given by $R_{p,t+1} = \boldsymbol{W}_t^T \boldsymbol{R}_{t+1}$. $\mathbf{0}$ and $\mathbf{1}$ denote the column vectors 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

$$\boldsymbol{W}_t^* = \arg\max_{\boldsymbol{W}_t} \rho_t(R_{p,t+1}), \quad \text{s.t. } \mathbf{1}^T \boldsymbol{W}_t = 1, \tag{2.14}$$

where the different candidates of $\rho(\cdot)$ result in different optimal portfolios. As we argued in previous sections, we are interested in building portfolios that take into account systemic risk. For this reason, we consider the performance measure CoSR that is defined as a ratio of a conditional reward measure over a conditional risk measure

$$\text{CoSR}_t(R_{p,t+1}) = \frac{\text{CoER}_t(R_{p,t+1})}{\text{CoSD}_t(R_{p,t+1})}, \tag{2.15}$$

with the conditional reward CoER defined as

$$\text{CoER}_t(R_{p,t+1}) = E_t(R_{p,t+1} - R_{m,t+1}|\text{SE}_{t+1}) = \boldsymbol{W}_t^T \boldsymbol{\mu}_{t|\text{SE}} - \mu_{m,t|\text{SE}}, \tag{2.16}$$

where $\text{SE}_{t+1} = \{R_{m,t+1} < C\}$ denotes a systemic event (SE) where the market return goes below a certain threshold $C$ over next month, $\boldsymbol{\mu}_{t|\text{SE}} = E_t(\boldsymbol{R}_{t+1}|\text{SE}_{t+1})$ is the vector of conditional expected returns on individual assets, and $\mu_{m,t|\text{SE}} = E_t(R_{m,t+1}|\text{SE}_{t+1})$ is the conditional expected market return. Analogously, the conditional risk measure CoSD is defined as the conditional second moment of the portfolio's excess return, that is:

$$\begin{aligned} \text{CoSD}_t(R_{p,t+1}) =& \big[Var_t(R_{p,t+1} - R_{m,t+1}|\text{SE}_{t+1})\big]^{1/2} \\ =& \left(\boldsymbol{W}_t^T \boldsymbol{\Sigma}_{t|\text{SE}} \boldsymbol{W}_t + \sigma^2_{m,t|\text{SE}} - 2\boldsymbol{W}_t^T \boldsymbol{\sigma}_{t|\text{SE}}\right)^{1/2}, \end{aligned} \tag{2.17}$$

where $\boldsymbol{\Sigma}_{t|\text{SE}} = Var_t(\boldsymbol{R}_{t+1}|\text{SE}_{t+1})$ denotes the conditional covariance matrix of asset returns, $\sigma^2_{m,t|\text{SE}} = Var_t(R_{m,t+1}|\text{SE}_{t+1})$ denotes the conditional variance of market return, and $\boldsymbol{\sigma}_{t|\text{SE}} = cov_t(\boldsymbol{R}_{t+1}, R_{m,t+1}|\text{SE}_{t+1})$ is the vector of conditional covariances between individual assets and the market portfolio. The portfolio selection problem under CoSR is given by

$$\boldsymbol{W}_t^* = \arg\max_{\boldsymbol{W}_t} \left\{\text{CoSR}_t(R_{p,t+1})\right\}, \quad \text{s.t. } \mathbf{1}^T \boldsymbol{W}_t = 1. \tag{2.18}$$

As pointed out by Lin et al. (2022), the optimization problem in (2.18) can be solved analytically under the absence of short-selling constraints ($\boldsymbol{W} \geq \mathbf{0}$). However, it is often the case that we want 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 in our later exercise. Unfortunately, under short-selling restrictions, the optimization problem in (2.18) cannot be solved analytically, and thus a numerical procedure must be employed.

As for benchmark portfolios, we consider the unconditional Sharpe ratio and the negative of portfolio variance as alternative objective functions for $\rho$ in (2.14) under short selling restrictions, where the resulting optimal portfolios are denoted by SR and MVP, respectively. In addition, DeMiguel et al. (2009) argued that the naive portfolio (1/N) should

be taken as a competitive benchmark to evaluate the performance of more sophisticated strategies. Hence we add it as a benchmark as well.

### 2.3.2 Simulation of return scenarios

Although CoSR has no closed-form expression in dynamic settings when non short-selling constraint is imposed, we can still use a Monte-Carlo simulation-based procedure to implement our ML and systemic risk-based portfolio. The dynamic SPNN-based CoSR can be estimated using its empirical analogue that we can calculate from simulated returns over the subset of simulated crisis 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.

#### 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, which is available at the end of 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 used to estimate the conditional density $p_{j,t} = p(R_{j,t+1}|\boldsymbol{X}_{j,t})$ following Cannon (2011). 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 $\tau_m$ and $\tau_{m+1}$ are such that $\hat{q}_{j,t+1}(\tau_m) \leq R_{j,t+1} < \hat{q}_{j,t+1}(\tau_{m+1})$, then

$$\hat{p}_{j,t} = \frac{\tau_{m+1} - \tau_m}{\hat{q}_{j,t+1}(\tau_{m+1}) - \hat{q}_{j,t+1}(\tau_m)}. \tag{2.19}$$

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

$$\hat{p}_{j,t} = z_1 \exp\Big(-\frac{|R_{j,t+1} - \hat{q}_{j,t+1}(\tau_1)|}{e_1}\Big), \tag{2.20}$$

  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

$$\hat{p}_{j,t} = z_M \exp\Big(-\frac{|R_{j,t+1} - \hat{q}_{j,t+1}(\tau_M)|}{e_M}\Big), \tag{2.21}$$

  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 above 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).

### Dependence modelling and scenario generation

Once the predictive marginal return distributions for individual assets and the market are obtained, next is to model joint return distribution via copula function. 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, in our case the multivariate distribution function of individual firm and market monthly returns, can be decomposed into univariate margins and a certain copula function, that is

$$F_{R_1,...,R_{N+1}}(u_1, ..., u_{N+1}) = C\left(F_{R_1}(u_1), ..., F_{R_{N+1}}(u_{N+1})\right), \tag{2.22}$$

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 use t-copula to model the dependence among monthly returns. The t-copula function is given by

$$C_{\nu,\boldsymbol{\mathcal{P}}}(u_1, ..., u_{N+1}) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_{N+1})} \frac{\Gamma(\frac{\nu+N+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^{N+1}|\boldsymbol{\mathcal{P}}|}} \Big(1 + \frac{x'\boldsymbol{\mathcal{P}}^{-1}x}{\nu}\Big)^{-\frac{\nu+N+1}{2}} dx, \tag{2.23}$$

where $\Gamma$ is the Gamma function, $\boldsymbol{\mathcal{P}}$ is a correlation matrix, and $\nu$ represents the degrees of freedom both for margins and copula function. We now generate future 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 CDF, say $\hat{F}_{\nu_{j,t}}$, of return series $\{R_{j,t}\}$ using a univariate t-location-scale distribution, 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 method of moment to estimate the degrees of freedom $\nu$ 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.

### 2.3.3 CoSR estimation

To estimate the performance measure CoSR based on simulated returns, we first estimate the elements of the vector of conditional expected returns on individual assets $\mu_{t|\text{SE}}$ using the average of the simulated arithmetic asset returns over one-month ahead period, that is

$$\hat{\mu}_{i,t|\text{SE}} = \frac{\sum_{s=1}^{S} R_{i,t+1}^{(s)} I\{R_{m,t+1}^{(s)} < C\}}{\#\text{SE}}, \tag{2.24}$$

where $S$ is the number of Monte Carlo simulations and $\#\text{SE} = \sum_{s=1}^{S} I\{R_{m,t+1}^{(s)} < C\}$ is the number of scenarios out of $S$ that represent a market distress. For each asset in the portfolio, the filtered mean vector (average of one-period ahead return conditional on a market distress episode) is given by $\hat{\boldsymbol{\mu}}_{t|\text{SE}} = (\hat{\mu}_{1,t|\text{SE}}, ..., \hat{\mu}_{N,t|\text{SE}})^T$. Similarly, the conditional expected market return $\mu_{m,t|\text{SE}}$ can be estimated as

$$\hat{\mu}_{m,t|\text{SE}} = \frac{\sum_{s=1}^{S} R_{m,t+1}^{(s)} I\{R_{m,t+1}^{(s)} < C\}}{\#\text{SE}}. \tag{2.25}$$

Thus, the estimator of CoER can be written as

$$\hat{\text{CoER}}_t = \boldsymbol{W}_t^T \hat{\boldsymbol{\mu}}_{t|\text{SE}} - \hat{\mu}_{m,t|\text{SE}}, \tag{2.26}$$

where $\boldsymbol{W}_t$ denotes the vector of portfolio weights that is known at month $t$. As for the CoSD, we first estimate the conditional covariance matrix of the vector of asset returns

$\boldsymbol{\Sigma}_{t|\text{SE}}$ using the Monte Carlo sample counterpart, with element $(i, j)$ defined as

$$\hat{\Sigma}_{t(i,j)|\text{SE}} = \frac{\sum_{s=1}^{S}\left(R_{i,t+1}^{(s)} - \hat{\mu}_{i,t}\right)\left(R_{j,t+1}^{(s)} - \hat{\mu}_{j,t}\right) I\{R_{m,t+1}^{(s)} < C\}}{\#\text{SE} - 1}, \tag{2.27}$$

for $i, j = 1, \ldots, N$. We then estimate the conditional variance of market return $\sigma_{m,t|\text{SE}}^2$ as

$$\hat{\sigma}_{m,t|\text{SE}}^2 = \frac{\sum_{s=1}^{S}\left(R_{m,t+1}^{(s)} - \hat{\mu}_{m,t}\right)^2 I\{R_{m,t+1}^{(s)} < C\}}{\#\text{SE} - 1}. \tag{2.28}$$

Analogously, for each asset $i$, an estimator of the conditional covariance between asset's $i$ and market returns $\sigma_{im,t|\text{SE}}$ is given by

$$\hat{\sigma}_{im,t|\text{SE}} = \frac{\sum_{s=1}^{S}\left(R_{i,t+1}^{(s)} - \hat{\mu}_{i,t}\right)\left(R_{m,t+1}^{(s)} - \hat{\mu}_{m,t}\right) I\{R_{m,t+1}^{(s)} < C\}}{\#\text{SE} - 1}, \tag{2.29}$$

thus the estimator of the vector of conditional covariances between individual assets and the market portfolio is $\hat{\boldsymbol{\sigma}}_{t|\text{SE}} = (\hat{\sigma}_{1m,t}, ..., \hat{\sigma}_{Nm,t})^T$. Combining the above estimators, we obtain the following estimator of CoSD at month $t$:

$$\widehat{\text{CoSD}}_t = \left(\boldsymbol{W}_t^T \hat{\boldsymbol{\Sigma}}_{t|\text{SE}} \boldsymbol{W}_t + \hat{\sigma}_{m,t|\text{SE}}^2 - 2\boldsymbol{W}_t^T \hat{\boldsymbol{\sigma}}_{t|\text{SE}}\right)^{1/2}. \tag{2.30}$$

## 2.4 Empirical analysis

### 2.4.1 Data

Our empirical analysis is conducted using monthly cross-sectional US market data spanning from January 1985 to December 2021, for a period of 37 years. In this section, we first provide details of the predictor set and then discuss the choice of portfolio assets.

#### Description of predictors

We use the 94 monthly stock-level explanatory variables considered in Gu et al. (2020).[2] The corresponding variable selection procedure was implemented by Green et al. (2013). We manually matched this dataset with monthly stock returns obtained from the CRSP database on the WRDS website. The equities presented in this original dataset are from

[2] We manually computed the value-weighted average of characteristics for the S&P 500 market index using the 500 highest market cap companies. The correlation coefficient between S&P 500 return and the constructed market return is beyond 0.99.

listed firms in NASDAQ, AMEX, and NYSE ranging from 1965 to 2021. The detailed specifications of variables are available from Online Appendix F of Gu et al. (2020).

In addition to stock-level characteristics, we also consider 14 macroeconomic variables. Among those, eight are used in 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); and six 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).

Lastly, we include industry dummies based on the first two digits of SIC code as in Gu et al. (2020). In summary, we have got 94 stock-specific variables, 14 macroeconomic variables, and 74 industry dummy variables. Throughout our empirical studies, the explanatory variables of use will be the following covariates as defined in Gu et al. (2020):

$$z_{i,t} = x_t \otimes c_{i,t}, \tag{2.31}$$

where $c_{i,t}$ denotes the vector of 94 characteristics for firm $i$, and $x_t$ represents the vector of macroeconomic variables with an added constant $C$. Thus, $z_{i,t}$ is the vector of predictors including interactions between macroeconomic variables and stock-level signals. The total number of covariates is $94 \times (14 + 1) + 74 = 1484$.

The original dataset used by Gu et al. (2020) spans from March 1957 to December 2016, covering 60 years of history. However, it includes a large number of missing variables.[3] After deleting the missing data, we obtain a dataset that starts in January 1985 and ends in December 2021. To alleviate the computational burden associated with neural 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, it is not possible to fill in those missing values with the month cross-sectional medians as done by Gu et al. (2020). Because of this missing data problem, filling with medians does not seem appropriate since it will negatively affect the explanatory power of some predictors. We, therefore, choose to only consider the sample period without missing observations.

**The choice of portfolio assets**

As argued by Lin et al. (2022), 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 objective function of CoSR is more relevant when the universe of portfolio assets covers large financial institutions that are potentially systemic, although not necessarily classified as Systemically Important Financial Institutions (SIFIs). Therefore, we first consider a set of portfolio assets that includes only large financial institutions.

In November of 2021, the Financial Stability Board (FSB), in consultation with Basel Committee on Banking Supervision and national authorities, identified a list of Global SIFIs (G-SIFIs). The total number of G-SIFIs contained in the FSB's list is 30, among which 5 are traded on the US market throughout our sample period. Besides, the Board of Governors of the US Federal Reserve System maintains a list of Domestic SIFIs (D-SIFIs). This list contains financial firms not big enough to be classified as G-SIFIs, but are still considered to be domestic systemically important. According to the list released by the Federal Reserve as of March 2014, 23 banks traded on the US stock market were identified as D-SIFIs. Among those D-SIFIs, 12 are traded throughout our sample period. Thus, from the above, we obtain a list of 17 SIFIs consisting of 5 G-SIFIs and 12 D-SIFIs.

Following Brownlees and Engle (2016), we select large financial firms with a market capitalization greater than 5 bln USD as of the end of June 2007. After applying this filter criterion to our dataset, we are left with a list of 38 assets that covers the aforementioned 17 SIFIs. Therefore, we finally obtain a list of 38 portfolio assets (hereafter set1) including 17 SIFIs and 21 non-SIFIs. These firms are listed in Table 2.1 within Appendix A.

To add robustness to our empirical findings, we also consider a set of portfolio assets which are randomly selected from our dataset. This allows us to explore the out-of-sample performance of our approach on portfolio assets that come from different sectors and with different sizes. Since the intensity of the computational simulation methods that we employ makes it difficult to work with high-dimensional portfolios, we restrict ourselves to a relatively moderate number of portfolio assets and set the dimension of the randomly chosen set to 50 (hereafter set2). Among those 4 belong to the mining sector, 19 belong to the manufacturing sector, 5 belong to the transportation sector, 4 belong to the wholesale sector, 4 belong to the retail sector, 10 belong to the finance sector, and 4 belong to the

service sector. Table 2.2 in Appendix A lists these assets.

### 2.4.2 Estimation and selection of SPNN model

#### Sample splitting

We predict conditional quantiles of asset returns over the evaluation period via a recursive estimation procedure. To achieve this, we first divide our original sample into two disjoint but consecutive subsamples. The first subsample - known as in-sample - is further divided 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 initial size of our recursive window is set to 180 monthly observations (from January 1985 to December 1999). The increment size of the window is one month, which results in an out-of-sample with 264 monthly observations starting from January 2000 and ending in December 2021.

It is well known that the ML models are prone to overfit the data, so it is crucial to go through a rigorous procedure of hyperparameter tuning. The choice of hyperparameters helps control the model's complexity and determine the model's predictive power as well. Following Gu et al. (2020), we use the validation subsample $\mathcal{L}_2$ to perform model selection. Specifically, for each iteration, we use as a validation subsample $\mathcal{L}_2$ the last 20% of cross-sectional data of each in-sample for all 577 firms and the market, with the first 80% of the observations as the training subsample. The neural network model is estimated several times using different sets of hyperparameters on $\mathcal{L}_1$. The subsequent $\mathcal{L}_2$ is then employed to determine the optimal tuning parameters by evaluating the quantile forecasts based on the model estimates obtained over $\mathcal{L}_1$ for the respective hyperparameter set. In particular, the quantile score (QS) is adopted for evaluating quantile forecasts, which takes into account both sharpness and reliability; see Hong et al. (2016). It is defined as the mean of pinball losses throughout the forecasting horizon and across all targeted quantile levels:

$$\text{QS} = \frac{1}{\#M \times H \times (N+1)} \sum_{m \in M} \sum_{t=1}^{H} \sum_{j=1}^{N+1} \rho\big(R_{j,t}, \hat{q}_{j,t}(\tau_m)\big), \tag{2.32}$$

where $M$ is the quantile set of interest (we set $M = \{1, 2, ..., 99\}$), $R_{j,t}$ is the realized return of individual firm or market, and $H$ indicates the forecast horizon ($H = 12$ in our case).

After selecting the best set of hyperparameters, we re-estimate our model using the

in-sample data on $\mathcal{L}_1 + \mathcal{L}_2$, based on which we obtain the final quantile forecasts of returns over the out-of-sample period $\mathcal{L}_3$. As for the data preprocessing, we standardize features by removing the mean and scaling to unit variance. The data is first normalized within each of the training subsamples during hyperparameter tuning and then normalized for observations within the whole in-sample when making final forecasts. Due to the computational intensity of ML-based approaches, we re-fit our model once a year and retain the corresponding estimates to obtain the quantile forecasts for that year; see Gu et al. (2020) and Kynigakis and Panopoulou (2021).

**SPNN configuration**

We consider neural networks with up to three hidden layers. In particular, we consider the following specifications: (1) SPNN model that has a single hidden layer with 32 neurons (hereafter SPNN1); (2) SPNN model that has two hidden layers with 32 and 16 neurons (hereafter SPNN2); and (3) SPNN model that has three hidden layers with 32, 16, and 8 neurons (hereafter SPNN3).

In practice, we adopt the Rectified Linear Unit (ReLU) $g(x) = \max(0, x)$ as the activation function of hidden layers, which promotes sparsity in the number of active neurons and allows for an efficient derivative computation as well; see Nair and Hinton (2010) among others. As for the output layer, we apply the identity activation function $g(x) = x$ following Hatalis et al. (2019). Following Gu et al. (2020), we implement the Adaptive moment estimation algorithm (Adam), which computes individual adaptive learning rates for the model parameters using estimates of the first and second moments of the gradients.

**Training and regularization methods**

The network training is time-consuming due to the high degree of computational complexity involved in tuning abundant parameters and processing mass data. To improve the generalization power of fitted SPNN models and reduce the training cost, in addition to applying $l_1$ penalization, we consider additional DL techniques including batch training, batch normalization, early stopping, and forecast averaging; see Gu et al. (2020) and Kynigakis and Panopoulou (2021) for implementations of these regularization methods.

**Hyperparameters**

We use a two-dimensional grid search approach to find the optimal set of 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 $\lambda$ and the learning rate of Adam optimizer $lr$. For the grid of values we keep following Gu et al. (2020) and set $\lambda \in [10^{-5}, 10^{-3}]$ and $lr \in [10^{-3}, 10^{-2}]$.

Our goal of model selection is modest in the sense of fixing a variety of hyperparameters ex ante, though tuning on a larger set of hyperparameters might help in terms of accuracy.[4] Note that unlike Gu et al. (2020) who set batch size to 10,000, we adopt 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. Keskar et al. (2016) argued that using a large batch tends to suffer from a generalization drop due to sharp minima, see also Masters and Luschi (2018) and others for the preference for small batch. For the remaining hyperparameters, we just follow Gu et al. (2020). Specifically, the number of maximum epochs is set to 100, the patience in early stopping is set to 5, and the number of ensemble models is set to 10.

### 2.4.3 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 Section 2.3.2. Combining the distributional forecasts with the fitted t-copula model, we generate 30,000 return scenarios at the beginning of each month over the out-of-sample period.

The portfolio optimization problem defined in (2.18) is solved on a monthly basis by maximizing the ex-ante CoSR measure. Specifically, we estimate the reward and risk measures by computing the first and second conditional moments based on filtered realizations that satisfy the SE condition. Following Acharya et al. (2017) and Brownlees and Engle (2016), we choose two different SE thresholds $C$: i) $C = VaR^m_{5\%}$ indicating the most that the financial market loses with 95% confidence over the next month, and ii) $C = -6.7\%$, which corresponds to a 40% decrease in the market index over a six-month period.

[4] 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.

For the comparison purpose, we assess the out-of-sample performance of our approach against three benchmark portfolios, namely sample-based SR portfolio, sample-based MVP, and 1/N portfolio.[5] In addition, we also consider S&P 500 Index as a fundamental benchmark. We assume that our investors have an initial wealth of $\mathrm{FW}_0 = 1$ and an initial cumulative log return $\mathrm{CR}_0 = 0$ at the beginning of the backtesting period (December 1999).

Three main steps are performed to calculate the ex-post final wealth and cumulative return at the $k$-th recalibration ($k = 0, 1, 2, ..., 263$). Firstly, we generate return scenarios based on the algorithms described in Section 2.3.2, and obtain the solution $\boldsymbol{W}_{k+1}^*$ to the optimization problem in (2.14) for each of the performance measures under consideration. This step is performed using the Matlab built-in function *fmincon*. Following Kresta et al. (2015), we randomly choose 20 starting points in order to approach the global optimum when solving (2.14). Secondly, the ex-post final wealth is calculated as

$$\mathrm{FW}_{k+1} = \mathrm{FW}_k(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1}), \tag{2.33}$$

where $\boldsymbol{R}_{k+1}$ is the ex-post vector of simple returns between $k$ and $k+1$. Thirdly, the ex-post cumulative log return is calculated as

$$\mathrm{CR}_{k+1} = \mathrm{CR}_k + \ln(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1}). \tag{2.34}$$

Note that the latter equation reports the cumulative performance of the portfolio net of wealth. That is, expression (2.33) implies that $\mathrm{FW}_{K+1} = \mathrm{FW}_0 \prod_{k=0}^{K}(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1})$. Taking logs from the left and right-hand 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 (2.34). By repeatedly computing $\mathrm{FW}_{k+1}$ and $\mathrm{CR}_{k+1}$, we obtain the wealth and cumulative return paths over the backtesting period.

### 2.4.4 Results

In this section, we first briefly illustrate the results of return quantile forecasts and examine the predictive power of candidate predictors using two variable importance measures namely mean squared sensitivity (MSS) and quantile causality measure (QC). We then provide the

[5] To reduce the estimation error of sample covariance matrix, we applied the shrinkage estimator proposed by Ledoit and Wolf (2004) to SR portfolio and MVP.

backtesting results with and without accounting for proportional transaction costs. Finally, we calculate the portfolio's long-run marginal expected shortfall (LRMES) to compare the level of systemic risk generated by different strategies under investigation.

### Quantile forecasts and variable importance

To present some insights on the return quantile forecasts obtained using SPNN models, in Figure 2.3 we display the realized returns and the prediction intervals obtained using SPNN1. To further save space, we only show results for the S&P 500 Index below.[6] From Figure 2.3, we see that the return quantile forecasts are able to capture most of the variation in the realized returns over the out-of-sample period, especially during crisis episodes.

Next we investigate the relative importance of individual predictors for the performance of SPNN model on both training and testing sets. Gu et al. (2020) highlighted the importance of quantifying the influence of each predictor as a way of interpreting ML-based models. Unlike Gu et al. (2020) and Kynigakis and Panopoulou (2021) who use the change in the out-of-sample $R^2$ to measure the variable importance in the context of mean regression, hereafter we adopt two measures that are more suitable for measuring performance related to quantile forecasts.

We first consider the Mean Squared Sensitivity (MSS), which measures the sensitivity of the output of the $m$-th neuron in the output layer with respect to the $p$-th input predictor (Zurada et al. 1994; Yeh and Cheng 2010):

$$\text{MSS}_{p,m} = \sqrt{\frac{\sum_{t\in(\mathcal{L}_1+\mathcal{L}_2)}\left(s_{p,m}|_{\boldsymbol{X}_t}\right)^2}{|\mathcal{L}_1|+|\mathcal{L}_2|}}, \tag{2.35}$$

with

$$s_{p,m}\big|_{\boldsymbol{X}_t} = \frac{\partial \hat{Q}_{R_{t+1}}(\tau_m|\boldsymbol{X}_t)}{\partial x_{p,t}}(\boldsymbol{X}_t), \tag{2.36}$$

where $\boldsymbol{X}_t = (x_{1,t}, ..., x_{P,t})^T$ refers to the $t$-th observation of the $P$ predictors in the in-sample $(\mathcal{L}_1+\mathcal{L}_2)$ on which we perform the sensitivity analysis, $s_{p,m}\big|_{\boldsymbol{X}_t}$ refers to the sensitivity of the output of the $m$-th neuron in the output layer (which in our case is the $\tau_m$-th conditional quantile) with respect to the input of the $p$-th neuron in the input layer evaluated at $\boldsymbol{X}_t$, and $|\mathcal{L}_i|$ denote the number of observations in set $\mathcal{L}_i$, for $i = \{1, 2\}$. The sensitivities defined in (2.36) can be calculated using the chain rule for the partial derivatives of the

[6] The corresponding results for portfolio assets are available upon request.

inner layers, 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

$$\tilde{\text{MSS}}_p = \frac{1}{M}\sum_{m=1}^{M} \text{MMS}_{p,m}, \tag{2.37}$$

which allows us to measure the variable importance across all quantiles of interest.

Next, we consider the QRNN causality measure developed by Lin and Taamouti (2022), 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 is defined as

$$\text{QC}_p(\tau) = \ln\left[\frac{E\left[\rho_\tau\left(R_{t+1} - Q_{R_{t+1}}(\tau|\overline{\boldsymbol{X}}_t)\right)\right]}{E\left[\rho_\tau\left(R_{t+1} - Q_{R_{t+1}}(\tau|\boldsymbol{X}_t)\right)\right]}\right], \tag{2.38}$$

where $\overline{\boldsymbol{X}}_t$ denotes the information set at time $t$ on all predictors, except the $p$-th predictor. $\text{QC}_p(\tau)$ measures the degree of causal effect from a certain predictor $p$ to the $\tau$-th quantile of the predictand given the past of the latter. As pointed out by Song and Taamouti (2021), QC can be viewed as a measure of the amount of information brought by the past of the $p$-th predictor to improve the prediction of the $\tau$-th quantile of asset return $R_{t+1}$. Similar to the average measure $\tilde{\text{MSS}}_p$, in our empirical analysis we compute the average QC for each predictor $x_p$ as

$$\tilde{\text{QC}}_p = \ln\left[\frac{\frac{1}{M|\mathcal{L}_3|}\sum_{m=1}^{M}\sum_{t\in\mathcal{L}_3}\rho_{\tau_m}\left(R_{t+1} - \hat{Q}_{R_{t+1}}(\tau_m|\overline{\boldsymbol{X}}_t)\right)}{\frac{1}{M|\mathcal{L}_3|}\sum_{m=1}^{M}\sum_{t\in\mathcal{L}_3}\rho_{\tau_m}\left(R_{t+1} - \hat{Q}_{R_{t+1}}(\tau_m|\boldsymbol{X}_t)\right)}\right], \tag{2.39}$$

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.

Figure 2.4 reports the variable importance measured by MSS for the 10 most influential firm-level predictors and for all macroeconomic variables under consideration based on the fitted SPNN1 model, while Figures 2.5 and 2.6 in Appendix B report the corresponding variable importance results measured by QC for set1 and set2, respectively.[7] The variable 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 are ranked such that those with the highest importance are at the top and the lowest are

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

at the bottom.

The top-10 most influential firm-level features measured by MSS as displayed in the top panel of Figure 2.4 can be grouped into five categories. The first group contains risk measures including the total and idiosyncratic return volatilities (retvol, idiovol); Next are liquidity variables including dollar volume (dolvol), debt capacity/firm tangibility (tang), bid-ask spread (baspread), turnover (turn), and number of zero trading days (zerotrade); A single momentum predictor constitutes the third group, which is the short-term reversal (mom1m); The fourth group includes a valuation ratio, which is the R&D expense-to-market ratio; The last group consists of industry dummy (sic2). As for the corresponding results of macroeconomic variables, we see from the bottom panel of Figure 2.4 that all contribute significantly to model training. Among those, the total financial uncertainty index (macro_TFU) is identified as the most influential macro-level predictor.

Analogously, the rankings based on QC measure as shown in Figures 2.5 and 2.6 within Appendix B draw a similar conclusion. The results for both portfolio set1 and set2 agree on a fairly small set of dominant predictive firm-level predictive signals, which covers the risk measures of total and idiosyncratic return volatilities (retvol, idovol), the liquidity variables of dollar volume (dolvol), industry-adjusted size (mve_ia), bid-ask spread (baspread) and turnover (turn), the short-term reversal (mom1m), and an accounting variable that indicates the number of years since first Compustat coverage (age). While for the macro state variables, the results again confirm their predictive power and place the greatest emphasis on the total financial uncertainty index (macro_TFU) in both cases.

To better illustrate the variable importance over recursive windows, we display the time-varying rankings of predictors in SPNN1 measured by MSS and QC in Figures 2.7 - 2.12 within Appendix B, consecutively. In particular, we rank the importance of individual predictors according to their average contribution over all quantiles of the returns and across all recursive in-sample or out-of-sample windows depending on the measure of use. Columns in these figures correspond to the start year in each window, and color gradient within each column indicates the most influential (dark blue) to least influential (light blue) predictors.

### Backtesting results

After obtaining the quantile forecasts of returns based on fitted SPNN models, we estimate the corresponding conditional marginal distributions which we combine with t-copula to

generate return scenarios that we use to solve the portfolio optimization problem. In this section, we form a backtesting analysis to assess the economic value of using our SPNN-based return forecasts for asset allocation under systemic risk. To do so, we compare the out-of-sample performance of SPNN-CoSR portfolios with those of benchmark portfolios under consideration. The optimized portfolios were built recursively using conditional/unconditional return moments estimated from simulated/historical return observations at each iteration starting in January 2000. The buy-and-hold portfolio returns are calculated for a one-month period, and portfolios are rebalanced monthly until the end of the out-of-sample period (i.e. December 2021).

The backtesting results are displayed in Figures 2.13 and 2.14. There are several noticeable features from these figures. Firstly, we observe that all candidate portfolios outperform the market S&P 500 portfolio. Secondly, all portfolios perform poorly during the 2007-2008 financial crisis. The SR, MVP and 1/N strategies lose almost all of their values during that period, while the SPNN-CoSR portfolios perform significantly better than others, even though they lost around half of their values since the last peak in 2007. In particular, among SPNN-CoSR portfolios, the SPNN1-based strategy for both SE thresholds deliver the best out-of-sample performance. Thirdly, all SPNN-based CoSR portfolios show a strong upward trend in profitability throughout the evaluation period, which can be mainly attributed to the relatively stable performance during market distress. The backtesting results confirm the benefits of combining SPNN-based return forecasts with the incorporation of systemic risk into traditional mean-variance framework when constructing optimal portfolios.

Tables 2.3 and 2.4 report the values of several performance metrics for set1 and set2, respectively.[8] The results vary among different strategies depending on the choice of performance measure, with the exception is 1/N portfolio which does not rely on any optimization problem or model estimation. For the case of portfolio set1, the SPNN-CoSR portfolios dominate all other benchmark strategies in terms of profitability, whichever model configuration is being considered. Among those the SPNN1-based CoSR portfolio with $C = -6.7\%$ delivers the highest value by the end of evaluation period. Besides, the GARCH-based CoSR portfolios are serious competitors which provide comparable profits over out-of-sample period, but still cannot beat our proposed approach. Specifically, our

[8] Note that in Tables 2.3, 2.4, 2.5 and 2.6 we also add the corresponding results obtained using the GARCH-Copula based approach (denoted by "GARCH-based CoSR") of Lin et al. (2022), which serves as an advanced statistical benchmark model.

investors would multiply their wealth by 20.145 and 21.216 using SPNN1-based CoSR portfolios with $C = VaR^m_{5\%}$ and $C = -6.7\%$, respectively, which is almost twice than that of GARCH-based CoSR portfolios with $C = VaR^m_{5\%}$ (12.797) and $C = -6.7\%$ (13.650). The MVP gives the lowest final wealth (5.348) and annual return (0.079), while the sample-based SR portfolio performs as the second-worst with final wealth of 6.973 and annual return of 0.092. Interestingly, the naive 1/N strategy outperforms all sample-based portfolios in terms of profitability, with the former exhibiting a final wealth of 9.541 and an annual return of 0.108. The backtesting results for portfolio set2 draw similar conclusions as those of set1, we thus omit the details for brevity.

The results of ex-post Sharpe ratio, Sortino ratio and Calmar ratio again demonstrate the superiority of proposed approach. In particular, the SPNN1-based CoSR portfolio with $C = -6.7\%$ delivers the highest values for all performance ratios among candidate portfolios. In addition, we also test the significance of the difference of Sharpe ratios between SPNN1-based CoSR portfolio and that of each benchmark strategy following Ledoit and Wolf (2008). The results are reported in Tables 2.7 and 2.8 for set1 and set2, respectively. According to bootstrap $p$-values, the null hypothesis of equal Sharpe ratios is rejected at significance level of 0.01 in all cases. The testing results further confirm the enhanced portfolio performance of our approach.

Besides the above mentioned performance ratios, investors may consider alternative statistics to gain deeper insights of their trading strategies. Therefore, we add Maximum Drawdown (MDD), Maximum One-Month loss (MOL), and average Turnover (TO) as additional performance metrics. Formally, the MDD is defined as

$$\text{MDD} = \max_{t_0 \le t_1 \le t_2 \le T_0} \left\{ r_{p,t_0:t_1} - r_{p,t_0:t_2} \right\}, \tag{2.40}$$

where $r_{p,t_0:t}$ denotes the cumulative portfolio return from time $t_0$ to $t_i$, for $i \in \{1, 2\}$, with $t_0$ and $T_0$ being the first and last months of evaluation period. MOL measures the largest decline in portfolio value over one-month period, and the average TO is defined as

$$\text{TO} = \frac{1}{T} \sum_{t=1}^{T} \left( \sum_{i=1}^{N} \left| \omega_{i,t+1} - \frac{\omega_{i,t}(1 + R_{i,t+1})}{1 + \sum_{j=1}^{N} \omega_{j,t} R_{j,t+1}} \right| \right), \tag{2.41}$$

where $\omega_{i,t}$ is the desired weight of portfolio asset $i$ at time $t$.

Tables 2.3 and 2.4 also report the values of these alternative measures. We first look

at the results for portfolio set1. While the SPNN1-based CoSR portfolio with $C = -6.7\%$ provides the highest profitability, it has the lowest MDD as well. Furthermore, all SPNN-based CoSR portfolios regardless of model configurations and SE thresholds outperform other competitors in terms of MDD, which demonstrates the better performance of our proposed approach during market distress. Next, the values of MOL of SPNN-based CoSR portfolios are lower than those of GARCH-based counterparts, while the sample-based MVP displays the lowest MOL since it focuses on the risk only. Lastly, as for the corresponding results on portfolio set2, our SPNN-based CoSR portfolios are doing slightly worse than the GARCH-based CoSR portfolios in terms of MOL, while the formers still outperform the latters in terms of MDD and other measures of profitability as well.

**Effects of transaction costs**

The estimation of transaction cost (TC) is based on TO as defined in (2.41). After accounting for a proportional TC of $c$, the portfolio return is now calculated as follows:

$$\tilde{R}_{p,t+1} = (1 + R_{p,t+1})\left(1 - c\sum_{i=1}^{N}\left|\omega_{i,t+1} - \frac{\omega_{i,t}(1 + R_{i,t+1})}{1 + \sum_{j=1}^{N}\omega_{j,t}R_{j,t+1}}\right|\right) - 1. \tag{2.42}$$

Gu et al. (2020) argued that, given the large role of price trend predictors employed by ML-based approaches, it is unsurprising that the ML-based trading strategies are characterized by relatively high TO. This also holds for our SPNN-based approach as shown in Tables 2.3 and 2.4. Although our SPNN-based CoSR portfolios show a higher TO than that of the sample-based benchmarks, their values are still much lower than those of the GARCH-based CoSR portfolios. Unsurprisingly, the 1/N portfolio delivers the lowest TO due to its well-diversified property.

Although the CoSR portfolios with relatively high TO are more flexible to adapt to the changes in market conditions than other benchmarks, their portfolio values are likely to decrease due to their higher TC during rebalancing. To analyze the effects of TC, we set a relatively high value of $c = 50$ basis points (bps) and recompute the ex-post paths of final wealth and cumulative return for all portfolios under consideration. Tables 2.5 and 2.6 report the values of the performance measures after taking into account proportional TC.[9] In short, the inclusion of TC does not change our main conclusions. The SPNN-based

[9] To save space, the Figures 2.15 and 2.16 that illustrate the backtesting results after considering transaction costs are removed to the Appendix B.

CoSR portfolios still outperform all other competitors in terms of profitability. Remarkably, the final wealth of SPNN-based CoSR portfolios is still more than twice that of other benchmarks excluding the 1/N portfolio.

**Portfolio-level systemic risk**

In this section, we measure the portfolio-level systemic risk using portfolio's LRMES proposed by Lin et al. (2022), which is defined as

$$\text{LRMES}_{p,t} = \sum_{i=1}^{N} \omega_{i,t} \, \text{LRMES}_{i,t}, \tag{2.43}$$

where $\text{LRMES}_{i,t}$ indicates the expected loss in equity value of asset $i$ over month $t$. The portfolio's LRMES can be interpreted as the expected percentage drop in portfolio value under stressed market conditions. Figure 2.17 illustrates the portfolio-level LRMES over the evaluation period.[10] Overall speaking, the SPNN-based CoSR portfolios give the best performance in terms of systemic risk measured by LRMES. The relatively low portfolio-level LRMES indicates less potential loss during crisis periods. Specifically, the SPNN-based CoSR portfolio with $C = -6.7\%$ presents the lowest LRMES than all other competitors throughout the out-of-sample period whichever portfolio set is being considered, while all benchmark strategies provide much higher and volatile LRMES values.

## 2.5 Conclusion

We explore whether using return forecasts generated via smooth pinball quantile regression neural network can add value to systemic risk-based portfolio selection. The optimal portfolio is constructed by maximizing an ex-ante conditional Sharpe ratio based on simulated return scenarios, and its out-of-sample performance is compared with that of tangency portfolio, minimum variance portfolio, and equally-weighted portfolio. The proposed approach outperforms all other benchmarks in terms of profitability and portfolio-level systemic risk. The testing results of the difference of Sharpe ratios further confirm its significant outperformance against benchmark strategies. Although our portfolio is characterized by a relatively high turnover rate, its superiority is still tenable after accounting for a consider-

[10] To save space, we only show the results obtained by SPNN1 here. However, the corresponding results for other SPNN configurations are available upon request.

able amount of proportional transaction costs. Another side contribution of our paper is the implementation of two variable importance measures, which we propose to rank the most influential predictors in SPNN models. The relevant results demonstrate the substantial predictive information brought by macroeconomic variables, whereas only a limited number of firm-level signals contribute to the training and prediction process.

# A - Figures

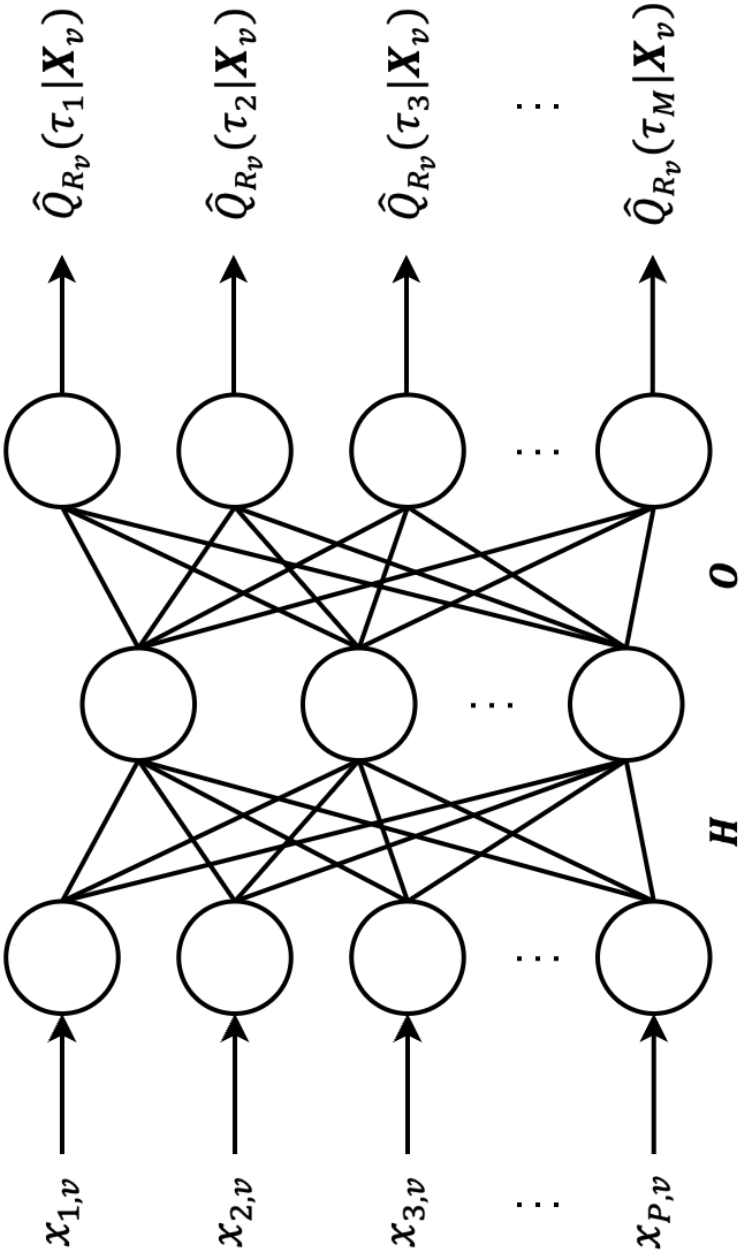



Figure 2.1: Schematic diagram of SPNN with a single hidden layer.

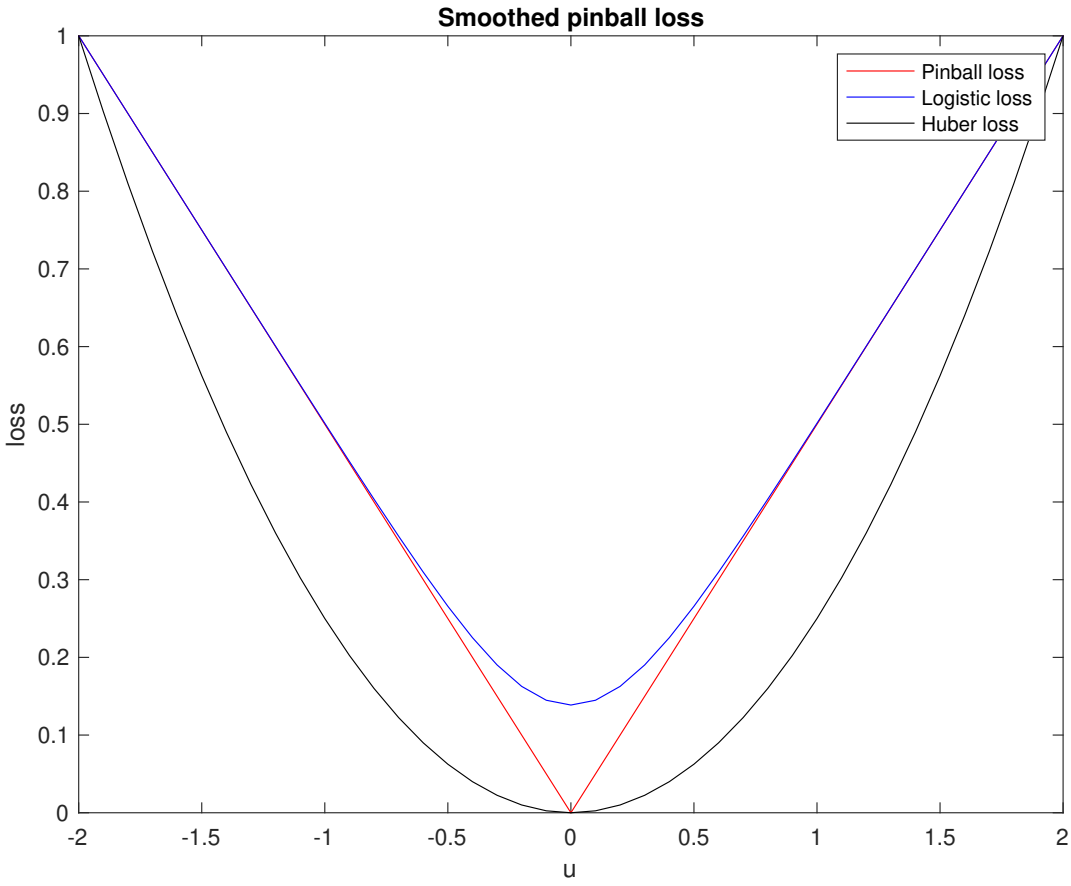



Figure 2.2: Pinball loss versus smoothed ones.

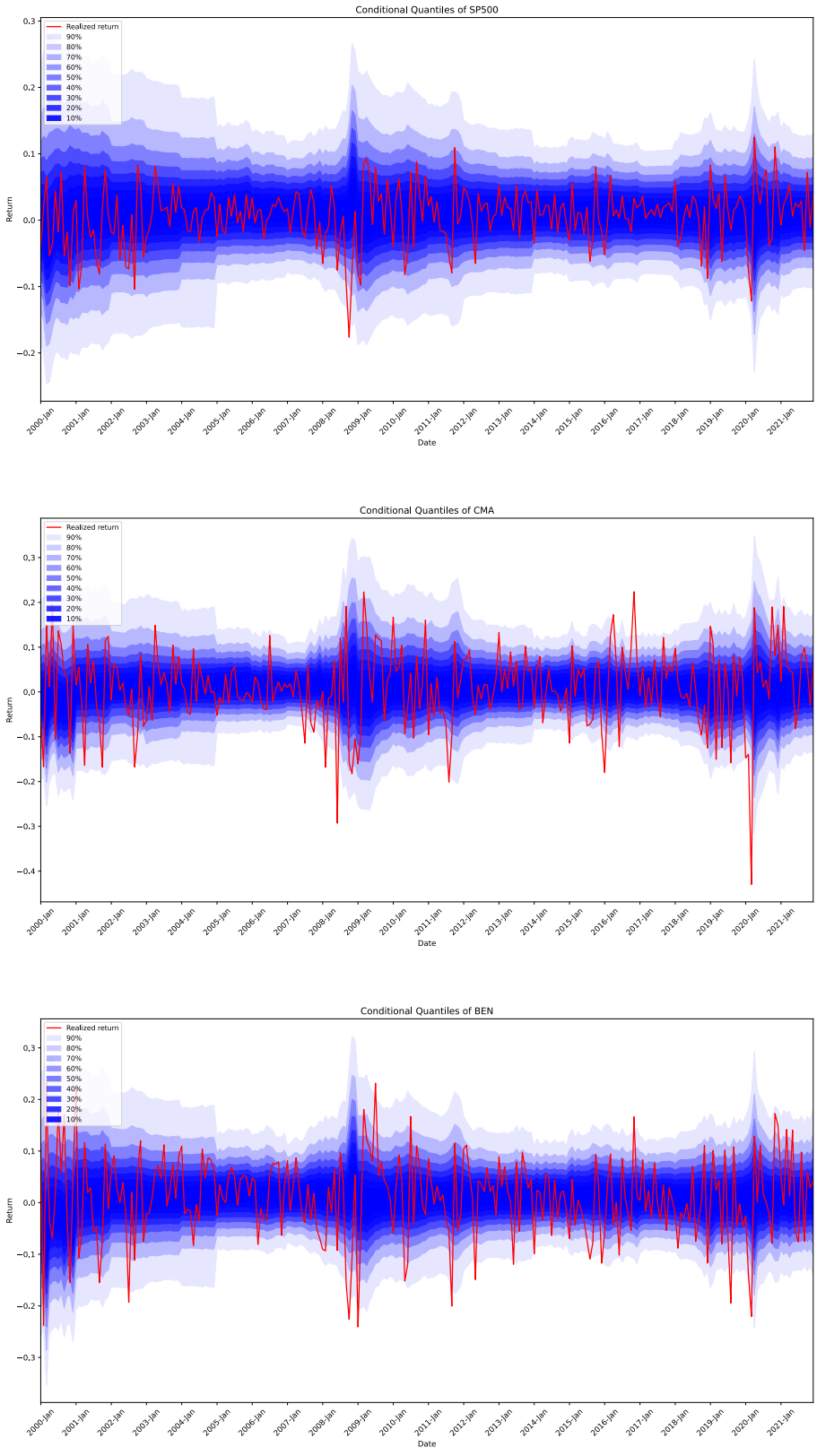
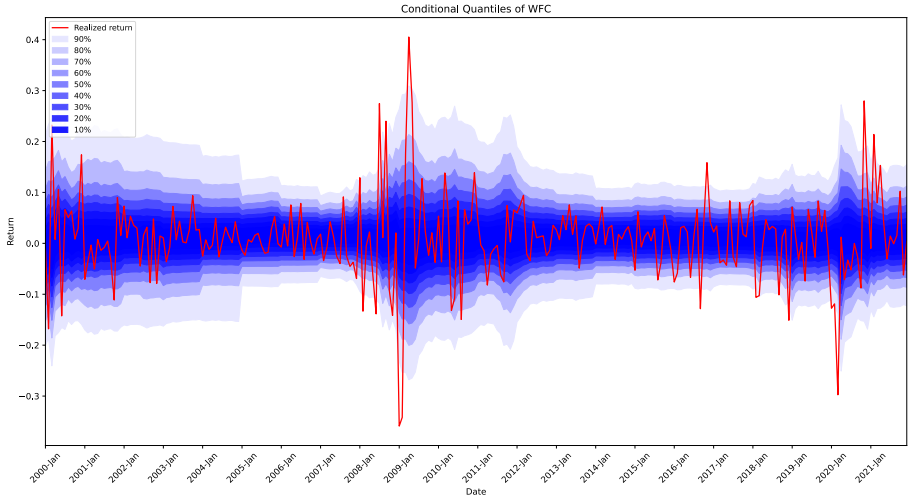
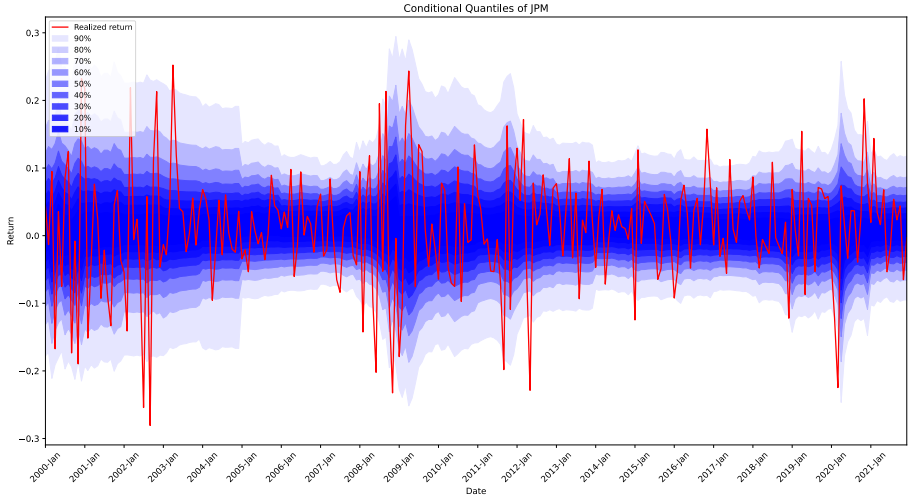



Figure 2.3: Conditional quantiles of returns on market index and a few portfolio assets obtained from SPNN1.



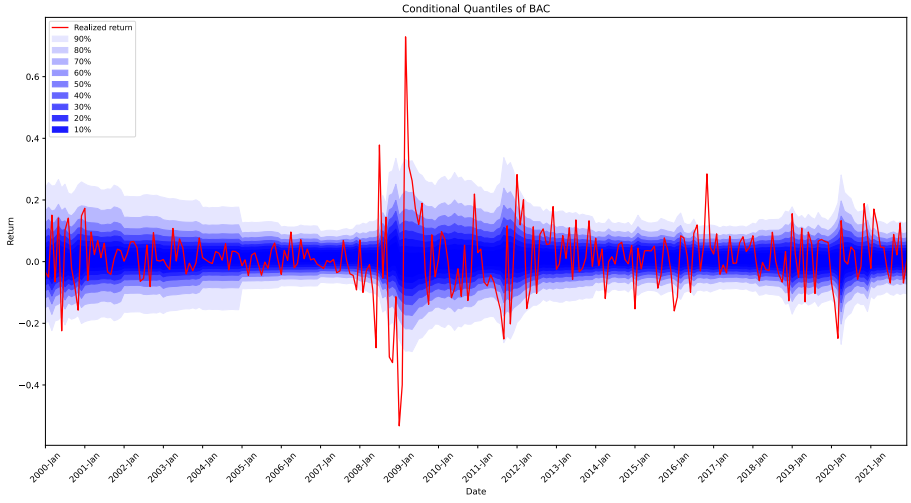




Figure 2.3: (continued)

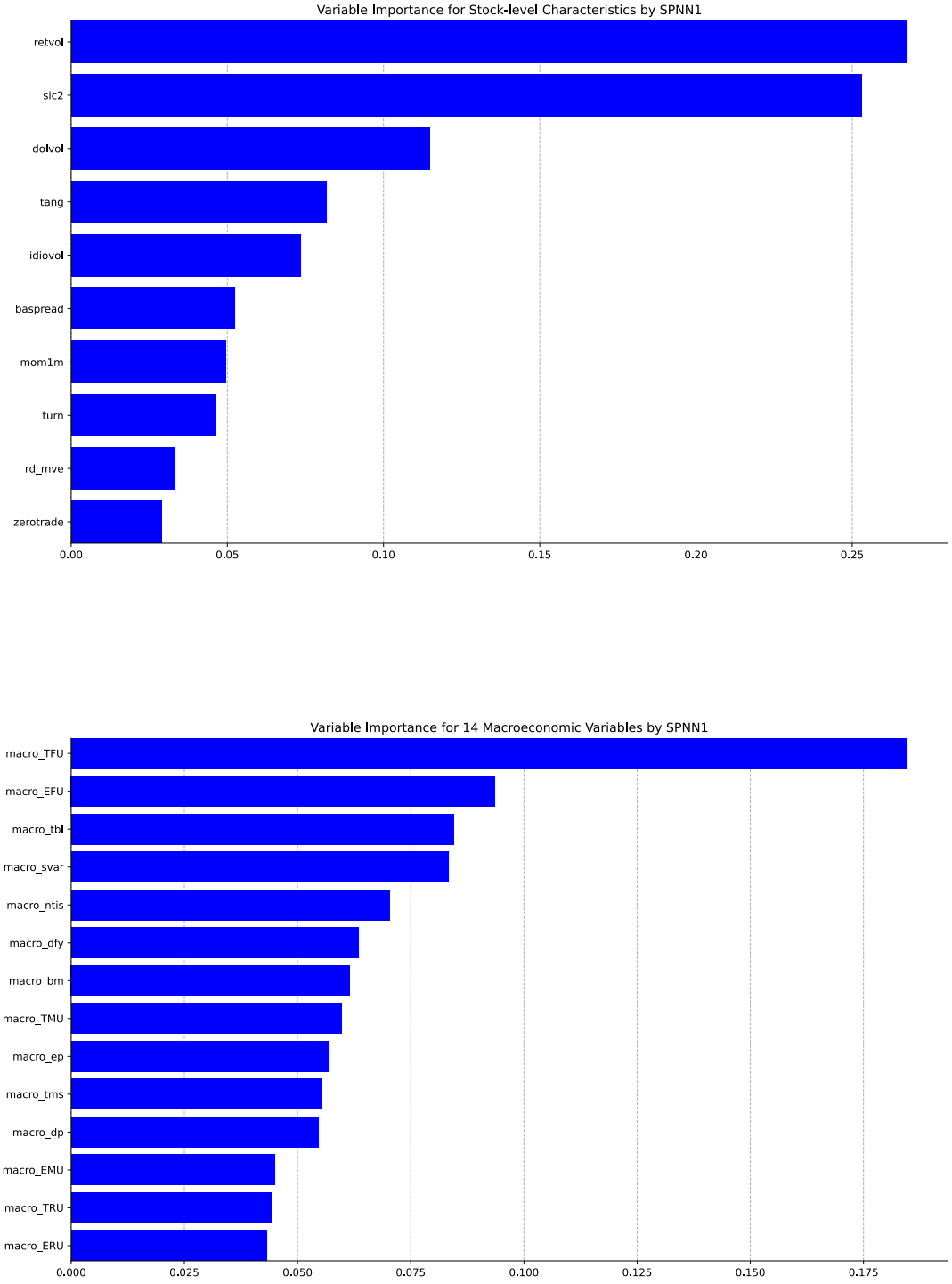



Figure 2.4: Top panel displays the top-10 most influential firm-level predictors in SPNN1 measured by MSS, while the bottom panel reports the corresponding results for all macroeconomic variables. Variable importance is an average over all quantiles and recursive in-sample windows. Variable importance is normalized to sum to one.

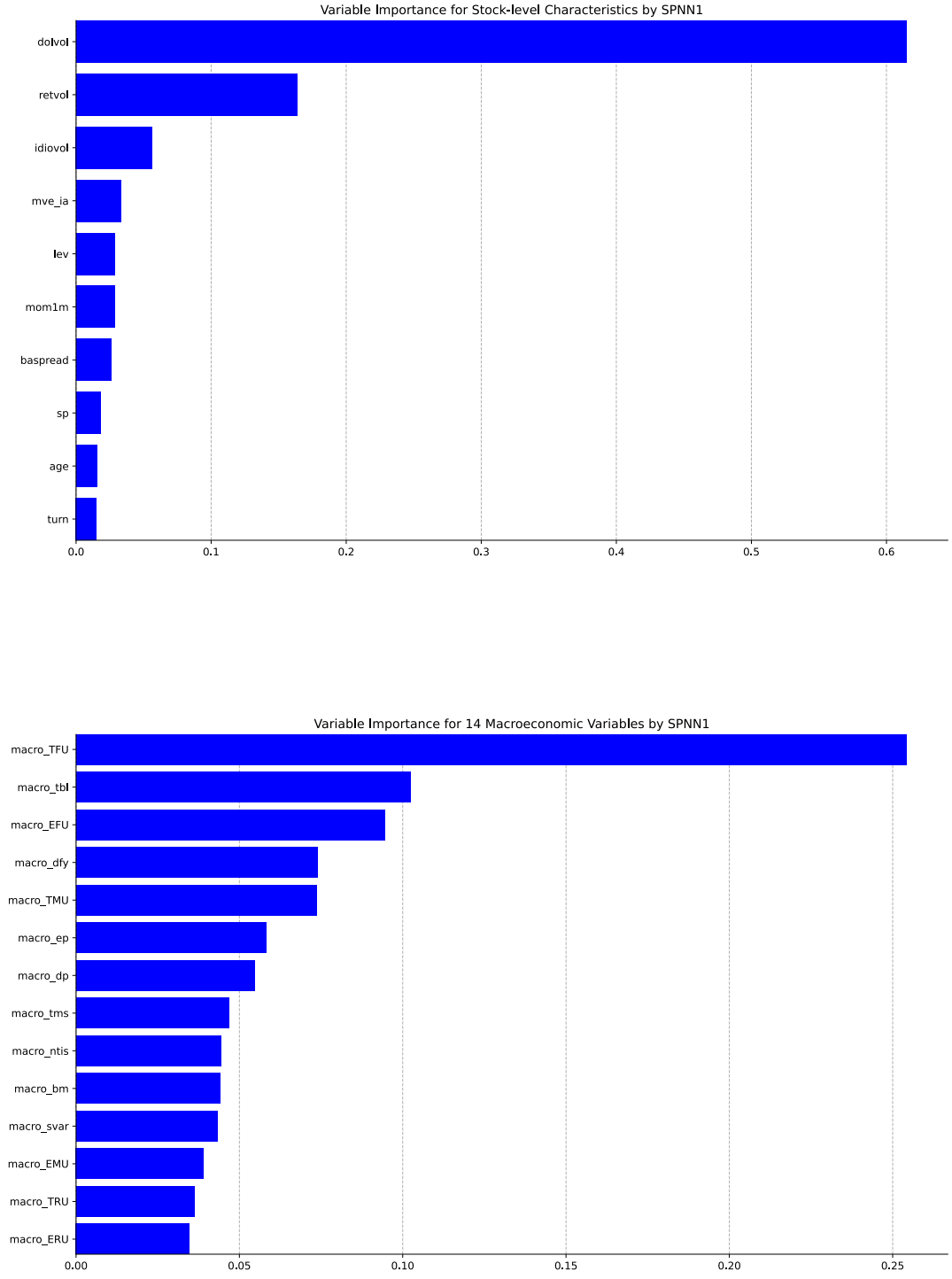



Figure 2.5: Top and bottom panels display the variable importance of top-10 most influential firm-level predictors and all macroeconomic variables measured by QC in SPNN1 for portfolio set1, respectively. Variable importance is an average over all quantiles and recursive in-sample windows. Variable importance is normalized to sum to one.

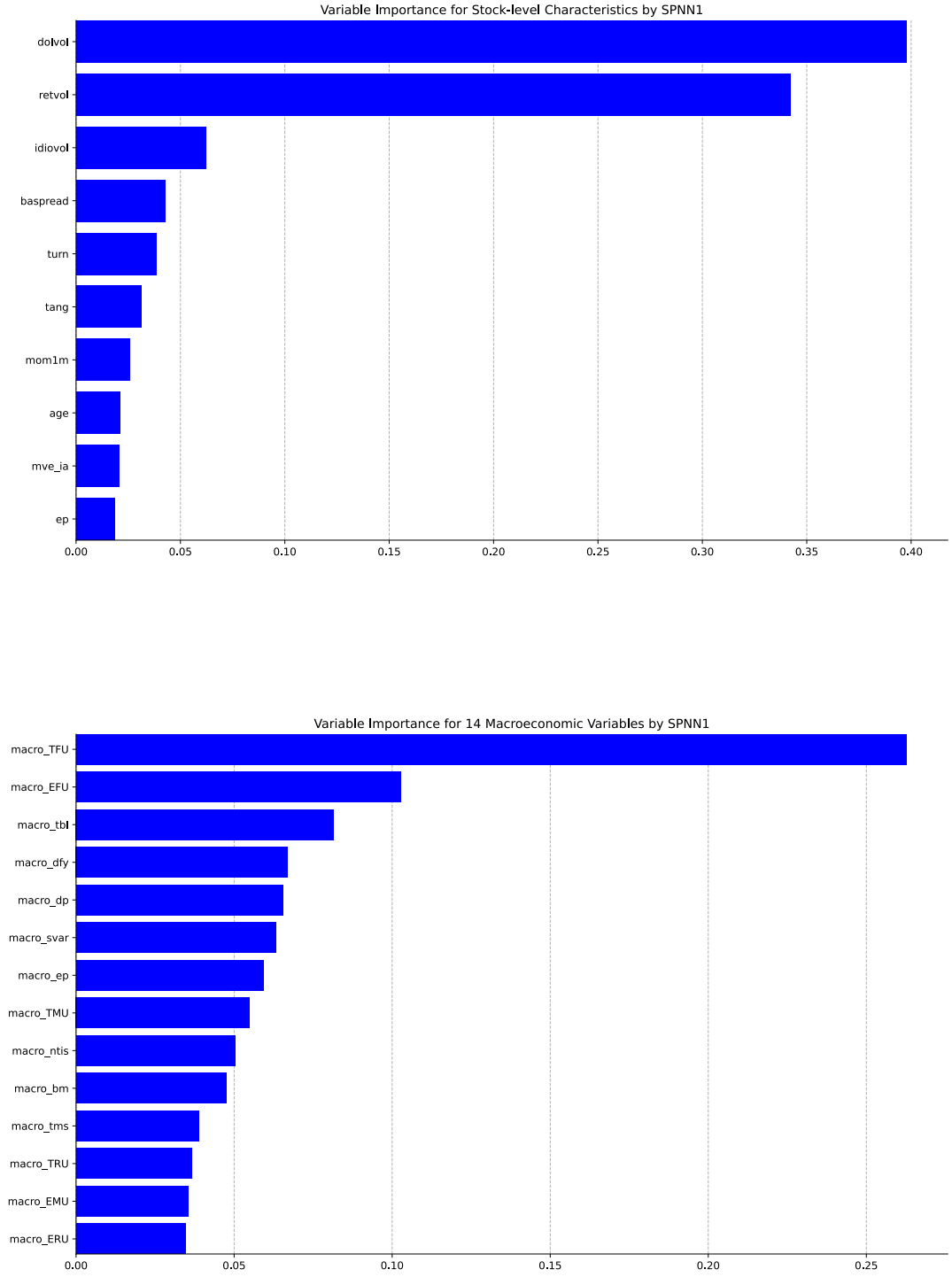



Figure 2.6: Top and bottom panels display the variable importance of top-10 most influential firm-level predictors and all macroeconomic variables measured by QC in SPNN1 for the portfolio set2, respectively. Variable importance is an average over all quantiles and recursive in-sample windows. Variable importance is normalized to sum to one.

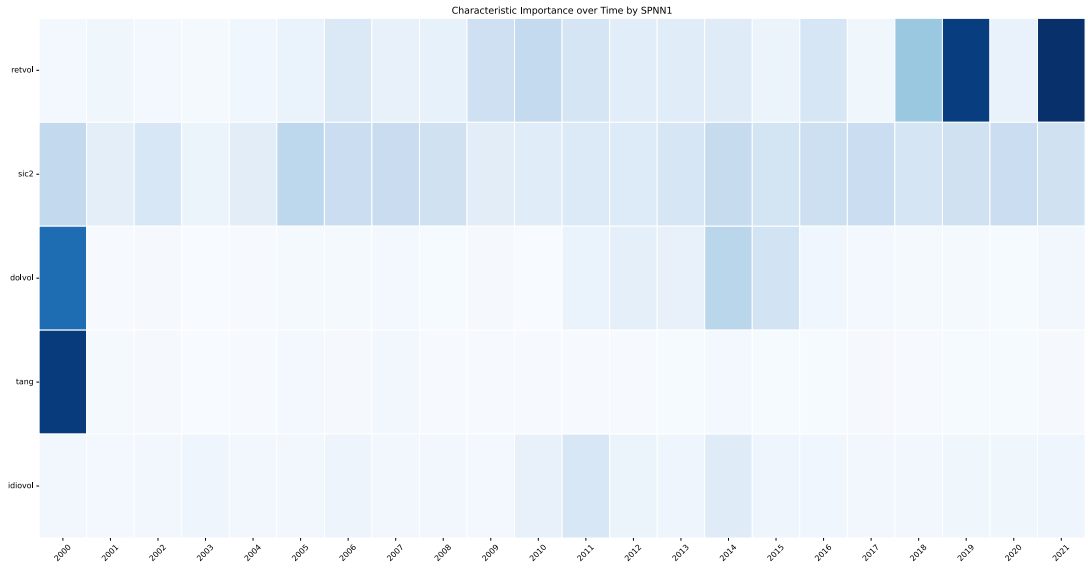



Figure 2.7: Time-varying variable importance of the top-5 most influential firm-level predictors measured by MSS. Predictors are ordered based on the average value of their MSS over recursive trainings, with the most influential features at the top and the least influential at the bottom. Columns correspond to the year end of each of the 22 in-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

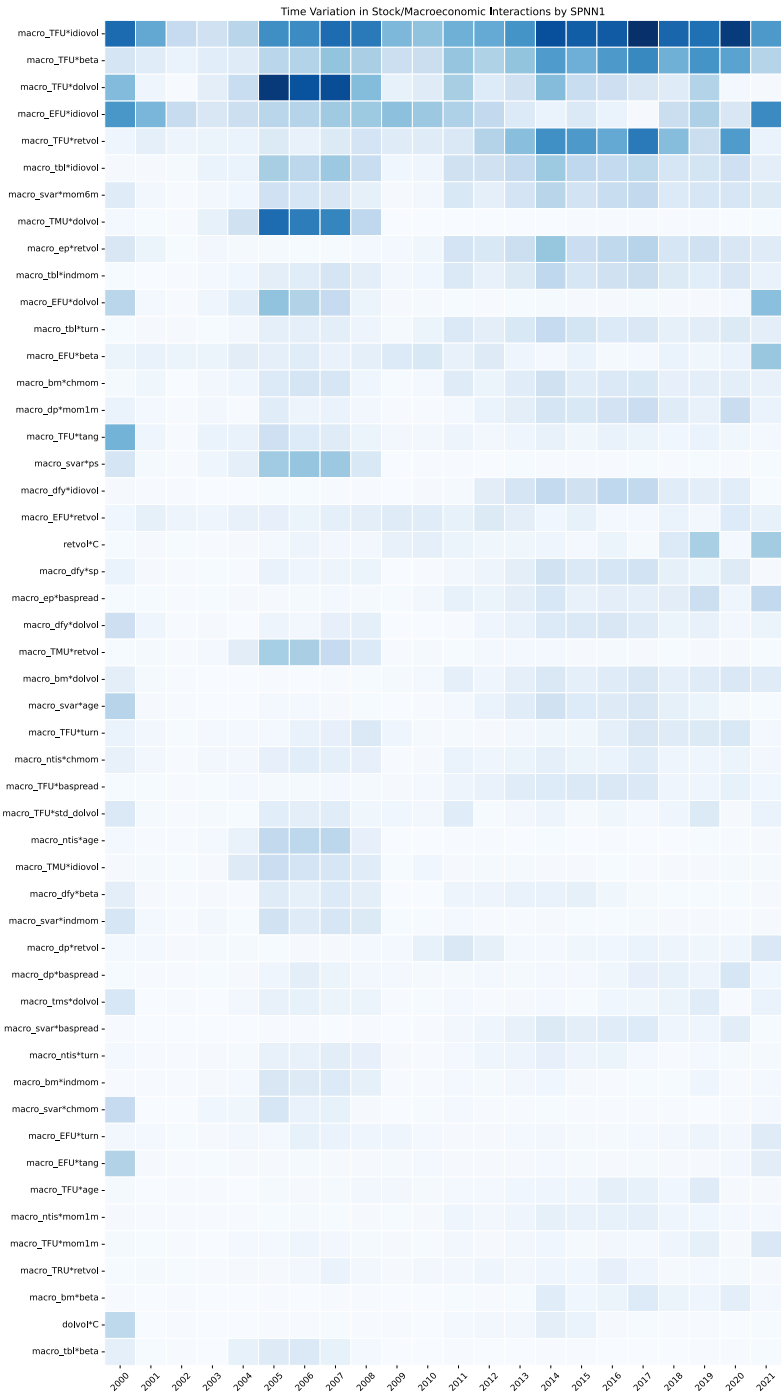



Figure 2.8: Time-varying variable importance of the top-50 most influential predictors of interactions between each firm characteristic with macroeconomic variables measured by MSS. Columns correspond to the year end of each of the 22 in-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

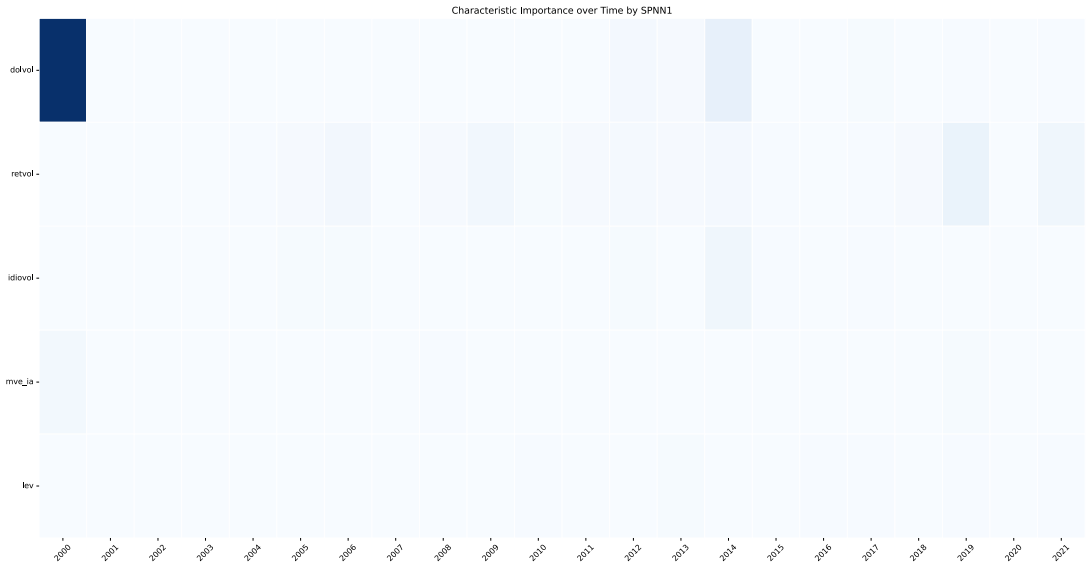



Figure 2.9: Time-varying variable importance of the top-5 most influential firm-level predictors measured by QC for portfolio set1. Columns correspond to the year start of each of the 22 out-of-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

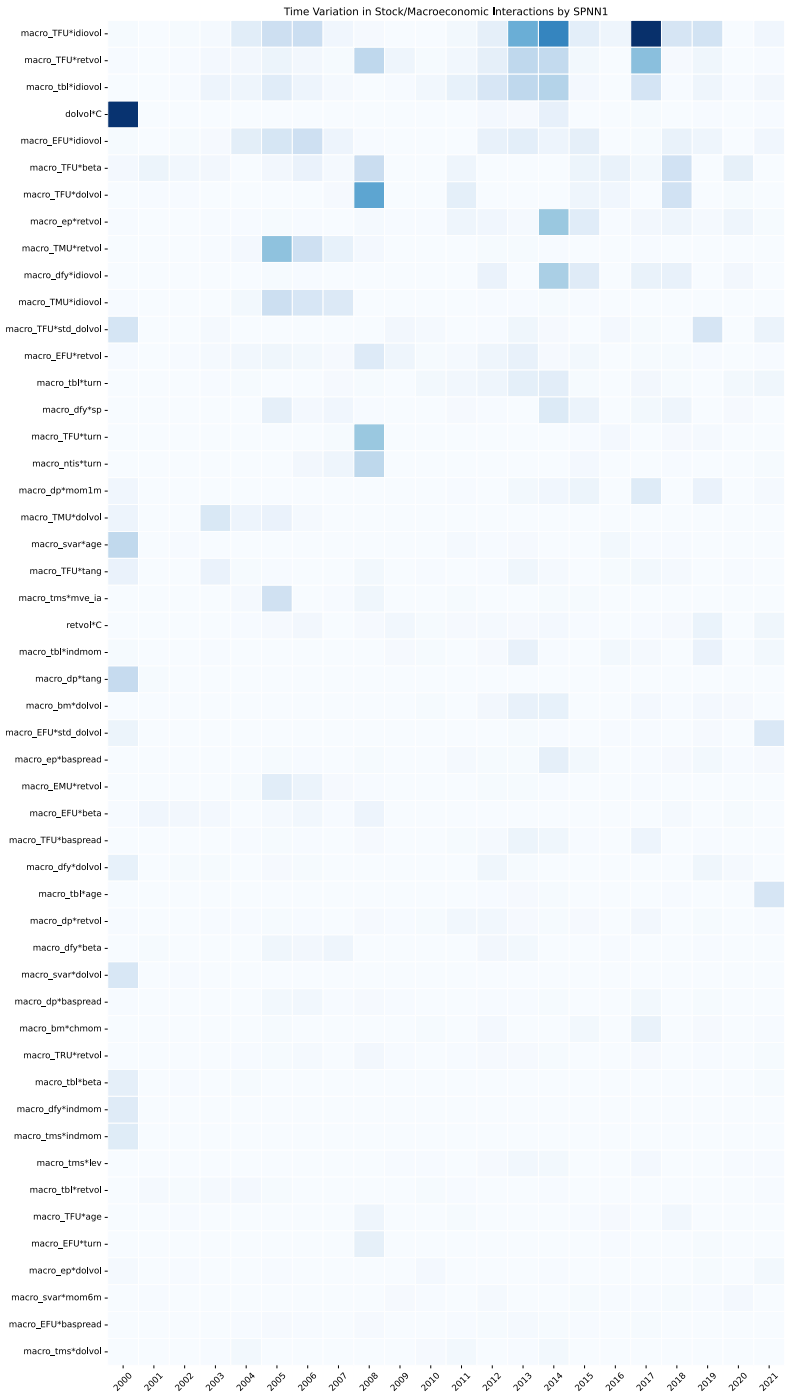



Figure 2.10: Time-varying variable importance of the top-50 most influential predictors of interactions between each firm characteristic with macroeconomic variables measured by QC for portfolio set1. Columns correspond to the year start of each of the 22 out-of-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

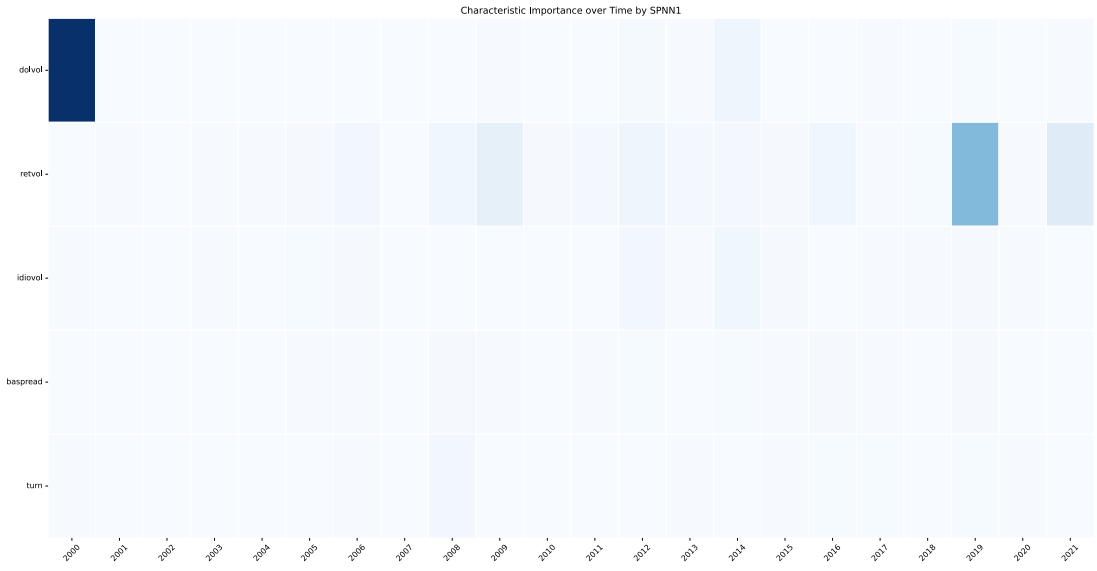



Figure 2.11: Time-varying variable importance of the top-5 most influential firm-level predictors measured by QC for portfolio set2. Columns correspond to the year end of each of the 22 out-of-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

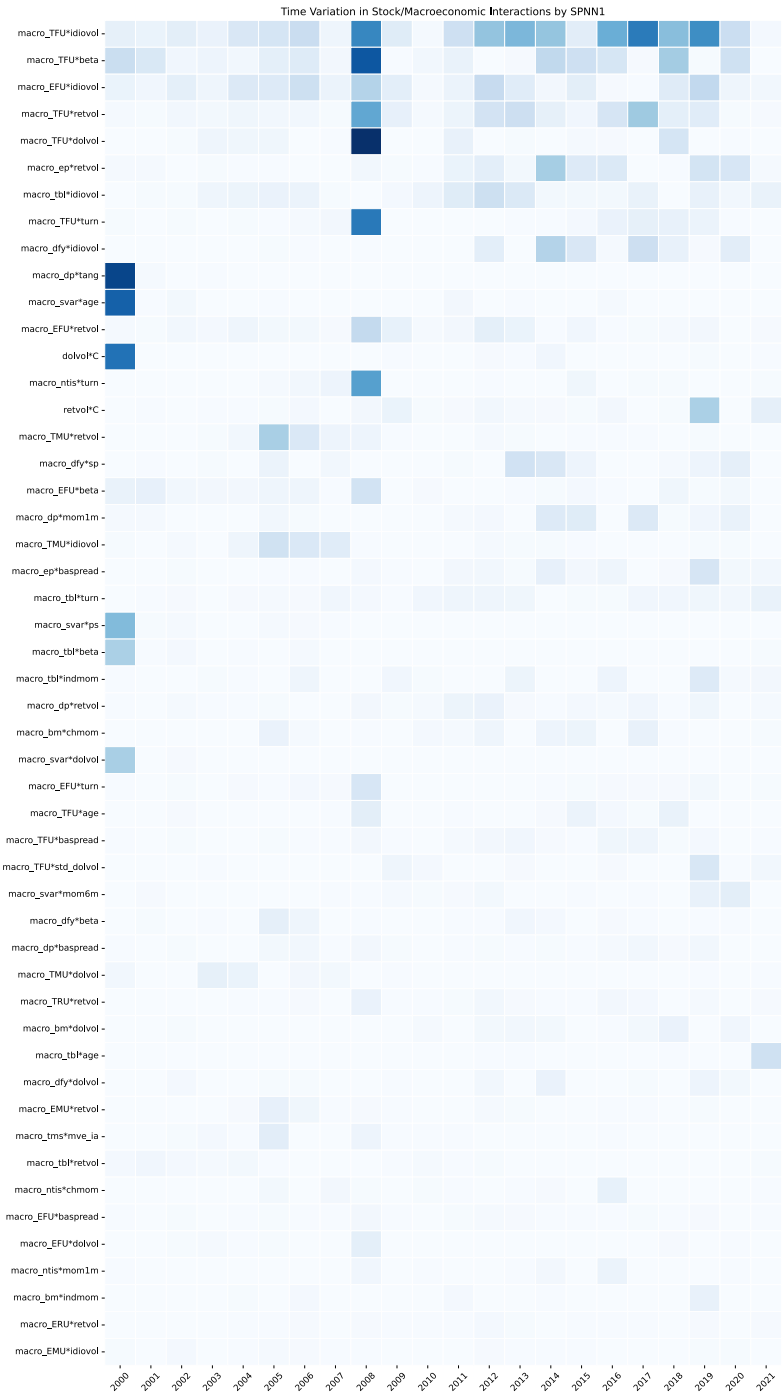



Figure 2.12: Time-varying variable importance of the top-50 most influential predictors of interactions between each firm characteristic with macroeconomic variables measured by QC for portfolio set2. Columns correspond to the year end of each of the 22 out-of-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

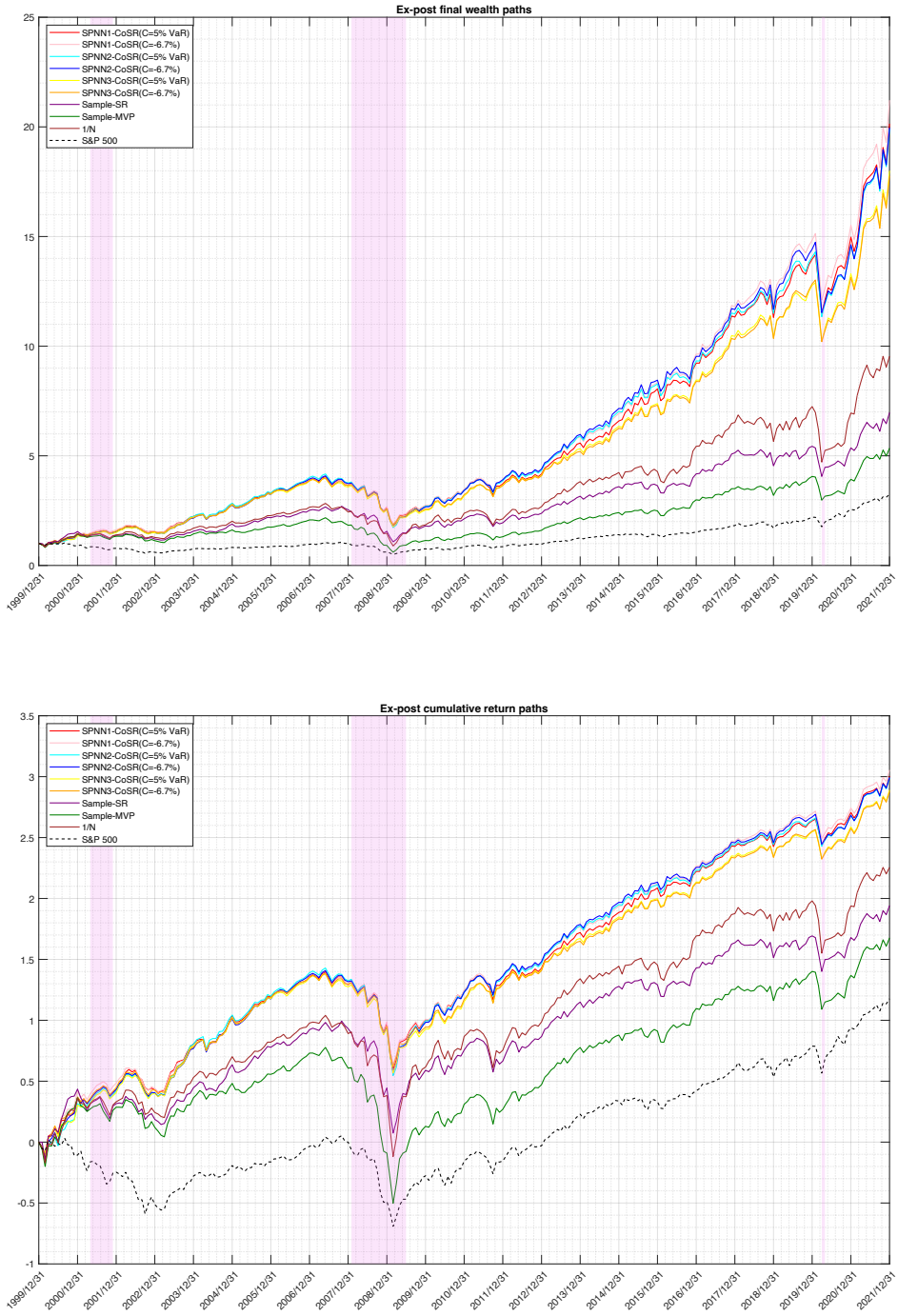



Figure 2.13: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies for portfolio set1. The shaded areas denote recession periods as defined by NBER.

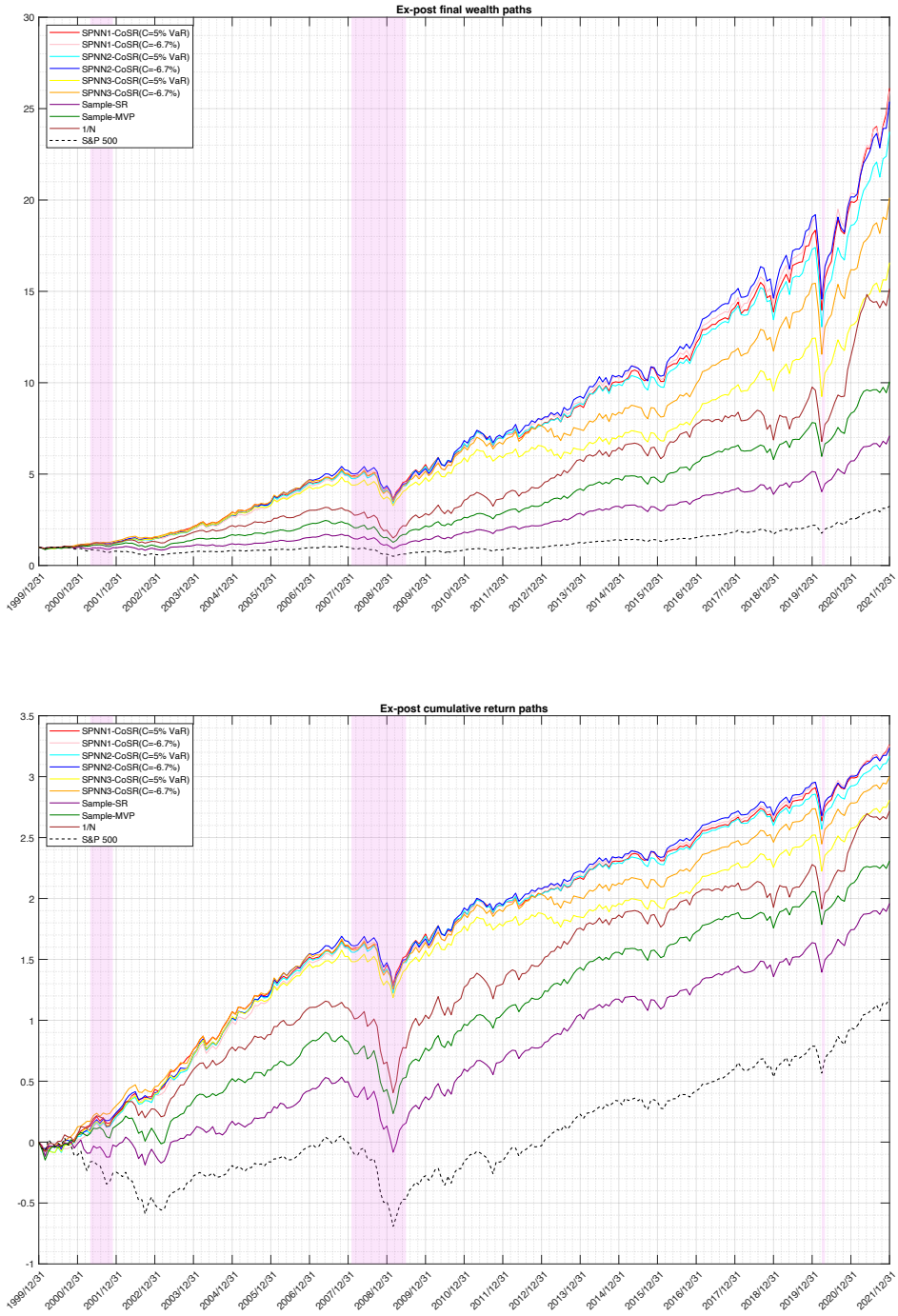



Figure 2.14: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies for portfolio set2. The shaded areas denote recession periods as defined by NBER.

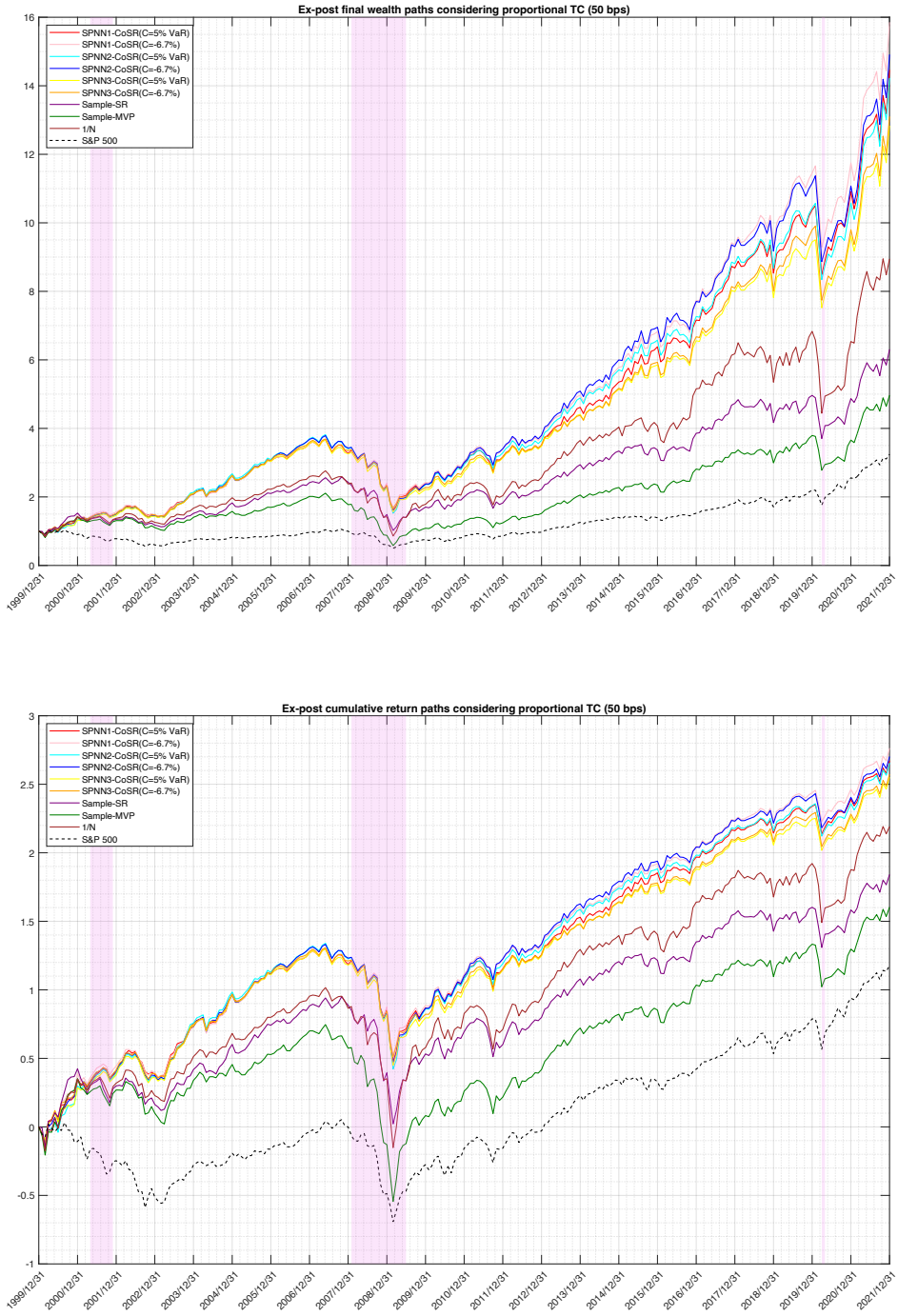



Figure 2.15: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies with 50 bps proportional TC for portfolio set1. The shaded areas denote recession periods as defined by NBER.

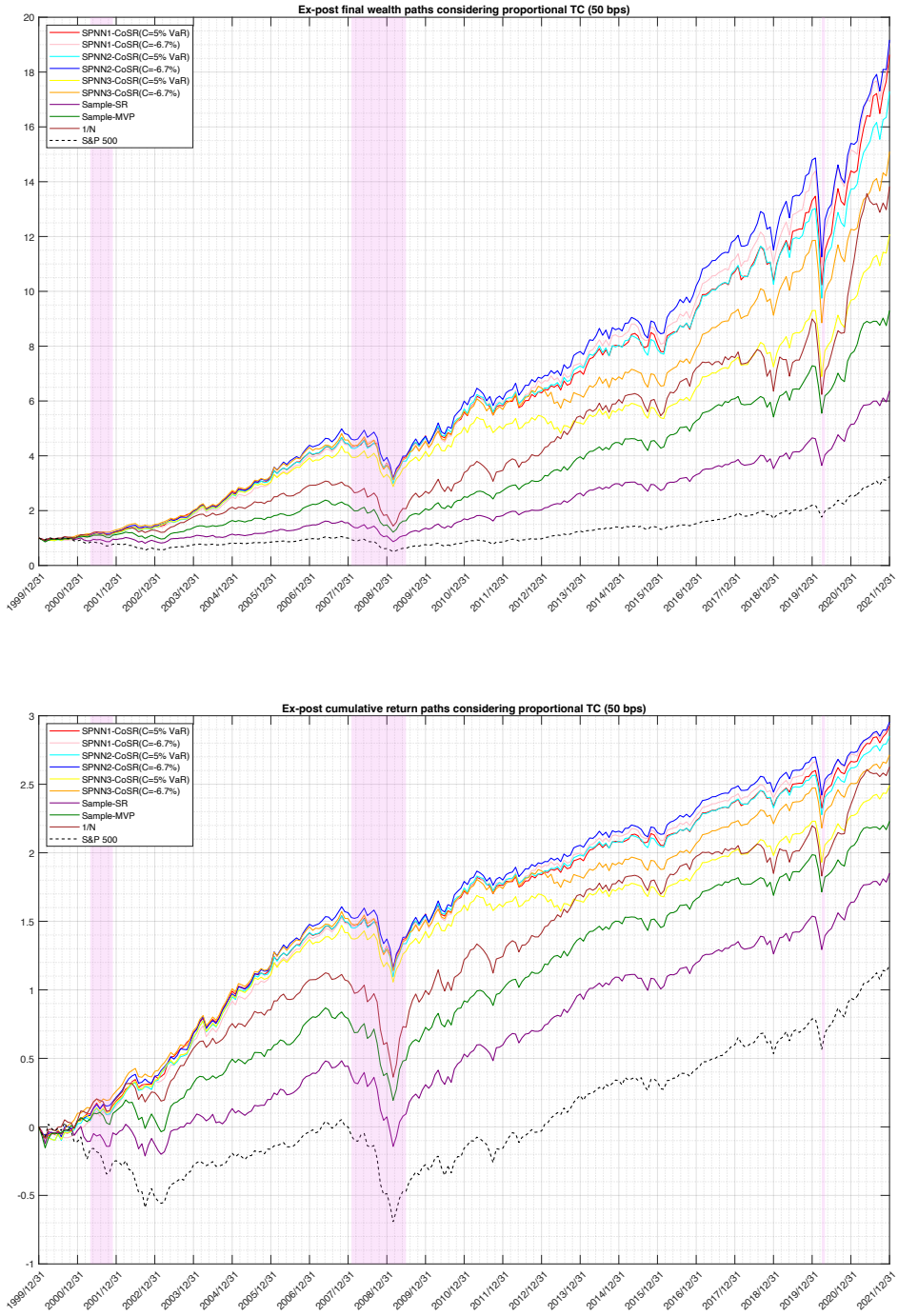



Figure 2.16: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies with 50 bps proportional TC for portfolio set2. The shaded areas denote recession periods as defined by NBER.

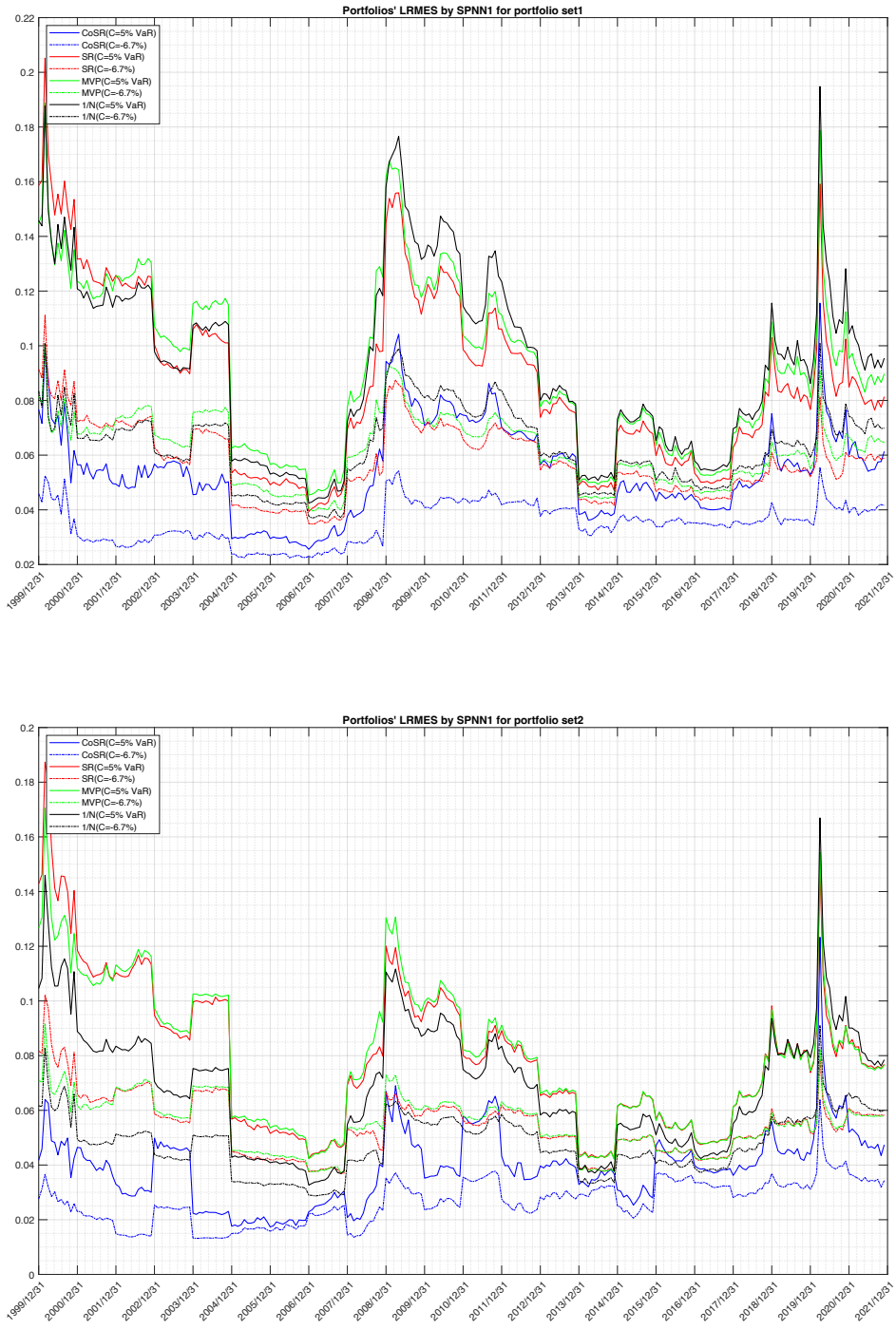



Figure 2.17: Portfolio-level LRMES by SPNN1 for portfolio set1 (top panel) and set2 (bottom panel).

# B - Tables

Table 2.1: Portfolio assets of set1

| **Firm name** | **Ticker** |
|---|---|
| Synovus Financial Corp. | SNV |
| Jefferies Financial Group Inc. | JEF |
| Cincinnati Financial Corporation | CINF |
| Comerica Incorporated | CMA |
| Loews Corporation | L |
| Vornado Realty Trust | VNO |
| Fifth Third Bancorp | FITB |
| Regions Financial Corporation | RF |
| M&T Bank Corporation | MTB |
| Franklin Resources, Inc. | BEN |
| Wells Fargo & Company | WFC |
| Huntington Bancshares Incorporated | HBAN |
| Marsh & McLennan Companies, Inc. | MMC |
| Host Hotels & Resorts, Inc. | HST |
| CNA Financial Corporation | CNA |
| JPMorgan Chase & Co. | JPM |
| Humana Inc. | HUM |
| Lincoln National Corporation | LNC |
| The Bank of New York Mellon Corporation | BK |
| Aflac Incorporated | AFL |
| Northern Trust Corporation | NTRS |
| American Express Company | AXP |
| Bank of America Corporation | BAC |
| The PNC Financial Services Group, Inc. | PNC |
| Aon plc | AON |
| Globe Life Inc. | GL |
| Cigna Corporation | CI |

Table 2.1: (continued)

| **Firm name** | **Ticker** |
|---|---|
| The Progressive Corporation | PGR |
| Public Storage | PSA |
| KeyBank | KEY |
| U.S. Bancorp | USB |
| SLM Corporation | SLM |
| American International Group, Inc. | AIG |
| SEI Investments Company | SEIC |
| Truist Financial Corporation | TFC |
| State Street Corporation | STT |
| Zions Bancorporation | ZION |
| UnitedHealth Group Incorporated | UNH |

Table 2.2: Portfolio assets of set2

| **Firm name** | **Ticker** |
|---|---|
| Coca-Cola Consolidated, Inc. | COKE |
| Apple Inc. | AAPL |
| Vulcan Materials Company | VMC |
| Associated Banc-Corp | ASB |
| Bel Fuse Inc. | BELFA |
| S&P Global Inc. | SPGI |
| FMC Corporation | FMC |
| Cardinal Health, Inc. | CAH |
| Johnson & Johnson | JNJ |
| Merck & Co., Inc. | MRK |
| Coeur Mining, Inc. | CDE |
| Communications Systems, Inc. | JCS |
| Fifth Third Bancorp | FITB |
| Rollins, Inc. | ROL |

Table 2.2: (continued)

| **Firm name** | **Ticker** |
|---|---|
| First Horizon Corporation | FHN |
| Franklin Electric | FELE |
| Weyco Group | WEYS |
| Barnes Group Inc. | B |
| Diebold Nixdorf | DBD |
| Hawkins, Inc. | HWKN |
| Barnwell Industries, Inc. | BRN |
| McDonald's Corporation | MCD |
| Safeguard Scientifics, Inc. | SFE |
| Rite Aid Corporation | RAD |
| PotlatchDeltic Corporation | PCH |
| Lee Enterprises, Inc. | LEE |
| Tenet Healthcare Corporation | THC |
| Methode Electronics, Inc. | MEI |
| PNM Resources | PNM |
| John Hancock Income Securities Trust | JHS |
| Nordstrom, Inc. | JWN |
| Southwest Airlines Co. | LUV |
| One Liberty Properties, Inc. | OLP |
| Otter Tail Corporation | OTTR |
| Owens & Minor, Inc. | OMI |
| PACCAR Inc | PCAR |
| Leggett & Platt | LEG |
| Newell Brands | NWL |
| Moog Inc. | MOG |
| Blackstone Mortgage Trust | BXMT |
| Luby's, Inc. | LUB |
| RLI Corp. | RLI |
| AT&T Inc. | T |
| Sasol Limited | SSL |

Table 2.2: (continued)

| **Firm name** | **Ticker** |
|---|---|
| Seacoast Banking Corporation of Florida | SBCF |
| Enbridge Inc. | ENB |
| Transcat, Inc. | TRNS |
| Trustco Bank | TRST |
| Agnico Eagle Mines Limited | AEM |
| Valmont Industries, Inc. | VMI |

Table 2.3: Backtesting results for portfolio set1

| Strategy | Final wealth | Annual return | MDD | MOL | TO | Sharpe ratio | Sortino ratio | Calmar ratio |
|---|---|---|---|---|---|---|---|---|
| SPNN1-based CoSR (C=$VaR^m_{5\%}$) | 20.145 | 0.146 | 0.549 | 0.203 | 0.125 | 0.754 | 1.237 | 0.266 |
| SPNN1-based CoSR (C=-6.7%) | 21.216 | 0.149 | 0.527 | 0.205 | 0.110 | 0.767 | 1.257 | 0.283 |
| SPNN2-based CoSR (C=$VaR^m_{5\%}$) | 19.908 | 0.146 | 0.587 | 0.205 | 0.127 | 0.742 | 1.199 | 0.248 |
| SPNN2-based CoSR (C=-6.7%) | 19.965 | 0.146 | 0.566 | 0.208 | 0.110 | 0.753 | 1.216 | 0.258 |
| SPNN3-based CoSR (C=$VaR^m_{5\%}$) | 18.002 | 0.140 | 0.557 | 0.202 | 0.128 | 0.734 | 1.204 | 0.252 |
| SPNN3-based CoSR (C=-6.7%) | 17.770 | 0.140 | 0.534 | 0.202 | 0.115 | 0.730 | 1.202 | 0.262 |
| GARCH-based CoSR (C=$VaR^m_{5\%}$) | 12.797 | 0.123 | 0.568 | 0.226 | 0.392 | 0.635 | 1.035 | 0.216 |
| GARCH-based CoSR (C=-6.7%) | 13.650 | 0.126 | 0.557 | 0.252 | 0.384 | 0.651 | 1.055 | 0.226 |
| Sample-based SR | 6.973 | 0.092 | 0.602 | 0.210 | 0.038 | 0.394 | 0.661 | 0.153 |
| Sample-based MVP | 5.348 | 0.079 | 0.722 | 0.196 | 0.028 | 0.317 | 0.546 | 0.110 |
| 1/N | 9.541 | 0.108 | 0.686 | 0.245 | 0.024 | 0.408 | 0.667 | 0.157 |

Table 2.4: Backtesting results for portfolio set2

| Strategy | Final wealth | Annual return | MDD | MOL | TO | Sharpe ratio | Sortino ratio | Calmar ratio |
|---|---|---|---|---|---|---|---|---|
| SPNN1-based CoSR (C=$VaR^m_{5\%}$) | 26.116 | 0.160 | 0.304 | 0.162 | 0.128 | 1.063 | 1.813 | 0.525 |
| SPNN1-based CoSR (C=-6.7%) | 25.949 | 0.160 | 0.325 | 0.159 | 0.117 | 1.043 | 1.777 | 0.491 |
| SPNN2-based CoSR (C=$VaR^m_{5\%}$) | 23.728 | 0.155 | 0.342 | 0.171 | 0.120 | 0.984 | 1.705 | 0.452 |
| SPNN2-based CoSR (C=-6.7%) | 25.382 | 0.158 | 0.351 | 0.162 | 0.106 | 1.013 | 1.743 | 0.451 |
| SPNN3-based CoSR (C=$VaR^m_{5\%}$) | 16.572 | 0.136 | 0.323 | 0.177 | 0.120 | 0.873 | 1.500 | 0.421 |
| SPNN3-based CoSR (C=-6.7%) | 20.120 | 0.146 | 0.335 | 0.171 | 0.109 | 0.945 | 1.620 | 0.436 |
| GARCH-based CoSR (C=$VaR^m_{5\%}$) | 12.435 | 0.121 | 0.305 | 0.117 | 0.181 | 1.012 | 1.985 | 0.398 |
| GARCH-based CoSR (C=-6.7%) | 12.137 | 0.120 | 0.319 | 0.120 | 0.190 | 0.988 | 1.878 | 0.377 |
| Sample-based SR | 7.093 | 0.093 | 0.461 | 0.141 | 0.041 | 0.513 | 0.907 | 0.202 |
| Sample-based MVP | 10.052 | 0.111 | 0.487 | 0.149 | 0.029 | 0.628 | 1.087 | 0.227 |
| 1/N | 15.145 | 0.131 | 0.529 | 0.192 | 0.035 | 0.635 | 1.071 | 0.249 |

Table 2.5: Backtesting results of set1 after considering proportional transaction costs (50 bps)

| Strategy | Final wealth | Annual return | MDD | MOL | Sharpe ratio | Sortino ratio | Calmar ratio |
|---|---|---|---|---|---|---|---|
| SPNN1-based CoSR (C=$VaR^m_{5\%}$) | 14.462 | 0.129 | 0.562 | 0.204 | 0.660 | 1.081 | 0.230 |
| SPNN1-based CoSR (C=-6.7%) | 15.855 | 0.134 | 0.539 | 0.206 | 0.685 | 1.119 | 0.248 |
| SPNN2-based CoSR (C=$VaR^m_{5\%}$) | 14.225 | 0.128 | 0.601 | 0.208 | 0.646 | 1.043 | 0.214 |
| SPNN2-based CoSR (C=-6.7%) | 14.912 | 0.131 | 0.580 | 0.211 | 0.668 | 1.078 | 0.225 |
| SPNN3-based CoSR (C=$VaR^m_{5\%}$) | 12.834 | 0.123 | 0.572 | 0.205 | 0.635 | 1.041 | 0.215 |
| SPNN3-based CoSR (C=-6.7%) | 13.103 | 0.124 | 0.548 | 0.204 | 0.641 | 1.055 | 0.226 |
| GARCH-based CoSR (C=$VaR^m_{5\%}$) | 4.534 | 0.071 | 0.601 | 0.230 | 0.338 | 0.586 | 0.118 |
| GARCH-based CoSR (C=-6.7%) | 4.941 | 0.075 | 0.591 | 0.256 | 0.362 | 0.618 | 0.127 |
| Sample-based SR | 6.309 | 0.087 | 0.605 | 0.210 | 0.370 | 0.624 | 0.144 |
| Sample-based MVP | 4.971 | 0.076 | 0.725 | 0.197 | 0.299 | 0.519 | 0.104 |
| 1/N | 8.950 | 0.105 | 0.689 | 0.246 | 0.394 | 0.645 | 0.152 |

Table 2.6: Backtesting results of set2 after considering proportional transaction costs (50 bps)

| Strategy | Final wealth | Annual return | MDD | MOL | Sharpe ratio | Sortino ratio | Calmar ratio |
|---|---|---|---|---|---|---|---|
| SPNN1-based CoSR (C=$VaR^m_{5\%}$) | 18.633 | 0.142 | 0.322 | 0.163 | 0.937 | 1.587 | 0.442 |
| SPNN1-based CoSR (C=-6.7%) | 19.058 | 0.143 | 0.334 | 0.161 | 0.929 | 1.575 | 0.429 |
| SPNN2-based CoSR (C=$VaR^m_{5\%}$) | 17.296 | 0.138 | 0.359 | 0.172 | 0.872 | 1.500 | 0.385 |
| SPNN2-based CoSR (C=-6.7%) | 19.174 | 0.144 | 0.367 | 0.163 | 0.912 | 1.558 | 0.391 |
| SPNN3-based CoSR (C=$VaR^m_{5\%}$) | 12.082 | 0.120 | 0.339 | 0.178 | 0.759 | 1.301 | 0.354 |
| SPNN3-based CoSR (C=-6.7%) | 15.097 | 0.131 | 0.350 | 0.173 | 0.841 | 1.434 | 0.375 |
| GARCH-based CoSR (C=$VaR^m_{5\%}$) | 7.699 | 0.097 | 0.321 | 0.122 | 0.788 | 1.528 | 0.303 |
| GARCH-based CoSR (C=-6.7%) | 7.336 | 0.095 | 0.346 | 0.125 | 0.755 | 1.427 | 0.274 |
| Sample-based SR | 6.367 | 0.088 | 0.465 | 0.142 | 0.479 | 0.850 | 0.189 |
| Sample-based MVP | 9.300 | 0.107 | 0.491 | 0.150 | 0.602 | 1.043 | 0.217 |
| 1/N | 13.826 | 0.127 | 0.533 | 0.193 | 0.610 | 1.029 | 0.238 |

Table 2.7: Statistics (set1)

| Strategy | $p$ value | $\hat{\Delta}$ | Original statistic | Block size |
|---|---|---|---|---|
| SPNN1-based CoSR (C=$VaR^m_{5\%}$) | | | | |
| Sample-based SR | 0.005 | 0.104 | 3.285 | 8 |
| Sample-based MVP | 0.000 | 0.126 | 4.468 | 10 |
| 1/N | 0.017 | 0.100 | 3.120 | 10 |
| SPNN1-based CoSR (C=-6.7%) | | | | |
| Sample-based SR | 0.005 | 0.108 | 3.727 | 10 |
| Sample-based MVP | 0.000 | 0.130 | 5.144 | 10 |
| 1/N | 0.009 | 0.104 | 3.336 | 10 |

Table 2.8: Statistics (set2)

| Strategy | $p$ value | $\hat{\Delta}$ | Original statistic | Block size |
|---|---|---|---|---|
| SPNN1-based CoSR (C=$VaR^m_{5\%}$) | | | | |
| Sample-based SR | 0.006 | 0.159 | 3.446 | 10 |
| Sample-based MVP | 0.008 | 0.126 | 3.128 | 10 |
| 1/N | 0.008 | 0.123 | 3.213 | 10 |
| SPNN1-based CoSR (C=-6.7%) | | | | |
| Sample-based SR | 0.009 | 0.153 | 3.249 | 10 |
| Sample-based MVP | 0.015 | 0.120 | 2.852 | 10 |
| 1/N | 0.013 | 0.118 | 2.881 | 10 |

# Chapter 3

# Machine learning based portfolio selection under systemic risk and asymmetry

## 3.1 Introduction

In the previous chapters, we focused on the implementation of the conditional Sharpe ratio (CoSR), which can be seen as a novel reward-risk performance measure defined over systemic event. The backtesting results have demonstrated the superiority of CoSR portfolio against popular benchmark strategies. However, one drawback of CoSR is that it cannot quantify the asymmetry in the estimated conditional return distribution. More precisely, CoSR measures conditional reward and risk through two-sided type measures (i.e. mean and standard deviation, respectively), where the positive and negative deviations from the benchmark are weighted in the same manner. It is well known that the asset return distribution is characterized by asymmetry and heavy tail (Biglova et al. 2014). These properties would be more pronounced when the market is in distress, where we describe from the generated return scenarios. In other words, the simulated asset return distributions are supposed to be more left-skewed under stressed market conditions in our case. Therefore, in order to account for the conditional asymmetry of portfolio return distribution in the presence of systemic risk, following the same spirit of Rachev ratio (RR) of Biglova et al. (2004), this chapter proposes a new performance measure, i.e., the conditional Rachev ratio (CoRR), which inherits the merits of both RR and CoSR measures. Using the same

machine learning (ML) model from chapter 2, we maximize an ex-ante CoRR measure to solve for optimal portfolios based on simulated returns in this chapter. The backtesting results are significantly improved against those from the CoSR-based approach.

### 3.1.1 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 (1966b) 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 been always 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.

Roughly speaking, systemic risk is defined as the possibility of the breakdown of a 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 a portfolio's systemic risk, see Biglova et al. (2014). Despite this urgent need, only a few studies have examined the implications of systemic risk for investment decisions; see Capponi and Rubtsov (2022) and references therein. Recently, Lin et al. (2022) studied the tradeoff between reward and risk under systemic risk by introducing a performance ratio that extends the classical Sharpe ratio. However, their measure is unable to 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.

Moreover, 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); see DeMiguel et al. (2009) and Tu and Zhou (2011) among others. 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.

### 3.1.2 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 (see below). Finally, we measure the out-of-sample performance of all portfolio candidates by various metrics in terms of both profitability and systemic risk.

### 3.1.3 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 portfolios obtained by maximizing this new measure are expected to deliver resilient performance over 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 optimization problem based on our proposed performance ratio. In that same section, we also describe the method for probabilistic forecasting, the algorithm for scenario generation, and the estimation of performance measure, consecutively. Section 4 uses a high-dimensional dataset on the US market to conduct a large-scale comparative study, in which we assess the out-of-sample performance of all portfolio candidates. Section 5 concludes. All figures and tables are reported in the Appendix at the end of this paper.

## 3.2 Quantile regression neural network

We start by briefly reviewing the traditional quantile regression (QR), which is one of the building blocks of the quantile regression neural network (QRNN). We then introduce the mathematical formulation of QRNN and its several advanced variants including smooth pinball neural network (SPNN). Before we describe our quantile models, let us first set some notations. Using the terminology of the literature on neural networks, we denote by $\boldsymbol{R} = (R_1, ..., R_V)$ the $1 \times V$ vector of monthly returns for $V$ 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 $P \times V$ matrix of $P$ covariates across $V$ training samples, 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), but we will do so in Section 3.3.

### 3.2.1 Model specification

The quantile regression (QR) proposed by Koenker and Bassett (1978) describes the relationship between conditional quantiles of the predictand given a set of predictors. Formally, the $\tau$-th conditional quantile of $R_v$ is given by

$$Q_{R_v}(\tau|\boldsymbol{X}_v) = \boldsymbol{X}_v^T \boldsymbol{\beta}(\tau), \quad \forall v \in \{1, ..., V\}, \quad \tau \in (0, 1), \tag{3.1}$$

where the column vector $\boldsymbol{\beta}(\tau) = [\beta_0(\tau), ..., \beta_P(\tau)]^T$ contains regression coefficients, and it can be estimated as

$$\hat{\boldsymbol{\beta}}(\tau) = Arg \min_{\boldsymbol{\beta}(\tau)} \frac{1}{V} \sum_{v=1}^{V} \rho_\tau \left[ R_v - \boldsymbol{X}_v^T \boldsymbol{\beta}(\tau) \right]. \tag{3.2}$$

The asymmetric loss $\rho_\tau$ (known as pinball loss or check function) is defined as

$$\rho_\tau(u) = \begin{cases} \tau u & u \geq 0 \\ (\tau - 1)u & u < 0 \end{cases}. \tag{3.3}$$

The fitted conditional quantile is expressed as

$$\hat{Q}_{R_v}(\tau | \boldsymbol{X}_v) = \boldsymbol{X}_v^T \hat{\boldsymbol{\beta}}(\tau). \tag{3.4}$$

QR provides a more complete picture of the conditional distribution of $\boldsymbol{R}$ than conditional mean regression and does not make any distributional assumption on the response variable. Moreover, QR is robust to outliers and can thus be estimated more accurately than traditional moments regression (Gonzalo and Taamouti 2017). The QR model defined in (3.1) is, however, unable to capture possible nonlinear relationships between $\boldsymbol{R}$ and $\boldsymbol{X}$. To overcome this issue, Taylor (2000) originally applied the quantile regression neural network (QRNN) that combines QR with ANN to depict the complex nonlinear relationships between predictors and the response variable without pre-specifying a functional form. Formally, the conditional $\tau$-th quantile of $R_v$ based on a QRNN model $f(\cdot)$ with a single hidden layer can be formulated as

$$Q_{R_v}(\tau | \boldsymbol{X}_v) = f\left(\boldsymbol{X}_v, \boldsymbol{H}(\tau), \boldsymbol{O}(\tau)\right) = g_2\Big[ \sum_{k=1}^{K} o_k(\tau) g_1\big( \sum_{j=1}^{P} h_{j,k}(\tau) x_j^v \big) \Big], \tag{3.5}$$

where $\boldsymbol{H}(\tau) = (h_{1,1}(\tau), ..., h_{P,K}(\tau))^T$ is the vector of weights that links the input layer with the hidden layer, $\boldsymbol{O}(\tau) = (o_1(\tau), ..., o_K(\tau))^T$ is the vector of weights responsible for connecting the hidden layer with the output layer, and $K$ is the number of hidden neurons. The activation functions $g_1(\cdot)$ and $g_2(\cdot)$ are generally specified as a sigmoid/rectifier function and a linear function, respectively. The set of parameters $\boldsymbol{\beta}(\tau) \equiv \left\{ \boldsymbol{H}(\tau), \boldsymbol{O}(\tau) \right\}$ can be

estimated as follows:

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\boldsymbol{\beta}(\tau)}{\text{Argmin}}\, L(\tau) = \underset{\boldsymbol{\beta}(\tau)}{\text{Argmin}}\, \frac{1}{V}\sum_{v=1}^{V} \rho_\tau\Big[\big(R_v - f(\boldsymbol{X}_v, \boldsymbol{\beta}(\tau))\big)\Big], \tag{3.6}$$

and the fitted conditional quantiles are obtained as $\hat{\boldsymbol{Q}}_{\boldsymbol{R}}(\tau|\boldsymbol{X}) = f(\boldsymbol{X}, \hat{\boldsymbol{\beta}}(\tau))$.

### 3.2.2 Smoothing pinball loss

Neural network parameters are typically determined via gradient-based nonlinear optimization algorithms by which the gradients are calculated using the backpropagation algorithm, see Cannon (2011). In particular, the gradient of (3.6) can be computed iteratively by updating the backpropagation equations based on the least absolute error function, see Hanson and Burr (1988). However, $\rho_\tau$ is non-differentiable at the origin ($u = 0$), which requests a smooth approximation of $\rho_\tau$ in order to apply gradient-based optimization methods.

To smooth $\rho_\tau$, one can resort to the Huber norm introduced by Huber (2004), which is defined as:

$$h(u) = \begin{cases} \dfrac{1}{2}u^2 & |u| \le \varepsilon \\ \varepsilon(|u| - \dfrac{1}{2}\varepsilon) & \text{otherwise} \end{cases}, \tag{3.7}$$

where $\varepsilon$ denotes a threshold value; see Chen (2007), Cannon (2011), Cannon (2018), and Xu et al. (2017) for empirical applications of Huber norm. The check function is approximated by

$$\rho_\tau^{(A)}(u) = |\tau - I\{u < 0\}|\, h(u), \tag{3.8}$$

where $I\{u < 0\}$ refers to an indicator function that is equal to one when $u < 0$ and zero otherwise. An alternative way to smooth the loss function was proposed by Zheng (2011), which smoothes $\rho_\tau$ using a logistic function:

$$\rho_\tau^{(A)}(u) = \tau u + \alpha\, \ln\big(1 + \exp(-\frac{u}{\alpha})\big), \tag{3.9}$$

where $\alpha > 0$ is the smoothing parameter. As argued by Arends et al. (2020), the loss function in equation (3.9) combines Huber loss and pinball loss together. Zheng (2011) has shown that $\rho_\tau^{(A)}(u) = \rho_\tau(u)$ as $\alpha \to 0^+$. By applying $\rho_\tau^{(A)}$ in (3.6), we obtain the following

updated objective function

$$L^{(A)}(\tau) = \frac{1}{V}\sum_{v=1}^{V} \rho_{\tau}^{(A)}\Big[\big(R_v - f(\boldsymbol{X}_v, \boldsymbol{\beta}(\tau))\big)\Big], \tag{3.10}$$

that we can now minimize using standard gradient-based optimization algorithms to obtain the estimate of $\boldsymbol{\beta}(\tau)$. In our empirical analysis, we adopt the logistic loss (3.9).[1]

### 3.2.3 Smooth pinball neural network

To further enhance the performance of estimating quantiles, Xu et al. (2017) extended the original QRNN model to composite quantile regression neural network (CQRNN), by which we can estimate multiple conditional quantiles (for different values of $\tau$) simultaneously and efficiently. CQRNN inherits one of the same capabilities as linear composite quantile regression (CQR) developed by Zou and Yuan (2008), i.e., combining multiple quantile regressions to better capture complex nonlinear relationships between the predictors and the predictand (Cannon 2018). Formally, CQRNN is similar to QRNN, and the only difference between the two lies in the objective function, which is now summed over $M$ values of $\tau$:

$$L_C^{(A)} = \frac{1}{M}\sum_{m=1}^{M} L^{(A)}(\tau_m), \tag{3.11}$$

where $\tau$ is equally spaced as $\tau_m = \frac{m}{M+1}$ for $m \in \{1, \cdots, M\}$. The expression in (3.11) is a composite version of the objective function in equation (3.10) since it evaluates multiple conditional quantiles synchronously. CQRNN is a flexible model not only because it allows us to uncover 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. As stated by Ouali et al. (2016), quantile crossing might result in an invalid predictive distribution of the predictand. Similarly, Bang et al. (2016) argued that this problem can make the estimation of regression quantiles less efficient and cause problems in the subsequent analysis. In

[1] We have also tried for the Huber loss and the backtesting results are similar to those of using logistic loss.

order to mitigate this issue, Cannon (2018) developed a monotonic CQRNN (MCQRNN) model that imposes monotonicity constraints on a standard MLP 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.

Recently, Hatalis et al. (2019) proposed an efficient alternative to MCQRNN namely smooth pinball neural network (SPNN) that introduces a set of constraints into the CQRNN framework. To prohibit the crossing between two neighbouring quantiles, the constraint $Q_{R_v}(\tau_1|\boldsymbol{X}_v) \leq \cdots \leq Q_{R_v}(\tau_M|\boldsymbol{X}_v),\ \forall v$ needs to be satisfied. However, it is difficult to solve the optimization problem via gradient-based methods with such constraints. To solve this, Hatalis et al. (2019) proposed to add a penalty term $p$ to the objective function (3.11), where $p$ is defined as

$$p = c\frac{1}{MV}\sum_{m=1}^{M}\sum_{v=1}^{V}\Big[\max\Big(0, \epsilon - \big(\hat{Q}_{R_v}(\tau_m|\boldsymbol{X}_v) - \hat{Q}_{R_v}(\tau_{m-1}|\boldsymbol{X}_v)\big)\Big)\Big]^2, \tag{3.12}$$

where $\hat{Q}_{R_v}(\tau_0|\boldsymbol{X}_v)$ is initialized to zero, $\epsilon$ 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 (3.11). Thus, the cost function of SPNN is defined as

$$L_S = L_C^{(A)} + p + \lambda||\boldsymbol{\beta}||_1, \tag{3.13}$$

where $\boldsymbol{\beta} \equiv \{\boldsymbol{H}, \boldsymbol{O}\} = \{\boldsymbol{H}(\tau_m), \boldsymbol{O}(\tau_m)\}_{m=1,\ldots,M}$ represents composite parameters of the neural network (i.e. parameters across all values of $\tau$), $||\cdot||_1$ refers to the $l_1$ norm, and $\lambda$ denotes the regularization parameter that controls model complexity and mitigates the overfitting problem. The training of SPNN can be conducted using standard gradient-based optimization algorithms. In our paper, we adopt SPNN for completing prediction tasks due to its virtues of simultaneously estimating multiple quantiles and preventing quantile crossing.

## 3.3 Portfolio selection under non-Gaussianity and systemic risk

In this section, we first review several existing performance ratios, in particular, the Rachev ratio, which we extend to propose our new measure that allows for non-Gaussianity and accounts for systemic risk. Next, we formulate the portfolio selection problem using the proposed measure. Furthermore, we discuss an algorithm for generating return scenarios, which we use to estimate performance measures (ours and the benchmark measures) and solve for the optimal portfolios.

### 3.3.1 Review of performance measures

#### Conditional Sharpe ratio under systemic risk

The new performance measure that we propose can be seen as an alternative to the conditional Sharpe ratio (CoSR) of Lin et al. (2022). The CoSR measure is defined as

$$\text{CoSR}(R_p) := \frac{\text{CoER}(R_p)}{\text{CoSD}(R_p)}, \tag{3.14}$$

where CoER denotes the conditional reward measure, which is defined as the conditional first moment of the portfolio's active return:

$$\text{CoER}(R_p) := E(R_p - R_b|\text{SE}), \tag{3.15}$$

where $R_p$ represents portfolio return, $R_b$ denotes benchmark rate (which we set as the market return $R_m$ in our empirical analysis), and $\text{SE} = \{R_m < C\}$ denotes some systemic event (SE) during which the market return over the next month goes below a certain threshold $C$. Analogously, the conditional risk measure CoSD is defined as the conditional second moment of the portfolio's active return:

$$\text{CoSD}(R_p) := \big[Var(R_p - R_b|\text{SE})\big]^{1/2}. \tag{3.16}$$

By maximizing CoSR in an ex-ante analysis, we are able to construct portfolios that perform relatively resilient during crisis periods.

### Conditional Rachev ratio under non-Gaussianity

Although by construction, the CoSR measure is able to account for systemic risk, it does not, unfortunately, allow for non-Gaussian (asymmetric) return distributions. One way to overcome this issue is to separately measure reward and loss using one-sided type parameter-dependent measures, e.g., the Rachev ratio and the Farinelli-Tibiletti ratio. By a proper choice of parameters, they can be adapted to investors' profiles expressing their preferences for asymmetric deviations away from the benchmark with non-Gaussian data; see Farinelli et al. (2009). To account for both non-Gaussianity and systemic risk, based on the unconditional Rachev ratio of Biglova et al. (2004), Biglova et al. (2014) propose a conditional Rachev ratio ($\text{CoRR}^{\text{Biglova}}$), which is defined as

$$\text{CoRR}^{\text{Biglova}}(R_p;\alpha,\beta) := \frac{E(R_p - R_b | R_1 \geq -\text{VaR}_{1-\alpha}(R_1), ..., R_N \geq -\text{VaR}_{1-\alpha}(R_N))}{-E(R_p - R_b | R_1 \leq -\text{VaR}_{\beta}(R_1), ..., R_N \leq -\text{VaR}_{\beta}(R_N))}, \tag{3.17}$$

where $R_i$ for $i = 1, ..., N$ denotes asset i's return and $\text{VaR}_\alpha(R) = -\inf\{x | Pr(R \leq x) > \alpha\}$ is the unconditional value-at-risk (VaR) of return. However, it is worth noting that the conditional information in this measure cannot be interpreted as a SE. Specifically, the $\text{CoRR}^{\text{Biglova}}$ measure does not connect systemic risk with the occurrence of market distress, instead, it evaluates portfolio performance conditional on the occurrence of idiosyncratic (individual) risk events for all portfolio assets under consideration. 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. 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. The same predicament also holds when estimating the risk measure of $\text{CoRR}^{\text{Biglova}}$. In our empirical analysis, we will provide more computational details related to this issue.

### CoES-based conditional Rachev ratio

Unlike Biglova et al. (2014), in our paper, a systemic event (SE) occurs when the market return goes below a certain threshold $C$ over a time horizon $h$. 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 systemic risk index, for example, the S&P 500 Index, that reflects broad market conditions. And the investors aim to maximize an ex-ante Rachev ratio conditional on the systemic

risk index being below $C$ between time $t$ and $t+h$. By implementing our investment strategy, one can find portfolios that deliver the best tradeoff between reward and risk of the portfolio's active return under stressed market conditions and non-Gaussian return distribution.

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 VaR of firm $i$'s return obtained conditioning on some systemic event $C(R_m)$ observed for the market portfolio, say $\text{CoVaR}_\alpha^{i|C(R_m)}$, is implicitly defined as

$$Pr(R_i \leq -\text{CoVaR}_\alpha^{i|C(R_m)}) = \alpha, \quad \alpha \in (0,1). \tag{3.18}$$

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:

$$\text{CoETL}(R_p;\alpha) := -E(R_p - R_b | R_p - R_b \leq -\text{CoVaR}_\alpha^{p|\text{SE}}). \tag{3.19}$$

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

$$\text{CoETP}(R_p;\alpha) := E(R_p - R_b | R_p - R_b \geq \text{CoVaR}_{1-\alpha}^{p|\text{SE}}), \tag{3.20}$$

which measures the mean gains that are greater than the $(1-\alpha)$-conditional percentile of $(R_p - R_b)$. Finally, based on the terms (3.19) and (3.20), our new measure is defined as

$$\text{CoRR}(R_p;\alpha,\beta) := \frac{\text{CoETP}(R_p;\alpha)}{\text{CoETL}(R_p;\beta)}, \tag{3.21}$$

where the two performance levels $\alpha$ and $\beta$ 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.,

$$\text{SE} = \left\{R_m < -\text{VaR}_\alpha(R_m)\right\}. \tag{3.22}$$

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. (2022) and consider $R_b = R_m$.[2]

### 3.3.2 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. 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

$$\boldsymbol{W}_t^* = \arg\max_{\boldsymbol{W}_t} \rho(R_{p,t+1}), \quad \text{s.t. } \mathbf{1}^T \boldsymbol{W}_t = 1, \tag{3.23}$$

where the different candidates of $\rho(\cdot)$ result in different optimal portfolios. In particular, the portfolio selection problem under CoRR is given by

$$\boldsymbol{W}_t^* = \arg\max_{\boldsymbol{W}_t} \ \text{CoRR}(R_{p,t+1}; \alpha, \beta), \quad \text{s.t. } \mathbf{1}^T \boldsymbol{W}_t = 1. \tag{3.24}$$

In practice, it is often the case for investors to place additional constraints on the optimiza-

[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 benchmark to the market index (i.e. $R_b = R_m$) with the proposed approach.

tion. 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 portfolios constructed based on historical return observations. Specifically, we consider the unconditional Sharpe ratio and the negatively signed portfolio variance as alternative objective functions for $\rho$ in (3.23) under short selling restrictions, where the resulting optimal portfolios are denoted by SR and MVP, respectively. The second contains the CoSR portfolio proposed by Lin et al. (2022), which we solve it using simulated return scenarios. The last consists of the well-diversified equal-weighted portfolio (1/N), which does not rely on any model estimation.

### 3.3.3 Simulation of return scenarios

Although CoSR has no closed-form expression in dynamic settings when the non-short-selling constraint is imposed, we can still use a Monte-Carlo simulation-based procedure to implement our ML and systemic risk-based portfolio. The dynamic SPNN-based CoSR can be estimated using its empirical analogue that we can calculate from simulated returns over the subset of simulated crisis 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.

#### 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 $p_{j,t} = p(R_{j,t+1}|\boldsymbol{X}_{j,t})$

following Cannon (2011). 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 $\tau_m$ and $\tau_{m+1}$ are such that $\hat{q}_{j,t+1}(\tau_m) \leq R_{j,t+1} < \hat{q}_{j,t+1}(\tau_{m+1})$, then

$$\hat{p}_{j,t} = \frac{\tau_{m+1} - \tau_m}{\hat{q}_{j,t+1}(\tau_{m+1}) - \hat{q}_{j,t+1}(\tau_m)}. \tag{3.25}$$

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

$$\hat{p}_{j,t} = z_1 \exp\Big(-\frac{|R_{j,t+1} - \hat{q}_{j,t+1}(\tau_1)|}{e_1}\Big), \tag{3.26}$$

  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

$$\hat{p}_{j,t} = z_M \exp\Big(-\frac{|R_{j,t+1} - \hat{q}_{j,t+1}(\tau_M)|}{e_M}\Big), \tag{3.27}$$

  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 above 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).

### 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, the multivariate distribution function of individual firm and market monthly returns, can be resolved into univariate margins and a certain copula function

$$F_{R_1,\ldots,R_{N+1}}(u_1,\ldots,u_{N+1}) = C\left(F_{R_1}(u_1),\ldots,F_{R_{N+1}}(u_{N+1})\right), \tag{3.28}$$

where $u_j \sim U(0,1)$ for $j = 1,\ldots,N+1$, $R_{N+1} = R_m$, 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

$$C_{\nu,\boldsymbol{\mathcal{P}}}(u_1,...,u_{N+1}) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_{N+1})} \frac{\Gamma(\frac{\nu+N+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^{N+1}|\boldsymbol{\mathcal{P}}|}} \Big(1+\frac{x'\boldsymbol{\mathcal{P}}^{-1}x}{\nu}\Big)^{-\frac{\nu+N+1}{2}} dx, \tag{3.29}$$

where $\Gamma$ denotes the Gamma function, $\boldsymbol{\mathcal{P}}$ represents the correlation matrix, and $\nu$ refers to the degrees of freedom both for margins and copula function. We now generate future 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 CDF, say $\hat{F}_{\nu_{j,t}}$, of return series $\{R_{j,t}\}$ using a univariate t-location-scale distribution, 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 $\nu$ and the correlation matrix $\boldsymbol{\mathcal{P}}$ of the t-copula, see McNeil et al. (2015).

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

- Convert $\boldsymbol{u}_{t+1}^{(s)}$ to return scenarios via quantile transformation: $R_{j,t+1}^{(s)} = \hat{F}^{-1}_{R_{j,t+1}}(u_{j,t+1}^{(s)})$, where $\hat{F}^{-1}_{R_{j,t+1}}$ 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.

### 3.3.4 CoRR estimation

Before starting the estimation, we set up some more notations. 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, $\#\text{SE} = \sum_{s=1}^{S} I\{R_{m,t+1}^s < -\widehat{\text{VaR}}_q(R_{m,t+1})\}$ 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 em-

pirical $q$th-quantile of the simulated market returns, say $\widehat{\text{VaR}}_q(R_{m,t+1})$, for $q = 1\%, 5\%$.[3] Analogously, the CoVaR of the portfolio return can be implicitly estimated by the $\alpha$-th empirical quantile of the conditional probability distribution of portfolio active return:

$$Pr(\tilde{\boldsymbol{R}}^{sim}_{p,t+1|SE} \leq -\widehat{\text{CoVaR}}^{p|\text{SE}}_{\alpha}) := Pr(\boldsymbol{R}^{sim}_{t+1|\text{SE}}\boldsymbol{W}_t - \boldsymbol{R}^{sim}_{m,t+1|\text{SE}} \leq -\widehat{\text{CoVaR}}^{p|\text{SE}}_{\alpha}) = \alpha, \quad (3.30)$$

where $\boldsymbol{R}^{sim}_{t+1|\text{SE}}$ and $\boldsymbol{R}^{sim}_{m|\text{SE}}$ denote $\#\text{SE} \times N$ matrix and $\#\text{SE} \times 1$ column vector of the simulated returns for portfolio assets and market portfolio that satisfy SE condition (hereafter we use the word "filtered" to refer to SE-truncated scenarios), respectively.

Let $\tilde{\boldsymbol{R}}^{sim}_{p,t+1|SE} = (\tilde{R}^1_{p,t+1|SE}, ..., \tilde{R}^{\#\text{SE}}_{p,t+1|SE})^T$ refer to the $\#\text{SE} \times 1$ vector of filtered return scenarios of portfolio active return, and $\#\text{TLE} = \sum_{s=1}^{\#\text{SE}} I\{\tilde{R}^s_{p,t+1|SE} \leq -\widehat{\text{CoVaR}}^{p|\text{SE}}_{\alpha}\}$ is the number of scenarios out of #SE that represents the conditional tail loss event (TLE). Using the above, the CoETL in (3.19) can be estimated as

$$\widehat{\text{CoETL}}_t(R_{p,t+1}; \alpha) = -\frac{\sum_{s=1}^{\#\text{SE}} \tilde{R}^s_{p,t+1|SE} I\{\tilde{R}^s_{p,t+1|SE} \leq \widehat{\text{CoVaR}}^{p|\text{SE}}_{\alpha}\}}{\#\text{TLE}}. \quad (3.31)$$

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

$$\widehat{\text{CoETP}}_t(R_{p,t+1}; \alpha) = \frac{\sum_{s=1}^{\#\text{SE}} \tilde{R}^s_{p,t+1|SE} I\{\tilde{R}^s_{p,t+1|SE} \geq \widehat{\text{CoVaR}}^{p|\text{SE}}_{1-\alpha}\}}{\#\text{TPE}}. \quad (3.32)$$

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

$$\widehat{\text{CoRR}}_t(R_{p,t+1}; \alpha, \beta) = \frac{\widehat{\text{CoETP}}_t(R_{p,t+1}; \alpha)}{\widehat{\text{CoETL}}_t(R_{p,t+1}; \beta)}. \quad (3.33)$$

## 3.4 Empirical analysis

### 3.4.1 Data

We perform our empirical analysis based on a monthly cross-sectional US dataset that spans from January 1985 to December 2021. In this section, we first provide details of the

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

predictor set and then discuss the choice of portfolio assets. All the tables in this section are included in the Appendix.

### Description of predictors

Following Gu et al. (2020), we adopt 94 monthly stock-level explanatory variables.[4] The original data can be retrieved from Dacheng Xiu's website. The corresponding variable screening procedure was implemented by Green et al. (2013). We manually matched this dataset with CRSP monthly returns. The equities presented in this original dataset cover the firms listed on NASDAQ, AMEX, and NYSE ranging from 1965 till 2021. The detailed variable definitions can be found in the Online Appendix F of Gu et al. (2020).

Except for the stock-level characteristics, we additionally 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).

Lastly, we include industry dummy variables following Gu et al. (2020). In summary, our predictor set contains 94 stock-specific variables, 14 macroeconomic variables, and 74 industry dummy variables. Throughout our empirical studies, the explanatory variables of use are the covariates defined as follows:

$$\boldsymbol{x}_{i,t} = \boldsymbol{m}_t \otimes \boldsymbol{f}_{i,t}, \tag{3.34}$$

where the vector $\boldsymbol{f}_{i,t}$ covers 94 characteristics for the $i$-th firm, and the vector $\boldsymbol{m}_t$ loads macroeconomic variables plus a constant value $C$. Therefore, $\boldsymbol{x}_{i,t}$ denotes the predictor vector including interaction terms between firm-specific and macro state variables. The resulting 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.

[4] We manually computed the value-weighted average of characteristics for the S&P 500 market portfolio using the 500 highest market cap companies. The correlation coefficient between S&P 500 return and the constructed market return is beyond 0.99.

However, their original data involves a large number of variables with missing values.[5] After deleting those missing data, the remaining sample spans from January 1985 to December 2021. To alleviate the computational burden associated with neural 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.

### The choice of portfolio assets

As argued by Lin et al. (2022), 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, although not necessarily classified as Systemically Important Financial Institutions (SIFIs). Therefore, we consider large financial firms as portfolio assets in our empirical analysis.[6]

In November of 2021, the Financial Stability Board announced a list of Global SIFIs (G-SIFIs) after negotiation with the Basel Committee on Banking Supervision. The total number of G-SIFIs contained in that list is 30, among which 5 are traded on the US market throughout our sample period. Besides, the Board of Governors of the US Federal Reserve System maintains a list of Domestic SIFIs (D-SIFIs). This list includes financial firms whose size is not large enough for being classified as G-SIFIs, but still considered to be systemically important in the domestic market. According to the list released by the Federal Reserve as of March 2014, 23 banks traded on the US stock market were identified as D-SIFIs. Among those D-SIFIs, 12 are traded throughout our sample period. Thus, from the above, we obtain a list of 17 SIFIs consisting of 5 G-SIFIs and 12 D-SIFIs.

Following Brownlees and Engle (2016), we select large financial institutions with a market capitalization bigger than 5 billion US dollars by June 2007. After applying this filter

[5] All data before January 1985 contains at least one variable with a large portion of missing observations. Thus, it is impractical to fill in those missing variables with the monthly cross-sectional medians as implemented by Gu et al. (2020). Because of this missing data problem, filling with medians might not be suitable since it will negatively affect the predictive power of some predictors. Therefore, we decide to only focus on the sample period without missing observations.

[6] Note that although the financial institutions might possess higher systemic risk than non-financial firms, our aim is not to only minimize the systemic risk of a portfolio but also maximize its profit under stressed market condition by maximizing a performance measure. It might be the case that systemic firms also exhibit a positive active return so it may be profitable to invest in them.

criterion to our dataset, we are left with a list of 38 assets that covers the aforementioned 17 SIFIs. Therefore, we finally obtain a list of 38 portfolio assets (hereafter set1) including 17 SIFIs and 21 non-SIFIs. These firms are listed in Table 3.1.

### 3.4.2 Estimation and selection of SPNN model

#### Sample splitting

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 carefully choose the optimal hyperparameters. The choice of hyperparameters helps control the model's complexity and determine the model's predictive power as well. 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 20% of cross-sectional data of each in-sample for all 577 firms and the market, with the first 80% of the samples included in the training subsample. 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 $\mathcal{L}_2$.[7] After choosing the best hyperparameters, we refit our model using the in-sample data on ($\mathcal{L}_1 + \mathcal{L}_2$) and the resulting estimates are used for obtaining final quantile forecasts over the out-of-sample $\mathcal{L}_3$. As for data preprocessing, we standardize covariates so each has a

[7] QS is used for assessing quantile forecasts, which accounts for both reliability and sharpness; see Hong et al. (2016). Formally, $\text{QS} = \frac{1}{\#M \times H \times (N+1)} \sum_{m \in M} \sum_{t=1}^{H} \sum_{j=1}^{N+1} \rho\left(R_{j,t}, \hat{q}_{j,t}(\tau_m)\right)$, where $M$ denotes the set of prespecified quantiles (we set $M = \{1, 2, ..., 99\}$), $R_{j,t}$ is the realized return of individual firm or market, and $H$ indicates the forecast horizon ($H = 12$ in our implementation).

zero mean and unit variance. 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; see Gu et al. (2020) and Kynigakis and Panopoulou (2021).

### 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).

Following the literature, we adopt the Rectified Linear Unit (ReLU) $g(x) = \max(0, x)$ as the activation function, which promotes sparsity in the weights and enables efficient computation of gradients as well (see Nair and Hinton 2010 and Glorot et al. 2011 among others). For the output layer, we apply the identity activation function $g(x) = x$ following Hatalis et al. (2019).

Neural network models are usually trained via Stochastic Gradient Descent (SGD) (Robbins and Monro 1951). Unlike the standard gradient descent that computes gradients within each iteration on the complete training samples, SGD carries out the computation on a random segment of the training data and performs backpropagation iteratively. As argued by Gu et al. (2020), this operation sacrifices the accuracy in exchange for substantially speeding up the training process.[8]

### 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 applying $l_1$ penalization, we consider additional DL techniques

[8] We adopt the Adam optimizer proposed by Kingma and Ba (2014) for model training.

including batch training, batch normalization, early stopping, and forecast averaging; see also Gu et al. (2020) and Kynigakis and Panopoulou (2021) for implementations of these regularization methods.[9]

#### Hyperparameters

We use a two-dimensional grid search approach to select optimal 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 $\lambda_1$ 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 $lr \in [10^{-3}, 10^{-2}]$.

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.[10] Note that unlike Gu et al. (2020) who set the batch size equal to 10,000, we choose to use 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. Keskar et al. (2016) stated that using a large batch tends to suffer from a generalization drop due to sharp minima, see also Bengio (2012) and Masters and Luschi (2018) for the preference of using a small batch. 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.

### 3.4.3 Portfolio formation

After fitting the SPNN models, we obtain quantile forecasts of monthly returns, based on which we estimate the conditional marginal return distributions following the method discussed in Section 3.3.3. 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 (3.24) is solved on a monthly basis by maximizing the ex-ante CoRR measure based on generated return scenarios, where the CoRR is

[9] 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 (3.13).

[10] 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.

estimated by (3.33). To obtain a robust estimator of our CoRR measure, we follow Biglova et al. (2014) and set relatively large parameters of $\alpha = \beta = 10\%$.

For the comparison purpose, we evaluate the out-of-sample performance of CoRR portfolios against those of four benchmark portfolios, namely the ML-based CoSR portfolio of Lin et al. (2022), the sample-based SR portfolio, the sample-based MVP, and the 1/N portfolio.[11] To reduce the estimation error of the sample covariance matrix, we applied the shrinkage estimator developed by Ledoit and Wolf (2004) to SR portfolio and MVP. Furthermore, we keep using active returns for all portfolios to ensure the comparability of backtesting results. In addition, we also consider S&P 500 market portfolio as a fundamental benchmark. To avoid the composition of portfolios that allocate large negative weights to all assets under SE conditions, we do not allow for short sales in our analysis. The initial values of wealth and cumulative return at the beginning of the backtesting period (December 1999) are set to one and zero respectively, i.e., $\mathrm{FW}_0 = 1$ and $\mathrm{CR}_0 = 0$.

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 Section 3.3.3, and obtain the optimal weights $\boldsymbol{W}_{k+1}^*$ for each of the performance measures under consideration. This step is performed using the Matlab built-in function *fmincon*.[12] Then, we compute the final wealth as

$$\mathrm{FW}_{k+1} = \mathrm{FW}_k(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1}), \tag{3.35}$$

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

$$\mathrm{CR}_{k+1} = \mathrm{CR}_k + \ln(1 + \boldsymbol{W}_k^{*T}\boldsymbol{R}_{k+1}). \tag{3.36}$$

Note that the latter equation reports the cumulative performance of the portfolio net of wealth. That is, expression (3.35) 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 (3.36). By repeatedly computing $\mathrm{FW}_{k+1}$ and $\mathrm{CR}_{k+1}$ for different

[11] We have considered adding the $\mathrm{CoRR}^{\mathrm{Biglova}}$ measure of Biglova et al. (2014) as an additional benchmark, however, following their model setting, we were not able to get enough subsets to estimate the measure based on our dataset. We have further increased the simulation sample size from 30,000 to 60,000, but the resulting subsamples are still very limited and thus cannot be used to obtain a robust estimator. We thus exclude it from our comparison set.

[12] Following Kresta et al. (2015), we randomly choose 20 starting points to approach the global optimum.

objective functions, we obtain the ex-post paths of final wealth and cumulative return over the evaluation period.

### 3.4.4 Results

In this section, we first illustrate return quantile forecasts and examine the predictive power of predictors using two variable importance measures namely mean squared sensitivity (MSS) and quantile causality measure (QC). Thereafter, we provide backtesting results with and without accounting for transaction costs. Finally, we calculate the portfolio's long-run marginal expected shortfall (LRMES) and CoES to compare the level of systemic risk generated by different strategies under investigation. All the figures and tables that are related to the empirical analysis can be found in a separate online appendix.

#### Quantile forecasts and variable importance

To present some insights on the return quantile forecasts using SPNN models, in Figure 3.1 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 assets (CMA, WFC and JPM).[13] From Figure 3.1, we see that the return quantile forecasts are able to capture most of the variation of realized returns, especially during crisis episodes.

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) and Kynigakis and Panopoulou (2021) 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):

$$\text{MSS}_{p,m} = \sqrt{\frac{\sum_{t\in(\mathcal{L}_1+\mathcal{L}_2)}\left(s_{p,m}|\boldsymbol{x}_t\right)^2}{|\mathcal{L}_1|+|\mathcal{L}_2|}}, \tag{3.37}$$

with

$$s_{p,m}\big|_{\boldsymbol{X}_t} = \frac{\partial \hat{Q}_{R_{t+1}}(\tau_m|\boldsymbol{X}_t)}{\partial x_{p,t}}(\boldsymbol{X}_t), \tag{3.38}$$

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

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 $\tau_m$-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 (3.38) 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

$$\widetilde{\text{MSS}}_p = \frac{1}{M} \sum_{m=1}^{M} \text{MMS}_{p,m}. \tag{3.39}$$

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 (2022), 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

$$\text{QC}_p(\tau) = \ln\left[\frac{E\left[\rho_\tau\left(R_{t+1} - Q_{R_{t+1}}(\tau|\overline{\boldsymbol{X}}_t)\right)\right]}{E\left[\rho_\tau\left(R_{t+1} - Q_{R_{t+1}}(\tau|\boldsymbol{X}_t)\right)\right]}\right], \tag{3.40}$$

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 $\tau$-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 $\tau$-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

$$\widetilde{\text{QC}}_p = \ln\left[\frac{\frac{1}{M|\mathcal{L}_3|}\sum_{m=1}^{M}\sum_{t\in\mathcal{L}_3}\rho_{\tau_m}\left(R_{t+1} - \hat{Q}_{R_{t+1}}(\tau_m|\overline{\boldsymbol{X}}_t)\right)}{\frac{1}{M|\mathcal{L}_3|}\sum_{m=1}^{M}\sum_{t\in\mathcal{L}_3}\rho_{\tau_m}\left(R_{t+1} - \hat{Q}_{R_{t+1}}(\tau_m|\boldsymbol{X}_t)\right)}\right], \tag{3.41}$$

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 fitted SPNN1 model, Figure 3.2 reports the variable importance measured by MSS for the 10 most influential firm-level predictors and all macroeconomic variables under consideration, while Figure 3.3 reports the results of the corresponding variable

importance measured by QC.[14] The variable 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 3.2 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 represents liquidity variables like dollar volume (dolvol), debt capacity/firm tangibility (tang), bid-ask spread (baspread), turnover (turn), and number of zero trading days (zerotrade); the third group corresponds to a single momentum predictor namely the short-term reversal (mom1m); the fourth group is given by the R&D expense-to-market ratio (valuation ratio); and the last group consists of industry dummy (sic2). As for the macroeconomic variables, from the bottom panel of Figure 3.2, we see that all of them contribute significantly to model training, but among those, we find that the total financial uncertainty index (macro_TFU) can be ranked as the most influential macro-level predictor.

Analogously, the rankings based on the QC measure as shown in Figure 3.3 draw similar conclusions. The results reveal a fairly small set of dominant firm-level predictors, which covers the risk measures total and idiosyncratic return volatility, the liquidity variables dollar volume, industry-adjusted size (mve_ia), bid-ask spread and turnover, the short-term reversal, the valuation ratio of total debt-to-capitalization ratio (lev), and an accounting variable that indicates the number of years since first Compustat coverage (age). For the macro variables, the results confirm again their predictive power and place the greatest emphasis on the total financial uncertainty index.

To further illustrate the variable importance, Figures 3.4 to 3.7 display the time-varying rankings of the predictors in the SPNN1 model as measured by MSS and QC. 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 recursive 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 predictors as identified before.

[14] 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.

### 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 performance measures that are estimated from either simulated or historical return observations over the evaluation period. All portfolios are monthly rebalanced.

The backtesting results are displayed in Figure 3.8.[15] There are several noticeable features from these figures. Firstly, we observe that all candidate portfolios outperform the market S&P 500 portfolio. Secondly, all portfolios perform less well during the 2007-2008 financial crisis. The SR, MVP and 1/N strategies lose almost all of their values during that period, while the SPNN-CoSR portfolios perform significantly better than others, even though they lost around half of their values since the last peak in 2007. In particular, within SPNN-CoSR portfolios, the SPNN1-based strategy delivers the best out-of-sample performance for both SE thresholds. Thirdly, all SPNN-based CoSR portfolios 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. Thus, the backtesting results confirm the benefits of combining SPNN-based return forecasts with the incorporation of systemic risk into the traditional mean-variance framework when constructing optimal portfolios.

Table 3.2 reports the values of several statistics that are used to measure portfolio performance. The results vary among different strategies depending on the estimators of the performance measure employed during portfolio optimization, with the exception being the 1/N portfolio which does not rely on any optimization or model estimation. Overall, the SPNN1-based CoRR portfolios perform the best in terms of out-of-sample profitability, whichever SE threshold is being considered. Moreover, the SPNN1-based

[15] We omit the backtesting results obtained by SPNN4 and SPNN5 since we found that the portfolio performance starts to deteriorate from SPNN3. Our findings are in agreement with recent studies, see, for example, Gu et al. (2020) and Kynigakis and Panopoulou (2021), where the authors argued that "shallow" learning outperforms "deep" learning. Increasing the model complexity does not necessarily benefit us in terms of economic gains. However, the remaining results using other SPNN configurations under consideration are available upon request.

CoRR portfolios outperform SPNN1-based CoSR portfolios by a wide margin, with the latter being considered advanced ML-based benchmarks. Specifically, using SPNN1-based CoRR portfolios with SE thresholds $C1$ and $C2$, investors would multiply their wealth by 46.385 and 35.031 respectively, which are near twice that of SPNN1-based CoSR portfolios (15.052 for $C1$ and 20.141 for $C2$). The sample-based MVP offers the lowest final wealth (5.348) and annual return (0.079), while the sample-based SR portfolio performs the second worst with a final wealth of 6.973 and an annual return of 0.092. Interestingly, the naive 1/N strategy outperforms all sample-based portfolios in terms of profitability, with the former exhibiting a final wealth of 9.541 and an annual return of 0.108. The results of Sharpe ratio, Sortino ratio and Calmar ratio demonstrate again the superiority of our proposed approach. The SPNN1-based CoRR portfolio with $C2$ delivers the highest values of Sharpe ratio (0.811) and Sortino ratio (1.368), while the SPNN1-based CoRR portfolio with $C1$ presents the highest Calmar ratio (0.413) 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 ratio (FT) as alternative metrics. Formally, the MDD is calculated as

$$\text{MDD} = \max_{t_0 \leq t_1 \leq t_2 \leq T_0} \left\{ r_{p,t_0:t_1} - r_{p,t_0:t_2} \right\}, \tag{3.42}$$

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

$$\text{TO} = \frac{1}{T} \sum_{t=1}^{T} \left( \sum_{i=1}^{N} \left| \omega_{i,t+1} - \frac{\omega_{i,t}(1 + R_{i,t+1})}{1 + \sum_{j=1}^{N} \omega_{j,t} R_{j,t+1}} \right| \right), \tag{3.43}$$

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

$$\text{FT}(R_p; p, q) = \frac{\left( E(R_p - R_b)_+^p \right)^{1/p}}{\left( E(R_b - R_p)_+^q \right)^{1/q}}, \tag{3.44}$$

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 3.2 reports the values of the above-mentioned alternative measures. Overall, both ML-based CoSR and CoRR portfolios provide lower MDD than the rest whichever systemic risk-based performance measure is used. In particular, the SPNN1-based CoRR portfolio with $C1$ presents the lowest MDD of 0.461, while the sample-based MVP delivers the highest MDD of 0.722. In terms of the FT ratio, the SPNN1-based CoRR portfolios dominate other benchmark strategies, regardless of the choice of SE threshold. This indicates that our proposed approach achieves better performance under different asymmetric preferences depending on different choices of partial moment orders.

**Effect of transaction costs**

The calculation of transaction cost (TC) is based on TO as defined in (3.43). After accounting for a proportional TC of $c$, the portfolio return is now calculated as follows:

$$\tilde{R}_{p,t+1} = (1 + R_{p,t+1})\left(1 - c\sum_{i=1}^{N}\left|\omega_{i,t+1} - \frac{\omega_{i,t}(1 + R_{i,t+1})}{1 + \sum_{j=1}^{N}\omega_{j,t}R_{j,t+1}}\right|\right) - 1. \tag{3.45}$$

Given the major role that momentum predictors play in ML models, it is perhaps unsurprising that our SPNN-based trading strategies are characterized by relatively high TO, see also Gu et al. (2020). As we can see from Table 3.2, the SPNN1-based CoSR portfolio with $C1$ has the highest TO of 0.210, while the SPNN1-based CoRR with $C2$ provides the second highest TO of 0.180. The sample-based portfolios possess much lower TO (0.038 for SR and 0.028 for MVP) at the cost of less profitability. Unsurprisingly, the 1/N portfolio delivers the lowest TO 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 3.9 illustrates the ex-post paths of

final wealth and cumulative return after taking into account TC, whereas Table 3.3 reports the relevant performance measures. In short, we find that the inclusion of proportional TC does not alter our main conclusions. Generally speaking, the SPNN1-based CoRR portfolios still outperform all other competitors in terms of profitability and performance metrics. Remarkably, the final wealth of SPNN-based CoRR portfolios is more than twice that of SPNN1-based CoSR portfolios and is five times more than that of sample-based portfolios.

### 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. 2022):

$$\text{LRMES}_p = \sum_{i=1}^{N} \omega_i \, \text{LRMES}_i, \tag{3.46}$$

where $\text{LRMES}_i$ indicates the expected loss of asset $i$ over next month. The $\text{LRMES}_p$ 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

$$\text{CoES}_\alpha^{p|\text{SE}} = \sum_{i=1}^{N} \omega_i \, \text{CoES}_\alpha^{i|\text{SE}}, \tag{3.47}$$

where $\text{CoES}_\alpha^{i|\text{SE}} = E(R_i | R_i \leq \text{CoVaR}_\alpha^{i|\text{SE}})$ refers to the expected tail loss of asset $i$ conditional on market distress.[16] 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 3.10 illustrates the time-varying portfolio's LRMES and CoES over the evaluation period.[17] Overall speaking, the SPNN-based CoRR portfolios offer the best performance in terms of systemic risk. The relatively low values of LRMES and CoES indicate that these portfolios will suffer from less potential losses during crisis periods. Specifically, the SPNN1-based CoRR portfolio with $C1$ provides the lowest LRMES over the first half of

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

[17] To save space, we only show the results obtained by SPNN1 here. However, the corresponding results for other SPNN configurations are available upon request.

the evaluation period, while the SPNN1-based CoRR portfolio with $C2$ becomes the winner over the second half. The SPNN1-based CoSR portfolio with $C2$ is a serious competitor that presents slightly higher LRMES in the middle of the evaluation period. Similarly, the SPNN1-based CoRR and CoSR portfolios with $C1$ deliver the lowest and the second-lowest CoES among all candidate competitors, respectively. Thus, one can conclude that both ML-based strategies are able to minimize the portfolio-level systemic risk compared to other benchmark strategies, among which the CoRR measure performs the best.

## 3.5 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 market events, instead of relying on historical returns only, 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 our proposed portfolio selection approach against popular benchmarks. Our backtesting results demonstrate the superiority of the SPNN-based CoRR portfolio in terms of profitability, with its outperformance staying robust after the inclusion of moderate transaction costs. Last but not least, we also compare the portfolio-level systemic risk among all candidates using the portfolio's LRMES and CoES. Our SPNN-based CoRR portfolio, while characterized by the highest profitability, it delivers the lowest systemic risk throughout the evaluation period.

# Appendix A - Figures

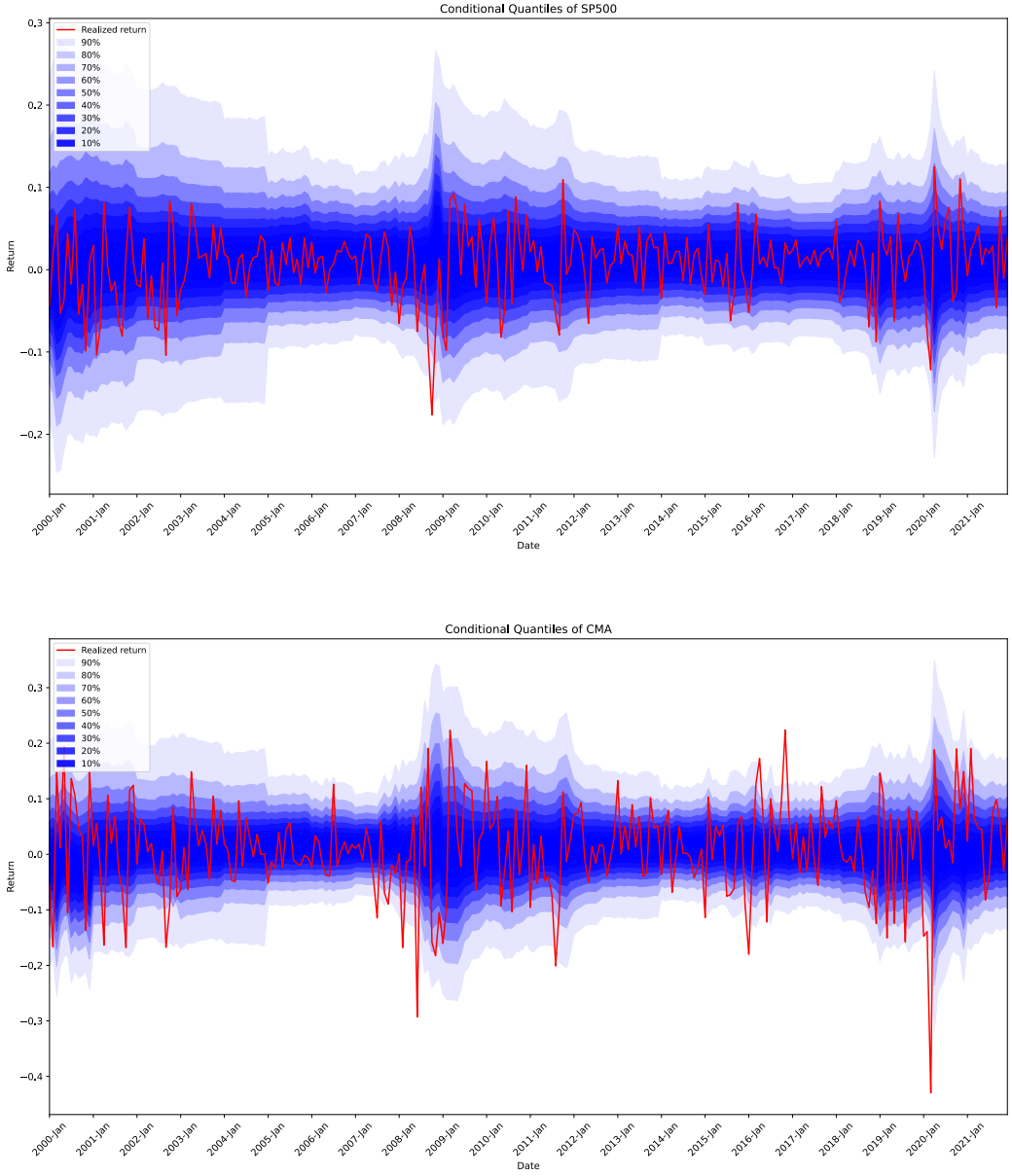



Figure 3.1: Conditional quantiles of returns on market index and a few individual firms obtained from SPNN1.

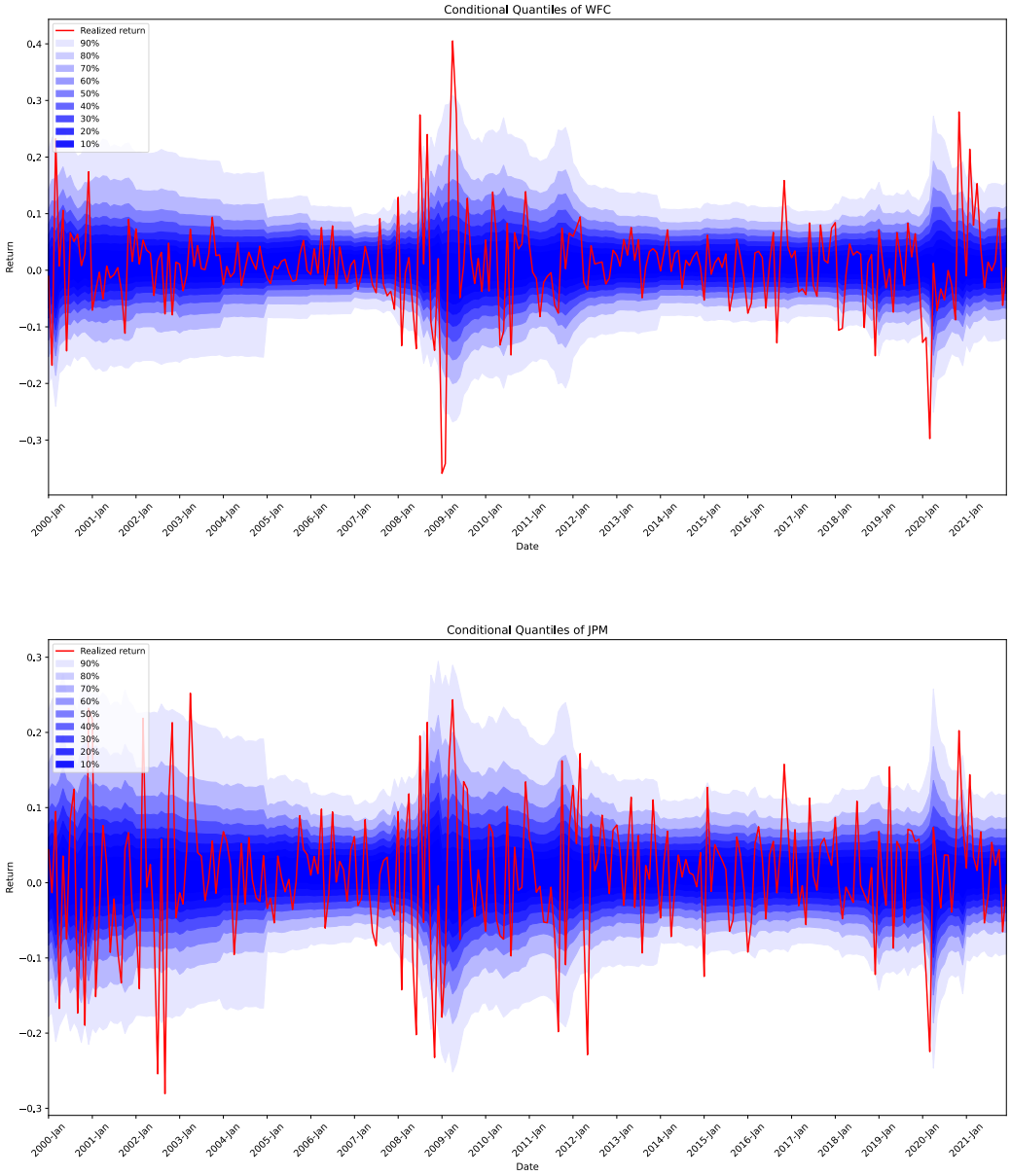



Figure 3.1: (continued)

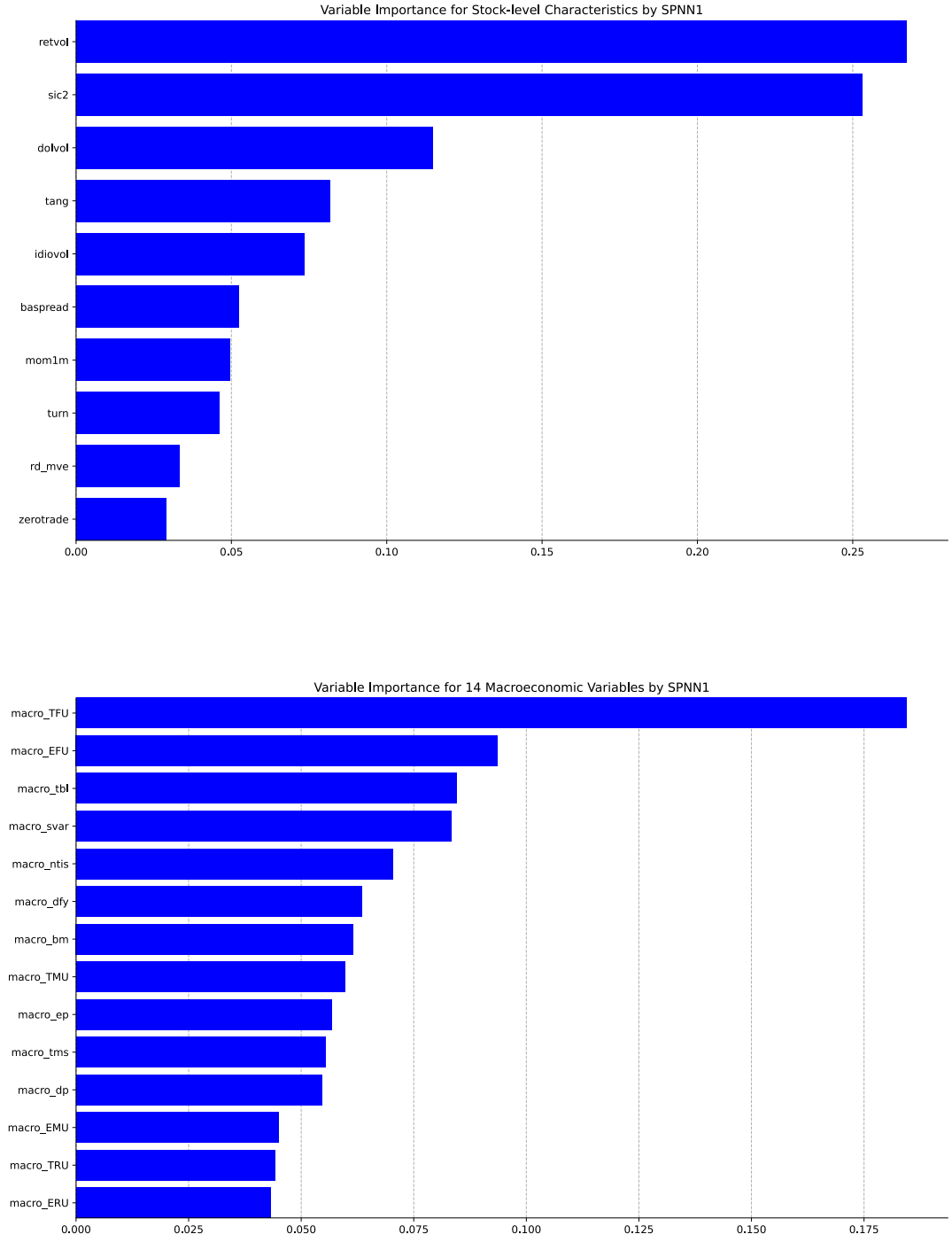



Figure 3.2: Top panel displays the top-10 most influential firm-level predictors in SPNN1 measured by MSS, while the bottom panel reports the corresponding results for all macroeconomic variables. Variable importance is an average over all quantiles and recursive in-sample windows. Variable importance is normalized to sum to one.

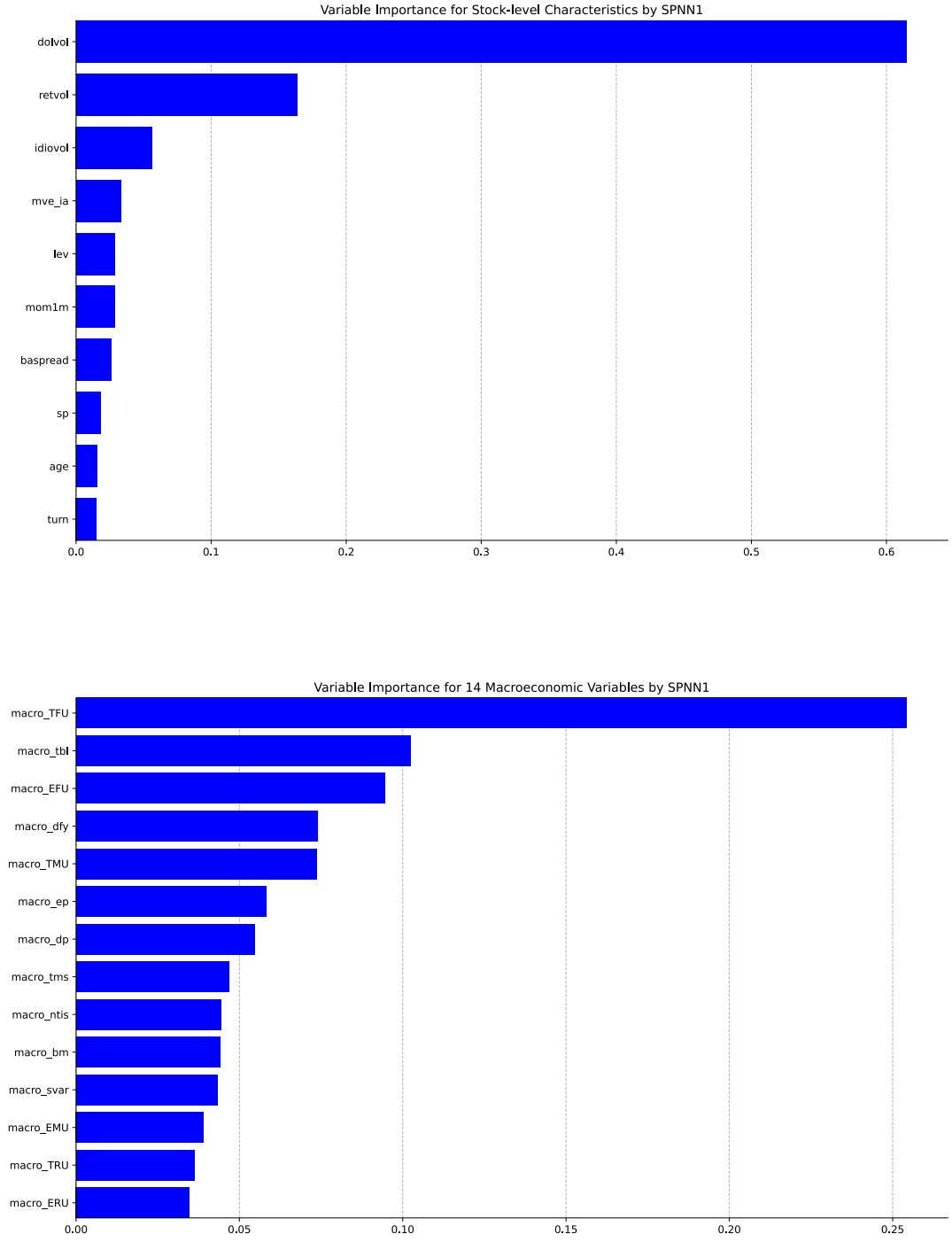



Figure 3.3: Top and bottom panels display the variable importance of top-10 most influential firm-level predictors and all macroeconomic variables measured by QC in SPNN1, respectively. Variable importance is an average over all quantiles and recursive in-sample windows. Variable importance is normalized to sum to one.

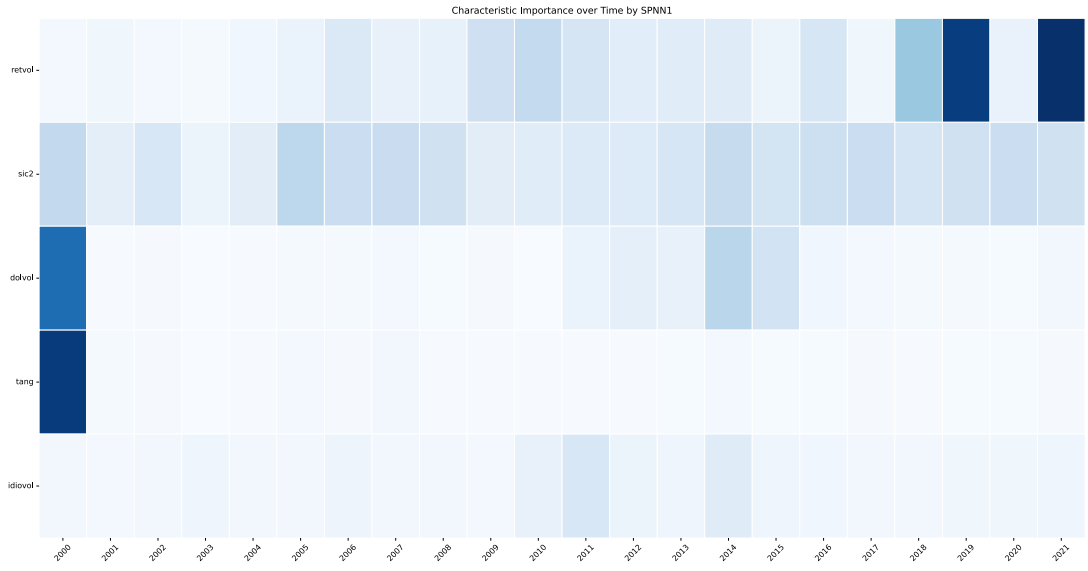



Figure 3.4: Time-varying variable importance of the top-5 most influential firm-level predictors measured by MSS. Predictors are ordered based on the average value of their MSS over recursive trainings, with the most influential features at the top and the least influential at the bottom. Columns correspond to the year end of each of the 22 in-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

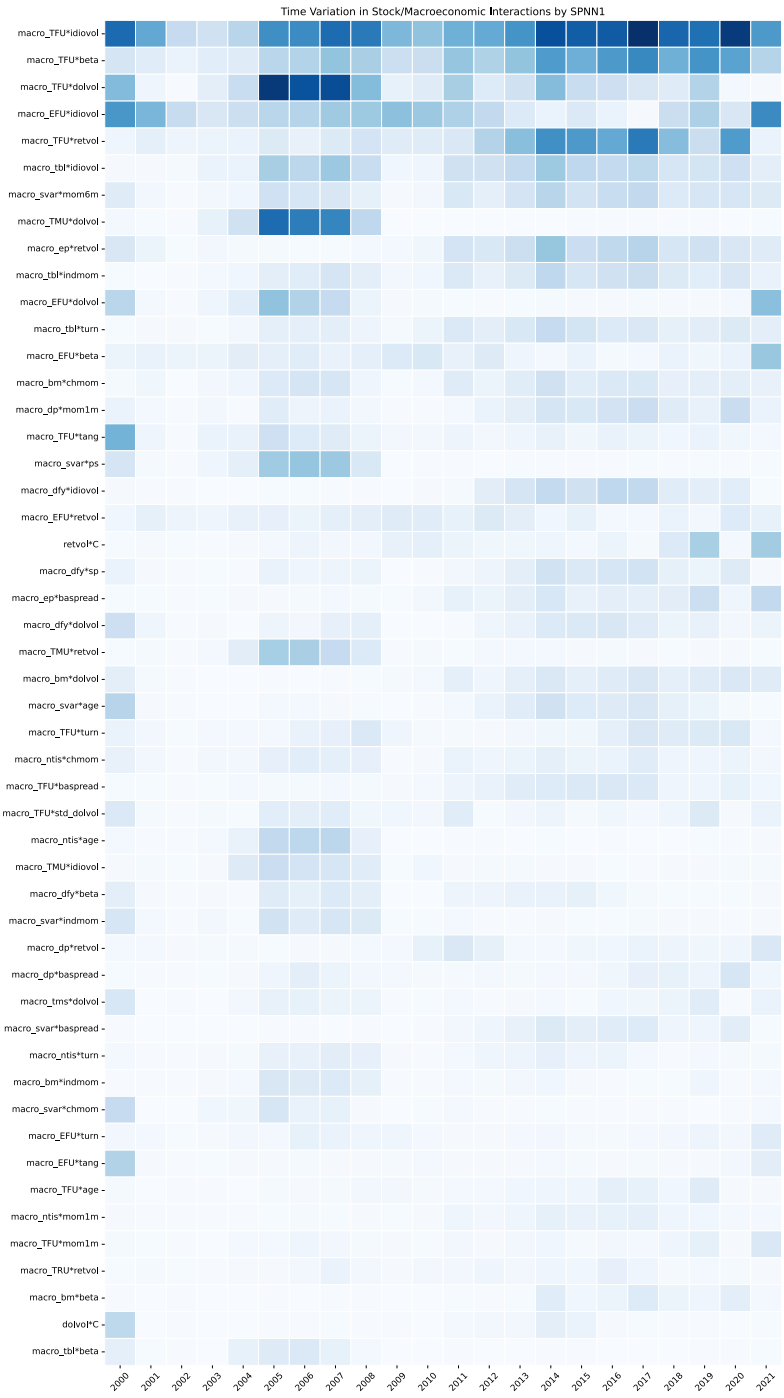



Figure 3.5: Time-varying variable importance of the top-50 most influential predictors of interactions between each firm characteristic with macroeconomic variables measured by MSS. Columns correspond to the year end of each of the 22 in-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

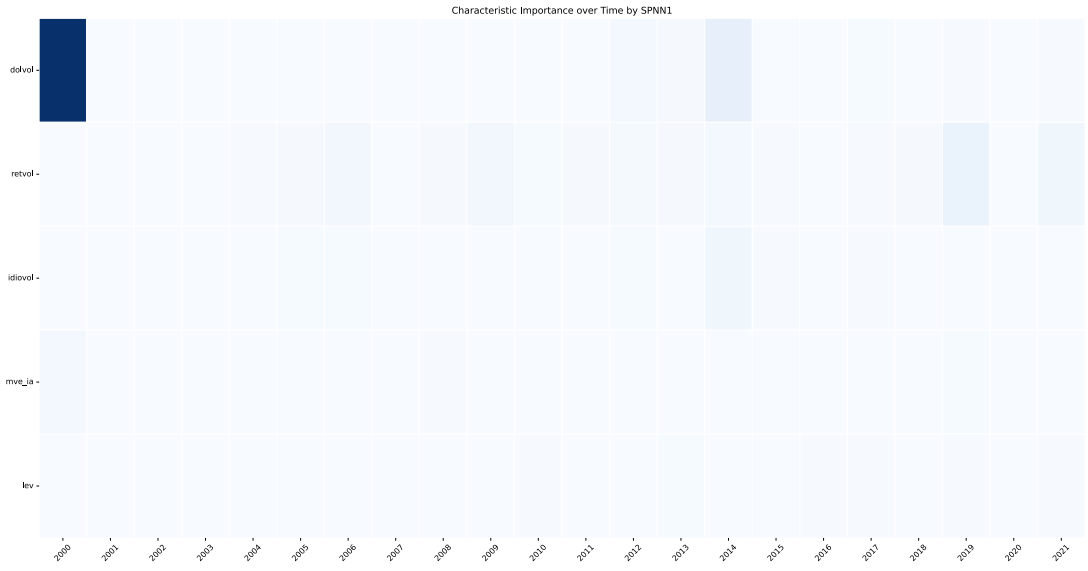



Figure 3.6: Time-varying variable importance of the top-5 most influential firm-level predictors measured by QC. Columns correspond to the year start of each of the 22 out-of-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

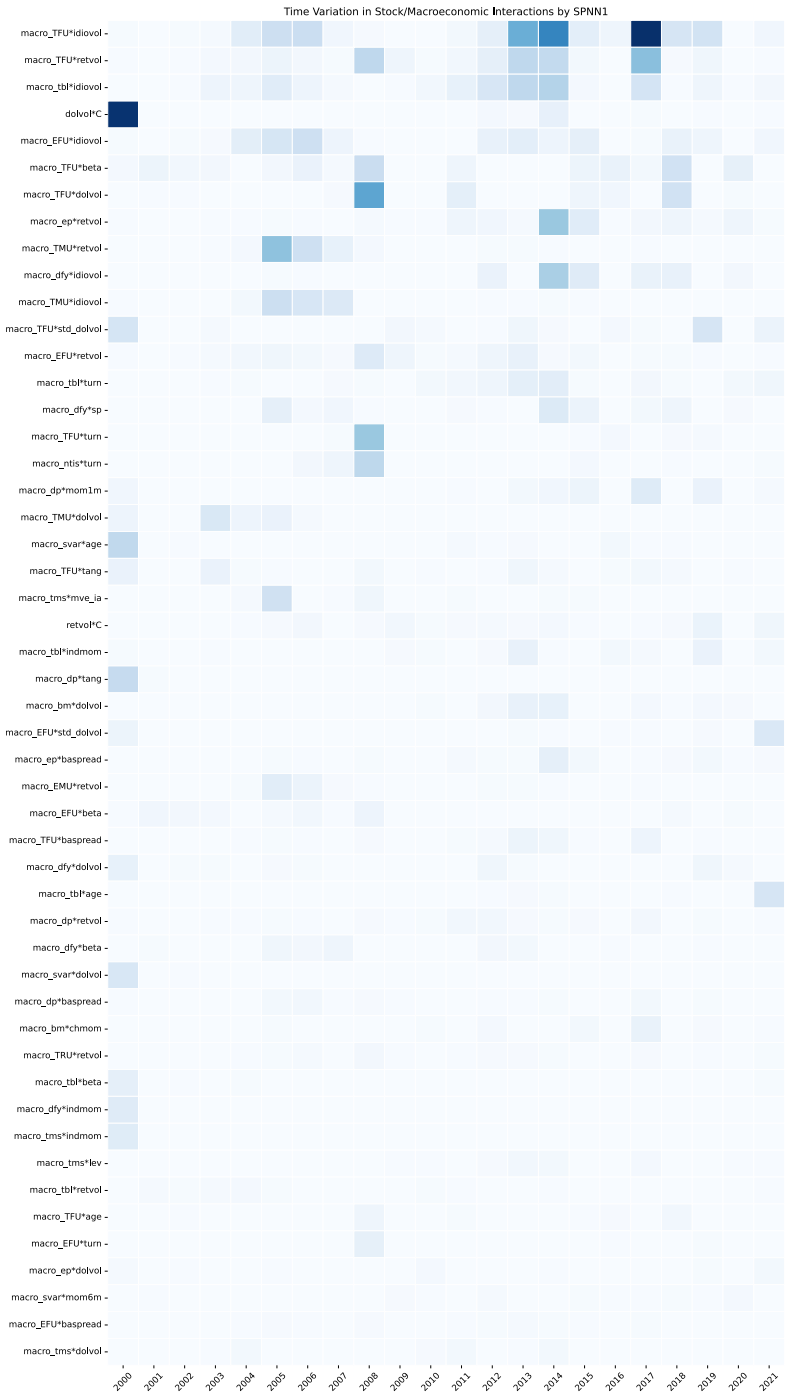



Figure 3.7: Time-varying variable importance of the top-50 most influential predictors of interactions between each firm characteristic with macroeconomic variables measured by QC. Columns correspond to the year start of each of the 22 out-of-sample windows, and color gradients within each column indicate the most influential (dark blue) to least influential (white) variables.

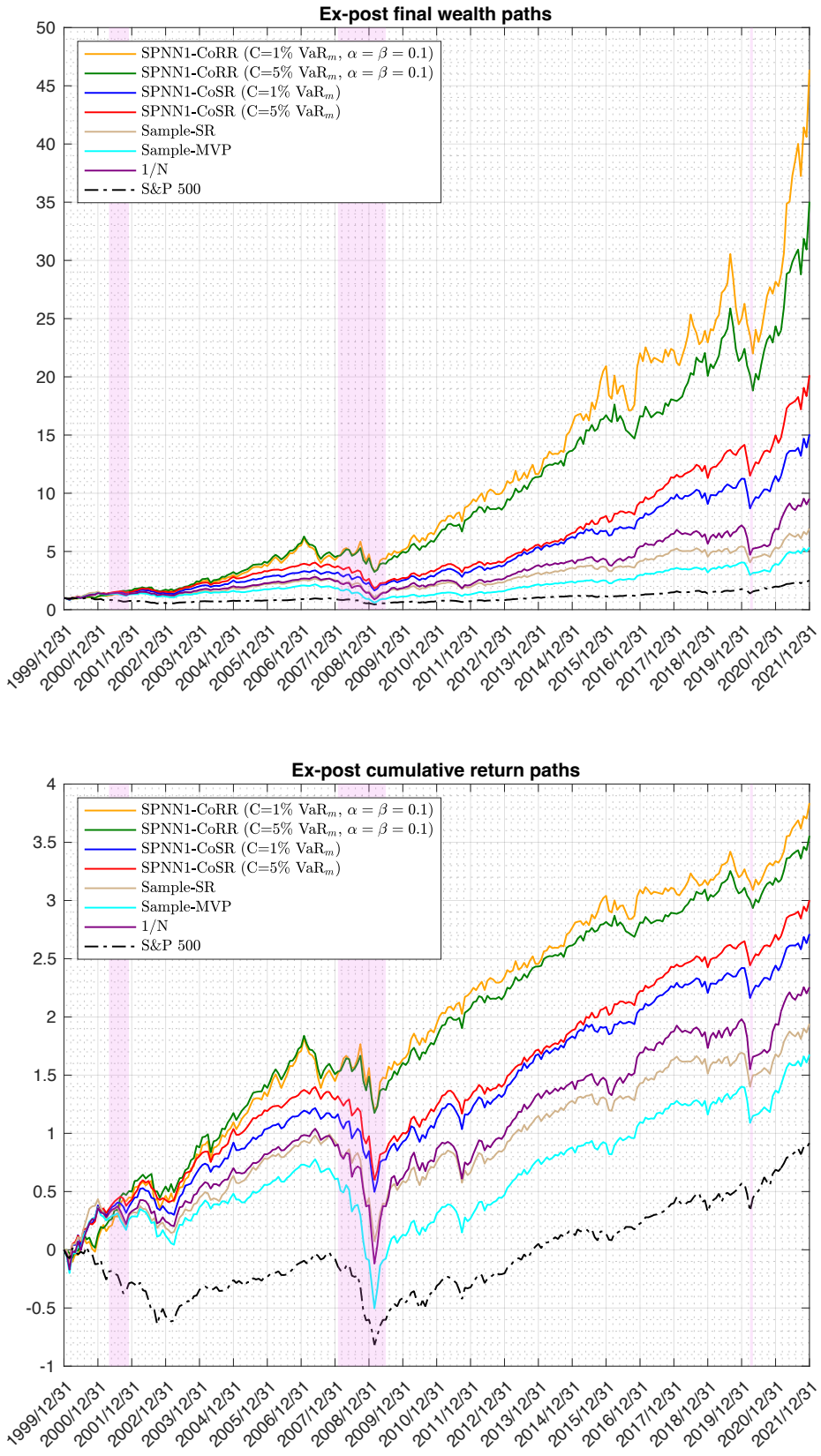



Figure 3.8: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies. The shaded areas denote recession periods as defined by NBER.

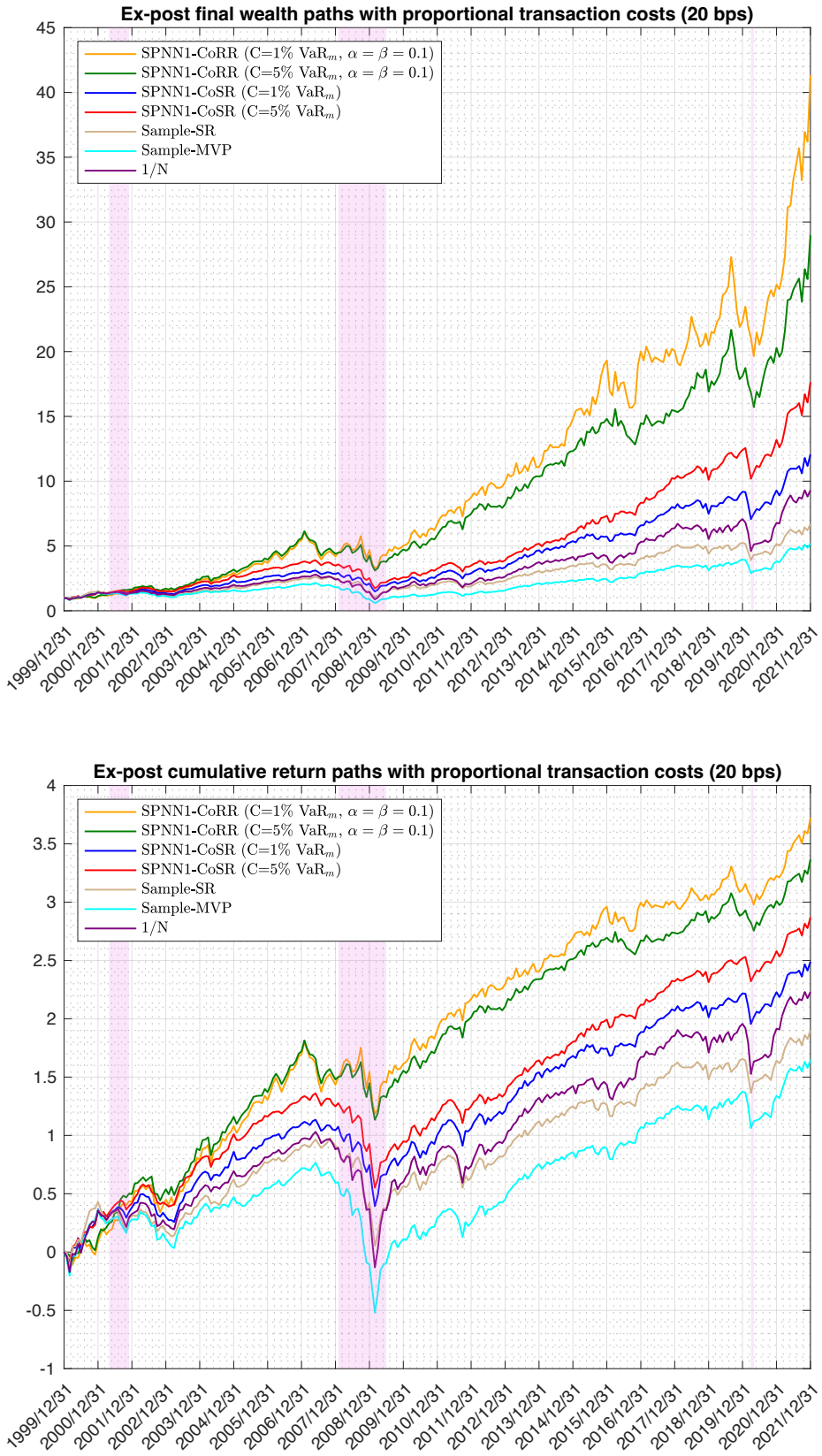



Figure 3.9: Ex-post final wealth (top panel) and ex-post cumulative return (bottom panel) paths obtained using different strategies with 20 bps proportional TC. The shaded areas denote recession periods as defined by NBER.

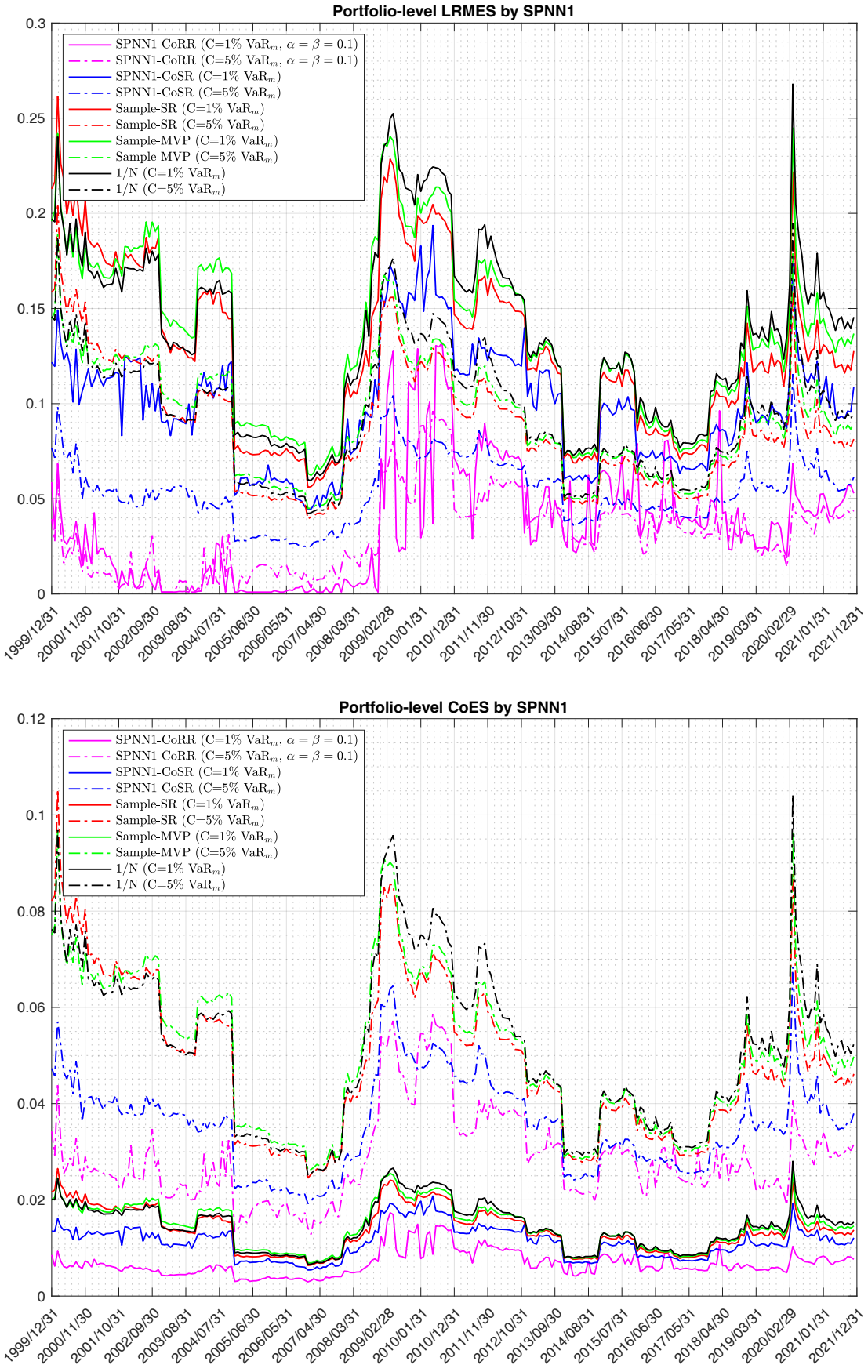



Figure 3.10: Portfolio-level LRMES and CoES estimated from simulated returns.

# Appendix B - Tables

Table 3.1: Portfolio assets

| **Firm name** | **Ticker** |
|---|---|
| Synovus Financial Corp. | SNV |
| Jefferies Financial Group Inc. | JEF |
| Cincinnati Financial Corporation | CINF |
| Comerica Incorporated | CMA |
| Loews Corporation | L |
| Vornado Realty Trust | VNO |
| Fifth Third Bancorp | FITB |
| Regions Financial Corporation | RF |
| M&T Bank Corporation | MTB |
| Franklin Resources, Inc. | BEN |
| Wells Fargo & Company | WFC |
| Huntington Bancshares Incorporated | HBAN |
| Marsh & McLennan Companies, Inc. | MMC |
| Host Hotels & Resorts, Inc. | HST |
| CNA Financial Corporation | CNA |
| JPMorgan Chase & Co. | JPM |
| Humana Inc. | HUM |
| Lincoln National Corporation | LNC |
| The Bank of New York Mellon Corporation | BK |
| Aflac Incorporated | AFL |
| Northern Trust Corporation | NTRS |
| American Express Company | AXP |
| Bank of America Corporation | BAC |
| The PNC Financial Services Group, Inc. | PNC |
| Aon plc | AON |
| Globe Life Inc. | GL |
| Cigna Corporation | CI |

Table 3.1: (continued)

| **Firm name** | **Ticker** |
|---|---|
| The Progressive Corporation | PGR |
| Public Storage | PSA |
| KeyBank | KEY |
| U.S. Bancorp | USB |
| SLM Corporation | SLM |
| American International Group, Inc. | AIG |
| SEI Investments Company | SEIC |
| Truist Financial Corporation | TFC |
| State Street Corporation | STT |
| Zions Bancorporation | ZION |
| UnitedHealth Group Incorporated | UNH |

Table 3.2: Backtesting results without transaction costs

| Performance measure | SPNN1-CoRR($C1$) | SPNN1-CoRR($C2$) | SPNN1-CoSR($C1$) | SPNN1-CoSR($C2$) | Sample-SR | Sample-MVP | 1/N |
|---|---|---|---|---|---|---|---|
| Final wealth | 46.385 | 35.031 | 15.052 | 20.141 | 6.973 | 5.348 | 9.541 |
| Annual return | 0.191 | 0.175 | 0.131 | 0.146 | 0.092 | 0.079 | 0.108 |
| MDD | 0.461 | 0.485 | 0.513 | 0.549 | 0.602 | 0.722 | 0.686 |
| MOL | 0.224 | 0.194 | 0.183 | 0.203 | 0.210 | 0.196 | 0.245 |
| TO | 0.109 | 0.180 | 0.210 | 0.125 | 0.038 | 0.028 | 0.024 |
| Sharpe ratio | 0.760 | 0.811 | 0.645 | 0.754 | 0.394 | 0.317 | 0.408 |
| Sortino ratio | 1.292 | 1.368 | 1.095 | 1.237 | 0.661 | 0.546 | 0.667 |
| Calmar ratio | 0.413 | 0.362 | 0.256 | 0.266 | 0.153 | 0.110 | 0.157 |
| FT ratio(p=1,q=1) | 1.070 | 1.090 | 1.040 | 0.878 | 0.907 | 0.916 | 0.886 |
| FT ratio(p=1,q=2) | 0.781 | 0.760 | 0.714 | 0.601 | 0.633 | 0.609 | 0.589 |
| FT ratio(p=1,q=3) | 0.614 | 0.589 | 0.554 | 0.450 | 0.492 | 0.466 | 0.443 |
| FT ratio(p=1,q=4) | 0.511 | 0.489 | 0.462 | 0.365 | 0.412 | 0.389 | 0.363 |

Table 3.3: Backtesting results with proportional transaction costs (20 bps)

| Performance measure | SPNN1-CoRR($C1$) | SPNN1-CoRR($C2$) | SPNN1-CoSR($C1$) | SPNN1-CoSR($C2$) | Sample-SR | Sample-MVP | 1/N |
|---|---|---|---|---|---|---|---|
| Final wealth | 41.336 | 28.975 | 12.062 | 17.641 | 6.700 | 5.194 | 9.300 |
| Annual return | 0.184 | 0.165 | 0.120 | 0.139 | 0.090 | 0.078 | 0.107 |
| MDD | 0.461 | 0.494 | 0.523 | 0.555 | 0.603 | 0.723 | 0.687 |
| MOL | 0.224 | 0.195 | 0.184 | 0.203 | 0.210 | 0.197 | 0.245 |
| Sharpe ratio | 0.736 | 0.763 | 0.584 | 0.716 | 0.384 | 0.310 | 0.403 |
| Sortino ratio | 1.248 | 1.284 | 0.993 | 1.174 | 0.646 | 0.535 | 0.658 |
| Calmar ratio | 0.400 | 0.335 | 0.229 | 0.251 | 0.150 | 0.108 | 0.155 |
| FT ratio(p=1,q=1) | 1.052 | 1.081 | 1.004 | 0.853 | 0.907 | 0.911 | 0.887 |
| FT ratio(p=1,q=2) | 0.771 | 0.754 | 0.698 | 0.588 | 0.629 | 0.606 | 0.590 |
| FT ratio(p=1,q=3) | 0.607 | 0.584 | 0.544 | 0.441 | 0.490 | 0.464 | 0.444 |
| FT ratio(p=1,q=4) | 0.505 | 0.484 | 0.454 | 0.358 | 0.410 | 0.388 | 0.364 |

# Bibliography


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., R. Engle, and M. Richardson (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review 102*(3), 59–64.

Acharya, V. V., C. Brownlees, R. Engle, F. Farazmand, and M. Richardson (2010). Measuring systemic risk. *Regulating Wall Street: The Dodd-Frank Act and the New Architecture of Global Finance*, 85–119.

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.

Allen, F. and E. Carletti (2013). What is systemic risk? *Journal of Money, Credit and Banking 45*(s1), 121–127.

Allen, L., T. G. Bali, and Y. Tang (2012). Does systemic risk in the financial sector predict future economic downturns? *The Review of Financial Studies 25*(10), 3000–3036.

Ang, A. and F. A. Longstaff (2013). Systemic sovereign credit risk: Lessons from the US and Europe. *Journal of Monetary Economics 60*(5), 493–510.

Arends, E. L., S. J. Watson, S. Basu, and B. Cheneka (2020). Probabilistic wind power forecasting combining deep learning architectures. In *2020 17th International Conference on the European Energy Market (EEM)*, pp. 1–6. IEEE.

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

Bang, S., S.-H. Eo, Y. M. Cho, M. Jhun, and H. Cho (2016). Non-crossing weighted kernel quantile regression with right censored data. *Lifetime Data Analysis 22*(1), 100–121.

Basak, S. and A. Shapiro (2001). Value-at-risk based risk management: Optimal policies and asset prices. *Review of Financial Studies 14*, 371–405.

Bassett, G. W., R. Koenker, and G. Kordas (2004). Pessimistic portfolio allocation and Choquet expected utility. *Journal of Financial Econometrics 2*, 477–492.

Battiston, S., M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli (2012). Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific reports 2*, 541.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade*, pp. 437–478. Springer.

Benoit, S., G. Colletaz, C. Hurlin, and C. Pérignon (2013). A theoretical and empirical comparison of systemic risk measures.

Benoit, S., J.-E. Colliard, C. Hurlin, and C. Pérignon (2017). Where the risks lie: A survey on systemic risk. *Review of Finance 21*(1), 109–152.

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. Rachev, and F. Fabozzi (2009). Modeling, estimation, and optimization of equity portfolios with heavy-tailed distributions. *Optimizing optimization: The next generation of optimization applications and theory*, 117–141.

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.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics 104*(3), 535–559.

Bisias, D., M. Flood, A. W. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *Annu. Rev. Financ. Econ. 4*(1), 255–296.

Board, F. S. (2016). 2016 list of global systemically important banks (g-sibs). *released November 21*.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, 498–505.

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

Branger, N., K. Lučivjanská, and A. Weissensteiner (2019). Optimal granularity for portfolio choice. *Journal of Empirical Finance 50*, 125–146.

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.

Brownlees, C. T., R. Engle, et al. (2012). Volatility, correlation and tails for systemic risk measurement. *Available at SSRN 1611229*.

Campbell, R., R. Huisman, and K. Koedijk (2001). Optimal portfolio selection in a value-at-risk framework. *Journal of Banking and Finance 25*, 1789–1804.

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*.

Chen, C. (2007). A finite smoothing algorithm for quantile regression. *Journal of Computational and Graphical Statistics 16*(1), 136–164.

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

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science 55*(5), 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies 22*(5), 1915–1953.

Diebold, F. X. and K. Yılmaz (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics 182*(1), 119–134.

Duffie, D. and J. Pan (1997). An overview of value at risk. *Journal of Derivatives 4*, 7–49.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R., E. Jondeau, and M. Rockinger (2014). Systemic risk in Europe. *Review of Finance 19*(1), 145–190.

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

Fang, H.-B., K.-T. Fang, and S. Kotz (2002). The meta-elliptical distributions with given marginals. *Journal of multivariate analysis 82*(1), 1–16.

Farinelli, S., M. Ferreira, D. Rossello, M. Thoeny, and L. Tibiletti (2008). Beyond Sharpe ratio: Optimal asset allocation using different performance ratios. *Journal of Banking & Finance 32*(10), 2057–2063.

Farinelli, S., M. Ferreira, D. Rossello, M. Thoeny, and L. Tibiletti (2009). Optimal asset allocation aid system: From "one-size" vs "tailor-made" performance ratio. *European Journal of Operational Research 192*(1), 209–215.

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

Farmer, L., L. Schmidt, and A. Timmermann (2019). Pockets of predictability. *Available at SSRN 3152386*.

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*.

Fishburn, P. C. (1977). Mean-risk analysis with risk associated with below-target returns. *67*, 116–126.

Girardi, G. and A. T. Ergün (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance 37*(8), 3169–3180.

Glorot, X., A. Bordes, and Y. Bengio (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323.

Gonzalo, J. and A. Taamouti (2017). The reaction of stock market returns to unemployment. *Studies in Nonlinear Dynamics & Econometrics 21*(4).

Gourieroux, C., J.-C. Héam, and A. Monfort (2012). Bilateral exposures and systemic solvency risk. *Canadian Journal of Economics/Revue canadienne d'économique 45*(4), 1273–1309.

Green, J., J. R. Hand, and X. F. Zhang (2013). The supraview of return predictive signals. *Review of Accounting Studies 18*(3), 692–730.

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.

Hanson, S. J. and D. J. Burr (1988). Minkowski-r back-propagation: Learning in connectionist models with non-euclidian error signals. In *Neural Information Processing Systems*, pp. 348–357.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *The Review of Financial Studies 29*(1), 5–68.

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*.

Hautsch, N., J. Schaumburg, and M. Schienle (2014). Financial network systemic risk contributions. *Review of Finance 19*(2), 685–738.

Helbing, D. (2013). Globally networked risks and how to respond. *Nature 497*(7447), 51–59.

Hong, T., P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting 32*(3), 896–913.

Hong, Y., Y. Liu, and S. Wang (2009). Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics 150*(2), 271–287.

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

Huber, P. J. (2004). *Robust statistics*, Volume 523. John Wiley & Sons.

Ibragimov, R. and D. Walden (2007). The limits of diversification when losses may be large. *Journal of Banking and Finance 31*, 2551–2569.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1683.

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.

Jorion, P. (2007). *Value at Risk. The New Benchmark for Managing Financial Risk*, Volume 81. McGraw-Hill. 3rd Edition.

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*.

Keskar, N. S., D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang (2016). On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kirby, C. and B. Ostdiek (2012). Optimizing the performance of sample mean-variance efficient portfolios. In *AFA 2013 San Diego Meetings Paper*.

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

Konno, H. and H. Yamazaki (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science 37*(5), 519–531.

Kresta, A. et al. (2015). Application of performance ratios in portfolio optimization. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis 63*(6), 1969–1977.

Kynigakis, I. and E. Panopoulou (2021). Does model complexity add value to asset allocation? Evidence from machine learning forecasting models. *Journal of Applied Econometrics*.

Ledoit, O. and M. Wolf (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Liang, N. (2013). Systemic risk monitoring and financial stability. *Journal of Money, Credit and Banking 45*(s1), 129–135.

Lin, W., J. Olmo, and A. Taamouti (2022). Portfolio selection under systemic risk. *Available at SSRN 3561153*.

Lin, W. and A. Taamouti (2022). 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.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Masters, D. and C. Luschi (2018). Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*.

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.

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

Nair, V. and G. E. Hinton (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807–814.

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.

Ouali, D., F. Chebana, and T. B. Ouarda (2016). Quantile regression in regional frequency analysis: a better exploitation of the available information. *Journal of Hydrometeorology 17*(6), 1869–1883.

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

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.

Robbins, H. and S. Monro (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 400–407.

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

Shalit, H. and S. Yitzhaki (1984). Mean-Gini, portfolio theory, and the pricing of risky assets. *The Journal of Finance 39*(5), 1449–1468.

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

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

Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management 21*(1), 49–58.

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.

Tente, N., N. V. Westernhagen, and U. Slopek (2019). M-press-creditrisk: Microprudential and macroprudential capital requirements for credit risk under systemic stress. *Journal of Money, Credit and Banking 51*(7), 1923–1961.

The MathWorks, I. (2019). *Statistics and Machine Learning Toolbox*. Natick, Massachusetts, United State.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies 67*, 65–86.

Tu, J. and G. Zhou (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics 99*(1), 204–215.

Von Neuman, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

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.

Young, M. R. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science 44*(5), 673–683.

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.