Deep Learning-Based CoVaR Forecasting

Dan Yang*

Chengdu University of Technology, Chengdu 610059, China

*yangdandeemail@163.com

Abstract. This paper examines the feasibility of deep learning for Conditional Value at Risk forecasting in international stock markets. Using daily data for the stock markets of China, the United States, Japan, Germany, and Brazil from 1 July 2010 to 1 July 2025, the study develops a CNN-Transformer quantile regression model for CoVaR prediction. The empirical analysis is based on log return series, representative forecast plots, and the Diebold-Mariano test against benchmark models. The results show that the predicted CoVaR series are clearly time-varying and become more negative during periods of market stress, indicating that the proposed model captures meaningful dynamics in conditional tail risk. The Diebold-Mariano test further shows that the proposed model outperforms both the CNN-QR benchmark and the Transformer-QR benchmark, while the overall test results remain positive for all market pairs. These findings suggest that combining local feature extraction with long-range dependency modeling helps improve CoVaR forecasting performance. The study provides empirical support for the application of deep learning to CoVaR prediction and contributes to the literature on tail risk forecasting in international stock markets.

Keywords: CoVaR Forecasting, Deep Learning, Tail Risk.

1 Introduction

In increasingly integrated equity markets, extreme shocks rarely remain confined to a single market. Stress in a major market can spread through cross-market linkages and amplify losses elsewhere, which makes tail risk measurement an important issue in financial risk management and market surveillance. Value at Risk, denoted as VaR, is widely used to measure potential losses at a given confidence level, but it does not directly describe how the distress of one market affects the risk of another. Conditional Value at Risk, denoted as CoVaR, extends risk measurement to a conditional setting and provides a useful framework for describing systemic spillovers under stress[1].

Existing studies have widely applied CoVaR to systemic risk measurement and cross-market tail spillovers[2]. Most of this literature is based on quantile regression, dynamic conditional correlation generalized autoregressive conditional heteroskedasticity models, or copula-based methods[3]. At the same time, deep learning has been increasingly used in financial forecasting. Recent surveys show that neural networks are widely applied to financial time series analysis, including Convolutional Neural

T. G. Guan et al. (eds.), *Proceedings of the 2026 4th International Conference on Digital Economy and Management Science (CDEMS 2026)*, Advances in Economics, Business and Management Research 392,
https://doi.org/10.2991/978-94-6239-699-9_37

Networks, Long Short-Term Memory networks, Transformer models, and hybrid structures[4]. Studies on neural network quantile methods further suggest that deep learning can improve Value at Risk estimation and systemic risk modeling when nonlinear structure is important[5],[6].

However, deep learning-based CoVaR forecasting remains relatively limited. Existing CoVaR studies often focus on spillover identification and economic interpretation, while deep learning studies in finance more often examine returns, volatility, or VaR[4-6]. Against this background, this study examines CoVaR forecasting for the stock markets of China, the United States, Japan, Germany, and Brazil from a deep learning perspective. The paper develops a deep learning model for CoVaR prediction and evaluates its out-of-sample performance through forecast plots and the Diebold-Mariano test against benchmark models. The aim is to provide empirical evidence on the feasibility of deep learning for CoVaR forecasting and to contribute to the literature on tail risk prediction in international stock markets.

2 Methodology

2.1 CNN-Transformer Quantile Regression Model for CoVaR Forecasting

Conditional Value at Risk, hereafter CoVaR, extends tail risk measurement from a single market to a conditional setting and is well suited to the analysis of extreme spillovers. Let $r_{i,t}$ denote the return of target market i at time t , and let $r_{j,t}$ denote the return of source market j. At quantile level τ, the lower-tail CoVaR of market i conditional on distress in market j can be written as

$$\Pr\left(r_{i,t} \leq CoVaR^{\tau}_{i|j,t} \mid r_{j,t} \leq VaR^{\tau}_{j,t}, \mathcal{F}_{t-1}\right) = \tau \tag{1}$$

where $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$. In this study, CoVaR forecasting is formulated as a conditional quantile prediction problem, so the model directly estimates the lower conditional quantile of the target market under the distress state of the source market[1],[3].

To estimate CoVaR, this study employs a Convolutional Neural Network, hereafter CNN, combined with a Transformer and quantile regression. The input is a lagged return sequence constructed from the target market and the source market. The CNN layer extracts local temporal patterns from the input sequence and captures short-horizon variation in financial returns. The extracted features are then passed to the Transformer encoder. Through self-attention, the Transformer models dependence across different time positions and strengthens the representation of longer-horizon temporal information. The final hidden representation is mapped to the target quantile through an output layer, which yields a direct CoVaR forecast at the chosen quantile level. This design reflects the idea that CoVaR forecasting requires both nonlinear feature extraction and conditional quantile estimation[7],[8].

Model estimation is based on the quantile loss function

$$L_\tau(y_t, \hat{y}_t) = \max\{\tau(y_t - \hat{y}_t), (\tau - 1)(y_t - \hat{y}_t)\} \tag{2}$$

where y_t denotes the realized return of the target market and $\hat{y}_t$ denotes the predicted conditional quantile. This loss assigns asymmetric penalties to overprediction and underprediction and therefore aligns the estimation target with quantile regression. In the present setting, the resulting forecast is interpreted as the CoVaR estimate for the target market conditional on the distress information of the source market[5].

2.2 Evaluation Methods for CoVaR Forecasts

The forecasting performance of the proposed model is evaluated by the Diebold-Mariano test, hereafter the DM test, which compares predictive accuracy across competing models. In this study, the DM test is used to examine whether the proposed deep learning model provides more accurate CoVaR forecasts than the benchmark models. Since CoVaR forecasting is formulated as a conditional quantile prediction problem, the comparison is based on the quantile loss, so the evaluation criterion remains consistent with the estimation objective[9].

Let $l_{1,t}$ and $l_{2,t}$ denote the forecast losses of two competing models at time t, and let the loss differential be defined as

$$d_t = l_{1,t} - l_{2,t} \tag{3}$$

The null hypothesis of the DM test is that the two models have equal predictive accuracy, which implies that the expected loss differential is zero. The test statistic is given by

$$DM = \frac{\bar{d}}{\sqrt{\widehat{Var}(\bar{d})}}, \quad \bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t \tag{4}$$

where $\bar{d}$ is the sample mean of the loss differential and $\widehat{Var}(\bar{d})$ is a consistent estimator of its variance. A statistically significant DM statistic indicates that the difference in forecast performance between the two models is not zero. In this paper, the DM test is used to compare the proposed model with the Convolutional Neural Network quantile regression benchmark and the Transformer quantile regression benchmark, thereby assessing whether the proposed model achieves superior predictive performance in CoVaR forecasting[9].

3 Empirical Results

3.1 Data and Descriptive Statistics

This study uses daily closing price data for five representative stock market indices, namely the CSI 300 Index for China, the S&P 500 for the United States, the Nikkei 225 for Japan, the DAX 40 for Germany, and the Ibovespa for Brazil. The sample period spans from 1 July 2010 to 1 July 2025. To ensure comparability across markets, the closing price series are transformed into log returns, which are calculated as the first difference of the logarithm of consecutive closing prices.

Table 1. reports the descriptive statistics of the log return series. The mean returns of all five markets are close to zero, while the standard deviations indicate clear differences in market volatility. Brazil shows the largest standard deviation, which suggests relatively stronger fluctuations during the sample period. The skewness values are all negative, indicating that the return distributions are tilted toward the left tail.

Table 1. Descriptive Statistics of Log Returns.

Country	Mean	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	J-B Test	ADF Test
China	0.0001	0.081	-0.092	0.014	-0.471	8.578	4670.61***	-56.998***
US	0.0005	0.091	-0.128	0.011	-0.622	17.301	31270.28***	-67.266***
Japan	0.0004	0.097	-0.145	0.013	-0.621	12.816	15051.04***	-40.881***
Germany	0.0004	0.109	-0.131	0.012	-0.447	11.018	10313.20***	-60.627***
Brazil	0.0002	0.130	-0.160	0.015	-0.797	15.718	26250.54***	-67.557***

The kurtosis values of all markets are far above 3, which reveals a pronounced leptokurtic pattern and suggests that extreme observations occur more frequently than under the normal distribution. The Jarque-Bera statistics are significant for all series, which further indicates that the log returns do not follow a normal distribution. Such distributional characteristics support the use of CoVaR as a tail risk measure and also provide motivation for applying a deep learning framework to CoVaR forecasting. In addition, the Augmented Dickey-Fuller test strongly rejects the null hypothesis of a unit root for all return series, which indicates that the log returns are stationary and suitable for subsequent modeling and forecast evaluation.

3.2 CoVaR Forecasting Results

Fig. 1 presents the predicted CoVaR series for four representative market pairs. Two pairs are selected with China as the target market, namely US → China and Japan → China. Two pairs are selected with China as the source market, namely China → US and China → Brazil. This design makes it possible to examine the predictive behavior of the model under both inbound and outbound risk transmission involving the Chinese stock market.

Several common features can be observed from the four panels. The predicted CoVaR series are clearly time-varying rather than mechanically stable, which indicates that the proposed model can adjust lower-tail risk forecasts to changing market conditions. The predicted CoVaR becomes more negative during periods of intensified market stress and moves back toward a relatively moderate level when conditions improve. This pattern is consistent with the economic meaning of CoVaR as a measure of downside risk under distress. In addition, the predicted CoVaR generally co-moves with large fluctuations in the corresponding return series, which suggests that the model responds to periods of elevated volatility and tail events.

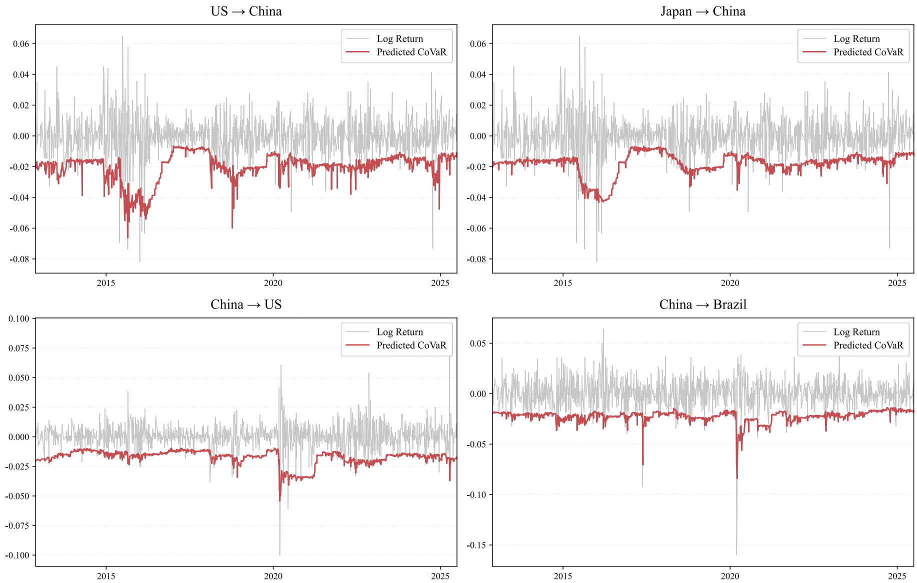

Fig. 1. CoVaR Forecasts for Selected Market Pairs.

Pronounced changes in CoVaR are concentrated in several shock periods. In the US → China and Japan → China panels, the predicted CoVaR declines sharply during major disturbance episodes and then gradually recovers, indicating that China's conditional tail risk increases when external market stress intensifies. In the China → US and China → Brazil panels, the CoVaR series also shows marked downward movements during turbulent periods, which suggests that the model captures time variation in China-related spillovers to foreign markets. Compared with China → US, the China → Brazil pair shows more abrupt downward jumps in some periods, implying that the magnitude of conditional tail risk differs across destination markets.

Overall, the forecasting results display plausible tail-risk dynamics across different market pairs. The predicted CoVaR series remain sensitive to stress episodes and reflect differences between China as a risk receiver and China as a risk transmitter. These patterns provide preliminary evidence that the proposed deep learning model can produce economically interpretable CoVaR forecasts.

3.3 DM Test Results

Table 2 reports the Diebold-Mariano test results for four representative market pairs, including two cases with China as the target market and two cases with China as the source market. Due to space limitations, only these representative pairs are presented in detail. They are sufficient to illustrate the overall pattern of model comparison in this study.

Table 2. DM Test Results for Selected Representative Market Pairs.

Pair	vs CNN-QR	vs Transformer-QR
US → China	2.804^{***}	2.381^{**}
China → US	1.991^{**}	2.779^{***}
Japan → China	2.608^{***}	2.554^{**}
China → Brazil	2.311^{**}	1.821^{*}

The results show that all reported DM statistics are positive and statistically significant. This indicates that the proposed model achieves lower forecast loss than both the CNN-QR benchmark and the Transformer-QR benchmark for the selected pairs. Therefore, the proposed model exhibits superior predictive performance in CoVaR forecasting. The advantage is observed both when China is influenced by external market stress and when China acts as the source of spillovers, which suggests that the proposed model performs consistently across different directions of risk transmission.

The full set of DM test results for all market pairs is not reported here in order to keep the presentation concise. Nevertheless, the overall results remain consistent with the representative evidence shown in Table 2. For all market pairs, the DM statistics are positive and statistically significant, which further supports the robustness of the comparative advantage of the proposed model.

Taken together, the DM test results indicate that the combination of local feature extraction and long-range dependency modeling contributes to improved tail risk forecasting. The CNN component helps capture short-horizon fluctuations in financial returns, while the Transformer component strengthens the modeling of temporal dependence across a longer horizon. Their joint use therefore provides a more effective framework for CoVaR prediction than either single-structure benchmark.

4 Conclusion

This paper investigates the feasibility of deep learning for CoVaR forecasting in international stock markets. Using daily data for China, the United States, Japan, Germany, and Brazil from 1 July 2010 to 1 July 2025, the study develops a CNN-Transformer quantile regression model and evaluates its forecasting performance through representative forecast plots and the Diebold-Mariano test.

The results show that the predicted CoVaR series exhibit clear time variation and respond sensitively to periods of market stress. The Diebold-Mariano test further indicates that the proposed model outperforms both the CNN-QR benchmark and the Transformer-QR benchmark. These findings suggest that combining local feature extraction with long-range dependency modeling helps improve CoVaR forecasting performance. The study provides empirical support for the application of deep learning to CoVaR prediction, although future research may extend the sample scope and examine alternative model structures.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Adrian, T., Brunnermeier, M.K.: CoVaR. American Economic Review 106(7), 1705–1741 (2016). https://doi.org/10.1257/aer.20120555.
2. Abuzayed, B., Bouri, E., Al-Fayoumi, N., Jalkh, N.: Systemic risk spillover across global and country stock markets during the COVID-19 pandemic. Economic Analysis and Policy 71, 180–197 (2021). https://doi.org/10.1016/j.eap.2021.04.010.
3. Koenker, R., Bassett Jr., G.: Regression Quantiles. Econometrica 46(1), 33–50 (1978). https://doi.org/10.2307/1913643.
4. Tang, Y., Song, Z., Zhu, Y., Yuan, H., Hou, M., Ji, J., Tang, C., Li, J.: A survey on machine learning models for financial time series forecasting. Neurocomputing 512, 363–380 (2022). https://doi.org/10.1016/j.neucom.2022.09.003.
5. Chronopoulos, I., Raftapostolos, A., Kapetanios, G.: Forecasting Value-at-Risk Using Deep Neural Network Quantile Regression. Journal of Financial Econometrics 22(3), 636–669 (2024). https://doi.org/10.1093/jjfinec/nbad014.
6. Keilbar, G., Wang, W.: Modelling systemic risk using neural network quantile regression. Empirical Economics 62(1), 93–118 (2022). https://doi.org/10.1007/s00181-021-02035-1.
7. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791.
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention Is All You Need. In: Advances in Neural Information Processing Systems 30, pp. 5998–6008 (2017).
9. Diebold, F.X., Mariano, R.S.: Comparing Predictive Accuracy. Journal of Business and Economic Statistics 13(3), 253–263 (1995). https://doi.org/10.1080/07350015.1995.10524599.