INTERNATIONAL MONETARY FUND

Forecasting Tail Risk via Neural Networks with Asymptotic Expansions

Yuji Sakurai and Zhuohui Chen

WP/24/99

2024
MAY

WORKING PAPER

 WP/24/99

IMF Working Paper
Monetary and Capital Markets Department

Forecasting Tail Risk via Neural Networks with Asymptotic Expansions
Prepared by Yuji Sakurai and Zhuohui Chen*

Authorized for distribution by Jihad Alwazir
May 2024

ABSTRACT: We propose a new machine-learning-based approach for forecasting Value-at-Risk (VaR) named CoFiE-NN where a neural network (NN) is combined with Cornish-Fisher expansions (CoFiE). CoFiE-NN can capture non-linear dynamics of high-order statistical moments thanks to the flexibility of a NN while maintaining interpretability of the outputs by using CoFiE which is a well-known statistical formula. First, we explain CoFiE-NN. Second, we compare the forecasting performance of CoFiE-NN with three conventional models using both Monte Carlo simulation and real data. To do so, we employ Long Short-Term Memory (LSTM) as our main specification of the NN. We then apply the CoFiE-NN for different asset classes, with a focus on foreign exchange markets. We report that CoFiE-NN outperfoms the conventional EGARCH-t model and the Extreme Value Theory model in several statistical criteria for both the simulated data and the real data. Finally, we introduce a new empirical proxy for tail risk named *tail risk ratio* under CoFiE-NN. We discover that the only 20 percent of tail risk dynamics across 22 currencies is explained by one common factor. This is contrasting to the fact that 60 percent of volatility dynamics across the same currencies is explained by one common factor.

RECOMMENDED CITATION: Sakurai, Y., Chen, Z., “Forecasting Tail Risk via Neural Networks with Asymptotic Expansions”

JEL Classification Numbers:	C45, C53, G17
Keywords:	Machine learning; Value-at-Risk; Neural Network
Authors’ E-Mail Addresses:	ysakurai@imf.org; zchen4@imf.org

* We would like to deeply thank Romain Veyrune for encouraging us to conduct this research. We are grateful to Christopher Erceg, Jason Wu, Tsendsuren Batsuuri, Michael Gorbanyov Gabriela Elizabeth Conde Vitureira, Istvan Mak, Tetsuo Kurosaki, Nikolaos Kourentzes, Toshinao Yoshiba, Keiichi Goshima, Kanji Suzuki, and the seminar participants at the IMF for their comments. All remaining errors are ours.

WORKING PAPERS

Forecasting Tail Risk via Neural Networks with Asymptotic Expansions

Prepared by Yuji Sakurai and Zhuohui Chen[1]

[1] Production assistance provided by Julie Vaselopulos.

Contents

Glossary

CAViaR	Conditional Autoregressive Value at Risk
CoFiE	Cornish-Fisher Expansions
EVT	Extreme Value Theory
FFNN	Feed-Forward Neural Network
GFC	Global-Financial-Crisis
GPD	Generalized Pareto Distribution
LSTM	Long Short-Term Memory Model
ML	Machine Learning
NN	Neural Network
PCA	Principal Component Analysis
Pre-GFC	Pre-Global-Financial-Crisis
TRR	Tail Risk Ratio
VaR	Value-at-Risk

I. Introduction

Forecasting tail risk has been one of the essential topics for both academics and practitioners. In real-world implementation, forecasting tail risk often takes the form of predicting Value-at-Risk (VaR) that is a measure of the potential loss for a given probability. Among many, predicting VaR is used in financial risk management, trading risk regulations, and early warning system.[1]

In this paper, we propose a new machine-learning-based approach for forecasting VaR named CoFiE-NN where a neural network (NN) is combined with Cornish-Fisher expansions (CoFiE).[2] Two advantages of CoFiE-NN are flexibility and interpretability. On one hand, the CoFiE-NN can represent nonlinear relationship between statistical moments because of the universal approximation nature of NNs and hence it is flexible to capture the nonlinear dynamics of moments. On the other hand, the CoFiE-NN explicitly links statistical moments with the percentile of distribution based on a well-known statical formula named Cornish-Fisher expansions and thus it is easy to interpret which moments impact VaR.

Cornish-Fisher expansions has been studied in the literature of financial risk management. The previous studies include, but not limited to, Jaschke (2002), Christoffersen and Gonçalves (2005), Giamouridis (2006), Lönnbark (2016). To the best of our knowledge, this is the first study to combine the Cornish-Fisher expansions with a NN and conduct the empirical applications.

Combining Cornish-Fisher expansions with a neural network has several advantages.[3] First, the Cornish-Fisher expansions are relatively simple to implement in any computational language.[4] Second, it allows us to examine the impact of skewness and kurtosis. For example, the formula shows that 99 percentile is larger than 2.33, which is the number based on the standard normal distribution, if the excess kurtosis is positive due to fat-tail feature. Third, and most important, it helps us predict VaR even with relatively small amounts of data. This feature is advantageous especially when we are interested in foreign exchange rates in developing countries or emerging markets where it is difficult to obtain sufficiently long historical data or the FX policy regime has changed recently.

In our empirical analyses, we test the performance of CoFiE-NN with two types of neural network models: Feed-Forward Neural Network (FFNN) and Long Short-Term Memory (LSTM). FFNN is the simplest NN. We are interested in whether CoFiE-NN outperforms conventional models even when the NN component is specified as a simple FFNN. LSTM is more sophisticated than FFNN. LSTM has been successfully used in time series modeling because it captures both short-term and long-term relationship between inputs and outputs. Our main specification is LSTM.

We compare the performance of VaR forecasting based on CoFiE-NN with three conventional models using both Monte Carlo simulated data and real data. The three conventional models are EGARCH-t model, CAViaR model, and Extreme Value Theory (EVT) model with the generalized Pareto distribution (GPD). We use these three models as benchmarks, following the previous studies such as Wu and Yan (2019). Testing with the

[1] Jorion (2006) discusses the use of VaR from the view of financial risk management and trading risk regulations.

[2] This asymptotic expansion was first derived by Cornish and Fisher (1938).

[3] Amédée-Manesme et al. (2019) point out that the fourth advantage is no assumption about the time horizon.

[4] For expository simplicity, we consider the Cornish-Fisher expansion up to the fourth order.

simulated data helps us cleanly investigate under what conditions CoFiE-NN beats the conventional models, while testing with real data is more directly relevant to practitioners' interest.

To conduct rigorous statistical analyses, we employ Kupiec (1995) test, Christoffersen (1998) test, and Lopez (1999) quadratic loss function. Kupiec (!995) test allows us to test whether breaches of VaR forecast are too many or too less. Christoffersen (1998) test allows us to analyze whether VaR breaches are independent or autocorrelated. Specifically, we conduct a joint test of Kupiec (1995) and Christoffersen (1998). Lopez (1999) quadratic loss function allows us to measure the magnitude of VaR breaches.

For the simulated data, we report that the CoFiE-NN with LSTM tends to outperform the EGARCH-t model when the sample period is relatively short. This is surprising because the simulated data is generated from the extended EGARCH-t model with stochastic parameters but resembling enough to the conventional EGHARCH-t model. By contrast, the CoFiE-NN underperforms CAViaR in terms of Kupiec (1995) and the joint test but outperform it in Lopez (1999) loss function under all settings of the training data size.

We then apply the CoFiE-NN for 30 assets across different asset classes, with a special emphasis on foreign exchange markets where high-order moments could be a key in forecasting VaR. We set up four sample periods to examine the out-of-sample forecasting performance in different market environments: Pre-Global-Financial-Crisis (Pre-GFC), Global-Financial-Crisis (GFC), Pre-COVID-19-Crisis (Pre-Covid), COVID-19-Crisis (Covid) periods. We consider that studying the performance of CoFiE-NN in both normal and crisis periods is important because there is a concern that Machine Learning (ML) approach performs poorly in the stressed market environment.

We find that the CoFiE-NN tends to show the better performance relative to the EGARCH-t model and the EVT in the real data in several statistical criteria. Specifically, CoFiE-NN with LSTM outperforms the EGARCH-t in terms of Kupiec (1995) test and the joint under the all four different sample periods except one case although it underperforms the EGARCH-t in terms of Lopez (1999) quadratic loss function. By contrast, CoFiE-NN underperforms CAViaR in all criteria under the four sample periods in general. Finally, we report that CoFiE-NN outperforms the EVT in all criteria for all four different sample periods except one case.

Finally, we discuss how the forecast of VaR under CoFiE-NN can be used to monitor tail risk. We introduce an empirical measure named *tail risk ratio* that is volatility-scaled VaR. The tail risk ratio allows us to make cross-country analysis easy because we can compare large VaR of high-volatility currency and small VaR of low-volatility currency using the tail risk ratio in a consistent way. We construct the tail risk ratios for 22 currencies during 2019/8-2023/7 period. We then conduct principal component analysis (PCA) to extract common factors in a similar manner to Longstaff et al. (2011).[5]

We discover that the only 20 percent of tail risk dynamics across 22 currencies is explained by one common factor. This is contrasting to the fact that 60 percent of volatility dynamics across the same set of currencies is explained by one common factor. We also look at coefficients of the top three PCA factors for both tail risk ratios and volatilities. The results indicate that: (i) the first PCA factor is a global factor for both the tail risk ratios and volatilities; (ii) the second PCA factor of the tail risk ratios is related to Chinese Yuan while the second PCA factor of volatilities is associated with Latin American countries; and (iii) the third PCA factors of

[5] Longstaff et al. (2011) conduct PCA of the changes in sovereign Credit Default Swap (CDS) spreads for 26 countries and find that most of the sovereign credit risk can be linked to global factors.

both tail risk ratios and volatilities are difficult to interpret but they are related to Eastern European countries although the third PCA factor of tail risk ratios is also related to Asian countries.

We also briefly discuss the implications of the tail risk ratio for portfolio optimization problems. We show that the tail risk ratio naturally arises in the context of optimal asset allocation under a VaR constraint.

Our primary focus in this paper is forecasting tail risks for the FX markets, but we emphasize that CoFiE-NN is applicable to various contexts where conventional models have been employed for forecasting VaR. The most relevant application is a FX intervention strategy based on VaR proposed by Lafarguette and Veyrune (2021). They employ VaR to define the non-intervention range of FX spot rate return. If the return goes above (below) the upper (lower) bound, a central bank has an option to intervene. They explain that the VaR-based FX intervention rule provides a hedge against tail risk to the market participants while it allows the exchange rate smoothly adjusting its new level.

In addition to the rule-based FX intervention strategy, we can apply VaR forecasting with CoFiE-NN for other purposes such as financial risk management and early warning system. Jorion (2006) describes how VaR is used for financial risk management. De Nicolò and Lucchetta (2017) employ VaR of macroeconomic variables and financial indicators as early warning system.

The rest of the paper is organized as follows. Section 2 reviews the literature. Section 3 explains the CoFiE-NN framework. Section 4 explains the data set. Section 5 reports our main results. Section 6 discusses an empirical application of the tail risk ratio under CoFiE-NN. Section 7 concludes.

II. Literature

This study contributes to three strands of the literature.

The first strand of the literature is application of machine learning for forecasting VaR. The literature on ML-based VaR forecasting is scarce. We find four relevant papers. Taylor (2000) is an early study that integrates a quantile regression with a feed-forward neural network with a single hidden layer. Similarly, Buczynski and Chlebus (2023) combine GARCH model with a neural network while Chronopoulos et al. (2023) integrates quantile regression with a neural network for forecasting VaR. Wu and Yan (2019) develop a conditional quantile model which is more closely related to our CoFiE-NN approach than Buczynski and Chlebus (2023) and Chronopoulos et al. (2023). They combine LSTM with a new heavy-tailed quantile function which has four inputs: mean, volatility, up-tail-controlling and down-tail-controlling parameters. They set up LSTM to generate these four parameters over time. Our work is different from these studies in that we develop a framework for forecasting value-at-risk by combining a modern ML approach with Cornish-Fisher expansions established in statistics. As we discuss below, the CoFiE-NN framework performs well even with relatively small size of data because Cornish-Fisher expansions reduce the problem of forecasting VaR to the problem of forecasting moments. Also, the CoFiE-NN framework allows us to forecast Expected Shortfall (ES) without additional calibration, although forecasting ES is not our focus in this paper.

The second strand of the literature is application of machine learning for forecasting VaR. There have been several recent studies on forecasting volatilities which is the second-order moment, including but not limited to Zhang et al. (2023), Zhu et al. (2023), and Niu et al. (2023). Zhang et al. (2023) employes LSTM for equity volatility forecasting. Zhu et al. (2023) employ deep NN for S&P 500 firms' equity volatility forecasting. Niu et al. (2023) employ different ML approaches such as Convolutional Neural Network (CNN), Random Forest, Support Vector Machine, for equity volatility forecasting Our paper is different from these studies because the CoFiE-NN is mainly designed to forecast VaR.

The third strand of the literature is the application of machine learning for foreign exchange markets. Amat et al. (2018) employ ridge regressions for forecasting the level of the exchange rates for 12 major industrial economies. They find that the machine learning approach outperforms OLS regressions when forecasting exchange rates at one month horizon. Rojas and Herman (2018) employ several machine learning techniques including a simple feed forward neural networks, but not including LSTM. They analyze the Mexican peso against the US dollar and find that Support Vector Machine shows the best performance. Taylor (2000) develops a quantile regression with a single-hidden-layer feed-forward neural network for forecasting VaR of exchange rates. Unlike the previous studies, our focus is forecasting the tail risk of the foreign exchange markets with LSTM which is a recently developed neural network model.

III. CoFiE-NN Framework

For expository simplicity, we explain the CoFiE-NN framework using the Cornish-Fisher expansions up to the fourth order. It means that mean, volatility, skewness, and kurtosis show up in the expansion. We can apply the same framework up to any high order of cumulants.

Let us denote time series of log return of the foreign exchange rate with r_t. We first compute statistical cumulants of the log return from the 1st order to 4-th order $\mathrm{K}_t = (\kappa_{1,t}, \kappa_{2,t}, \kappa_{3,t}, \kappa_{4,t})$. Note that skewness is $\frac{\kappa_3}{\kappa_2^{1.5}}$ and excess kurtosis is $\frac{\kappa_4}{\kappa_2^2}$. We then estimate a neural network specified as

$$\mathrm{K}_{t+1} = f(\mathrm{K}_t, \mathrm{K}_{t-1}, .. \mathrm{K}_{t-s}, Z_t) \tag{1}$$

Where f is the neural network to model the dynamics of moments or equivalently, cumulants. Z_t is a vector of exogenous variables. In our empirical applications, we employ LSTM as our main specification. K_{t-s} is a vector of the cumulants at time $t-s$.

The building blocks of the CoFiE-NN is outlined as follows:

1. Compute the mean, variance, skewness, and kurtosis using the historical data of the asset prices. For example, one can compute 10-day mean, variance, skewness, and kurtosis.

2. Train the neural network using the historical data of the cumulants defined as training data.

3. Forecast the cumulants with the neural network.

4. Compute VaR based on Cornish-Fisher expansion using the forecast of the cumulants.

Cornish-Fisher expansion is represented as

$$R_{\alpha,t} = \mu_t + \sigma_t w_{\alpha,t} \tag{2}$$

$$w_{\alpha,t} = x_\alpha + \frac{\kappa_{3,t}}{\kappa_{2,t}^{1.5}} \cdot \frac{1}{6} H_2(x_\alpha) + \frac{\kappa_{4,t}}{\kappa_{2,t}^2} \cdot \frac{1}{24} H_3(x_\alpha) - \left(\frac{\kappa_{3,t}}{\kappa_{2,t}^{1.5}}\right)^2 \cdot \frac{1}{36} (2H_3(x_\alpha) + H_1(x_\alpha)) + \cdots \tag{3}$$

Where $R_{\alpha,t}$ is the alpha-percentile of VaR. μ_t is the mean of log return. σ_t is the volatility. x_α is the alpha-percentile of the standard normal distribution. $H_n(x)$ is n-th order Hermite polynomials.[6] $w_{\alpha,t}$ is the alpha-percentile of the distribution which volatility is standardized to one. We discuss the implications of $w_{\alpha,t}$ with more details in Section 6. In our simulation analyses and empirical applications below, we set the alpha equal to 97.5 percentile.

It is noteworthy that if skewness and kurtosis are zero, the approximation leads to $w_{\alpha,t} = x_\alpha$ which means that the VaR is the same as the one under the assumption of normal distribution, $R_{\alpha,t} = \mu_t + \sigma_t x_{\alpha,t}$.

In our empirical applications, we set a vector of exogenous variables $Z_t = (r_t, r_t^2)$ to capture the relationship between the contemporary return and future moments. For example, the impact of the contemporary return of volatility known as "leverage effect" in the literature on equity volatility can be captured by including log return to a list of exogenous variables. We also include the squared contemporary return based on an analogy of GARCH (1,1) model.[7]

Figure 1 shows the overview of the CoFiE-NN framework.

We make a few comments on potential extensions of CoFiE-NN framework. The first extension is Expected Shortfall (ES).[8] Our primary focus is predicting VaR but the CoFiE-NN framework can also predict ES without re-estimating parameters. To predict ES, we use the expansion proposed by Giamouridis (2006) which is analogous to Cornish-Fisher expansions. Following Giamouridis (2006), the alpha-percentile ES is given by

$$ES_{\alpha,t} = \mu_t + \sigma_t e_{\alpha,t} \tag{4}$$

$$e_{\alpha,t} = \frac{\phi(w_{\alpha,t})}{\alpha}\left\{1 + \frac{\kappa_3}{\kappa_2^{1.5}} \cdot \frac{1}{6} w_{\alpha,t}^3 + \frac{\kappa_4}{\kappa_2^2} \cdot \frac{1}{24}\left(w_{\alpha,t}^4 - 2w_{\alpha,t}^2 - 1\right)\right\} \tag{5}$$

where $\phi(x)$ is the density function of the standard normal distribution.

The second extension is multivariate case. As discussed in Lamb et al. (2019), there are two approaches. To illustrate these approaches, let us consider a portfolio with N number of risky assets. We denote the portfolio return with r_t^p.

$$r_t^p = \sum_{i=1}^N \omega_i r_{i,t}, \tag{6}$$

where r_i and ω_i are the return and the weight for the i-th asset, respectively. When estimating the VaR of this portfolio, one approach is to compute the cumulants of the portfolio return directly by constructing the time series of the portfolio return. Another approach is to compute the cumulants of the portfolio return by using those of the constituent asset returns $r_{i,t}$. For example, the mean and the variance of the portfolio return are computed by using the following relationships.

$$E[r_t^p] = \sum_{i=1}^N \omega_i E\left[r_{i,t}\right], \tag{7}$$

[6] For example, $H_1(x) = x$, $H_2(x) = x^2 - 1$, and $H_3(x) = x^3 - 3x$.

[7] We confirm that the CoFiE-NN with LSTM outperforms conventional models in general even when $Z_t = r_t$.

[8] Yamai and Yoshiba (2005) compare advantages and disadvantages of VaR and ES from a practical perspective.

$$Var[r_t^p] = \sum_{i=1}^{N} \omega_i^2 Var[r_{i,t}] + \sum_{i \neq j,} \omega_i \omega_j Cov[r_{i,t}, r_{j,t}], \tag{8}$$

We can compute the skewness and the kurtosis in a similar manner. Under the CoFiE-NN framework, we first predict the cumulants of the returns for each asset and cross terms such as covariance, co-skewness, and co-kurtosis. We then compute the cumulants of the portfolio return using the cumulants of the constituent asset returns.

Compared to the first approach, the advantage of the second approach is that there is no need to re-calibrate the parameters of the neural network even when the weights are changed. The disadvantage of the second approach is that the number of the cumulants that need to be predicted increases quickly when the number of the assets increases.

The third extension is forecasting VaR beyond t+1. Let us consider that we forecast VaR at time $t+u$. To do so, we iteratively compute the pre-estimated nonlinear equation below.

$$\mathrm{K}_{t+1+q} = f(\mathrm{K}_{t+q}, \mathrm{K}_{t+q-1}, .. \mathrm{K}_{t+q-s}, Z_t) \text{ for } q = 0,1, \cdots u-1. \tag{9}$$

Once we obtain K_{t+u}, we apply the Cornish-Fisher expansions for calculating VaR at time $t+u$.

A. FFNN

Our main neural network model is LTSM but it is natural to ask whether a simple neural network can beat conventional models in forecasting VaR. To do so, we employ a Feed-Forward Neural Network (FFNN) as another neural network. FFNN consists of several components: an input layer, hidden layers, and output layers. Suppose that the number of hidden layers is L and x_t is a vector of the cumulants at time t.

$$z_{1,t} = f(W_0 x_t + b_0) \tag{10}$$

$$z_{l+1,t} = f(W_l z_{l,t} + b_l) \text{ for } l = 1, \dots, L \tag{11}$$

$$x_{t+1} = f(W_{L+1} z_{L+1,t} + b_{L+1}) \tag{12}$$

where f is the activation function. W_l is the weight matrix and b_l is the bias vector. In our empirical applications, we employ a sigmoid function as the activation function and the number of hidden layers equal to 5 as we find that $L = 5$ provides the best performance on average. The number of units in hidden layers are the same as the dimension of the inputs. We calibrate parameters to minimize the loss function based on backpropagation algorithm. The loss function is defined as the sum of the squared errors.

B. LSTM

We employ LSTM as our main neural network model because it has been successfully applied for time series modeling. LSTM is specified as

$$f_t = \sigma_g(W_f u_t + U_f h_{t-1} + b_f), \tag{13}$$

$$i_t = \sigma_g(W_i u_t + U_i h_{t-1} + b_i), \tag{14}$$

$$o_t = \sigma_g(W_o u_t + U_o h_{t-1} + b_o), \tag{15}$$

$$\tilde{c_t} = \sigma_c(W_c u_t + U_c h_{t-1} + b_c), \tag{16}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c_t}, \tag{17}$$

$$h_t = o_t \odot \sigma_h(c_t), \tag{18}$$

$$y_t = \sigma_h\left(W_y h_j + b_y\right), \tag{19}$$

where u_t is the input vector, f_t is the activation vector of forget gate. i_t is the activation vector of update gate. o_t is the activation vector of the output gate. $\tilde{c_t}$ is the activation vector of the cell input. c_t is the cell state vector. h_t is the hidden state vector. y_t is the output vector For more details, please refer to Hochreiter and Schmidhuber (1997).

C. Calibration

LSTM is implemented using machine learning package in Matlab. Table 1 shows the setting of calibrations of LSTM. As a preliminary analysis, we have tested LSTM under several different settings. We discovered that a small-scale LSTM is sufficient for our empirical applications. We also found that a key hyperparameter is the number of hidden units. If it is set to a large number (e.g., 250), the out-of-sample performance tends to be poor.

D. Ensuring Monotonicity

One well-known problem of Cornish-Fisher expansions is the possibility of non-monotonicity. It does not ensure that the VaR is a monotonously increasing function of the percentile. The previous studies have proposed several solutions for recovering monotonicity.

Maillard (2012) analyzes under which conditions the monotonicity is ensured. Notice that the VaR is a cubic function of the percentile where coefficients are functions of volatility, skewness, and kurtosis if we truncate the Cornish-Fisher expansions up to the fourth order. Maillard (2012) obtains theoretical conditions that ensure the monotonicity of the cubic function. The condition is obtained by deriving the first order derivative of the cubic function and explicitly solving when the first-order derivative is always positive. The condition is represented as two inequalities.

$$|s_t| < \sqrt{2} - 1, \tag{20}$$

$$9k_t^2 - (3 + 33s^2)k + 30s^4 + 7s^2 \leq 0, \tag{21}$$

where $s_t = S_t/6$ and $k_t = K_t/24$ where S_t and K_t are skewness and excess kurtosis, respectively.

We apply these inequalities to skewness and kurtosis before computing VaR using the Cornish-Fisher expansions. If combination of skewness and kurtosis is outside of the region specified by the inequalities, we adjust skewness and kurtosis to be inside of the region by changing a vector of skewness and kurtosis to the closest one within in the region.

Chernozhukov et al. (2016) apply a rearrangement method which is used in the mathematical research areas of functional analysis and optimal transportation. Different from the solution of Maillard (2012), the rearrangement method changes the shape of entire distribution and applicable to Cornish-Fisher expansions with any order of moments.

We employ Maillard (2012)'s solution as our main solution for non-monotonicity issue because of intuitiveness and computational tractableness.

E. EGARCH-t Model

We compare the performance of CoFiE-NN with three conventional models. The first conventional model is EGARCH-t model for volatility σ_t. We employ the EGARCH-t model as a benchmark model for three reasons.[9] First, there is no parameter restriction in the EGARCH model unlike the original GARCH model. Hence, it is easy to estimate the EGARCH model for large number of assets without failure of estimating the model parameters. Second, it can capture both fat-tail nature of conditional distribution which is important for emerging markets' FX as well as volatility clustering. Third, the EGARCH-t model is used as a benchmark model in the previous studies of ML-based VaR forecasting such as Wu and Yan (2019).

The EGARCH-t model is represented as

$$r_t = \sigma_t \cdot \sqrt{\frac{\nu-2}{\nu}}\epsilon_t \tag{22}$$

$$\log\sigma_t^2 - \log\bar{\sigma}^2 = \rho(\log\sigma_{t-1}^2 - \log\bar{\sigma}^2) + \beta(|\epsilon_{t-1}| - E[|\epsilon_{t-1}|]) + \gamma\,\epsilon_{t-1} \tag{23}$$

Where the log return of the exchange rate with r_t. ρ is the auto-coefficient, and $\bar{\sigma}$ is the mean-reverting level of the volatility and ν is volatility of volatility. γ is the parameter to capture the feedback effect. ϵ_t is sampled from Student t distribution with degree of freedom ν. The EGARCH-t model is estimated based on the maximum likelihood estimation.

F. CAViaR

The second conventional model is Conditional Autoregressive Value at Risk (CAViaR) model proposed by Engle and Manganelli (2004). CAViaR model is a conditional autoregressive specification for VaR. CAViaR is represented as

$$VaR_{\alpha,t} - \overline{VaR_\alpha} = \rho\left(VaR_{\alpha,t-1} - \overline{VaR_\alpha}\right) + \beta\left(\alpha - f_{CAViaR}\left(r_{t-1}, VaR_{\alpha,t-1}\right)\right), \tag{24}$$

$$f_{CAViaR}\left(r_{t-1}, VaR_{\alpha,t-1}\right) = \left(1 + \exp\left(G(r_{t-1} - VaR_{\alpha,t-1})\right)\right)^{-1} \tag{25}$$

where $\overline{VaR_\alpha}$ is the mean-reverting level of α-percentile VaR. The autoregressive coefficient is ρ. β is the volatility of VaR. G is the speed of the adjustment to the breach of VaR. If G is very large, then f_{CAViaR} is

[9] There have been academic studies on modeling FX volatility using GARCH-type models. For example, Abdullah et al. (2017) find that GARCH-t model works best for Bangladesh Taka against the US Dollar. Theodossiou (1994) employ EGARCH-M model for modeling Canadian Dollar.

reduced to an indicator function where it takes zero if the return is larger than the VaR and otherwise it is equal to one. The intuition is that if the return exceeds the VaR today, the next day VaR is increased.

Parameters of the CAViaR model is estimated based on minimizing the following objective function.

$$\min_{G,\rho,\beta} \frac{1}{T} \sum_{t=1}^{T} \left[I\left(r_t < VaR_{\alpha,t}\right) - \alpha \right] \left[r_t - VaR_{\alpha,t} \right]. \qquad (26)$$

Note that the CAViaR model is designed to predict the VaR with pre-determined percentile and does not predict volatility or other statistical moments. By contrast, CoFiE-NN is designed to predict the distribution.

G. EVT

The third conventional model is the Extreme Value Theory (EVT). We adopt a generalized Pareto distribution (GDP) to model extremely large shocks which exceed a pre-determined threshold. Following McNeil and Frey (2000), we combine EVT with a GARCH-type model. Specifically, we use it as a part of the distribution of the shock term in the EGARCH model as described as follows.

$$r_t = \sigma_t \cdot \epsilon_t,$$

$$F_{CDF}(\epsilon_t) = \begin{cases} 1 - \frac{N_u}{N}\left(1 - F_{GPD}(\epsilon_t, \gamma_u, \theta_u, \beta_u)\right), & if\ \epsilon_t > \theta_u, \\ \Phi(\epsilon_t), & if\ \theta_d \le \epsilon_t \le \theta_u, \\ \frac{N_d}{N}\left(1 - F_{GPD}(-\epsilon_t, \gamma_d, \theta_d, \beta_d)\right), & if\ \epsilon_t < \theta_d, \end{cases} \qquad (27)$$

where F_{CDF} is the cumulative distribution function of ϵ_t. Φ is the cumulative distribution function of the standard normal distribution. θ_u and θ_d are pre-determined threshold levels. N_u and N_d are the number of samples which are in the corresponding region.

$F_{GPD}(\gamma, \beta)$ is the cumulative distribution function of the generalized Pareto distribution with shape parameter γ and scale parameter β and the shift parameter θ, which is represented as

$$F_{GPD}\ (\epsilon, \gamma, \theta, \beta) = 1 - \left(1 + \gamma \frac{\epsilon - \theta}{\beta}\right)^{-1/\gamma}. \qquad (28)$$

The volatility σ_t is modeled as the EGARCH process. In our empirical application below, we assume that $\theta_u = \Phi(0.05)$ and $\theta_d = \phi(0) = -\infty$ because our main interest is large positive shocks which mean large depreciation of the domestic currency against USD in the case of FX returns.

IV. Data

We apply our machine-learning-based VaR forecasting model for time series of 30 assets across four major asset classes (FX, commodity, interest rates, and equity). Specifically, we obtain historical data of 20 currencies (EUR, JPY, CNY, KRW, GBP, MXN, INR, CAD, BRL, AUD, CHF, THB, MYR, ZAR, TWD, SGD, NOK, SEK, NZD, DKK) and two stock indices (NASDAQ 100 and Nikkei 225) and three commodities (WTI oil price, Brent oil price, Henry Hub natural gas) and two interest rates (two-year and five-year yields of the US Treasury) from the website of the Federal Reserve Economic Data (FRED). Regarding the remaining three currencies, we download the historical data of Hungarian Forint (HUF) and Polish Zloty (PLN) and Czech Koruna (CZK) against Euro from the website of ECB and convert them into the prices against the US dollar.

Regarding our empirical study, we study the performance of CoFiE-NN under four different settings of sample period. For each sample period, we set up training period and test period. Time frequency is daily. Specifically, the four sample period settings are as follows:

(1) Pre-Global-Financial-Crisis period (Pre-GFC) in which training data is from 2004/7/2 to 2005/6/30 and test data is from 2005/7/1 to 2007/6/29.

(2) Global-Financial-Crisis period (GFC) in which training data is from 2003/7/1 to 2007/6/29 and test data is from 2007/7/2 to 2009/12/31.

(3) pre-COVID-19 Crisis period (Pre-COVID) in which training data is from 2015/1/6 to 2015/12/31 and test data is from 2016//1/4 to 2019/12/31.

(4) COVID-19 Crisis period (COVID) in which training data is from 2018/3/1 to 2020/2/28 and test data is from 2020/3/2 to 2023/7/27.

We construct these four settings for two reasons. First, it is important to study the performance of models under both normal period and crisis period, especially given the concern that machine learning approach might overfit to data and thus vulnerable to stressed market environment. Second, there might be structural changes in market dynamics after the 2008–2009 global financial crisis and/or the COVID-19 crisis. Hence, it is useful to analyze the performance of VaR forecasting models both before and after these crises.

Table 2 shows descriptive statistics for 30 assets across different asset classes. We make two observations. First, the three commodities show extremely large kurtosis but several currencies such as Swiss Franc and Korean Won also show large kurtosis. Another observation is skewness. Some currencies such as Swiss Franc shows negative skewness, which means large appreciation with small probability while other currencies such as Mexican Peso show positive skewness, which means large depreciation with small probability. These observations indicate that foreign exchange rate changes cannot be generated from normal distributions. Hence, we need to use distributions which capture the asymmetry and the fat-tail behavior.

A. Testing VaR Forecasting Performance

The statistical test of Kupiec (1995) for VaR forecasting performance is based on the likelihood ratio test and defined as

$$LR_{uncond} = 2\log\left(\left(\frac{1-\hat{\alpha}}{1-\alpha}\right)^{T-I(\alpha)}\left(\frac{\hat{\alpha}}{\alpha}\right)^{I(\alpha)}\right), \tag{29}$$

$$\hat{\alpha} = \frac{1}{T}I(\alpha), \tag{30}$$

$$I(\alpha) = \sum_{t}^{T} I_t(\alpha), \tag{31}$$

where $I_t(\alpha)$ is the indicator function when the breach occurs at time t. The distribution of the likelihood ratio under the null hypothesis is chi squared distribution with 1 degree of freedom. This test is referred to as the Kupiec (1995)'s unconditional coverage test. Following Campbell (2005), if we apply an approximation based on normal distribution, the Kupiec (1995) test statistic is given by

$$z = \frac{\sqrt{T}(\hat{\alpha}-\alpha)}{\sqrt{\alpha(1-\alpha)}}. \tag{32}$$

Second, the statistical test of Christoffersen (1998) for VaR forecasting performance analyzes whether VaR beaches are independent or autocorrelated. Christoffersen (1998)'s independence test is defined as follows:

$$L_0 = (1-\pi_2)^{(n_{00}+n_{10})}\pi_2^{(n_{01}+n_{11})}, \tag{33}$$

$$L_1 = \pi_{00}^{n_{00}}\pi_{01}^{n_{01}}\pi_{10}^{n_{10}}\pi_{11}^{n_{11}}, \tag{34}$$

where $\pi_2 = \frac{(n_{01}+n_{11})}{n_{00}+n_{10}+n_{01}+n_{11}}$ and $\pi_{ij} = \Pr\left(I_t = j | I_{t-1} = i\right)$. The likelihood ratio is given by

$$LR_{ind} = -2\log\frac{L_1}{L_0}. \tag{35}$$

The distribution of the likelihood ratio under the null hypothesis is chi squared distribution with 1 degree of freedom.

The statistic for the joint test is formulated as $LR_{joint} = LR_{uncond} + LR_{ind}$ where the distribution of LR_{joint} under the two null hypotheses is chi squared distribution with 2 degrees of freedom.

Third, Lopez (1999)'s quadratic loss function for VaR forecasting performance is given as

$$QLF = 1_{r_t > VaR_{\alpha,t}}\left\{1 + \left(r_t - VaR_{\alpha,t}\right)^2\right\}. \tag{36}$$

The quadratic loss function is designed to quantify how much forecast of VaR is underestimated. Using both the quadratic loss function and the Kupiec (1995) test, we can study not only how often VaR breaches occur but also the magnitude of VaR breaches.

V. Results

A. Testing CoFiE-NN using Monte Carlo Simulation

We generate artificial time series data of foreign exchange spot rates using the EGARCH-t model with stochastic degree of freedom and asymmetry-controlling parameter as described below.

$$r_t = \sigma_t \cdot \sqrt{\frac{\nu_t - 2}{\nu_t}}\epsilon_{r,t}, \tag{37}$$

$$\log\sigma_t^2 - \log\bar{\sigma}^2 = \rho_\sigma(\log\sigma_{t-1}^2 - \log\bar{\sigma}^2) + \beta_\sigma(|\epsilon_{t-1}| - E[|\epsilon_{t-1}|]) + \gamma_t\,\epsilon_{r,t-1}, \tag{38}$$

$$\gamma_t - \bar{\gamma} = \rho_\gamma(\gamma_{t-1} - \bar{\gamma}) + \beta_\gamma\,\epsilon_{\gamma,t}, \tag{39}$$

$$\nu_t - \bar{\nu} = \rho_\nu(\nu_{t-1} - \bar{\nu}) + \beta_\nu\,\epsilon_{\nu,t}, \tag{40}$$

where the first and second equations are the same as those above in the description of the EGARCH-t model except that the asymmetry-controlling parameter γ_t and the degree of freedom ν_t have time index. The third equation indicates that the asymmetry-controlling parameter follows a mean-reverting stochastic process. The fourth equation indicates that the degree of freedom also follows a mean-reverting stochastic process. $\epsilon_{r\,t}$ is sample from Student's t distribution while $\epsilon_{\gamma,t}$ and $\epsilon_{\nu,t}$ are sampled from the standard normal distribution.

We adopt the EGARCH-t model with these two stochastic features described above because it produces nonlinear dynamics of skewness and kurtosis but still resemble enough the conventional EGARCH-t model which CoFiE-NN competes against.

We set the parameters of this EGARCH-t model with stochastic skewness and kurtosis as follows. $\bar{\sigma} = 0.02$, $\rho_\sigma = 0$, $\beta_\sigma = 0$, $\bar{\gamma} = -0.2$, $\rho_\gamma = 0.99$, $\beta_\gamma = 0.05$, $\bar{\nu} = 10$, $\rho_\nu = 0.99$, $\beta_\nu = 0.2$. We consider that these parameter values are reasonable.

We estimate the CoFiE-NN using four different settings of the sample period length: (1) T=250 days, (2) T=500 days, (3) T=1000 days, (4) T=2000 days. We can study how the size of training data impacts the performance by looking at the results under these different settings.

Table 3 (a) reports the results for all models. CoFiE-NN with LSTM is denoted with LSTM and CoFiE-NN with FFNN is denoted with FFNN. To make the comparison of the performance easy, we compute winning rate of CoFiE-NN against each conventional model under each criterion. Tables 3 (b) and 3 (c) show the winning rates of CoFiE-NN with FFNN and with LSTM, respectively.

Table 3 (b) shows that CoFiE-NN with FFNN outperforms the EGARCH-t in all three criteria under all four settings of the training data size except one when T=500 in terms of Lopez (1999) quadratic loss function. CoFiE-NN with FFNN also outperforms CAViaR in terms of the loss function but underperforms it in terms of Kupiec (1995) test and the joint test. CoFiE-NN with FFNN outperforms the EVT in all criteria.

Table 3 (c) shows that CoFiE-NN with LSTM outperforms the EGARCH-t in terms of Kupiec (1995) test and the joint test when the training data size is relatively small (T=250 and T=500) but underperform it when the data size is relatively large (T=1000 and T=2000). CoFiE-NN with LSTM also outperforms the EGARCH-t in the loss function except one case (T=250). Similar to the case above, CoFiE-NN with LSTM outperforms CAViaR in the loss function but underperforms it in Kupiec (1995) test and the joint test. CoFiE-NN with LSTM outperforms the EVT in all criteria.

These results indicate that CoFiE-NN works successfully even when the training data is small. It is also noteworthy that CoFiE-NN tends to show the better performance than the EGARCH-t even though the data is generated from the extended EGARCH-t model resembling the conventional EGARCH-t model.

B. Comparison of VaR Forecasts for 30 Assets

Figure 2 shows the accumulated breaches of VaR based on CoFiE-NN approach and two conventional models, EGARCH-t model and CAViaR model for 30 assets including 23 currencies during the COVID-19 crisis period. As described above, the training data is from 2018/3/1 to 2020/2/28 and test data is from 2020/3/2 to 2023/7/27. The black-solid line denotes CoFiE-NN. The red-dotted line denotes the EGARCH-t model. The blue-dashed line denotes CAViaR model. The green-dashed line shows the ideal number of breaches which is computed as probability multiplied by the number of days up to time t. Hence, it linearly increases with time. If the black-solid line is closer to the green-dashed line than the red-dotted one, it means that the CoFiE-NN outperforms the EGARCH-t model in forecasting VaR.

Tables 4 (a) (b) (c) (d) report the results of two specifications of CoFiE-NN and the two conventional models across 30 asset classes under the four cases of sample periods.[10] To facilitate the comparison of the performance, we compute winning rate of CoFiE-NN against each conventional model under each criterion. Tables 4 (e) and (f) show the winning rates of CoFiE-NN with FFNN and with LSTM, respectively.

For CoFiE-NN with FFNN, we find that it outperforms the EVT in all criteria for all four sample periods except one case (Pre-GFC). However, CoFiE-NN with FFNN underperform the EGARCH-t and the EVT in general. The result suggests that FFNN may not be powerful enough to beat these conventional models.

For CoFiE-NN with LSTM, we make three observations. First, it outperforms the EGARCH-t in Kupiec (1995) test and the joint test under the all four different sample periods except one case (Covid) while it underperforms the EGARCH-t in terms of the loss function. Second, CoFiE-NN underperforms CAViaR in general. Third, CoFiE-NN outperforms the EVT in all criteria for all four different sample periods except one case (Pre-GFC).

In summary, CoFiE-NN with LSTM tends to outperform the EGARCH-t and the EVT but tends to underperform CAViaR.

VI. Empirical Application: Tail Risk Ratio

A. Defining Tail Risk Ratio

We introduce a new concept named *tail risk ratio* which measures the severity of tail event after controlling for volatility. Specifically, we define the α-percentile tail risk ratio $TRR_{\alpha,i,t}$ as volatility-scaled α-percentile VaR.

$$TRR_{\alpha,i,t} = \frac{\mathrm{VaR}_{\alpha,i,t} - \mu_{i,t}}{\sigma_{i,t}}, \tag{41}$$

where μ_{it} and $\sigma_{i,t}$ are the mean and the volatility of the i-th asset at time t. $\mathrm{VaR}_{\alpha,i,t}$ is the alpha-percentile of VaR for the i-th asset.

We make a few comments on the definition of the tail risk ratio. First, the tail risk ratio is aimed to capture the size of unexpectedly large shock to an asset relative to its volatility. We consider that scaling by volatility is necessary because the increase in VaR can be simply due to the increase in volatility, which is misleading to interpret as increase in tail risk. In other words, level of VaR may differ across assets due to differences in volatility but scaling by volatility allows us to quantify the pure tail risk. One practical advantage is that we can make cross-country comparison of tail risk in the foreign exchange markets by accounting for difference in the level of VaR.

Second, the definition of the tail risk ratio analogous to the definition of the Sharpe ratio that is volatility-scaled excess return. Sharpe ratio is understood as the risk-adjusted return. In this analogy, the tail risk ratio is interpreted as the severity of tail event after the adjustment of volatility.

[10] Results for EVT are not report in these tables but we confirm that the EVT underperform CoFiE-NN as shown in Table 4 (e) (f). Also, note that "-" denotes that the statistic is not obtained as a finite number.

Under CoFiE-NN framework, the tail risk ratio is the percentile of the volatility-standardized distribution which is introduced in the description of CoFiE-NN.

$$TRR_{i,t} = \frac{VaR_{i,t}-\mu_{it}}{\sigma_{i,t}} = w_{\alpha,t} = x_\alpha + \frac{\kappa_{3,t}}{\kappa_{2,t}^{1.5}} \cdot \frac{1}{6} H_2(x) + \frac{\kappa_{4,t}}{\kappa_{2,t}^{2}} \cdot \frac{1}{24} H_3(x) - \left(\frac{\kappa_{3,t}}{\kappa_{2,t}^{1.5}}\right)^2 \cdot \frac{1}{36}(2H_3(x) + H_1(x)) + \cdots . \quad (42)$$

It is easy to make two observations. First, if both skewness and kurtosis are zero, the tail risk ratio is reduced to the percentile of the cumulative density function of the normal distribution x_α. Second, if skewness and kurtosis are time-varying, the tail risk ratio is also time-varying even if volatility is constant.

It is noteworthy that the tail risk ratio is constant under conventional GARCH-type models. To see this, let us compute the tail risk ratio under a specific GARCH-type model. For example, under the EGARCH-t model, the tail risk ratio is given by

$$TRR_{\alpha,i,t} = \frac{VaR_{\alpha,i,t}-\mu_{it}}{\sigma_{i,t}} = \sqrt{(\nu - 2\,)/\nu}\, F_\nu^{-1}(\alpha). \quad (43)$$

where $F_\nu^{-1}(\cdot)$ is the inverse function of the cumulative distribution function of Student's t distribution with the degree of freedom ν. The formula above shows that the tail risk ratio is constant under the EGARCH-t model.

It is also worth mentioning that tail risk ratio cannot be defined under the CAViaR model because it does not explicitly model the volatility dynamics. It is practically possible to use the volatility extracted from another model (e.g., a specific GARCH-type model) but it creates internal inconsistency between the VaR forecasting model versus the volatility forecasting model.

B. Analyzing Tail Risk Dynamics in Foreign Exchange Markets

Surprisingly, to the best of our knowledge, no academic study explicitly measures the stress in foreign exchange markets. This is contrasting to the large number of academic and policy markers' studies on measuring the stress in stock markets (e.g., Kelly and Jiang (2014)). Therefore, it is natural to ask what the tail risk ratio can tell us when and how the foreign exchange markets are stressed.

To do so, we apply principal component analysis (PCA) for the tail risk ratios of 22 currencies during the 2019/8-2023/7 period.[11] Specifically, we compute the VaR for these currencies based on CoFiE-NN. The training period is the 2014/8-2019/7 period. We then apply PCA for the VaR for 22 currencies for extracting common factors. For comparison, we do the same exercise for the one-month historical volatilities of the same set of currencies using the same sample period.

Table 5 shows the proportion of variance of tail risk ratios and volatilities explained by each PCA factor. It shows that the only 20 percent of tail risk dynamics across 22 currencies is explained by the first PCA factor which is interpreted as a global factor. By contrast, 60 percent of volatility dynamics across the same set of currencies is explained by the first PCA factor. These two results indicate that tail risks in the foreign exchange markets are more country-specific compared to the volatilities.

Figures 3 (a) and (b) show coefficients of the first PCA factor extracted from tail risk ratios and volatilities, respectively. The first PCA factor of tail risk ratios of 22 currencies is interpreted as a global factor as almost all

[11] We drop Danish Krone in this analysis because Danish Krone (DKK) is pegged to Euro.

coefficients are positive and in the same magnitude except a few currencies. Similarly, the first PCA factor of volatilities of 22 currencies also is considered as a global volatility factor as all coefficients are positive.

Figures 3 (c) and (d) show coefficients of the second PCA factor extracted from tail risk ratios and volatilities, respectively. The result is contrasting to the first PCA factor case. The second PCA factor of tail risk ratios is understood as a Chinese-Yuan-specific factor as the coefficient is largest for Chinese Yuan and small for other countries. By contrast, the second PCA factor of volatilities captures the factor related to Latin American countries as the coefficients are large for Brazilian Real and Mexican Peso.

Figures 3 (e) and (f) show coefficients of the third PCA factor extracted from tail risk ratios and volatilities, respectively. For both tail risk ratio and volatility, the third PCA factor is difficult to interpret. Yet it appears that the third PCA factor of tail risk ratios is related to Eastern European countries as the coefficients are high for Czech Koruna, Polish Zloty, and Hungarian Forint. Similarly, the third PCA factor of volatilities is related to Eastern European countries, too. However, it can also be interpreted as Asia-specific factor as the coefficients are negative and high for Japanese Yen, Chinese Yuan, South Korean Won, Indian Rupee, and Malaysian Ringgit.

Figures 4 (a) and (b) show time evolution of the first PCA factor extracted from tail risk ratios and volatilities, respectively. As mentioned above, we consider that the first PCA factor is the global factor. There is a stark difference between them. The global volatility factor increased dramatically at the onset of the COVID-19 crisis. It also increased gradually after Russian invasion to Ukraine. By contrast, the global tail risk factor shows spikes more frequently than the global volatility factor. The global tail risk factor jumped to the highest level in June 2021 after when the Federal Reserve published updated dot plot indicating more hawkish policy stance.[12] These results show that the global tail risk factor captures the market stress which the global volatility factor does not necessarily capture.

C. Portfolio Optimization under VaR constraint and Tail Risk Ratio

Tail risk ratio naturally arises in the context of portfolio optimization under a VaR constraint. To see this, consider that there are N assets. We denote a vector of the stochastic returns and the expected returns with $r_t = \left(r_{t,1}, r_{t,2}, \cdots, r_{t,N}\right)'$ and $r_t^e = E[r_t]$, respectively. The asset allocation is denoted with $\omega_t = \left(\omega_{t,1}, \omega_{t,2}, \cdots, \omega_{t,N}\right)'$. Suppose that an investor maximizes the expected return of her portfolio under the VaR constraint.[13]

$$\max_{\omega_t} \omega_t' r_t^e, \tag{44}$$

$$\mathrm{VaR}_\alpha \left(-\omega_t' \tilde{r}_t\right) \leq \bar{L} \tag{45}$$

[12] See the article from Bloomberg: https://www.bloomberg.com/opinion/articles/2021-06-17/powell-surprise-marks-start-of-fed-s-wtreacherous-retreat#xj4y7vzkg.

[13] Portfolio optimization problems under a VaR constraint have been studied in finance. The previous studies include, but not limited to, Emmer et al. (2001), Basak and Shapiro (2001), and Yiu (2004). Related to these studies, Adrian and Shin (2014) provide microfoundations to explain why VaR-based rules are widely used in financial risk management using a contracting model. Miranda-Agrippino and Rey (2020) describe a global bank as the investor who optimizes the expected return of his portfolio subject to a VaR constraint essentially under the assumption of the multivariate normal distribution.

where $\widetilde{r_t}$ is defined as the unexpected component of the asset returns ($\widetilde{r_t} = r_t - r_t^e$). $\bar{L}$ is the maximum VaR budget. The second inequality means that the investor limits the unexpected loss below the pre-determined threshold L.

When the asset returns are generated from a multivariate normal distribution, the VaR constraint is transformed to the constraint of the portfolio return volatility. Let us denote the covariance matrix of the asset return with Σ. We obtain

$$x_\alpha \cdot \omega_t' \Sigma \omega_t \leq \bar{L}. \tag{46}$$

The solution of the asset allocation under the assumption of the multivariate normal distribution is given by

$$\omega_t^* = \frac{1}{2\lambda x_\alpha} \Sigma^{-1} r_t^e, \tag{47}$$

where λ is the Lagrange multiplier. The solution above indicates that mean-VaR portfolio optimization problem is reduced to the classic mean-variance portfolio optimization problem when the asset returns are generated from a multivariate normal distribution.

In general, this does not hold. Let us denote the tail risk ratio of the portfolio return with $w_{p,\alpha,t}$. The VaR constraint is represented as

$$w_{p,\alpha,t} \cdot \omega_t' \Sigma \omega_t \leq \bar{L}. \tag{48}$$

The key difference between (46) and (48) is that x_α is a constant given α while $w_{p,\alpha,t}$ is a non-linear function of the asset allocation ω_t. Specifically, the tail risk ratio for the portfolio return depends on two factors: (1) skewness and kurtosis and higher moments of each asset return and (2) the asset allocation ω_t. The VaR-constrained investor may reduce her exposure to the asset returns when the moments increase even when the volatility of the portfolio return remains the same.

VII. Conclusion

In this paper, we proposed a new machine-learning-based approach for forecasting Value-at-Risk named CoFiE-NN where a neural network (NN) is combined with Cornish-Fisher expansions (CoFiE). The new approach has two advantages. It can capture non-linear dynamics of high-order statistical moments thanks to the universal approximation property of a NN while maintaining interpretability of the outputs by explicitly linking moments with the percentile of distribution via Cornish-Fisher expansions.

We employ Long Short-Term Memory (LSTM) as our main specification of a NN because LSTM has been successful in time series modeling. Note that any NN can be employed in the CoFiE-NN.

We compared the performance of VaR forecasting based on CoFiE-NN with three conventional models using both Monte Carlo simulation and real data. We show that CoFiE-NN with LSTM tends to outperform the EGARCH-t model even when the sample size of training data is small using the simulated data. We then apply the CoFiE-NN for 30 assets across different asset classes. We find that CoFiE-NN tends to outperform the EGARCH-t in terms of Kupiec (1995) test but not in Lopez (1999) quadratic loss function. CoFiE-NN underperforms CAViaR, but it outperforms the EVT model in general for the real data.

Finally, we introduce a new empirical proxy for tail risk named *tail risk ratio* under CoFiE-NN. We discover that the only 20 percent of tail risk dynamics across 22 currencies is explained by one common factor. This contrasts with the fact that 60 percent of volatility dynamics across the same set of currencies is explained by one common factor.

There are several topics for future research. First, it would be useful to extend CoFiE-NN framework to multivariate case and multiple period forecasting, as briefly discussed in Section 3. In particular, the multivariate CoFiE-NN will allow us to generate VaR of multiple assets so that we can investigate the economic benefit of CoFiE-NN for constructing portfolio allocations using real data. Second, it would be interesting to see whether including the fifth or even higher order moments would improve VaR forecasting under CoFiE-NN framework. Third, although we study 30 assets across different asset classes, more comprehensive performance analyses of CoFiE-NN would be needed.

Annex I. Tables

Table 1. Setting of Calibration and Hyperparameters of LSTM

Variable	Value
Dimension	6
Number of units	2
Calibration method	Adams
Maximum epochs	250
Gradient thresholds	1
Learn rate drop factor	0.2

Table 2. Descriptive Statistics

	Mean	Stdev	Skewness	Kurtosis
EUR	0.00%	0.6%	0.1	2.5
JPY	0.01%	0.6%	-0.4	4.4
CNY	0.00%	0.2%	0.0	15.8
KRW	0.00%	0.7%	-0.6	46.9
GBP	0.00%	0.6%	-0.6	9.5
MXN	0.01%	0.7%	0.7	10.4
INR	0.00%	0.6%	-0.6	9.5
CAD	0.00%	0.6%	-0.1	5.3
BRL	0.02%	1.0%	0.0	7.7
AUD	0.00%	0.8%	-0.6	10.8
CHF	-0.01%	0.7%	-1.2	34.1
THB	0.00%	0.4%	0.1	9.2
MYR	0.00%	0.4%	-0.4	8.5
ZAR	0.02%	1.1%	0.2	4.2
TWD	0.00%	0.3%	-0.3	14.5
SGD	0.00%	0.3%	0.0	4.7
NOK	0.00%	0.8%	0.2	4.1
SEK	0.00%	0.7%	-0.1	3.6
NZD	0.00%	0.8%	-0.4	4.5
DKK	0.00%	0.6%	-0.2	4.5
CZK	-0.01%	0.7%	0.3	3.8
HUF	0.01%	0.9%	0.3	4.0
PLN	0.00%	0.8%	0.5	5.9
NASDAQ	0.03%	1.8%	0.0	6.6
US 2Y	0.00%	4.8%	0.0	6.5
US 10Y	-0.01%	2.5%	0.0	25.2
WTI	0.02%	5.6%	-14.2	1718.1
Brent	0.02%	2.7%	-2.0	75.0
Nikkei	0.01%	1.5%	-0.4	6.2
Henry Hub	0.00%	5.4%	-0.2	48.8

Table 3 (a). Performance of VaR Forecasting using Simulated Data

	Kupiec (1995) test					Joint test					Lopez (1999) loss function				
Size	LSTM	FFNN	EGARCH-t	CAViaR	EVT	LSTM	FFNN	EGARCH-t	CAViaR	EVT	LSTM	FFNN	EGARCH-t	CAViaR	EVT
T=250	6.6	4.7	10.4	4.8	92.7	8.0	5.7	9.8	6.0	103.5	26.6	23.4	22.4	25.5	65.7
T=500	7.6	3.2	9.9	2.8	108.9	8.8	4.4	8.8	4.0	118.9	16.8	22.2	20.7	25.1	72.9
T=1000	12.4	2.2	6.5	1.7	124.0	13.1	3.4	7.7	2.9	137.6	11.5	22.2	21.8	24.6	81.4
T=2000	18.5	2.1	3.9	1.5	114.0	18.9	3.3	5.1	2.9	123.7	7.7	21.6	22.6	25.5	76.4

Table 3 (b). VaR Forecasting using Simulated Data: Winning rates of CoFiE-NN with FFNN

CoFiE-NN (FFNN)	Winning rate (Kupiec test)			Winning rate (Joint test)			Winning rate (Lopez function)		
Size	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT
T=250	66%	49%	89%	67%	50%	90%	50%	51%	61%
T=500	70%	39%	92%	66%	38%	94%	46%	64%	73%
T=1000	60%	41%	96%	63%	39%	96%	56%	65%	72%
T=2000	55%	35%	92%	55%	39%	92%	56%	74%	69%

Table 3 (c). VaR Forecasting using Simulated Data: Winning rates of CoFiE-NN with LSTM

CoFiE-NN (LSTM)	Winning rate (Kupiec test)			Winning rate (Joint test)			Winning rate (Lopez function)		
Size	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT
T=250	63%	49%	88%	62%	48%	91%	46%	53%	60%
T=500	55%	39%	91%	54%	35%	89%	65%	80%	75%
T=1000	22%	41%	89%	27%	11%	88%	80%	94%	75%
T=2000	6%	35%	81%	7%	1%	81%	95%	99%	77%

Table 4 (a). Performance of VaR Forecasting: Real Data (Pre-Global Financial Crisis)

Test	Kupiec (1995) test				Joint test				Lopez (1999) loss function			
Asset	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR
EUR	0.3	4.0	7.7	9.1	0.9	0.1	0.1	11.4	14.0	6.0	4.0	24.0
JPY	4.0	10.2	24.8	0.4	4.1	0.0	-	0.9	6.0	3.0	0.0	10.0
CNY	24.8	24.8	24.8	13.4	-	-	-	13.4	0.0	0.0	0.0	2.0
KRW	0.3	1.0	17.7	1.0	1.0	0.3	0.0	1.3	14.0	9.0	1.0	9.0
GBP	1.7	0.1	1.7	15.4	2.0	1.4	0.3	20.9	8.0	11.0	8.0	28.0
MXN	0.1	9.1	0.6	5.3	1.5	0.5	1.0	6.4	11.0	24.0	15.0	21.0
INR	0.1	0.1	1.7	15.4	0.6	1.4	0.3	20.9	11.0	11.0	8.0	28.0
CAD	4.0	10.2	24.8	0.0	4.1	0.0	-	0.5	6.0	3.0	0.0	12.0
BRL	0.0	4.3	2.7	3.3	0.9	7.4	0.2	11.5	13.0	20.0	7.0	19.0
AUD	4.0	24.8	24.8	0.0	7.7	-	-	1.3	6.0	0.0	0.0	12.0
CHF	10.2	24.8	24.8	5.6	10.2	-	-	5.7	3.0	0.0	0.0	5.0
THB	19.0	24.9	10.2	0.0	23.4	8.1	0.0	0.6	30.0	33.0	3.0	12.0
MYR	24.8	24.8	24.8	0.3	-	-	-	1.1	0.0	0.0	0.0	14.0
ZAR	10.2	10.2	13.4	1.7	17.0	6.8	0.0	4.7	3.0	3.0	2.0	8.0
TWD	13.4	7.7	5.3	2.7	13.4	0.0	3.7	6.1	2.0	4.0	21.0	7.0
SGD	1.7	10.2	10.2	10.2	2.0	0.0	0.0	10.2	8.0	3.0	3.0	3.0
NOK	13.4	13.4	7.7	1.0	13.4	0.0	0.1	1.3	2.0	2.0	4.0	9.0
SEK	17.7	7.7	24.8	4.0	17.7	0.0	-	4.1	1.0	4.0	0.0	6.0
NZD	1.7	7.7	7.7	3.3	1.9	0.1	0.1	4.8	8.0	4.0	4.0	19.0
DKK	0.6	17.7	24.8	2.7	1.6	0.0	-	2.9	15.0	1.0	0.0	7.0
CZK	1.0	7.7	7.7	1.7	1.3	0.0	0.0	1.9	9.0	4.0	4.0	8.0
HUF	5.6	0.0	2.7	0.3	5.7	0.7	0.2	1.0	5.0	13.0	7.0	14.0
PLN	1.0	5.6	0.0	5.6	1.3	0.1	1.3	5.7	9.0	5.0	12.0	5.0
NASDAQ	0.1	4.0	0.1	1.1	0.6	3.7	0.5	1.5	11.0	6.0	11.0	16.0
USTTwo	13.4	24.8	24.8	13.4	13.4	-	-	13.4	2.0	0.0	0.0	2.0
USTTen	1.7	7.7	24.8	0.1	4.7	0.0	-	0.5	8.0	4.0	0.0	11.0
WTI	0.3	5.6	4.0	0.6	0.9	0.1	0.1	1.4	14.0	5.0	6.0	15.0
BRENT	4.0	10.2	13.4	0.0	4.1	0.0	0.0	0.6	6.0	3.0	2.0	12.0
Nikkei	1.6	4.2	93.0	0.0	1.9	1.7	1.7	0.6	17.0	20.0	58.0	12.0
Henry Hub	0.0	1.7	13.4	2.4	1.3	0.3	0.0	2.6	12.0	17.1	2.0	18.1

Table 4 (b). Performance of VaR Forecasting: Real Data (Global Financial Crisis)

Test	Kupiec (1995) test				Joint test				Lopez (1999) loss function			
Asset	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR
EUR	0.2	12.5	0.9	0.8	1.1	3.3	0.5	0.5	17.0	31.0	12.0	19.0
JPY	2.5	20.5	0.9	0.4	10.2	0.4	0.5	0.5	22.0	36.0	12.0	18.0
CNY	31.3	31.3	31.3	2.5	-	-	-	-	0.0	0.0	0.0	22.0
KRW	15.5	170.2	4.2	0.4	20.1	5.4	6.4	6.4	33.0	88.0	24.0	18.0
GBP	3.3	28.0	2.2	0.9	7.0	2.1	0.3	0.3	23.0	40.0	10.0	12.0
MXN	20.5	74.2	0.2	0.4	32.3	7.0	3.0	3.0	36.0	59.0	17.0	18.0
INR	0.2	28.0	2.2	0.9	1.1	2.1	0.3	0.3	17.0	40.0	10.0	12.0
CAD	9.7	94.9	3.3	0.8	10.0	1.5	3.7	3.7	29.0	66.0	23.0	19.0
BRL	0.8	38.6	23.8	0.4	3.1	6.0	0.0	0.0	19.0	45.0	1.0	18.0
AUD	6.2	71.4	1.5	0.0	8.7	0.0	0.4	0.4	26.0	58.0	11.0	16.0
CHF	0.0	11.1	0.4	0.0	0.8	0.2	0.6	0.6	15.0	30.0	13.0	15.0
THB	2.2	6.2	31.3	0.4	4.4	12.7	-	-	10.0	26.0	0.0	18.0
MYR	31.3	31.3	31.3	4.2	-	-	-	-	0.0	0.0	0.0	24.0
ZAR	9.8	6.2	2.2	0.9	14.7	0.7	2.1	2.1	5.0	26.0	10.0	12.0
TWD	0.9	5.1	2.2	2.2	1.3	2.9	0.3	0.3	12.0	25.0	10.0	10.0
SGD	0.9	26.1	0.9	0.0	2.3	6.7	5.6	5.6	12.0	39.0	12.0	16.0
NOK	3.3	71.4	4.2	4.2	3.3	1.3	0.0	0.0	23.0	58.0	24.0	24.0
SEK	7.3	88.8	1.3	0.4	7.8	6.3	2.0	2.0	27.0	64.0	20.0	18.0
NZD	2.5	58.0	0.0	0.8	4.2	1.4	0.9	0.9	22.0	53.0	16.0	19.0
DKK	0.4	28.0	0.8	0.0	1.0	2.1	2.3	2.3	13.0	40.0	19.0	15.0
CZK	7.3	77.1	8.5	4.2	12.0	4.7	0.4	0.4	27.0	60.0	28.0	24.0
HUF	4.2	104.3	12.5	5.1	5.2	0.3	3.3	3.3	24.0	69.0	31.0	25.0
PLN	8.5	85.9	0.2	2.5	8.5	2.2	1.0	1.0	28.0	63.0	17.0	22.0
NASDAQ	0.8	71.4	0.9	0.4	1.0	0.0	0.5	0.5	19.0	58.0	12.0	18.0
USTTwo	8.5	174.0	3.2	0.0	8.5	1.7	0.3	0.3	28.1	89.3	9.1	16.1
USTTen	0.0	151.9	2.2	0.0	3.5	0.0	2.1	2.1	16.0	83.0	10.0	15.0
WTI	1.8	15.5	38.6	0.8	6.4	0.8	6.0	6.0	21.0	33.1	45.1	19.0
BRENT	0.4	12.5	12.5	4.2	1.0	1.2	1.2	1.2	13.0	31.0	31.0	24.0
Nikkei	0.5	72.8	3.1	0.0	6.7	1.2	0.3	0.3	18.0	58.0	9.0	16.0
Henry Hub	1.7	0.0	8.1	0.5	2.0	14.0	0.1	0.1	11.0	15.1	6.0	13.1

Table 4 (c). Performance of VaR Forecasting: Real Data (Pre COVID-19 Crisis)

Test	Kupiec (1995) test				Joint test				Lopez (1999) loss function			
Asset	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR
EUR	0.5	41.0	41.0	2.8	3.6	0.0	0.0	0.0	21.0	1.0	1.0	33.0
JPY	1.3	0.3	9.5	0.5	2.1	2.8	8.1	8.1	19.0	22.0	11.0	21.0
CNY	84.5	36.2	8.0	16.9	87.5	0.7	2.2	2.2	81.0	59.0	12.0	47.0
KRW	15.3	13.1	5.4	0.9	15.4	0.1	2.0	2.0	8.0	9.0	14.0	20.0
GBP	1.9	2.8	0.8	1.3	6.1	2.4	3.5	3.5	18.0	33.0	29.0	19.0
MXN	0.3	7.6	1.3	0.5	4.5	8.4	0.8	0.8	27.0	39.0	19.0	28.0
INR	4.1	2.8	0.8	1.3	11.4	2.4	3.5	3.5	35.0	33.0	29.0	19.0
CAD	13.1	23.4	6.6	0.3	16.5	0.1	0.4	0.4	9.0	5.0	13.0	22.0
BRL	26.8	30.7	9.5	0.9	26.8	0.0	0.2	0.2	4.0	3.0	11.0	20.0
AUD	3.4	35.4	35.4	3.4	3.9	0.0	0.0	0.0	16.0	2.0	2.0	16.0
CHF	26.8	41.0	41.0	Inf	26.8	0.0	0.0	0.0	4.0	1.0	1.0	448.0
THB	0.9	6.6	49.5	4.3	1.5	0.4	-	-	20.0	13.0	0.0	15.0
MYR	23.4	41.0	49.5	0.3	30.2	0.0	-	-	5.0	1.0	0.0	22.0
ZAR	0.9	0.8	3.4	0.8	1.5	3.5	1.3	1.3	20.0	29.0	16.0	29.0
TWD	20.4	30.7	49.5	5.4	20.4	0.0	-	-	6.0	3.0	0.0	14.0
SGD	6.6	30.7	26.8	0.3	8.6	0.0	0.0	0.0	13.0	3.0	4.0	22.0
NOK	4.3	35.4	30.7	0.5	5.8	0.0	0.0	0.0	15.0	2.0	3.0	21.0
SEK	52.0	15.3	8.0	0.0	53.3	0.1	2.2	2.2	67.0	8.0	12.0	24.0
NZD	30.7	20.4	35.4	6.6	30.8	0.1	0.0	0.0	3.0	6.0	2.0	38.0
DKK	11.2	35.4	30.7	1.3	11.4	0.0	0.0	0.0	10.0	2.0	3.0	19.0
CZK	20.4	30.7	23.4	9.5	20.4	0.0	0.1	0.1	6.0	3.0	5.0	11.0
HUF	4.3	20.4	11.2	5.4	5.8	5.0	2.9	2.9	15.0	6.0	10.0	14.0
PLN	23.4	15.3	5.4	2.6	23.4	4.1	1.8	1.8	5.0	8.0	14.0	17.0
NASDAQ	2.8	4.3	5.4	0.9	3.4	0.5	0.4	0.4	33.0	15.0	14.0	20.0
USTTwo	26.8	41.0	35.4	30.7	26.8	0.0	0.0	0.0	4.0	1.0	2.0	3.0
USTTen	41.0	15.3	0.5	0.9	41.0	0.1	0.5	0.5	1.0	8.0	21.0	20.0
WTI	6.6	9.5	0.9	0.3	6.9	2.6	0.6	0.6	13.0	11.0	20.0	27.0
BRENT	6.6	4.3	4.3	1.9	8.6	5.5	1.5	1.5	13.0	15.0	15.0	18.0
Nikkei	0.0	5.4	0.0	18.3	1.2	1.7	0.2	0.2	24.0	14.0	25.0	48.0
Henry Hub	26.7	4.8	28.5	5.1	28.1	10.2	6.9	6.9	55.2	37.4	4.3	15.3

Table 4 (d). Performance of VaR Forecasting: Real Data (COVID-19 Crisis)

Test	Kupiec (1995) test				Joint test				Lopez (1999) loss function			
Asset	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR	LSTM	FFNN	EGARCH-t	CAViaR
EUR	1.3	57.9	0.0	0.3	1.4	2.4	0.9	0.9	26.0	63.0	20.0	23.0
JPY	36.4	85.5	4.4	6.4	39.8	10.0	0.4	0.4	22.0	74.0	12.0	33.0
CNY	2.5	3.8	41.9	0.0	3.0	0.7	-	-	0.0	30.0	0.0	20.0
KRW	3.4	10.7	7.0	0.9	3.8	0.3	0.2	0.2	33.0	37.0	10.0	25.0
GBP	1.8	7.4	0.0	0.3	6.7	6.4	0.4	0.4	23.0	34.0	20.0	23.0
MXN	0.5	18.9	24.2	0.1	2.3	10.6	0.0	0.0	36.0	43.0	3.0	19.0
INR	1.3	7.4	0.0	0.3	12.5	6.4	0.4	0.4	17.0	34.0	20.0	23.0
CAD	0.1	23.7	33.8	0.1	0.3	0.8	0.0	0.0	29.0	46.0	1.0	22.0
BRL	0.5	22.0	0.0	0.1	0.6	0.1	0.3	0.3	19.0	45.0	21.0	22.0
AUD	5.4	91.0	99.3	0.9	7.9	2.6	1.8	1.8	26.0	76.0	79.0	25.0
CHF	2.4	34.4	0.9	0.0	2.4	0.0	0.1	0.1	15.0	52.0	25.0	21.0
THB	0.1	36.4	8.6	0.4	0.7	5.4	0.2	0.2	10.0	53.0	9.0	18.0
MYR	20.4	48.8	8.6	2.4	23.1	3.2	0.2	0.2	0.0	59.0	9.0	28.0
ZAR	17.5	0.1	0.5	1.2	17.6	2.3	0.1	0.1	5.0	22.0	24.0	16.0
TWD	2.5	5.4	41.9	0.0	3.0	0.1	-	-	12.0	32.0	0.0	20.0
SGD	1.8	30.7	0.0	0.1	5.3	4.5	1.0	1.0	12.0	50.0	20.0	22.0
NOK	9.5	116.8	60.2	1.3	12.5	4.8	1.9	1.9	23.0	85.0	64.0	26.0
SEK	0.9	40.4	0.1	0.5	2.4	0.5	1.2	1.2	27.0	55.0	22.0	24.0
NZD	1.3	72.5	1.8	0.1	1.3	0.0	0.6	0.6	22.0	69.0	15.0	19.0
DKK	0.0	34.4	0.3	3.0	2.9	5.8	0.2	0.2	13.0	52.0	23.0	29.0
CZK	0.0	46.7	2.4	1.3	2.6	5.4	0.0	0.0	27.0	58.0	28.0	26.0
HUF	40.4	77.6	20.4	2.4	43.4	7.7	7.1	7.1	24.0	71.0	44.0	28.0
PLN	20.4	44.5	1.8	1.3	30.4	8.0	1.1	1.1	28.0	57.0	27.0	26.0
NASDAQ	2.4	48.8	6.4	0.3	2.4	0.2	0.1	0.1	19.0	59.0	33.0	23.0
USTTwo	178.1	437.6	105.0	3.0	179.1	0.1	4.8	4.8	28.1	169.2	81.3	29.2
USTTen	1.3	141.5	20.4	0.4	8.5	2.3	0.2	0.2	16.0	93.3	44.1	18.2
WTI	6.4	27.1	0.4	24.2	13.2	17.9	18.9	18.9	21.0	53.0	18.0	6.8
BRENT	0.4	28.8	1.8	10.4	19.3	16.2	15.0	15.0	13.0	49.3	27.1	8.2
Nikkei	3.6	9.3	15.7	0.0	6.2	0.2	1.7	1.7	18.0	36.0	41.0	20.0
Henry Hub	5.0	87.1	43.3	1.5	7.0	1.7	-	-	11.0	77.5	0.0	16.9

Table 4 (e). VaR Forecasting using Real Data: Winning rates of CoFiE-NN with FFNN

CoFiE-NN (FFNN)	Winning rate (Kupiec test)			Winning rate (Joint test)			Winning rate (Lopez function)		
Sample period	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT
Pre-GFC	47%	17%	63%	47%	17%	63%	20%	77%	40%
GFC	10%	3%	73%	7%	3%	77%	3%	7%	70%
Pre-Covid	27%	13%	63%	27%	20%	67%	47%	77%	80%
Covid	23%	3%	70%	20%	0%	77%	10%	0%	63%

Table 4 (f). VaR Forecasting using Real Data: Winning rates of CoFiE-NN with LSTM

CoFiE-NN (LSTM)	Winning rate (Kupiec test)			Winning rate (Joint test)			Winning rate (Lopez function)		
Sample period	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT	EGARCH-t	CAViaR	EVT
Pre-GFC	70%	17%	80%	67%	47%	80%	20%	77%	43%
GFC	53%	7%	93%	50%	33%	93%	23%	37%	80%
Pre-Covid	63%	17%	67%	57%	27%	70%	30%	70%	77%
Covid	53%	3%	93%	47%	27%	97%	33%	27%	73%

Table 5. Proportion of Variance Explained by Each PCA Factor

	Tail risk ratio		Volatility	
Factor	Variance explained	Accumulation	Variance explained	Accumulation
1	19.9%	19.9%	60.7%	60.7%
2	9.7%	29.6%	13.6%	74.3%
3	7.8%	37.4%	6.3%	80.6%
4	6.6%	44.0%	4.9%	85.5%
5	6.4%	50.4%	3.7%	89.1%
6	5.3%	55.6%	2.0%	91.1%
7	5.0%	60.7%	1.9%	93.0%
8	4.5%	65.2%	1.6%	94.6%
9	4.3%	69.5%	0.9%	95.5%
10	3.9%	73.3%	0.8%	96.4%
11	3.4%	76.7%	0.7%	97.1%
12	3.2%	80.0%	0.6%	97.6%
13	3.0%	83.0%	0.5%	98.1%
14	2.6%	85.6%	0.4%	98.5%
15	2.4%	88.0%	0.4%	98.9%
16	2.3%	90.3%	0.3%	99.2%
17	2.2%	92.5%	0.3%	99.5%
18	2.1%	94.5%	0.2%	99.7%
19	1.9%	96.5%	0.2%	99.8%
20	1.7%	98.1%	0.1%	99.9%
21	1.5%	99.6%	0.1%	100.0%
22	0.4%	100.0%	0.0%	100.0%

Annex II. Figures

Figure 1. Overview of CoFiE-NN

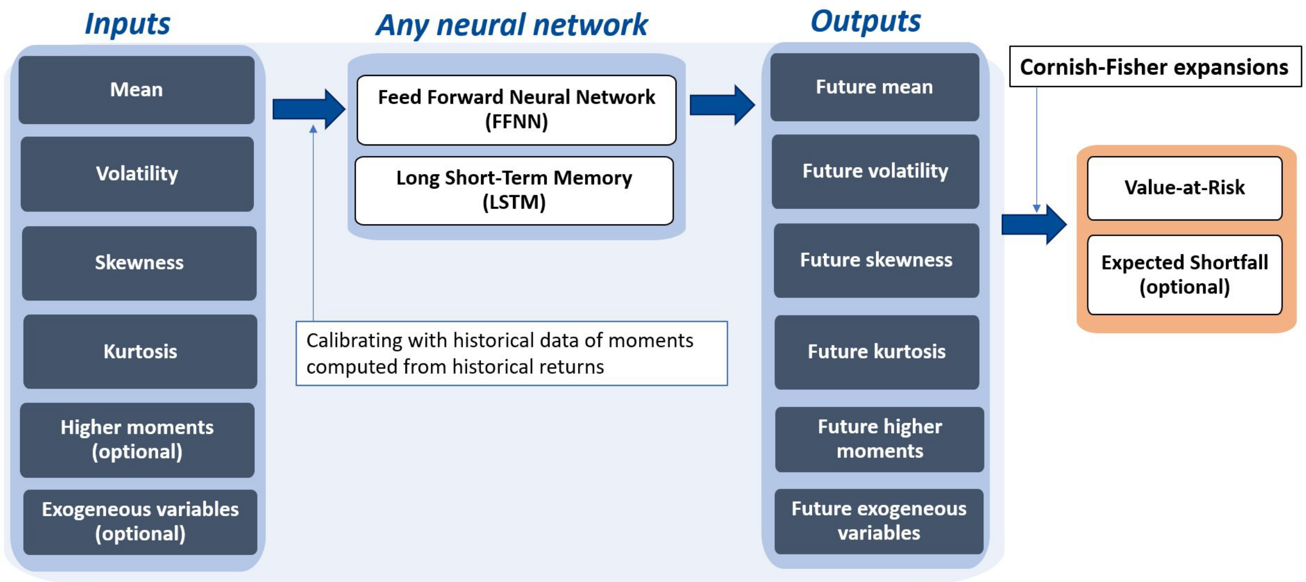


Figure 2. Accumulation of VaR Breaches for 30 Assets

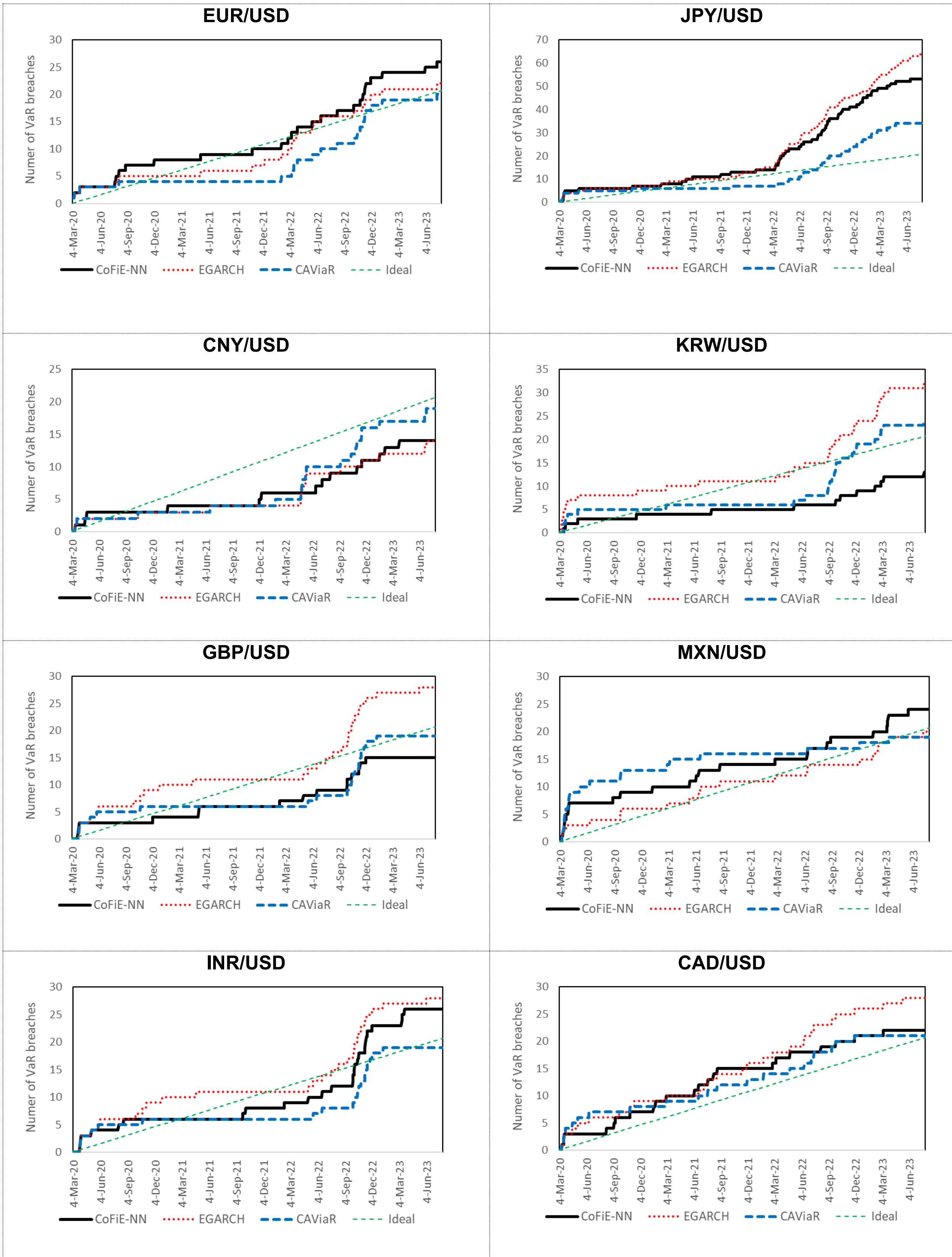


Figure 2. Accumulation of VaR Breaches for 30 Assets (continued)

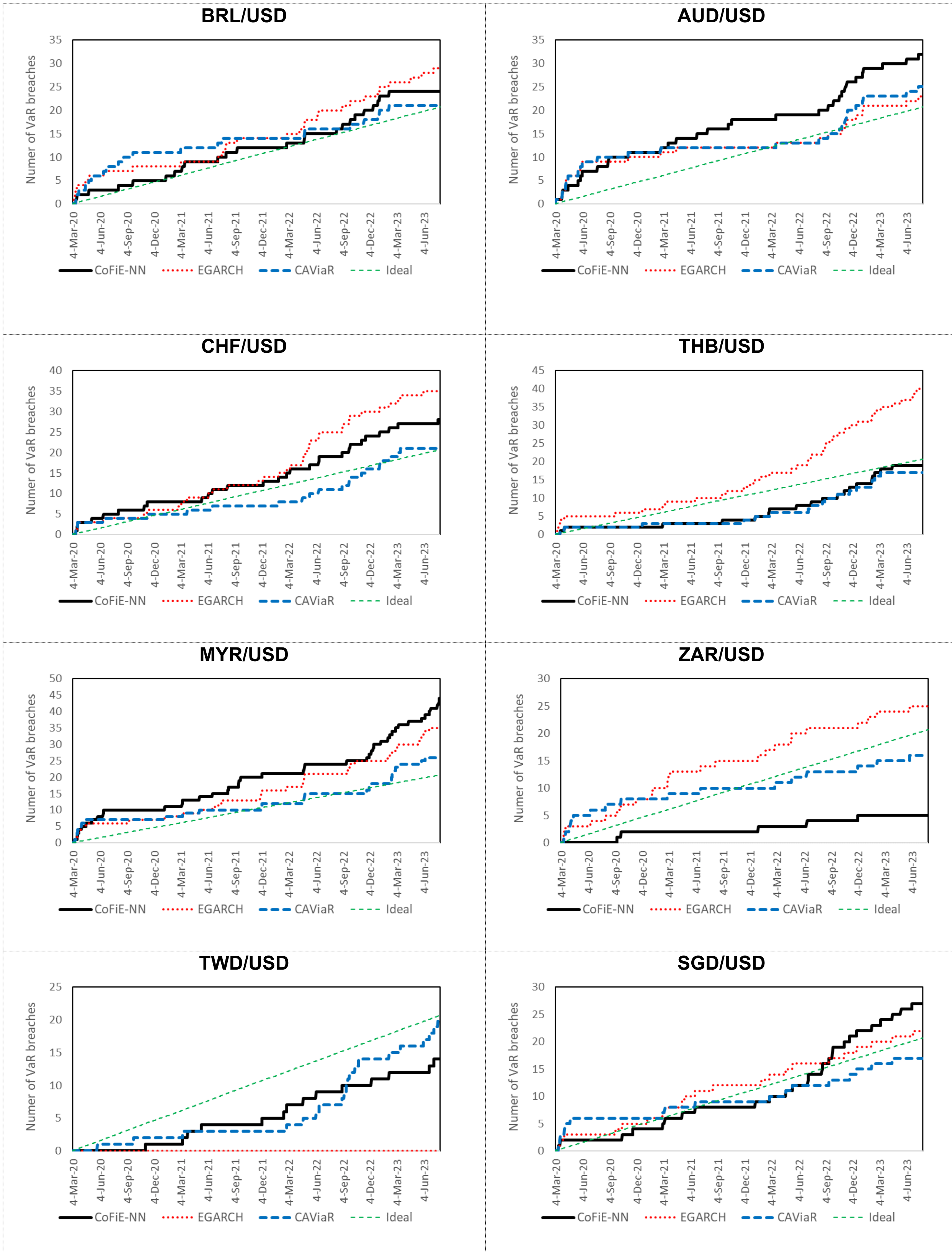


Figure 2. Accumulation of VaR Breaches for 30 Assets (continued)

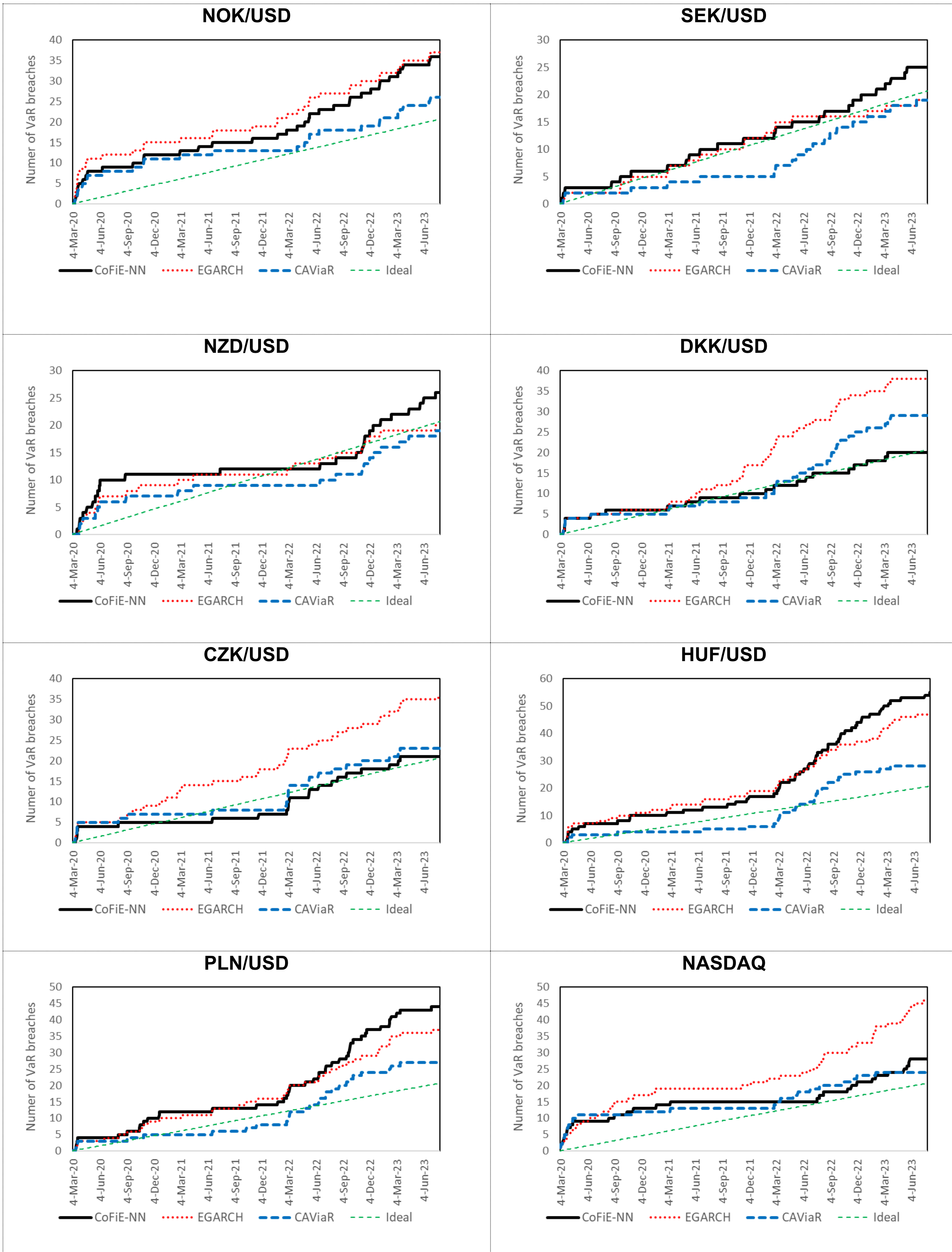


Figure 2. Accumulation of VaR Breaches for 30 Assets (concluded)

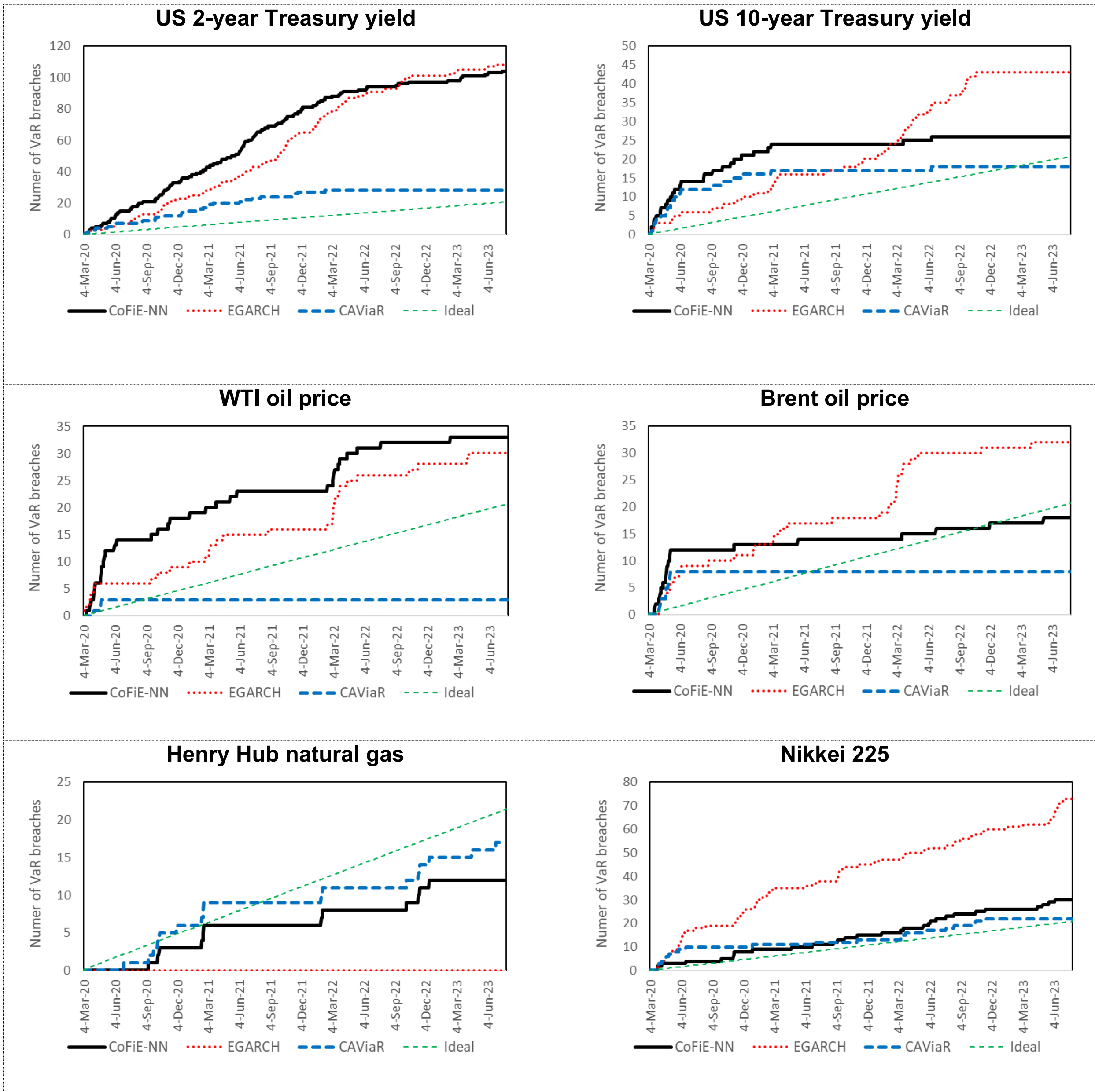


Figure 3. Coefficients of PCA Factors

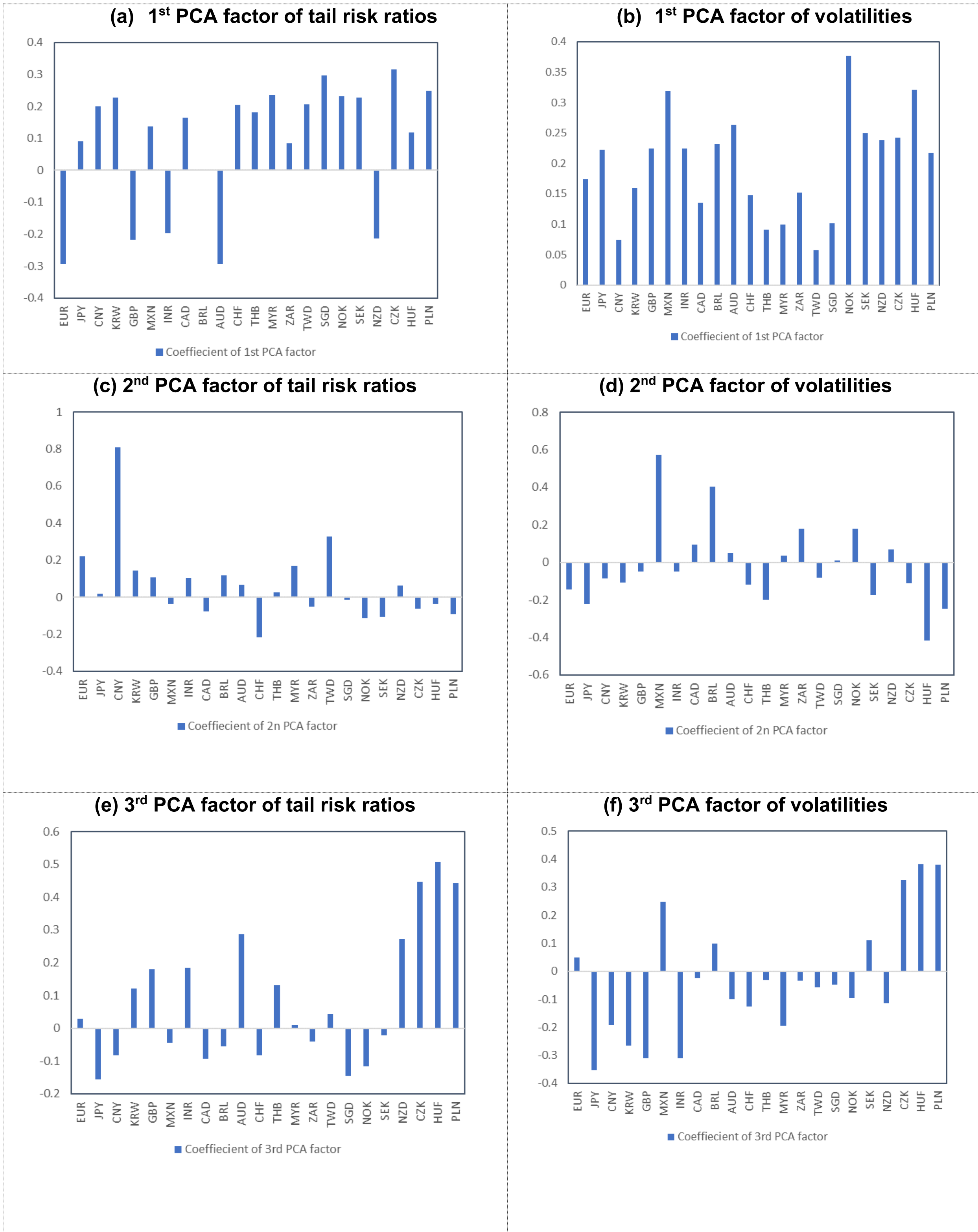
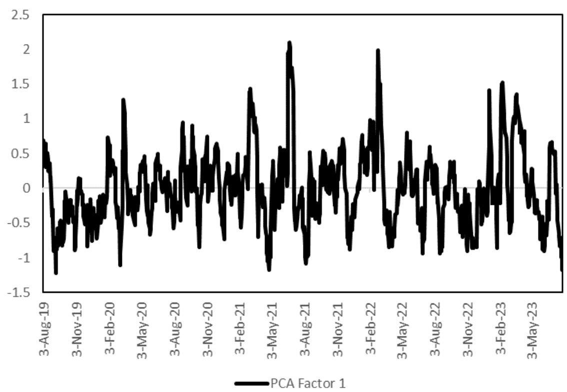
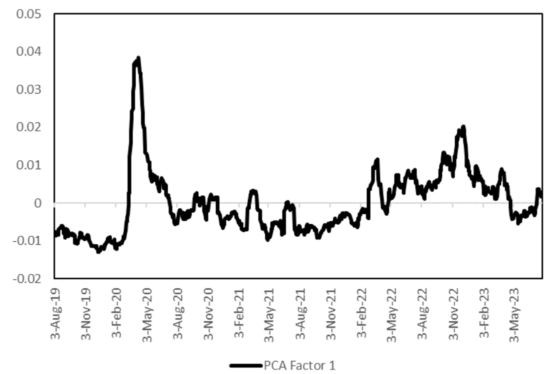


Figure 4. Evolution of Global Factors: Tail risk ratio and Volatility

(a) 1st PCA factor of tail score ratios



(b) 1st PCA factor of volatilities



References

Abdullah, A., Siddiqua, S., Siddiquee, M.S.H., Hossain, N., (2017). Modeling and forecasting exchange rate volatility in Bangladesh using GARCH models: A comparison based on normal and Student's t-error distribution. Financial Innovation, 3(18)

Adrian, T., Shin, H.S., (2014). Procyclical Leverage and Value-at-Risk. Review of Financial Studies, 27(2), 373-403.

Amat, C., Michalski, T., Stolz, G., (2018). Fundamentals and exchange rate forecastability with simple machine learning methods. Journal of International Money and Finance, 88, 1-24.

Amédée-Manesme, C., Barthélémy, F., Maillard, D., (2019) Computation of the corrected Cornish-Fisher expansion using the response surface methodology. Annals of Operations Research, 281(1), 423-453.

Basak, S., Shapiro, A., (2001).Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices. Review of Financial Studies, 14(2), 371-405.

Buczynski, M., Chlebus, M., (2023). GARCHNet: Value-at-Risk Forecasting with GARCH Models Based on Neural Networks. Computational Economics.

Campbell, S.D. (2005) "A Review of Backtesting and Backtesting Procedures," Board of Governors of the Federal Reserve System Working Paper.

Chernozhukov, V., Fernández-Val, I., Galichon, A., (2010). Rearranging Edgeworth–Cornish–Fisher expansions. Economic Theory, 42, 419-435.

Christoffersen, P. (1998). Evaluating Interval Forecasts. International Economic Review, 39, 841-862.

Christoffersen, P., Gonçalves, S. (2005). Estimation risk in financial risk management. Journal of Risk, 7, 1-28.

Chronopoulos, I., Raftapostolos, A., Kapetanios., G. (2023). Forecasting Value-at-Risk Using Deep Neural Network Quantile Regression. Journal of Financial Econometrics, 1-34.

Cornish, E. A., Fisher, Ronald A. (1938). Moments and Cumulants in the Specification of Distributions. Review of the International Statistical Institute, 5(5), 307-320.

De Nicolò, G., Lucchetta, M., (2017). Forecasting Tail Risks. Journal of Applied Econometrics, 32, 159-170.

Emmer, S., Klüppelberg, C., Korn, R., (2001). Optimal Portfolios with Bounded Capital at Risk. Mathematical Finance, 11(4) 365-384.

Engle, R., Manganelli, S., (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. Journal of Business and Economic Statistics, 22(4) 367-381.

Giamouridis, D. (2006). Estimating risk in financial risk management: a correction. Journal of Risk, 8(4), 121-125.

Hochreiter, S., Schmidhuber, J., (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

Jaschke, S.R., (2002). The Cornish–Fisher expansion in the context of Delta–Gamma-normal approximations. Journal of Risk, 4(4), 33-52.

Jorion, P., (2006). Value at Risk: The new benchmark for managing financial risk, 3rd edition, McGraw Hill.

Kelly, B., Jiang, H., (2014). Tail Risk and Asset Prices. Review of Financial Studies, 27(10), 2841-2871.

Kupiec, P., (1995). Techniques for Verifying the Accuracy of Risk Measurement Models, Journal of Derivatives, 3, 73-84.

Lafarguette, R., Veyrune, R. (2021) "Foreign Exchange Intervention Rules for Central Banks: A Risk-based Framework," IMF Working Paper.

Lamb, J.D., Monville, M. E., Tee, K.H., (2019) "Making Cornish-Fisher fit for risk measurement," Journal of Risk, 21(5), 53-81

Longstaff, F., Pan, J., Pedersen, L.H., Singleton, K.J., (2011). How Sovereign is Sovereign Credit Risk. American Economic Journal of Macroeconomics, 3(2), 75-103.

Lönnbark, C., (2016). Approximation methods for multiple period Value at Risk and Expected Shortfall prediction. Quantitative Finance, 16(6), 947-968.

Lopez, J. A., (1999). Methods for Evaluating Value-at-Risk Estimates, Economic Review. Federal Reserve Bank of San Francisco, 2, 3-17.

McNeil, A., Frey, R., (2000). Estimating of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance, 7, 271-300.

Miranda-Agrippino, S., Rey, H., (2020). U.S. Monetary Policy and the Global Financial Cycle. Review of Economic Studies, 87, 2754-2776.

Niu, Z., Wang, C., Zhang, H., (2023). Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models. International Review of Financial Analysis, 89, 102738.

Rojas, C., Herman, M. (2018) "Foreign Exchange Forecasting via Machine Learning," Stanford University Working Paper.

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

Theodossiou, P., (1994). The Stochastic Properties of Major Canadian Exchange Rates. The Financial Review, 29(2), 193-221

Wu, Q., Yan, X., (2019). Capturing deep tail risk via sequential learning of quantile. Journal of Economic Dynamics & Control, 109,103771.

Yamai, Y., Yoshiba, T., (2005). Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking & Finance, 29, 997-1015.

Yiu, K.F.C, (2004). Optimal portfolios under a value-at-risk constraint. Journal of Economic Dynamics and Control, 28, 1317-1334.

Zhang, C., Zhang, Y., Cucuringu, M., Qian, Z., (2023). Volatility Forecasting with Machine Learning and Intraday Commonality. Journal of Financial Econometrics, 73, 1-39.

Zhu, H., Bai, L., He, L., Liu, Z., (2023). Forecasting realized volatility with machine learning: Panel data perspective. Journal of Empirical Finance, 73, 251-271.

PUBLICATIONS

Forecasting Tail Risk via Neural Networks with Asymptotic Expansions

Working Paper No. WP/2024/099