

Neural network copula portfolio optimization for exchange traded funds

YANG ZHAO[†], CHARALAMPOS STASINAKIS^{*‡}, GEORGIOS SERMPINIS[‡], YUKUN SHI[§]

[†] International Institute for Financial Studies and RCFMRP, Jiangxi University of Finance and Economics, East Shuanggang Road, Nanchang, 330013, China

[‡] Adam Smith Business School, University of Glasgow, Adam Smith Building, Glasgow, G12 8QQ, UK

[§] School of Business, University of Leicester, Ken Edwards Building, Leicester, LE1 7RH, UK

This paper attempts to investigate if adopting accurate forecasts from Neural Network (NN) models can lead to statistical and economically significant benefits in portfolio management decisions. In order to achieve that, three NNs, namely the Multi-Layer Perceptron (MLP), Recurrent Neural Network (RNN) and the Psi Sigma Network (PSN), are applied to the task of forecasting the daily returns of three Exchange Traded Funds (ETFs). The statistical and trading performance of the NNs is benchmarked with the traditional Autoregressive Moving Average (ARMA) models. Next, a novel dynamic asymmetric copula model (NNC) is introduced in order to capture the dependence structure across ETF returns. Based on the above, weekly re-balanced portfolios are obtained and compared by using the traditional mean-variance and the mean-CVaR portfolio optimization approach. In terms of the results, PSN outperforms all models in statistical and trading terms. Additionally, the asymmetric skewed t copula statistically outperforms symmetric copulas when it comes to modelling ETF returns dependence. The proposed NNC model leads to significant improvements in the portfolio optimization process, while forecasting covariance accounting for asymmetric dependence between the ETFs also improves the performance of obtained portfolios.

Keywords: Copulas; Neural networks; Portfolio optimization; ETF

JEL Classification: G11, G17

Corresponding author. Email: Charalampos.Stasinakis@glasgow.ac.uk

1. Introduction

The basic premise of the modern portfolio theory is that portfolio diversification benefits originate from investing in financial assets that are not highly correlated. In other words, financial returns' dependence is explained by the linear correlation coefficient, while efficient portfolio frontiers are approximated by the mean-variance optimization (Markowitz, 1952). The baseline assumption of the theory is that financial returns follow a joint normal distribution. Once researchers relax or depart from this assumption of normality, they usually examine the skewness and the kurtosis of the financial returns' distribution. The logic behind this is that negative skewness implies that negative financial returns are more probable, while excess kurtosis suggests extreme observations are more likely to appear than it would be expected in normality conditions. There are other observed properties in financial return series that go against the Markowitz theory, such as time-varying skewness and kurtosis (Harvey and Siddique 1999; Jondeau and Rockinger 2003) or the long-term persistence of their mean and variance (Saqudique and Silvapulle 2001). For that reason, portfolio optimization methods that extend the traditional mean variance approach must be explored in order to achieve maximum risk reduction for a given level of expected return.

The early literature shows that the diversification benefits mainly depend on accurate predictions of the asset return moments. Nonetheless, there are two streams of research. One stream focuses on *asset allocation* and provides solid evidence that the forecasts of asset returns are important inputs for the mean-variance optimization. Best and Grauer (1991) show that the weights, mean, and variance of the mean-variance efficient portfolio are extremely sensitive to changes in asset means. Chopra *et al.* (1993) find that adjusted-input portfolios can achieve higher expected return, less variance and greater terminal wealth than unadjusted-input portfolios. Chopra and Ziemba (1993) also demonstrate that using inaccurate forecasts of asset returns can substantially degrade the performance of mean-variance optimization. Another strand of studies investigates the importance of forecasting the second moments (i.e. covariance structure) of asset returns on portfolio optimization. For instance, Chan *et al.* (1999) evaluate the out-of-sample performance of optimized portfolio based on the different models of covariances. They provide evidence that predictions of variance and covariance of asset returns are key inputs for the practitioner. Menchero *et al.* (2012) find that the risk of optimized portfolios tends to be underestimated by sample covariance and they show that the adjusted covariance can effectively reduce the out-of-sample volatilities of optimized portfolios.

Portfolio practitioners focusing more on obtaining accurate forecasts of financial returns face the difficult task of screening optimal models from the voluminous financial forecasting literature. The models available are characterized by linear or non-linear estimations and constant or time-varying parameterization processes. Neural networks (NNs) is a popular class of non-linear computation models when it comes to forecasting financial market variables, because of their data-adaptive learning and

clustering abilities (McNeilis 2005). Over the past decade, NNs have provided extensive empirical evidence for their high financial forecasting performance. For example, Wang (2009) show that using NNs to forecasting volatility increases the predictability of option-pricing models. Ebrahimpour *et al.* (2011) apply successfully a mixture of Multi-Layer Perceptron (MLP) experts in trend prediction of time series on the Tehran stock exchange. Dunis *et al.* (2011) apply Psi Sigma Networks (PSNs) to the task of forecasting the EUR/USD exchange rate. Their results indicate the superiority of PSN over traditional MLPs and Recurrent Neural Networks (RNNs) models. Finally, Guresen *et al.* (2011) provides an extensive survey of the successful applications of NNs in stock market index predictions, including applications of MLP, RNN and Higher Order Neural Networks (HONNs).

When it comes to predicting the covariance matrix among financial assets, three models are normally considered in the literature, namely the Dynamic Conditional Correlation (DCC) model (Engle, 2002), the Asymmetric Dynamic Conditional Correlation (ADCC) model (Cappiello *et al.* 2006) and the Generalized Autoregressive Score (GAS) model (Creal *et al.* 2013). The DCC is probably the most widely used econometric technique to estimate and predict the covariance of asset returns (see Andersen *et al.* 2006, Jondeau and Rockinger 2006b, Christoffersen *et al.* 2012, etc.). The ADCC model is a generalized version of the DCC model, which permits conditional asymmetries in correlations (see Syriopoulos and Roumpis 2009, Fei *et al.* 2010, etc.). Finally, the GAS model is a more recent technique that can be used to model the dynamic dependence of asset returns. The GAS framework uses the score of the conditional density function to drive the dynamics of the time-varying parameters (see Lucas *et al.* 2014, Creal *et al.* 2014, Salvatierra and Patton 2015, etc.).

It is a stylized fact that equity returns are more correlated during market downturns than market upturns (see Longin and Solnik 2001, Ang and Chen 2002, Hong *et al.* 2007, amongst others). This characteristic, known as asymmetric dependence, violates the assumption of modern portfolio theory that the financial returns follow joint normal distribution and their dependence can be fully described by the linear correlation coefficient as suggested by Markowitz (1952). Several empirical studies show that the asymmetric dependence can be well captured by copulas and taking into account this characteristic can produce economic gains for the investors with no short selling constraints (Patton 2004, Garcia and Tsafack 2011, Chu 2011).

The classical mean-variance optimization uses variance as a risk proxy, however earlier literature criticizes that assumptions. In other words, variance is not a perfect measure because it is symmetric and treats downside risk and upside risk in the same way (Ang *et al.* 2006). Several downside risk measures have been introduced in portfolio optimization practice, such as semi-variance (Markowitz 1959), Value-at-Risk (VaR) (Gaivoronski and Pflug 2005) and Conditional Value-at-Risk (CVaR) (Rockafellar and Uryasev 2000, Rockafellar and Uryasev 2002). The criticism against VaR originates from its lack

of subadditivity and convexity (Artzner *et al.* 1999). In addition, VaR is not easy to optimize when calculated using scenarios (Rockafellar and Uryasev 2000). For these reasons, CVaR, can be thought as a coherent risk measure, which has been widely applied in optimization practice (see amongst others Rockafellar and Uryasev (2002), Quaranta and Zaffaroni (2008) and He and Gong (2009)).

Following Boubaker and Sghaier (2013) and Low *et al.* (2013), this study proposes a novel Neural Network Copula (NNC) portfolio optimization approach. It is assumed that the investor has a one-period horizon and aims to construct a dynamically rebalanced portfolio based on three Exchange Traded Funds (ETFs). The time-varying weights of this portfolio require one-step-ahead forecasts of the expected return and the conditional variance–covariance matrix. The expected asset returns are predicted by the superior PSN model in a forecasting exercise over the period of 2011-2015. The forecasting performance of the PSN is benchmarked against the traditional Autoregressive Moving Average (ARMA) model and two NN structures, namely the MLP and RNN. The NNC process is able to extract the time-varying variance-covariance matrix based on a copula-based GAS model that captures the asymmetric dependence between the respective ETFs along with using the CVaR as a measure of risk. In terms of the results, the proposed asymmetric copula model statistically outperforms symmetric copulas in dependence modelling. The study provides evidence that the NNC process leads to significant improvements in portfolio optimization. Traditional ARMA-based portfolios perform worse than portfolios based on NN models, while the PSN portfolios deliver higher out-of-sample risk-adjusted returns. Finally, it is shown that forecasting covariance taking into account asymmetric dependence can improve the performance of optimization, however the magnitude of this improvement is relatively small. This finding is in line with the relevant strand of the literature suggesting the forecasts of asset returns are the most important inputs for the mean-variance optimization.

The rest of the paper is organized as follows. Section 2 provides a detailed description of the ETFs' dataset used in this paper. All forecasting models are described in section 3, while their performance is evaluated in section 4. The proposed NNC portfolio optimization process is explained in detail in section 5. The final portfolio optimization results are summarized in section 6, while some concluding remarks are given in Section 7. Finally, the appendix and online supplementary appendix provide technical and mathematical details essential for the understanding of this study.

2. Dataset

The advantages of ETFs over traditional trading are well documented (Avellaneda and Lee 2010, Dolvin 2010, Marshall *et al.* 2013). The main one is that they offer investors the opportunity to trade stock market indices at very low transaction costs with high level of diversification*. In this study, we examine

* The transaction costs for the three ETFs tracking their respective benchmarks do not exceed 0.5% per annum for medium size investors (see, for instance, www.interactive-brokers.com). Before the expansion of ETFs, traders had to pay a separate

three ETFs over the period of 2011-2015, namely the SPDR S&P 500 ETF Trust (SPY), SPDR Dow Jones Industrial Average ETF Trust (DIA) and PowerShares QQQ Trust (QQQ). These are designed to replicate major stock indices from US, while they are characterised by high liquidity and high volume of assets. It should be noted here that methods performing well on these highly scrutinized and arbitrated ETFs are expected to perform even better on other that are less liquid and less covered. Therefore, using these ETFs can also be considered as a tough to beat benchmark. Their details are presented in table 1 below.

Table 1: The ETFs under study

ETF	TRACKING INDEX	TICKER
SPDR S&P 500 ETF Trust	S&P 500	SPY
SPDR Dow Jones Industrial Average ETF Trust	Dow Jones Industrial Average	DIA
PowerShares QQQ Trust	NASDAQ-100	QQQ

All models in this study are applied in the task of forecasting the one day ahead arithmetic returns of the three ETFs. The descriptive statistics and correlation matrix of the return series are shown in the following table:

Table 2: Descriptive statistics and correlation matrix

Panel A: Descriptive Statistics			
Ticker	SPY	DIA	QQQ
Mean	0.00052	0.00045	0.00069
Standard deviation	0.00963	0.00885	0.01051
Skewness	-0.43942	-0.42037	-0.30708
Kurtosis	7.64749	7.01179	5.98211
Jarque-Bera (p value)	0.0000***	0.0000***	0.0000***
ADF (p value)	0.0000***	0.0000***	0.0000***
Panel B: Correlation Matrix			
Ticker	SPY	DIA	QQQ
SPY	1		
DIA	0.829 [0.798]	1	
QQQ	0.958 [0.946]	0.906 [0.887]	1

Note: Panel B reports the linear correlation and Spearman's rank correlation (bracket).

The three returns series exhibit slight negative skewness and positive kurtosis. The Jarque-Bera statistic confirms that the return series under study are non-normal at the 99% confidence level. The Augmented Dickey-Fuller (ADF) reports that the null hypothesis of a unit root is rejected at the 99% confidence level for all ETFs. The period under study and the relevant datasets are presented in table 3.

Table 3: The total dataset

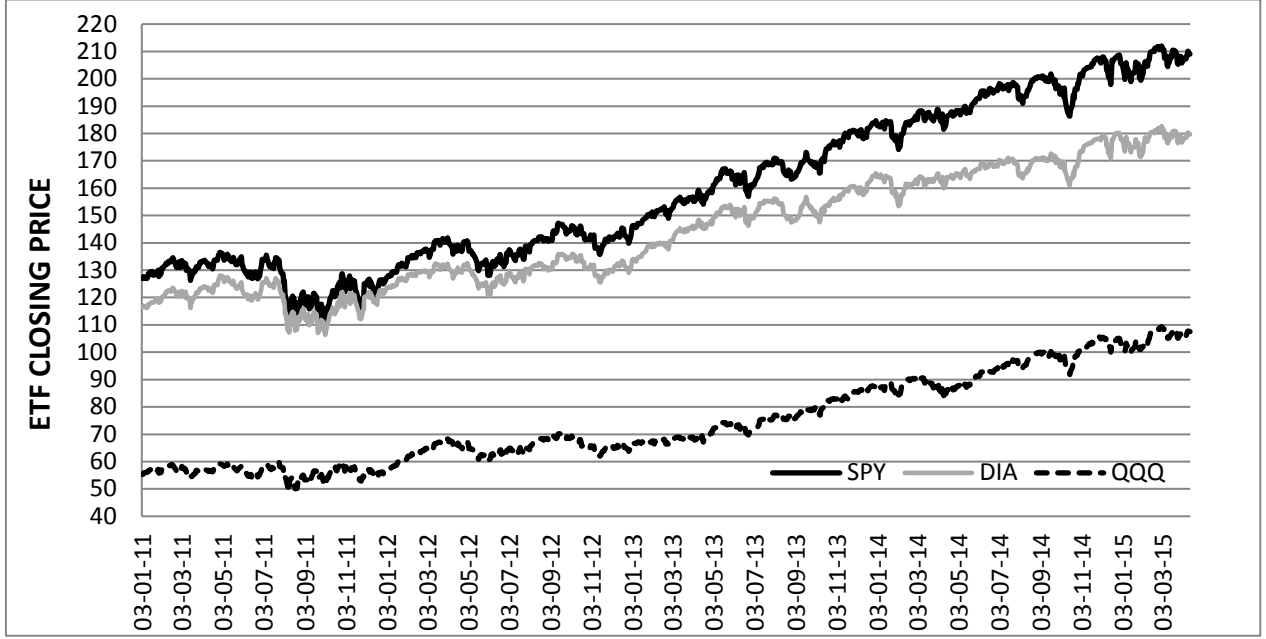
Datasets	Trading Days	Start Date	End Date
Total Dataset	1075	03/01/2011	13/04/2015
Training Dataset	502	03/01/2011	31/12/2012
Test Dataset	252	02/01/2013	31/12/2013
Out-of-sample Dataset	321	02/01/2014	13/04/2015

Note: The in-sample period is the sum of the training and test datasets.

commission for each individual stock of an industry-specific portfolio. Now there are sector-specific ETFs, which allow traders to pay only one commission to buy or sell short an entire group of stocks.

All models are trained in the in-sample and their forecasts are evaluated in the out-of-sample. Figure 1 presents the performance of the three ETFs during the period of 3rd January 2011 to 13th April 2015.

Figure 1: The ETFs under study[†]



3. Forecasting models

This section summarizes the models applied to the task of forecasting the one-day head return of the SPY, DIA and QQQ series under study. We implement in total four forecasting models, namely a baseline Autoregressive Moving Average Model (ARMA) and three traditional NN techniques.

3.1 Autoregressive moving average model (ARMA)

ARMA models are traditionally used in similar applications, as they are based on the assumption that the current value of a time-series can be approximated with a linear combination of its previous values plus a combination of current and previous values of the residuals (Brooks, 2008). Generally, an ARMA is be specified as below:

$$Y_t = \hat{\varphi}_0 + \hat{\varphi}_1 Y_{t-1} + \hat{\varphi}_2 Y_{t-2} + \dots + \hat{\varphi}_{\hat{p}} Y_{t-\hat{p}} + \hat{\varepsilon}_t - \hat{w}_1 \hat{\varepsilon}_{t-1} - \hat{w}_2 \hat{\varepsilon}_{t-2} - \dots - \hat{w}_{\hat{q}} \hat{\varepsilon}_{t-\hat{q}} \quad (1)$$

where:

- Y_t is the dependent variable at time t
- $Y_{t-1}, Y_{t-2}, \dots, Y_{t-\hat{p}}$ are the lagged dependent variables
- $\hat{\varphi}_0, \hat{\varphi}_1, \dots, \hat{\varphi}_{\hat{p}}$ are the regression coefficients
- $\hat{\varepsilon}_t$ is the residual term

[†] From the figure it is obvious that in- and out-of-sample period mainly cover a bull market. Nonetheless, bear markets are also covered in a way, since the 2011 drop is included in the in-sample.

- $\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \dots, \hat{\varepsilon}_{t-\bar{q}}$ are the previous values of the residual terms
- $\hat{w}_1, \hat{w}_2, \dots, \hat{w}_{\bar{q}}$ are the residual weights

Based on the in-sample correlogram (training and test subsets), the restricted ARMA (8,8), ARMA(10,10) and ARMA(7,7) model are chosen for the out-of-sample estimation of SPY, DIA and QQ respectively.

3.2 Neural Networks

In our study, we apply three traditional NN architectures as forecasting techniques. The first model is the most popular NN architecture, namely the Multi-Layer Perceptron (MLP). A standard MLP has at least three layers. The first layer is called the input layer (the number of its nodes corresponds to the number of explanatory variables). The last layer is called the output layer (the number of its nodes corresponds to the number of response variables). An intermediary layer of nodes, the hidden layer, separates the input from the output layer. Its number of nodes defines the amount of complexity the model is capable of fitting. In addition, the input and hidden layer contain an extra node called the bias node. This node has a fixed value of one and has the same function as the intercept in traditional regression models. Normally, each node of one layer has connections to all the other nodes of the next layer.

The network processes information as follows: the input nodes contain the value of the explanatory variables. Since each node connection represents a weight factor, the information reaches a single hidden layer node as the weighted sum of its inputs. Each node of the hidden layer passes the information through a non-linear activation function and passes it on to the output layer if the calculated value is above a threshold. The training of the network (which is the adjustment of its weights in the way that the network maps the input value of the training data to the corresponding output value) starts with randomly chosen weights and proceeds by applying a learning algorithm called back-propagation of errors (Shapiro 2000). The maximum number of the allowed back-propagation iterations is optimized by maximizing a fitness function in the test dataset (see table 3) through a trial and error procedure. More specifically, the learning algorithm tries to find those weights which minimize an error function (normally the sum of all squared differences between target and actual values). Since networks with sufficient hidden nodes are able to learn the training data (as well as their outliers and their noise) by heart, it is crucial to stop the training procedure at the right time to prevent overfitting (this is called ‘early stopping’). This is achieved by dividing the dataset into 3 subsets respectively called the training and test sets used for simulating the data currently available to fit and tune the model and the validation set used for simulating future values. The network parameters are then estimated by fitting the training data using the backpropagation of errors. The iteration length is optimized by maximizing the forecasting

accuracy for the test dataset. Then, the predictive value of the model is evaluated applying it to the validation dataset (out-of-sample dataset).

In addition to the classical MLP network, the Recurrent Neural Network (RNN) is also applied. A simple RNN has an activation feedback which embodies short-term memory. In other words, the RNN architecture can provide more accurate outputs because the inputs are (potentially) taken from all previous values. Although RNN require substantially more computational time (Tenti 1996), they can yield better results in comparison with simple MLPs due to the additional memory inputs. The third NN model included in the feature space is Psi Sigma Network (PSN). PSNs are considered as a class of feed-forward fully connected Higher Order Neural Network (HONN). First introduced by Ghosh and Shin (1991), the PSN structure is motivated by the need to create a network combining the fast learning property of single layer networks with the powerful mapping capability of HONNs, while avoiding the combinatorial increase in the required number of weights. The order of the network in the context of PSN is represented by the number of hidden nodes. In a PSN the weights from the hidden to the output layer are fixed to 1 and only the weights from the input to the hidden layer are adjusted, something that greatly reduces the training time. The description of each NN and their technical characteristics (input selection and parametrization) are presented in Online Appendix A.

4. Forecasting models' statistical and trading performance

In order to evaluate statistically the forecasts, the RMSE, the MAE, the MAPE and the Theil-U statistics are computed. For all four of the error statistics retained the lower the output, the better the forecasting accuracy of the model concerned. Their mathematical formulas are presented in Online Appendix B. The following table presents out-of-sample statistical performance of the models.

Table 4: Out-of-sample statistical performance					
ETF	Statistic	ARMA	MLP	RNN	PSN
SPY	MAE	0.0056	0.0058	0.0057	0.0055
	MAPE	167.44%	164.52%	151.86%	141.21%
	RMSE	0.0078	0.0076	0.0075	0.0071
	THEIL-U	0.9025	0.8286	0.8049	0.7598
DIA	MAE	0.0057	0.0055	0.0053	0.0051
	MAPE	162.07%	161.88%	130.99%	128.81%
	RMSE	0.0075	0.0073	0.0071	0.0068
	THEIL-U	0.9256	0.7635	0.7348	0.7086
QQQ	MAE	0.0085	0.0068	0.0067	0.0062
	MAPE	148.55%	128.74%	123.04%	119.22%
	RMSE	0.0088	0.0082	0.0081	0.0075
	THEIL-U	0.9077	0.8322	0.8279	0.7980

The above results show that the models' statistical ranking is consistent across all ETFs series. In general, the baseline ARMAs are found to have the worst statistical results compared to all models. The PSN appears to be consistently the superior model in statistical terms against all NNs and ARMAs.

Finally, RNN is the second best model. This statistical ranking is consistent with other similar studies on NNs (Sermpinis *et al.* 2014, Stasinakis *et al.* 2016).

In order to further validate the above findings, we perform two additional tests, namely the Pesaran-Timmermann (PT) (1992) and the Diebold Mariano (DM) (1995) test. The PT test is used to examine whether the directional movements of the real and forecast values are in step with one another. The PT test's null hypothesis is that the model under study has no power on forecasting the relevant ETF return series. The DM statistic tests the null hypothesis of equal predictive accuracy between two forecasts. In this case, the DM test is applied to couples of out-of-sample forecasts (best model vs. other model) using the MSE loss function. In our case, a negative realization of the DM value would indicate that the PSN forecast is more accurate than the competing forecast. The results of the two tests are provided in table 5.

Table 5: PT and DM statistics.

Test	ETF	ARMA	MLP	RNN	PSN
PT	SPY	(6.58)***	(7.25)***	(8.69)***	(9.12)***
	DIA	(7.56)***	(8.95)***	(9.05)***	(9.87)***
	QQQ	(6.84)***	(7.63)***	(8.15)***	(8.93)***
DM	SPY	(-8.12)***	(-6.93)***	(-6.06)***	-
	DIA	(-9.51)***	(-8.42)***	(-7.15)***	-
	QQQ	(-10.66)***	(-9.14)***	(-7.38)***	-

*Note: The values in the parentheses are the calculated PT and DM statistics. *** denotes that the null hypothesis is rejected at 1% significance level.*

From the above table, the PT statistics indicate that all models are capable of capturing the directional movements of the three ETF return series in the out-of-sample[‡]. Additionally, the null hypothesis of equal predictive accuracy is rejected for all comparisons at 1% significance level. Moreover, the statistical superiority of the PSN forecasts is further validated, as all the DM statistic realizations are negative. Additionally, RNN is found to have the closest forecasts to the superior PSN. All the statistical findings indicate that PSN provides the most accurate forecasts. It would be interesting to see if this superiority is translated also into higher trading performance.

Therefore, the competing forecasting models are compared also in terms of trading efficiency. In this application, the trading performance of the models is evaluated with a simple trading strategy. The position is 'long' and 'short', when the forecast return is positive and negative respectively. A 'long' or 'short' position means that we buy or sell respectively the ETF under study at the current price. As mentioned before, the low transaction costs make ETF very attractive to traders. This is the case especially in daily trading applications (as of this study), where high transaction costs can vastly decrease

[‡] Similar results are obtained also in the in-sample period. In-sample results are not provided within text for the sake of space and are available upon request.

profitability (Wyart *et al.* 2008). The trading performance measures are presented in Online Appendix B. Table 6 summarizes the out-of-sample trading results for the respective model and ETF series after transaction costs.

Table 6: Out-of-sample trading performance after transaction costs

ETF	Measure	ARMA	MLP	RNN	PSN
SPY	Annualised Return	5.94%	7.44%	11.59%	17.17%
	Sharpe Ratio	0.50	0.63	1.33	1.47
	Maximum Drawdown	-6.76%	-10.78%	-8.92%	-8.23%
DIA	Annualised Return	3.80%	12.81%	14.87%	18.37%
	Sharpe Ratio	0.33	1.63	1.94	2.24
	Maximum Drawdown	-12.48%	-9.54%	-8.28%	-8.14%
QQQ	Annualised Return	4.12%	8.75%	10.45%	16.52%
	Sharpe Ratio	0.69	1.32	1.71	2.37
	Maximum Drawdown	-8.75%	-8.11%	-5.77%	-5.89%

From the table above we note that the trading efficiency ranking coincides with the statistical one (as per table 4). The PSN delivers the best trading performance for all series under study. On average, PSN achieves 17.35% annualized returns and 2.08 Sharpe ratio after transaction costs in the out-of-sample period. The second best model in terms of the same trading performance measures is RNN. It projects on average profits and Sharpe ratio after transaction costs at the level of 12.30% and 1.66 respectively. The MLP remains the worse performing NN. Another interesting finding is that NNs present consistently lower maximum drawdown figures, which is a relative proxy for their trading risk.

Overall, the results indicate that PSN is the best performing model in the three forecasting exercises. The above evidence is interesting from a forecasting and model competing point of view, but they do not necessarily convince traders or practitioners with different backgrounds and preferences to drop simpler techniques such as ARMA models.[§] These traditional models, although not superior to the NNs, still project profits while being well established and easy to implement. Therefore, it is even more exciting to evaluate whether adopting superior ETF forecasts from the best NN can lead to further improved portfolio management decisions, that are worth the ‘technical and computational trouble’. This issue is explored with the analysis of the following section.

5. Neural network-based portfolio optimization

[§] In order to consider a nonlinear benchmark, we also experiment with a Smooth Transition Autoregressive (STAR) model which is a nonlinear extension of autoregressive models. Nonetheless, the out-of-sample statistical and trading performance was found inferior to the ARMA specifications. As such, it is logical to retain for the portfolio optimization the less complex and better performing linear ARMA. Nonetheless, the STAR results are not included for the sake of space and are available upon request.

This section provides the summary of the portfolio optimization procedures applied in this study. Initially, the traditional Mean-Variance (M-V) approach is described. Then, the proposed copula-based mean-CVaR optimization method is explained in detail.

5.1 Traditional mean-variance portfolio optimization with NN models

Modern portfolio theory suggests that there are two important inputs for the M-V portfolio optimization: expected returns and forecasts of covariance. The results of section 4 indicate that the NN forecasts are the more accurate approximations for the daily expected returns across all ETF series. As a next step, we want to investigate whether we can achieve significantly better performance in the M-V optimization by using their forecasts instead of the traditional ARMA.

Following Markowitz (1952), we assume that investors wish to find portfolios that have the best expected return-risk trade-off. The optimal portfolio weights can be obtained by minimizing the variance of the portfolio for a given level of expected return^{**}. Therefore, the optimization problem in our study can be expressed as:

$$\min_{\mathbf{w}_t} \sigma_{p,t}^2(\mathbf{w}_t) = \mathbf{w}_t^T \Sigma_t \mathbf{w}_t \text{ subject to } r_{p,t} = \mathbf{w}_t^T \mathbf{r}_{i,t} \text{ and } \mathbf{w}_t^T \mathbf{1} = 1 \quad (2)$$

where $\sigma_{p,t}^2$ denotes portfolio variance at time t , $r_{p,t}$ denotes the expected return of the portfolio, \mathbf{w}_t denotes the vector of portfolio weights, and Σ_t denotes the covariance matrix of ETF returns at time t . In this study, the weekly algorithmic returns are calculated as:

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}) \quad (3)$$

where $P_{i,t}$ and $r_{i,t}$ denote the price and logarithm return of ETF i at week t . Σ_t is predicted by three different models, namely the DCC-GARCH, the ADCC-GARCH and the GAS model.

5.2 Neural network copula-based mean-CVaR portfolio optimization (NNC)

Although variance is straightforward to calculate and widely used in financial practice, it is not a satisfactory risk measure from the risk measurement perspective. As a symmetric risk measure, variance penalizes profits and losses in an equal way. Thus, this study considers the CVaR as an alternative risk measure, it is easily interpretable and it satisfies several attractive mathematical properties. Minimization of portfolio CVaR is closely related to the minimization of portfolio VaR, while CVaR minimization

^{**} Equivalently, the optimal portfolio can be obtained by maximizing portfolio expected return for a given level of risk as measured by portfolio variance.

can be easily solved by a linear programming (Rockafellar and Uryasev 2002). Based on this rational, we adopt the Neural Network Copula-based Mean-CVaR (NNC) portfolio optimization.

The starting point of the process is to model the marginal distributions of ETF returns. The details of this analysis are explained in detail in Appendix A. It is also very crucial to select the appropriate copula for the datasets at hand. Although the Gaussian copula and t copula are the most widely used copulas in finance as they are convenient to use, neither of them are able to capture multivariate asymmetry. Some Archimedean copulas, such as the Clayton, the Gumbel and the Joe-Clayton specifications, allow asymmetry in the bivariate distribution, however, they are not easily generalized to high-dimensional applications. Following Christoffersen *et al.* (2012) and Christoffersen and Langlois (2013), we use the skewed t copula implied by the skewed t distribution discussed in Demarta and McNeil (2005) to overcome these problems. To further verify this selection, we compare the performance of the skewed t copula with nine alternatives and we find that it is indeed the best performing copula. The details of this analysis are provided in Appendix B. The superiority of the skewed t copula is intuitively reasonable since it can capture the upper and lower tail dependence along with the multivariate asymmetry. In other words, it is safe to assume that this copula can describe the ‘true’ dependence structure among the ETF index returns. The skewed t copula-based GAS model can now be used to obtain the dynamics of correlation (covariance) and apply the estimated copula to implement a Monte Carlo simulation to obtain portfolio CVaR for the optimizations. The CVaR optimization strategy allows the minimization of the downside tail risk of the portfolio (portfolio CVaR^{††}) for a given level of return. This strategy is suitable for investors who have a utility function characterized by the minimization of downside tail risk and are indifferent to (or might even prefer) upside variance.

Specifically, the proposed NNC approach can be separated into two stages. In the first stage, we calculate the expected weekly returns of the three ETF indices using the daily forecasts of ARMA and three NN models. In order to incorporate asymmetric dependence in our model, we use the skewed t copula to describe the dependence structure between asset returns. The time-varying correlation matrix for the skewed t copula model is predicting using the DCC, ADCC and GAS model. All the forecasts are obtained by using a “rolling window” approach. Then, we re-estimate the skewed t copula at each week t using a 1-year rolling window^{‡‡}. Given the estimated skewness parameter and degrees of freedom, as well as the correlation matrix predicted by the DCC, ADCC and GAS, a Monte Carlo simulation based on the skewed t copula is done to predict the VaR and CVaR for the ETF portfolio. Based on the above, given a series of target returns, it is possible to obtain an efficient frontier of optimal

^{††} CVaR is the abbreviation of the Conditional Value-at-Risk, which is also known as the Expected Shortfall.

^{‡‡} We use a rolling window instead of the full sample period and set a window size at 250 (one trading year) for all the data sets. We conduct rolling forecast by moving forward a day at a time and end with the forecast for 13/04/2015.

risk-return portfolios at each week t . Despite of being able to calculate the whole efficient frontier of every ETF portfolio at each time t , our decision to rebalance the portfolios is based on the optimal weights of the tangency portfolio, meaning the **one with the highest Sharpe ratio or Return/CVaR ratio**.

Following Rockafellar and Uryasev (2000, 2002), β -VaR and β -CVaR of the portfolio at time t in integral form are given by:

$$\alpha_{\beta}(\mathbf{w}_t) = \min \{ \alpha \in \mathbb{R} : \Psi(\mathbf{w}_t, \alpha) \geq \beta \} \quad (4)$$

and

$$\phi_{\beta}(\mathbf{w}_t) = (1 - \beta)^{-1} \int_{f(\mathbf{w}_t, \mathbf{r}_t) \geq \alpha_{\beta}(\mathbf{w}_t)} f(\mathbf{w}_t, \mathbf{r}_t) p(\mathbf{r}_t) d\mathbf{r}_t \quad (5)$$

where Ψ is the cumulative distribution for the loss associated with \mathbf{w}_t , the probability that \mathbf{r}_t occurs is $p(\mathbf{r}_t)$ and the loss function is presented by $f(\mathbf{w}_t, \mathbf{r}_t)$ as:

$$f(\mathbf{w}_t, \mathbf{r}_t) = -[w_{1,t}r_{1,t} + \dots + w_{n,t}r_{n,t}] = -\mathbf{w}_t^T \mathbf{r}_t \quad (6)$$

Rockafellar and Uryasev (2000) show that β -CVaR of portfolio in integral form can be well approximated using a Monte Carlo simulation. Therefore, the following equation is a suitable approximation that can be used to minimize CVaR for a given level of portfolio return:

$$\min_{(\mathbf{w}_t, \alpha)} F_{\alpha}(\mathbf{w}_t, \beta) = \alpha + \frac{1}{q(1 - \beta)} \sum_{m=1}^q [-\mathbf{w}_t^T \mathbf{r}_{m,t} - \alpha]^+ \text{ subject to } \mu(\mathbf{w}_t) = -\mathbf{w}_t^T \mathbf{r}_t \leq -R \text{ and } \mathbf{w}_t^T \mathbf{1} = 1 \quad (7)$$

where q denotes the number of samples generated by the skewed t copula-based Monte Carlo simulation, α denotes VaR at β level and $\mathbf{1}$ is a vector of ones and $\mathbf{r}_{m,t}$ is the m^{th} vector of simulated returns. The vector of portfolio weights, \mathbf{w}_t , can be obtained from the optimization procedure to generate the portfolio that minimizes CVaR for a given portfolio return R . A pseudo-algorithm of the NNC process is presented in Appendix D.

6. Final portfolio optimization results

In this study, we follow Jondeau and Rockinger (2006a) to construct weekly rebalanced portfolios. The reason that our portfolios are not rebalanced in a daily basis is that the benefits generated from portfolio optimization may be significantly offset by the sharp increase of transaction costs. The calculation of the portfolio weights for the M-V optimization depends on the predictions of the ETF returns from the NN models and the covariance matrix predicated by the DCC, ADCC and GAS model. Similarly, the

calculation of the portfolio weights for the mean-CVaR optimization also depends on the forecasts from the NN models and the CVaR from simulation. Specifically, for the estimation of portfolio CVaR, we re-estimate our model each week and use a Monte Carlo simulation to generate 10000 observations. The covariance matrix between assets is predicted by the same models as the M-V optimization. The portfolio performances are evaluated based on the achieved realized returns, Sharpe and Sortino^{§§} ratios and maximum drawdowns. Table 7 presents the results obtained by the traditional M-V approach.

Table 7: Performances of different trading strategies (Traditional M-V)

Panel A: ETF indices and 1/N portfolio				
	Realized Return	Sharpe Ratio	Sortino Ratio	Max Drawdown
SPY	10.233%	0.9155	1.3519	7.590%
DIA	6.773%	0.6053	0.8995	7.041%
QQQ	16.477%	1.2681	2.2277	8.822%
1/N	11.161%	0.9850	1.4937	7.415%
Panel B: Mean-Variance optimization without short-selling				
	Realized Return	Sharpe Ratio	Sortino Ratio	Max Drawdown
ARMA-DCC	11.404%	0.9660	1.5222	6.650%
ARMA-ADCC	11.760%	0.9973	1.6006	6.650%
ARMA-GAS	12.128%	1.0261	1.6485	6.650%
RNN-DCC	26.108%	2.1904	3.7299	7.197%
RNN-ADCC	26.791%	2.1982	3.4489	8.368%
RNN-GAS	27.122%	2.2697	3.7321	7.357%
PSN-DCC	26.829%	2.2420	3.6209	7.185%
PSN-ADCC	26.849%	2.2429	3.6235	7.185%
PSN-GAS	27.362%	2.2767	3.6979	7.185%
Panel C: Mean-Variance optimization with short-selling				
	Realized Return	Sharpe Ratio	Sortino Ratio	Max Drawdown
ARMA-DCC-S	11.889%	0.9183	1.4545	7.958%
ARMA-ADCC-S	11.837%	0.9127	1.4718	7.958%
ARMA-GAS-S	12.135%	0.9313	1.5012	7.785%
RNN-DCC-S	39.538%	3.0225	5.2083	8.232%
RNN-ADCC-S	40.112%	3.0113	5.4423	8.268%
RNN-GAS-S	40.359%	3.1502	5.2116	7.785%
PSN-DCC-S	40.200%	3.1219	5.0452	7.785%
PSN-ADCC-S	40.388%	3.1157	5.2154	7.785%
PSN-GAS-S	40.555%	3.1860	5.0328	7.785%

Note: The table presents the out-of-sample performances over the period January 2014 to March 2015 (68 weekly observations). Panel A reports performances of the three ETF indices and the 1/N portfolio (equally weighted buy-and-hold portfolio). Panel B reports performances of different M-V portfolios without short-selling. All the portfolios are weekly rebalanced tangency portfolios obtained by the M-V optimization based on various model combinations. For example, ARMA-DCC refers to the performance of the tangency portfolio of the efficient frontier of the three ETF assets, where the expected returns are obtained through ARMA forecasts, while the variance-covariance matrix is predicted by

^{§§} The Sortino ratio is a modification of the Sharpe ratio, but it only penalizes those returns falling below a user-specified target or required rate of return, while the Sharpe ratio penalizes both upside and downside volatility equally. Both ratios measure the risk-adjusted returns, but they frequently lead to differing conclusions as to the true nature of the investment's return.

DCC. Panel C reports performances of different M-V portfolios with short-selling. ‘-S’ denotes optimizations allowing short-selling.

The above table provides interesting findings. The QQQ ETF yields the highest annualized return (16.48%), Sharpe ratio (1.268) and Sortino ratio (2.228). Nevertheless, it also suffers the largest maximum drawdown (8.82%). It is also shown that the optimized portfolio from the ARMA model does not achieve significantly higher risk adjusted returns than the equally weighted portfolio (1/N). In contrast, portfolios from NN models clearly outperform both ETFs and the 1/N portfolio^{***}. The average Sharpe ratio and Sortino ratio (2.669 and 4.417, respectively) of portfolios from NN models are significantly higher than the average of portfolios from the ARMA model (0.997 and 1.590, respectively). In general, we find that the optimized portfolios from the ARMA models do not outperform equally weighted portfolio, whereas the portfolios from NN models achieve clearly superior performance. Thus, it can be inferred that the gain from optimal diversification of ARMA portfolios can more than offset by the out-of-sample estimation error from the ARMA model. The benefits of the optimized portfolio from NN models mainly originate from the accurate return predictions generated by the NN models.

Additionally, this is further investigated by allowing for short-selling. From panel C, we find that portfolios from the ARMA yield similar results as in Panel B, which indicates that allowing short-selling does not significantly improve the performance of ARMA portfolios. This can be attributed to the fact that large estimation errors are generated from the ARMA model. Meanwhile, the Sharpe and Sortino ratios of the NN portfolios are significantly improved, when there is no short-selling constraint. The average Sharpe ratio and Sortino ratio increase around 38.66% and 42.57%. This remarkable improvement implies that more accurate forecasts can lead to higher economic profits through portfolio optimization, especially where short-selling is allowed. Interesting empirical evidence are collected also when asymmetric dependence is taken into account. Panel B and C show that when asymmetric dependence properties are considered, portfolios consistently provide better performance in terms of annualized returns and Sharpe/Sortino ratios. As we discussed earlier, both the ADCC and the skewed t copula-based GAS model can capture asymmetric dependence between asset returns. This is important as the equity returns are normally more correlated when market goes down.

Finally, table 7 can differentiate the results from the best two NN model. It should be noted that the optimized portfolios from the PSN outperform portfolios from the RNN in terms of Sharpe ratio. Specifically, the average annualized return (Sharpe ratio) of PSN portfolios is 33.70% (2.696), which is around 1% (2.16%) higher than the average of RNN portfolios. The maximum drawdown of PSN portfolios is also around 5% lower than RNN portfolios. This is intuitively reasonable since the PSN model outperforms the RNN model in the predictions of single ETF. The results of these performance measures indicate that more accurate predictions of asset returns can lead to higher benefit of M-V optimization. The average Sortino ratio of PSN portfolios appears 2% lower than the average of RNN portfolios. This is not surprising since the M-V optimization uses the variance as a risk proxy, which treats downside risk and upside risk in the same way. Here we should note that as main results we

^{***} For the sake of space, we present results of portfolios based on ARMA forecasts (benchmark) and PSN and RNN forecasts (best and second best NN model respectively). MLP results are similar and are not presented for the sake of space

consider those obtained from equally weighted three-asset portfolios, as these are generally thought as more realistic applications than the two-asset ones. Similar trend in the results is obtained, when the two-asset cases are analysed (see Online Appendix C).

Therefore, it would be more interesting to investigate if NN models can provide significant improvement in the mean-CVaR optimization, which minimizes the tail risk instead of the variance. Table 8 below provides the equivalent results for the proposed NNC optimization method. Because of the nature of the NNC approach, the Sharpe ratio is replaced with the ratio of return over CVaR. The other performance measures remain the same.

Table 8: Performance of different trading strategies (Mean-95% CVaR)

Panel A: ETF indices and 1/N portfolio				
	Realized Return	Return/CVaR	Sortino	Max Drawdown
SPY	10.233%	2.9636	1.3519	7.590%
DIA	6.773%	1.8335	0.8995	7.041%
QQQ	16.477%	4.2314	2.2277	8.822%
1/N	11.161%	3.4367	1.4937	7.415%
Panel B: Mean-CVaR optimization without short selling				
	Realized Return	Return/CVaR	Sortino	Max Drawdown
ARMA-DCC- SKT	13.114%	4.0400	1.774	6.650%
ARMA-ADCC- SKT	12.869%	3.9714	1.735	6.650%
ARMA-GAS- SKT	13.116%	4.0721	1.786	6.650%
RNN-DCC- SKT	28.757%	9.3672	3.988	6.975%
RNN-ADCC- SKT	28.884%	9.4085	4.005	6.975%
RNN-GAS- SKT	28.940%	9.5513	4.013	7.015%
PSN-DCC- SKT	30.491%	9.8103	4.136	8.312%
PSN-ADCC- SKT	30.608%	9.8482	4.152	8.317%
PSN-GAS- SKT	30.726%	9.8860	4.168	8.404%
Panel C: Mean-CVaR optimization with short selling				
	Realized Return	Return/CVaR	Sortino	Max Drawdown
ARMA-DCC-SKT-S	13.873%	3.9422	1.692	7.958%
ARMA-ADCC-SKT-S	14.190%	4.0384	1.731	7.958%
ARMA-GAS-SKT-S	15.022%	4.2556	1.828	7.958%
RNN-DCC-SKT-S	42.609%	13.5624	5.544	7.835%
RNN-ADCC-SKT-S	43.151%	13.6425	5.507	8.931%
RNN-GAS-SKT-S	42.991%	13.9732	5.598	7.785%
PSN-DCC-SKT-S	43.081%	13.6203	5.503	8.890%
PSN-ADCC-SKT-S	43.307%	13.7847	5.639	7.785%
PSN-GAS-SKT-S	44.294%	14.0037	5.662	9.033%

Note: The table presents the out-of-sample performances over the period January 2014 to March 2015 (68 weekly observations). Panel A reports performances of the three ETF indices and the 1/N portfolio (equally weighted buy-and-hold portfolio). Panel B reports performances of different mean-CVaR portfolios without short-selling. All the portfolios are weekly rebalanced tangency portfolios obtained by the different mean-CVaR optimization based on various model combinations. For example, ARMA-DCC refers to the performance of the tangency portfolio of the efficient frontier of the three ETF assets, where the expected returns are obtained through ARMA forecasts, while the variance-covariance matrix is predicted by DCC. Panel C reports performances of different different mean-CVaR portfolios with short-selling. 'SKT' represents that the 95% CVaR is predicted using a Monte-Carlo simulation with the skewed t copulas to allow for asymmetric tail dependence '-S' denotes optimizations allowing short-selling.

From the table above, QQQ continues to yield the higher Return/CVaR and Sortino ratios along with the highest maximum drawdown. Similar to the results of the M-V optimization, table 8 shows that NN portfolios clearly outperform the ARMA ones. During the NNC process, the target is to minimize the CVaR of the respective portfolio. Therefore, the tangency portfolio is the one with the largest Return/CVaR ratio. The average Return/CVaR ratio and Sortino ratio (11.762 and 4.843 respectively) of portfolios from NN models are remarkably higher than the average of portfolios from the ARMA model (4.053 and 1.758 respectively). As in the M-V case, the gains from NN portfolios should be attributed to the more accurate return predictions generated by the NN models. Meanwhile, portfolios based on PSN forecasts continue to achieve around 2% higher average Return/CVaR ratio and Sortino ratio than the average of the RNN ones. This further confirms that NN portfolios and PSN ones, in particular, provide significant benefits to the investor, when the portfolio optimization is based on a tail risk measure, rather than a symmetric one. Additionally, the short-selling results are improved with the NNC process. The capital gains are not strong in the case of ARMA-based portfolios, but are maximized when it comes to PSN portfolios (above 3% average annualized returns).

Comparing the results between tables 7 and 8, we find that the average Sortino ratio (3.803) of the mean-CVaR portfolio at 95% confidence level is 10% higher than the Sortino ratio of the M-V portfolio (3.456). This result is consistent with our expectation, since the Sortino ratio is a measure of downside risk and fits well with the objective of the CVaR optimization (minimizing the tail risk of portfolio). Additionally, taking into account the asymmetric dependence yields slightly higher benefits than the asymmetric models that treat dependence structure symmetrically. It turns out that the average Return/CVaR ratio and Sortino ratio (11.854 and 4.860 respectively) of portfolios from the skewed t copula models are higher than the average of portfolios from the DCC model (11.590 and 4.793 respectively). The performance of the mean-CVaR portfolios at 99% confidence level is also investigated as an extra robustness test. These results are summarized in table 9 below.

Table 9: Performance of different trading strategies (Mean-99% CVaR)

Panel A: ETF indices and 1/N portfolio				
	Realized Return	Return/CVaR	Sortino Ratio	Max Drawdown
SPY	10.233%	2.9636	1.3519	7.590%
DIA	6.773%	1.8335	0.8995	7.041%
QQQ	16.477%	4.2314	2.2277	8.822%
Naïve	11.161%	3.4367	1.4937	7.415%
Panel B: Mean-CVaR optimization without short-selling				
	Realized Return	Return/CVaR	Sortino Ratio	Max Drawdown
ARMA-DCC-SKT	13.516%	3.6590	1.817	6.650%
ARMA-ADCC-SKT	13.751%	3.7224	1.853	6.650%
ARMA-GAS-SKT	13.533%	3.6636	1.823	6.650%
RNN-DCC-SKT	28.327%	7.2744	3.928	6.857%
RNN-ADCC-SKT	29.469%	7.5679	3.997	6.988%
RNN-GAS-SKT	29.495%	7.5746	4.001	6.971%
PSN-DCC-SKT	29.527%	7.5827	4.005	7.037%

PSN-ADCC-SKT	29.548%	7.5880	4.008	6.988%
PSN-GAS-SKT	29.608%	7.6035	4.016	6.971%
Panel C: Mean-CVaR optimization with short-selling				
	Realized Return	Return/CVaR	Sortino Ratio	Max Drawdown
ARMA-DCC-SKT-S	13.992%	3.2978	1.704	7.958%
ARMA-ADCC-SKT-S	14.336%	3.3789	1.754	7.958%
ARMA-GAS-SKT-S	14.877%	3.5062	1.806	7.958%
RNN-DCC-SKT-S	42.095%	9.3658	5.464	7.690%
RNN-ADCC-SKT-S	41.706%	9.2794	5.319	7.785%
RNN-GAS-SKT-S	42.097%	9.3663	5.363	7.633%
PSN-DCC-SKT-S	42.101%	9.3673	5.380	7.593%
PSN-ADCC-SKT-S	42.558%	9.4689	5.516	7.741%
PSN-GAS-SKT-S	42.731%	9.4510	5.550	7.785%

Note: The table presents the out-of-sample performances over the period January 2014 to March 2015 (68 weekly observations). Panel A reports performances of the three ETF indices and the 1/N portfolio (equally weighted buy-and-hold portfolio). Panel B reports performances of different mean-CVaR portfolios without short-selling. All the portfolios are weekly rebalanced tangency portfolios obtained by the different mean-CVaR optimization based on various model combinations. For example, ARMA-DCC refers to the performance of the tangency portfolio of the efficient frontier of the three ETF assets, where the expected returns are obtained through ARMA forecasts, while the variance-covariance matrix is predicted by DCC. Panel C reports performances of different different mean-CVaR portfolios with short-selling. 'SKT' represents that the 99% CVaR is predicted using a Monte-Carlo simulation with the skewed t copulas to allow for asymmetric tail dependence '-S' denotes optimizations allowing short-selling.

In general, the results at 99% CVaR are consistent with the ones at 95% CVaR. Table 9 shows additionally that when the objective is to minimize CVaR at 99% confidence level, maximum drawdown decreases significantly compared to the 95% case (on average around 6.23%). As mentioned before, the above results are similar also for two-asset portfolios for the respective ETFs and can be found in Online Appendix C.

6. Conclusions

The motivation of this paper is to investigate whether using more accurate ETF forecasts from superior NN models can lead to statistical and economically significant benefits in portfolio management decisions. Firstly, we apply three NN models, namely the MLP, RNN and PSN to the task of forecasting the daily returns of three ETFs. The statistical and trading performance of the NNs is benchmarked with the traditional ARMA. Secondly, a novel dynamic asymmetric copula model (NNC) is introduced in order to capture some of the well-documented features of the dependence structure across ETF returns. Thirdly, weekly re-balanced portfolios are obtained and compared by using the traditional M-V approach and mean-CVaR optimization.

This study provides several interesting findings related to the NN and portfolio optimization literature. In terms of the forecasting and trading performance of the individual models, the results suggest that PSN outperforms MLP and RNN models, while all NN structures provide more accurate forecasts and higher profitability over the ARMA models. The results related to model selection and

goodness of fit tests indicate that the asymmetric skewed t copula statistically outperforms symmetric copulas when it comes to modelling ETF returns dependence. Our empirical study also shows solid evidence that when the most accurate forecasts are exploited, the NNC model leads to significant improvements in the portfolio optimization process. Compared with portfolios from the classical ARMA model, portfolios from NN models can deliver significantly higher out-of-sample risk adjusted returns, when evaluate through the Sharpe and Sortino ratio or the novel Return/CVaR ratio. The results further confirm that NN portfolios and PSN ones, in particular, provide significant benefits to the investor, when the portfolio optimization is based on a tail risk measure (CVaR), rather than on a symmetric one (variance).

In addition, we find that forecasting covariance taking into account asymmetric dependence improves the performance of optimization, however the magnitude of this improvement is relatively small. In that sense, it is implied that the portfolio optimization benefits are driven mainly from the accurate ETF predictions, particularly the PSN ones, rather than the variance-covariance matrix estimates. This findings is in line with the relevant strand of the literature suggesting the forecasts of asset returns are the most important inputs for the mean-variance optimization. Overall, this study provides insight in the ETF market and sheds light to the difficult quest of optimal portfolio optimization procedures.

Acknowledgements

We wish to thank the two anonymous referees for useful comments.

Funding

The work is partially supported by Project of Key Research Institute of Humanities & Social Science in Jiangxi Province [grant number JD16063].

References

- Andersen, T.G., Bollerslev, T., Christoffersen, P.F., and Diebold, F.X., Volatility and correlation forecasting. In *Handbook of Economic Forecasting*, edited by G. Elliott and A. Timmermann, pp. 777-878. 2006 (North Holland, Netherlands).
- Ang, A. and Chen, J., Asymmetric correlations of equity portfolios. *J. Financ. Econ.*, 2002, **63**(3), 443-494.
- Ang, A., Chen, J., and Xing, Y., Downside risk. *Rev. Financ. Stud.*, 2006, **19**(4), 1191-1239.
- Artzner, P., Delbaen, F., Eber, J. M., and Heath, D., Coherent measures of risk. *Math. Finance*, 1999, **9**(3), 203-228.
- Avellaneda, M. and Lee, J.H., Statistical arbitrage in the US equities market. *Quant. Finance*, 2010, **10**(7), 761-782.
- Best, M.J. and Grauer, R.R., On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Rev. Financ. Stud.*, 1991, **4**(2), 315-342.
- Boubaker, H. and Sghaier, N., Portfolio optimization in the presence of dependent financial returns with long memory: A copula based approach. *J. Bank. Financ.*, 2013, **37**(2), 361-377.
- Brooks, C., *Introductory Econometrics for Finance*, 2nd Ed., 2008 (Cambridge University Press: Cambridge).
- Cappiello, L., Engle, R.F. and Sheppard, K., Asymmetric dynamics in the correlations of global equity and bond returns. *J. Financ. Econ.*, 2006, **4**(4), 537-572.
- Chan, L.K., Karceski, J. and Lakonishok, J., On portfolio optimization: Forecasting covariances and choosing the risk model. *Rev. Financ. Stud.*, 1999, **12**(5), 937-974.
- Chopra, V.K., Hensel, C.R. and Turner, A.L., Massaging mean-variance inputs: returns from alternative global investment strategies in the 1980s. *Mgmt. Sci.*, 1993, **39**(7), 845-855.
- Chopra, V.K. and Ziemba, W.T., The effect of errors in means, variances, and covariances on optimal portfolio choice. *J. Portfol. Mgmt.*, 1993, **19**, 51-62.
- Christoffersen, P., Errunza, V., Jacobs, K. and Langlois, H., Is the potential for international diversification disappearing? A dynamic copula approach. *Rev. Financ. Stud.*, 2012, **25**(12), 3711-3751.
- Christoffersen, P. and Langlois, H., The joint dynamics of equity market factors. *J. Financ. Quant. Anal.*, 2013, **48**(05), 1371-1404.
- Chu, B., Recovering copulas from limited information and an application to asset allocation. *J. Bank. Financ.*, 2011, **35**(7), 1824-1842.
- Creal, D., Koopman, S.J. and Lucas, A., Generalized autoregressive score models with applications. *J. App. Econometr.*, 2013, **28**(5), 777-795.
- Creal, D., Schwaab, B., Koopman, S.J. and Lucas, A., Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Rev. Econ. Stat.*, 2014, **96**(5), 898-915.
- Demarta, S. and McNeil, A.J., The t copula and related copulas. *Inter. Stat. Rev.*, 2005, **73**(1), 111-129.
- Diebold, F.X. and Mariano, R.S., Comparing predictive ability, *J. Bus. Econ. Stat.*, 1995, **13**(3), 253-263.
- Dolvin, S. D., S&P ETFs: arbitrage opportunities and market forecasting, *J. Index Invest*, 2010, **1**(1), 107-116.
- Dunis, C.L., Laws, J., and Sermpinis, G., Higher order and recurrent neural architectures for trading the EUR/USD exchange rate. *Quant. Finance*, 2011, **11**(4), pp. 615-629.
- Ebrahimpour, R., Nikoo, H., Masoudnia, S., Yousefi, M.R. and Ghaemi, M.S., Mixture of MLP-experts for trend forecasting of time series: A case study of the Tehran stock exchange. *Int. J. Forecasting*, 2011, **27**(3), pp. 804-816.
- Elman, J.L., Finding structure in time, *Cog. Sci.*, 1990, **14**(2), 179-211.
- Engle, R., Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Stat.*, 2002, **20**(3), 339-350.
- Fei, P., Ding, L. and Deng, Y., Correlation and volatility dynamics in REIT returns: performance and portfolio considerations. *J. Portfol. Mgmt.*, 2010, **36**(2), 113.
- Garcia, R., and Tsafack, G., Dependence structure and extreme comovements in international equity and bond markets. *J. Bank. Financ.*, 2011, **35**(8), 1954-1970.
- Gaivoronski, A.A. and Pflug, G., Value-at-risk in portfolio optimization: properties and computational approach. *J. Risk*, 2005, **7**(2), 1-31.
- Ghosh, J. and Shin, Y., The Pi-Sigma Network: An efficient Higher-order Neural Networks for Pattern Classification and Function Approximation, Proceedings of International Joint Conference of Neural Networks, 1, 13-18, 1991.
- Guresen, E., Kayakutlu, G., and Daim, T.U., Using artificial neural network models in stock market index prediction. *Expert Syst. Appl.*, 2011, **38**(8), pp. 10389-10397.
- Hansen, B.E., Autoregressive conditional density estimation. *Inter. Econ. Rev.*, 1994, **1**, 705-730.
- Harvey, C.R., and Siddique, A., Autoregressive conditional skewness. *J. Financ. Quant. Anal.*, 1999, **34**(4), 465-487.
- He, X. and Gong, P., Measuring the coupled risks: A copula-based CVaR model. *J. Comput. App. Math.*, 2009, **223**(2), 1066-1080.
- Hong, Y., Tu, J. and Zhou, G., Asymmetries in stock returns: statistical tests and economic evaluation. *Rev. Financ. Stud.*, 2007, **20**(5), 1547-1581.

- Jondeau, E., and Rockinger, M., Conditional volatility, skewness and kurtosis: Existence, persistence, and comovements, *J. Econ. Dynam. Control*, 2003, **27**(10), pp. 1699-1737.
- Jondeau, E. and Rockinger, M., Optimal portfolio allocation under higher moments, *Eur. Financ. Mgmt.*, 2006a, **12**(1), 29-55.
- Jondeau, E. and Rockinger, M., The copula-garch model of conditional dependencies: An international stock market application. *J. Int. Money Financ.*, 2006b, **25**(5), 827-853.
- Longin, F. and Solnik, B., Extreme correlation of international equity markets. *J. Finance*, 2001, **56**(2), 649-676.
- Low, R. K. Y., Alcock, J., Faff, R. and Brailsford, T., Canonical vine copulas in the context of modern portfolio management: Are they worth it? *J. Bank. Financ.*, 2013, **37**(8), 3085-3099.
- Lucas, A., Schwaab, B. and Zhang, X., Conditional euro area sovereign default risk. *J. Bus. Econ. Stat.*, 2014, **32**(2), 271-284.
- Markowitz, H., Portfolio selection. *J. Finance*, 1952, **7**(1), 77-91.
- Markowitz, H., *Portfolio Selection, Efficient Diversification of Investments*, 1959 (Wiley: New York, NY).
- Marshall, B.R., Nguyen, N.H. and Visaltanachoti, N., ETF arbitrage: Intraday evidence, *J. Bank. Financ.*, 2013, **37**(9), pp. 3486-3498.
- McNeil, A.J., Frey, R. and Embrechts, P., *Quantitative Risk Management: Concepts, Techniques, and Tools*, 2005 (Princeton University Press: Princeton, NJ).
- McNelis, P.D., *Neural Networks in Finance: Gaining Predictive Edge in the Market*, 2005 (Elsevier Academic Press: London).
- Menchero, J., Wang, J. and Orr, D.J., Improving risk forecasts for optimized portfolios. *Financial Anal. J.*, 2012, **68**(3), 40-50.
- Patton, A.J., On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *J. Financial Econometr.*, 2004, **2**(1), 130-168.
- Pesaran, M.H. and Timmerman, A.G., A simple nonparametric test of predictive performance, *J. Bus. Econ. Stat.*, 1992, **10**(4), pp. 461-465.
- Quaranta, A.G. and Zaffaroni, A., Robust optimization of conditional value at risk and portfolio selection. *J. Bank. Financ.*, 2008, **32**(10), 2046-2056.
- Rockafellar, R.T. and Uryasev, S., Optimization of conditional value-at-risk. *J. Risk*, 2000, **2**, 21-42.
- Rockafellar, R. T., & Uryasev, S., Conditional value-at-risk for general loss distributions. *J. Bank. Financ.*, 2002, **26**(7), 1443-1471.
- Salvatierra, I. D.L. and Patton, A.J., Dynamic copula models and high frequency data. *J. Emp. Financ.*, 2015, **30**, 120-135.
- Sermpinis, G., Stasinakis, C. and Dunis, C., Stochastic and genetic neural network combinations in trading and hybrid time-varying leverage effects. *J. Inter. Financial Mark. Instit. Mon.*, 2014, **30**, 21-54.
- Saqdique, S. and Silvapulle, P., Long-term memory in stock market returns: International evidence. *Inter. J. Financ. Econ.*, 2001, **6** (1), 59-67.
- Shapiro, A.F., A Hitchhiker's guide to the techniques of adaptive nonlinear models, *Insur. Math. Econ.*, 2000, **26**(2-3), 119-132.
- Stasinakis, C., Sermpinis, G., Psaradellis, I. and Verousis, T., Krill-Herd Support Vector Regression and heterogeneous autoregressive leverage: evidence from forecasting and trading commodities. *Quant. Finance*, 2016, **16**(12), 1-15.
- Syriopoulos, T. and Roumpis, E., Dynamic correlations and volatility effects in the Balkan equity markets. *J. Int. Financial Mark. Instit. Mon.*, 2009, **19**(4), 565-587.
- Tenti, P., Forecasting foreign exchange rates using recurrent neural networks, *Appl. Arti. Intelligence*, 1996, **10**(6), 567-581.
- Wang, Y. H., Nonlinear neural network forecasting model for stock index option price: Hybrid GJR-GARCH approach. *Expert Syst. Appl.*, 2009, **36**(1), 564-570.
- Wyart, M., Bouchaud, J.P., Kockelkoren, J., Potters, M. and Vettorazzo, M., Relation between bid-ask spread, impact and volatility in order-driven markets, *Quant. Finance*, 2008, **8**(1), 41-57.
- Zhang, G., Patuwo, B.E. and Hu, M.Y., Forecasting with artificial neural networks: The state of the art, *Int. J. Forecasting*, 1998, **14** (1), 35-62.

Appendix

A. Statistical and trading performance measures.

The statistical and trading performance measures of the forecasting models are calculated as shown in table B.1 and B.2 respectively.

Table B.1: Statistical performance measures

Statistical performance measures	Description
Mean absolute error	$MAE = \left(\frac{1}{n}\right) \sum_{\tau=t+1}^{t+n} \left \hat{Y}_{\tau} - Y_{\tau} \right $ with Y_{τ} being the actual value and \hat{Y}_{τ} the forecasted value
Mean absolute percentage error	$MAPE = \frac{1}{n} \sum_{\tau=t+1}^{t+n} \left \frac{Y_{\tau} - \hat{Y}_{\tau}}{Y_{\tau}} \right $
Root mean squared error	$RMSE = \sqrt{\frac{1}{n} \sum_{\tau=t+1}^{t+n} (\hat{Y}_{\tau} - Y_{\tau})^2}$
Theil-U	$Theil-U = \frac{\sqrt{\left(\frac{1}{n} \sum_{\tau=t+1}^{t+n} (\hat{Y}_{\tau} - Y_{\tau})^2\right)}}{\sqrt{\frac{1}{n} \sum_{\tau=t+1}^{t+n} \hat{Y}_{\tau}^2 + \frac{1}{n} \sum_{\tau=t+1}^{t+n} Y_{\tau}^2}}$

Table B.2: Trading performance measures

Trading performance measures	Description
Annualised return after transaction costs	$R^A = 252 * \frac{1}{N} * \left(\sum_{t=1}^N R_t\right) - TC^A$ where R_t the daily return and TC^A the annualized transaction cost
Annualised volatility	$\sigma^A = \sqrt{252} * \sqrt{\frac{1}{N-1} * \sum_{t=1}^N (R_t - \bar{R})^2}$
Sharpe ratio	$SR = \frac{R^A - R^f}{\sigma^A}$
Maximum drawdown	Maximum negative value of $\sum(R_t)$ over the period $MD = \underset{i=1, \dots, t; j=1, \dots, N}{Min} \left(\sum_{j=i}^t R_j \right)$

B. Modelling marginal density

The Ljung-Box test and the Engle's LM test show that the ETF return series exhibit some degree of autocorrelation and heteroscedasticity. In order to compensate for autocorrelation, the conditional mean is modelled with a simple ARMA model:

$$r_{i,t} = c + \sum_{j=1}^p \phi_j r_{i,t-j} + \sum_{k=1}^q \theta_k \varepsilon_{i,t-k} + \varepsilon_{i,t} \quad (C.1)$$

where $\varepsilon_{i,t} = \sigma_{i,t} z_{i,t}$. To capture the heteroscedasticity and asymmetric volatility clustering of ETF returns, we model the conditional variance using the GJR-GARCH dynamics:

$$\sigma_{i,t}^2 = \omega + \sum_{j=1}^p \alpha_j \varepsilon_{i,t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{i,t-k}^2 + \sum_{k=1}^q \gamma_k \varepsilon_{i,t-k}^2 I[\varepsilon_{i,t-k} < 0] \quad (\text{C.2})$$

The indicator function $I[\varepsilon_{i,t-k} < 0]$ equals 1 if $\varepsilon_{i,t-k} < 0$, and 0 otherwise. Applying this, allows us to capture the “leverage effect”, which implies lower returns than expected are followed with higher levels of volatility. Using these models, we construct the estimated standardized residuals as:

$$z_{i,t} = \frac{r_{i,t} - c - \sum_{j=1}^p \varphi_j r_{i,t-j} - \sum_{k=1}^q \theta_k \varepsilon_{i,t-k}}{\sigma_{i,t}} \quad (\text{C.3})$$

Descriptive statistics of ETF indices returns show that all the indices exhibit significant skewness and the hypothesis of normality is rejected by the Jarque–Bera test. In order to compensate for the skewness, we use the univariate skewed t distribution of Hansen (1994) to model the standardized residuals of each ETF index. Assuming $z_{i,t} \sim F_{skt}(\eta_i, \lambda_i)$, then

$$u_{i,t} = F_{skt}(z_{i,t}; \eta_i, \lambda_i), \quad \eta_i \in (2, \infty), \quad \lambda_i \in [-1, 1] \quad (\text{C.4})$$

where $u_{i,t}$ is the probability integral transform of $z_{i,t}$, λ_i is the skewness parameter and η_i is the degrees of freedom.

C. Copula modelling: multivariate non-normality and asymmetry

As mentioned in section 5, this study applies the skewed t copula proposed by Demarta and McNeil (2005). The cumulative distribution function of this skewed t copula is given by:

$$\mathbf{C}_{skt}(u_{1,t}, \dots, u_{n,t}; \Sigma, \lambda, \nu) = \mathbf{F}_{skt}(F_1^{-1}(u_{1,t}), \dots, F_n^{-1}(u_{n,t})) \quad (\text{D.1})$$

where λ is the parameter of skewness, ν is the parameter of degree of freedom, \mathbf{F}_{skt} is the cumulative distribution function of the multivariate skewed t density with correlation matrix Σ , and F_i^{-1} is the inverse cumulative distribution function of the univariate skewed t distribution.

From Patton (2006), if the joint distribution function \mathbf{F}_{skt} is n -times differentiable, the following equation is obtained by taking the n^{th} cross-partial derivative:

$$\begin{aligned}
\mathbf{f}_{skt}(u_{1,t}, \dots, u_{n,t}) &= \frac{\partial^n}{\partial x_1 \dots \partial x_n} \mathbf{F}_{skt}(z_{1,t}, \dots, z_{n,t}) \\
&= \prod_{i=1}^n f_i(z_{i,t}) \cdot \frac{\partial^n}{\partial x_1 \dots \partial x_n} \mathbf{C}_{skt}(F_1(z_{1,t}), \dots, F_n(z_{n,t})) \\
&= \prod_{i=1}^n f_i(z_{i,t}) \cdot \mathbf{c}_{skt}(F_1(z_{1,t}), \dots, F_n(z_{n,t}))
\end{aligned} \tag{D.2}$$

The equation suggests that the joint density \mathbf{f}_{skt} is equal to the product of the marginal densities and the skewed t copula density \mathbf{c}_{skt} . Thus, the joint log-likelihood is equal to the sum of univariate log-likelihood and the skewed t copula log-likelihood:

$$\log \mathbf{f}_{skt}(z_{1,t}, \dots, z_{n,t}) = \sum_{t=1}^T \sum_{i=1}^n \log f_{i,t}(z_{i,t}) + \sum_{t=1}^T \log \mathbf{c}_{skt}(F_1(z_{1,t}), \dots, F_n(z_{n,t})) \tag{D.3}$$

More details on the implementation of the skewed t copula can be found in Christoffersen *et al.*, (2012). Our choice to use the skewed t copula is supported by the literature as mentioned in section 5. Nonetheless, we also perform a comparative analysis between ten common copula alternatives to verify this selection. The analysis is based on three information criteria, the value of log-likelihood (LL), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The results of this analysis are presented in the table below.

Table D.1: Estimations for Different Copulas

Panel A: SPY - DIA						
	P1	P2	P3	LL	AIC	BIC
Gaussian	0.933			123.38	-244.76	-242.54
Clayton	5.838			129.15	-256.31	-254.09
Survival Clayton	3.364			79.20	-156.40	-154.18
Plackett	68.852			121.69	-241.37	-239.15
Frank	9.000			103.85	-205.70	-203.48
Gumbel	3.798			106.81	-211.62	-209.40
Survival Gumbel	4.663			125.07	-248.14	-245.92
Symmetrized Joe–Clayton	0.861	0.664		128.48	-252.97	-248.53
Student's t	0.900	2.158		125.80	-247.60	-243.17
Skewed Student's t	0.929	2.261	-0.063	131.53	-257.06	-250.40
Panel B: SPY - QQQ						
	P1	P2	P3	LL	AIC	BIC
Gaussian	0.864			82.96	-163.91	-161.69
Clayton	2.936			73.60	-145.21	-142.99
Survival Clayton	2.375			60.60	-119.19	-116.97
Plackett	29.252			78.64	-155.28	-153.06
Frank	9.000			74.90	-147.81	-145.59
Gumbel	2.789			75.79	-149.58	-147.36
Survival Gumbel	2.972			83.39	-164.77	-162.55

Symmetrized Joe–Clayton	0.759	0.653		83.58	-163.17	-158.73
Student's t	0.873	8.796		84.02	-164.04	-159.60
Skewed Student's t	0.850	9.389	-0.040	88.50	-171.00	-164.34
Panel C: DIA - QQQ						
	P1	P2	P3	LL	AIC	BIC
Gaussian	0.738			47.68	-93.35	-91.13
Clayton	2.024			50.65	-99.30	-97.08
Survival Clayton	1.194			27.07	-52.14	-49.92
Plackett	12.186			44.14	-86.28	-84.06
Frank	6.346			44.15	-86.31	-84.09
Gumbel	1.929			37.33	-72.65	-70.43
Survival Gumbel	2.190			51.98	-101.96	-99.74
Symmetrized Joe–Clayton	0.699	0.261		51.51	-99.03	-94.59
Student's t	0.753	44.970		57.74	-111.47	-107.03
Skewed Student's t	0.750	38.889	-0.050	59.82	-113.64	-106.98

Note: Columns P1 to P3 report the estimated parameters. Columns LL, AIC and BIC report the values of log-likelihood, the Akaike information criterion and the Bayesian information criterion, respectively. For each column of information criterion, the best three models are in bold.

From the table above, it is obvious that the skewed t copula consistently provides maximum LL and minimum AIC and BIC values. This implies that for the respective set of data the skewed t copula is always preferred.

D. The NNC portfolio optimization pseudo-algorithm

The following steps summarize the proposed NNC process:

- **Step 1:** Using different NN models and a benchmark model (i.e. ARMA model) to forecast one-step-ahead asset returns. Since the PSN and RNN models provide the best and second best out-of-sample statistical and trading performance respectively, we use them to calculate the weekly expected returns for each ETF index in the optimization part.
- **Step 2:** Consider an investor who has a one-period horizon and constructs a dynamically rebalanced portfolio. The time-varying weights of this portfolio requires one-step-ahead forecasts of the expected return and the conditional variance–covariance matrix. We use the conditional mean predicted by the ARMA, RNN and PSN models and the conditional variance–covariance matrix predicted by DCC, ADCC and GAS models to perform weekly rebalance portfolio optimization.
- **Step 3:** Copula modelling using in-sample data. Initially, we need to characterize individually the distribution of returns of each asset. Specifically, we use ARMA and GJR-GARCH to estimate the conditional mean and conditional volatility, respectively, and apply the skewed t distribution of Hansen (1994) to get the probability integral transforms of the standardized residuals.

- **Step 4:** Using the probability integral transforms estimated from the last step, the GAS model for the skewed t copula is estimated in order to obtain the time-varying correlation matrix. The time-varying correlation matrix is also obtained for comparison purposes by the DCC or ADCC GARCH.
- **Step 5:** Given the time-varying correlation matrix, the jointly-dependent uniform variates can be simulated by the skewed t copula random number generator.
- **Step 6:** Following Rockafellar and Uryasev (2000, 2002), a linear programming technique is applied to find optimal weights of ETFs that calculate a minimum CVaR for a certain level of return at time t
- **Step 7:** Repeat Step 1 to Step 6 using rolling window for weekly rebalance.