DSSR

Discussion Paper No. 141

Deep learning for multivariate volatility forecasting
in high-dimensional financial time series.

Rei Iwafuchi and Yasumasa Matsuda

May, 2024

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Deep learning for multivariate volatility forecasting in high-dimensional financial time series.

Rei Iwafuchi[1] and Yasumasa Matsuda[1]

[1]Tohoku University

May 2024

Abstract

The market for investment trusts of large-scale portfolios, including index funds, continues to grow, and high-dimensional volatility estimation is essential for assessing the risks of such portfolios. However, multivariate volatility models suitable for high-dimensional data have not been extensively studied. This paper introduces a new framework based on the Spatial AR model, which provides fast and stable estimation, and demonstrates its application through simulations using historical data from the S&P 500.

1 Introduction

In high-dimensional multivariate GARCH modeling, the paramount challenge is to construct a model that is both expressive and feasible in terms of the number of parameters that can be effectively estimated. Many models impose strong constraints to facilitate estimation; however, these constraints do not necessarily reflect the true nature and volatility characteristics of the actual data.

Traditionally, volatility matrix estimation has been attempted using two main approaches. The first approach involves a natural extension of ARCH by Engle [1] and GARCH models by Bollerslev [2], including the VECH model by Bollerslev, Engle, and Wooldridge [3], BEKK model by Engle and Kroner [4], and Constant Conditional Correlation (CCC) model by He and Teräsvirta [5]. The second approach attempts to simplify modeling by reducing the original data, with Factor-ARCH by Ding [6] and Orthogonal-GARCH models by Alexander [7] being prime examples.

However, each approach has its problems when estimating the volatility of high-dimensional data involving hundreds of dimensions. In models like the VECH and BEKK, the number of parameters increases exponentially with dimensionality, making estimation exceedingly difficult. Even more parsimonious models, such as the CCC process, are highly restrictive. Its relaxed counterpart, the Dynamic Conditional Correlation (DCC) model by Engle [8], frequently involves the cumbersome computation of large matrix determinants. Additionally, models utilizing factors often struggle to adequately capture the variability in high-dimensional data with large sample sizes using only a few factors. Moreover, volatility matrices estimated using a factor model in high dimensions almost invariably lack positive definiteness.

This paper introduces a robust and fast method for estimating high-dimensional volatility matrices using a Spatial Autoregressive (SAR) model combined with Lasso. Specifically, it focuses on the relationships between variables, removing conditional correlations with the SAR model before estimating volatility with the GARCH model. Furthermore, we present a flexible extension to models utilizing neural networks.

2 SAR + Volatility Framework

2.1 Notation

In the field of multivariate volatility modeling, we consider the volatility matrix in a context where past information has been observed. Specifically, there is interest in the conditional covariance matrix at time t, given the information available up to time $t-1$, which is typically denoted by $\mathcal{F}_{t-1}$. Let $\boldsymbol{r}_t$ be a random vector representing stock returns at time t. Based on this understanding, multivariate volatility models are represented as follows:

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$
$$\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1} \sim N(0, V_t)$$

In most cases, μ is assumed to be zero. The primary focus in this area is how to model V_t

The famous univariate GARCH(p,q) model, which we are going to utilize later, is described as:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$$
$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k \varepsilon_{t-k}^2 + \sum_{\ell=1}^{q} \beta_\ell \sigma_{t-\ell}^2$$

This paper utilizes the GARCH model to estimate V_t, elaborating on how the model accounts for the dynamic nature of volatility based on historical data and how it can be adapted to capture the complexities of multivariate time series. By employing the GARCH model framework, we aim to provide a robust method for estimating the conditional variances that are crucial for accurate financial forecasting and risk management.

2.2 SAR GARCH

The key assumption of the Spatial Autoregressive GARCH model is the following:

Assumption 1 *The structure of conditional correlations among stocks can be represented by the SAR model, and these relationships are sparse:*

$$\boldsymbol{r}_t = W\boldsymbol{r}_t + \boldsymbol{u}_t, \quad W_{ii} = 0 \tag{1}$$

It would be intuitively natural to assume that any given stock is related to a few others while unrelated to many. Under this assumption, we can estimate volatility for each univariate separately, significantly reducing the number of parameters.

Assumption 2 *The residuals of the SAR model $\boldsymbol{u}_t$ are conditionally and cross-sectionally uncorrelated, and follow a GARCH process.*

$$\boldsymbol{u}_t|\mathcal{F}_t \sim N(0, S_t), \quad S_{t,ij} = 0 \quad if \quad i \neq j \tag{2}$$

$$S_{t,ii} = \alpha_0 + \sum_{k=1}^{p} \alpha_k u_{t-k,i}^2 + \sum_{\ell=1}^{q} \beta_\ell S_{t-\ell,ii} \tag{3}$$

For estimating the volatility of residuals, the GARCH model is utilized. If the volatility of the residuals is correctly estimated, the conditional covariance matrix of $\{\boldsymbol{r}_t\}$ can be obtained as follows:

$$V_t = (I - W)^{-1} S_t [(I - W)^{-1}]^\top$$

2.3 Extension Using Deep Learning

Here, we introduce a method for extending the SAR+GARCH framework using deep learning. In order to capture the complex dependencies of time series, LSTM(Hochreiter and Schmidhuber [9] Gers, Schmidhuber, and Cummins [10]) is employed. Let NN be a neural network we employ. Then the formulation of the extended model using deep learning is as follows.

$$\begin{aligned}
\boldsymbol{r}_t &= W\boldsymbol{r}_t + \boldsymbol{u}_t, \quad W_{ii} = 0 \\
\boldsymbol{u}_t|\mathcal{F}_{t-1} &\sim N(0, S_t), \quad S_t = diag(\boldsymbol{s}_t) \\
\boldsymbol{s}_t &= \boldsymbol{s}_t^{garch} \odot \boldsymbol{s}_t^{nn} \\
\boldsymbol{s}_t^{garch} &= \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 \odot \boldsymbol{u}_{t-1} + \boldsymbol{\alpha}_2 \odot \boldsymbol{s}_{t-1}^{garch} \\
\log \boldsymbol{s}_t^{nn} &= NN(\frac{\boldsymbol{u}_{t-1}}{\boldsymbol{s}_{t-1}^{garch}})
\end{aligned}$$

The key idea here is that the important factors of volatility fluctuations can be captured by the GARCH model, and neural networks are used as an aid to capture nonlinear dependencies. GARCH model is extremely effective, and we utilize it to the fullest extent.

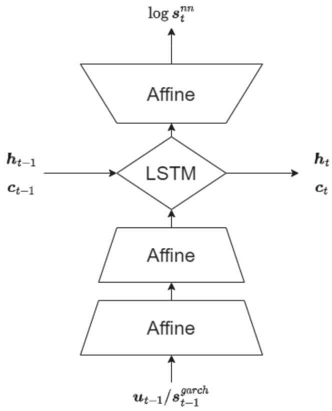


Figure 1: Brief Description of the Neural Network Structure

Figure 1 describe the structure of NN. In the neural network, the input is first compressed to an appropriate dimension using two Affine Layers. Then, complex dependencies are captured using an LSTM Layer, followed by another Affine Layer to restore the dimensions to their original state.

3 Estimation

In this study, we repeatedly predict the volatility of the following day using one month's worth of data. Then we minimize the negative log likelihood. For simplicity, we omit the constant term included in the negative log likelihood and double its value as bellow.

$$Loss = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{u}_t^\top S_t^{-1}\boldsymbol{u}_t - 2\log|I - W|$$

Furthermore, we utilize mini-batch training to seek global minimum, where the optimization method employed is Adam.

4 Empirical Example

To evaluate the performance of the proposed model, we conduct the experiments using price data of 331 stocks included in the S&P 500. The test period from June 22, 2018, to April 2, 2021, is divided into batches of 20 samples along the timeline. Training is conducted using samples from T days immediately preceding each test period, and losses of one to five periods ahead prediction are calculated to compare multiple models. As benchmark models, we adopt the Factor Model + POET(See Fan, Fan, and Lv [11], Fan, Liao, and Mincheva [12] and Fan, Liao, and Mincheva [13] for the details.) and the univariate GARCH model. To observe the differences caused by the sample size of the training data, experiments are conducted in three scenarios: T=250, T=1000, and T=3000.

Figure 2, Figure 3 and Figure 4 describe the mean loss of each test period in each scenario. Here, SARNN refers to the neural network-based extended model of the SAR GARCH model.

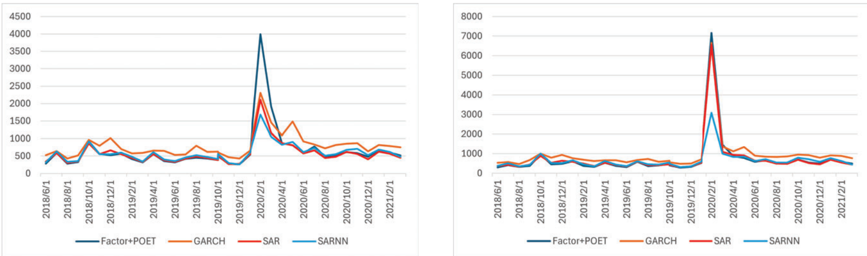
In the scenarios with T=250 and T=1000, both the SAR GARCH model and the SARNN model outperform the two benchmark models. Under these scenarios, the one-step-ahead forecasts of the two proposed models are almost identical in performance. However, for the five-steps-ahead forecast, SARNN shows exceptionally good performance. During shocks, sudden high-frequency trading creates complex dependencies between stocks and past prices, making multiple stepsd ahead predictions extremely challenging. However, the nonlinearity of SARNN is thought to have helped capture these relationships.

In the scenario with T=3000, the performance of the one- step-ahead forecast remains good for both models. However, for the five-steps-ahead forecast, SARNN's performance significantly deteriorated during the shock. This is thought to be due to the fact that there were no major shocks in the past 3000 samples, leading the highly expressive SARNN to overlearn the dependencies during ordinary period and neglect the complex dependencies necessary for the five-steps-ahead forecast.

5 Conclusion

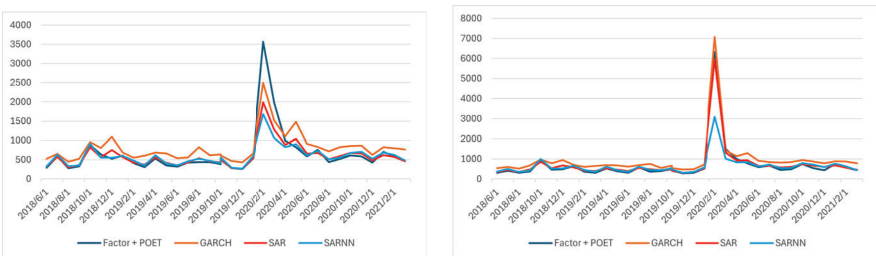
In this paper, we introduced the SAR + volatility framework as a new method for high-dimensional volatility estimation. This framework is a fast and stable model that does not require frequent matrix inversion calculations during optimization iterations. Additionally, the utilization of Lasso and neural networks allows for stable predictions even when the sample size is not sufficiently large compared to the dimension of the inputs.

The SAR GARCH model and its extension using neural networks, the SARNN model, demonstrated superior performance over benchmark models in simulations. Furthermore, it was found that with proper setting of the training data, the SARNN model significantly outperforms others in multi-steps-ahead forecasting during financial shocks. This is believed to be due to the nonlinearity of neural networks, which contributes to capturing the complex dependencies between stocks or with past returns during shocks. We believe that the results of this research can be applied to various issues, including the risk assessment of large-scale portfolios, and will contribute to the development of the field of finance.



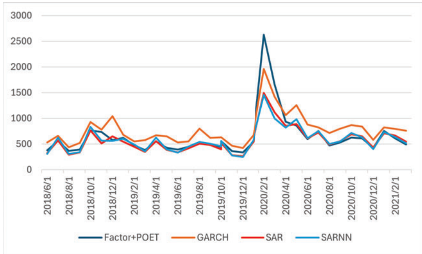
(a) One step ahead forecast (b) Five steps ahead forecast

Figure 2: Mean loss of each test period with T=250

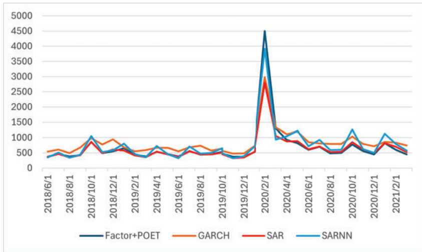


(a) One step ahead forecast (b) Five steps ahead forecast

Figure 3: Mean loss of each test period with T=1000



(a) One step ahead forecast



(b) Five steps ahead forecast

Figure 4: Mean loss of each test period with T=3000

References

[1] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation,” *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982, ISSN: 00129682, 14680262.

[2] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986, ISSN: 0304-4076.

[3] T. Bollerslev, R. F. Engle, and J. M. Wooldridge, “A capital asset pricing model with time-varying covariances,” *Journal of Political Economy*, vol. 96, no. 1, pp. 116–131, 1988, ISSN: 00223808, 1537534X.

[4] R. F. Engle and K. F. Kroner, “Multivariate simultaneous generalized arch,” *Econometric Theory*, vol. 11, no. 1, pp. 122–150, 1995, ISSN: 02664666, 14694360.

[5] C. He and T. Teräsvirta, “An extended constant conditional correlation garch model and its fourth-moment structure,” *Econometric Theory*, vol. 20, no. 5, pp. 904–926, 2004, ISSN: 02664666, 14694360.

[6] Z. Ding, *Time series analysis of speculative returns*. UMI, 1994.

[7] C. Alexander, “Orthogonal methods for generating large positive semi-definite covariance matrices,” *University of Reading Working Paper No. 2000-06*, 2000.

[8] R. Engle, “Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models,” *Journal of Business & Economic Statistics*, vol. 20, no. 3, pp. 339–350, 2002, ISSN: 07350015.

[9] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, ISSN: 0899-7667.

[10] F. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with lstm,” in *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, vol. 2, 1999, 850–855 vol.2.

[11] J. Fan, Y. Fan, and J. Lv, "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, vol. 147, no. 1, pp. 186–197, 2008, Econometric modelling in finance and risk management: An overview, ISSN: 0304-4076.

[12] J. Fan, Y. Liao, and M. Mincheva, "High-dimensional covariance matrix estimation in approximate factor models," *The Annals of Statistics*, vol. 39, no. 6, pp. 3320–3356, 2011.

[13] J. Fan, Y. Liao, and M. Mincheva, "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 75, no. 4, pp. 603–680, 2013, ISSN: 13697412, 14679868.