

Reducing Overestimating and Underestimating Volatility via the Augmented Blending-ARCH Model

Jun Lu ¹, & Shao Yi ¹

¹JJ Capital Fund

Correspondence: Jun Lu, jun.lu.locky@gmail.com.

Received: March 17, 2022 Accepted: April 20, 2022 Available online: April 23, 2022

doi:10.11114/aef.v9i2.5507 URL: <https://doi.org/10.11114/aef.v9i2.5507>

Abstract

SVR-GARCH model tends to “backward eavesdrop” when forecasting the financial time series volatility in which case it tends to simply produce the prediction by deviating the previous volatility. Though the SVR-GARCH model has achieved good performance in terms of various performance measurements, trading opportunities, peak or trough behaviors in the time series are all hampered by underestimating or overestimating the volatility. We propose a blending ARCH (BARCH) and an augmented BARCH (aBARCH) model to overcome this kind of problem and make the prediction towards better peak or trough behaviors.

The method is illustrated using real data sets including SH300 and S&P500. The empirical results obtained suggest that the augmented and blending models improve the volatility forecasting ability.

Keywords: Volatility prediction, Blending-ARCH, Peak and trough behaviors of volatility forecasting, Neural networks, Effective ratio.

1. Introduction

In quantitative finance, the asset returns are usually modeled by the normal distribution since its form is normal (hence the name normal distribution) (Levy & Duchin, 2004; Wirjanto & Xu, 2009; Lu, 2022b). Due to its normal form, the normal distribution cannot model the fat tails (leptokurtosis) and asymmetry (skewness) of the asset returns. Choi & Nam (2008) suggested the SU-normal distribution to describe the two non-normal features embedded in financial time series. And recently, a polynomial adjusted Student-t distribution is proposed to model asset returns with heavy-tailed and skewed distributions (León & Níguez, 2021).

Volatility is a measure of the degree of fluctuation of financial return and is a proxy for risk which is incorporated into the famous Sharpe ratio and information ratio to measure the performance of a quantitative strategy on the asset (Sharpe, 1966; 1994; Goodwin, 1998). And even the volatility is used to quantify the premium (option price) in both option call and option put; the higher the volatility, the higher the option price (Haug, 2007; Natenberg, 2014). Conditional heteroscedastic models such as the autoregressive conditional heteroscedasticity (ARCH) (Engle, 1982), generalized ARCH (GARCH) (Bollerslev, 1986), exponential GARCH (EGARCH) (Nelson, 1991), and GJR (Glosten et al., 1993) can model the volatility time series that exhibit time-varying volatility and volatility clustering, i.e., periods of swings interspersed with periods of relative calm. However, empirical studies show that these models have low forecasting accuracy (Jorion, 1995; Choudhry & Wu, 2008; Bezerra & Albuquerque, 2017).

In recent years, machine learning methods have been employed to provide superior results in many time series problems, e.g., flight ticket prediction (Lu, 2017b; Rajankar & Sakharkar, 2019), and metabolic pathway dynamics (Costello & Martin, 2018). Machine learning-based methods have also been proposed to improve the performance since it can capture nonlinear features hidden in the time series such as leptokurtosis, asymmetry, volatility clustering, and momentum, e.g., empirical test shows support vector regression (SVR) has superior results than GARCH models (Pérez-Cruz et al., 2003; Chen et al., 2010; Li, 2014; Bezerra & Albuquerque, 2017).

Empirical evidence shows that there are oscillations between several regimes in the financial market, in which the overall distribution of returns is a mixture of two or more than two states of normal (Guidolin, 2011; Levy & Kaplanski, 2015). Bezerra & Albuquerque (2017) proposed an SVR-GARCH model that captures the regime-switching behavior and performs better than existing methods since SVR is a kernel-based method and the Gaussian kernel can be decomposed into an infinite mixture of polynomials (Lu, 2021). However, a pictorial analysis on the volatility series via the SVR-GARCH model shows that the model tends to “backward eavesdrop”, that is, the model simply reports the volatility by the previous value (in a sense of deviation). Moreover, a pictorial analysis on the predicted series shows that the SVR-GARCH model tends to underestimate the volatility in the peak areas, and overestimate the volatility in the trough areas. While, in real quantitative strategies, trading opportunities always happen in these areas. For example, in option trading, traders tend to sell option if the implied volatility (IV)¹ is higher than the predicted volatility (i.e., the proxy of realized volatility (RV)) during the peak area of the volatility sequence; if the predicted volatility is underestimated, the trading can lose money. On the other hand, if the IV is lower than the predicted volatility during the trough area of the volatility series, the traders will buy option; if the predicted volatility is overestimated, the strategy will make less money.

The emphasis of this paper is on addressing practical problems that arise in implementing ARCH family models and the SVR-GARCH model. In such settings, it is common knowledge that the overestimation/underestimation can be too large, leading to a lack of interpretability, and other issues. For these reasons, it is well motivated to develop an augmented method to constrain the predicted volatility closer to the realized volatility during the peak and trough areas and towards less “backward eavesdropping”. With this goal in mind, we propose the blending and augmented algorithm on traditional ARCH model methods.

2. Augmented Blending Volatility Forecasting

2.1 Parametric and Semi-parametric Volatility Models

Let P_t be asset price at time t where $t \in [1, 2, \dots, T]$, the return of the asset at time t can be obtained by the following equation:

$$r_t = \ln \left(\frac{P_t}{P_{t-1}} \right), \quad \forall t \in [2, 3, \dots, T]. \quad (1)$$

ARCH. The ARCH model is simply a linear model having the following form for ARCH(p):

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i h_{t-i} + \eta_t, \quad (2)$$

where h_t is the conditional variance at time t , and η_t is the error term (Engle, 1982). The conditional variance is postulated to be a linear function of the past p innovations. The reference model for volatility is the base upon which the rests of the ARCH family models are built.

GARCH. The GARCH model can capture volatility clustering and improves the ARCH model making it extensively used to model real financial time series problem (Bollerslev, 1986). The GARCH(p, q) model is defined as follows:

$$\begin{aligned} r_t &= a + a_t, \\ a_t &= \sqrt{h_t} \cdot \epsilon_t, \quad \epsilon_t \sim i.i.d.(0, 1), \\ h_t &= \alpha_0 + \sum_{i=1}^p \beta_i h_{t-i} + \sum_{i=1}^q \alpha_i a_{t-i}^2, \end{aligned} \quad (3)$$

where the innovation ϵ_t is an independent identically distributed (i.i.d.) random variable with zero mean and unit variance and it is common to set $a = 0$.

EGARCH. Nelson (1991) developed the EGARCH to model the skewness of financial returns whilst the variance is positive. The EGARCH(p, q) model is defined as follows:

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^p \beta_i \ln(h_{t-i}) + \sum_{i=1}^q \alpha_i \frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sqrt{h_{t-i}}} \quad (4)$$

where γ_i is the asymmetric response parameter.

¹In financial mathematics, the implied volatility (IV) of an option contract is the value of the volatility of the underlying instrument which will return a theoretical value equal to the current market price of the said option when we input in an option pricing model, e.g., the Black-Scholes or Black-Scholes-Merton models (Merton et al., 1978).

GJR. To capture asymmetric response of volatility, Glosten et al. (1993) introduced the $GJR(p, q)$ model:

$$h_t = \alpha_0 + \sum_{i=1}^q (\alpha_i + \gamma_i S_{t-i}^-) a_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}, \quad (5)$$

where

$$S_{t-i}^- = \begin{cases} 1, & a_{t-i} < 0; \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where the parameters are nonnegative.

SVR-GARCH. Machine learning aided method has been explored to forecast the volatility more precisely. Bezerra & Albuquerque (2017) showed the SVR-GARCH with mixture of Gaussian kernels can achieve better performance. In the SVR-GARCH model, the output variable is h_t and the input vector is $\mathbf{x}_t = [a_{t-1}^2, h_{t-1}]^T$:

$$r_t = f(r_{t-1}) + a_t, \quad (7)$$

where $f(\cdot)$ is the decision function estimated by SVR for the mean equation. Then, the variance is estimated by

$$\tilde{h}_t = g(\tilde{h}_{t-1}, a_{t-1}^2), \quad (8)$$

where $g(\cdot)$ is the decision function estimated by SVR again.

2.2 Performance Measure

In real applications, the measurement of h_t is not observed directly and Andersen & Bollerslev (1998) showed the following realized volatility (RV) is closer to the theoretical volatility:

$$\tilde{h}_t = \frac{1}{5} \sum_{i=0}^4 r_{t-i}^2. \quad (9)$$

In the following development of our methods, we only consider this volatility proxy while other proxies may alter the results. However, this issue is beyond the scope of this paper since any volatility proxy is an imperfect estimator of the true conditional variance (Patton, 2011). To measure the proposed methods numerically in the next section, we use the root mean square error (RMSE) to evaluate the prediction performance which is given by

$$\text{RMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2}, \quad (10)$$

where y_t denote the observation at time t and \hat{y}_t represent the prediction of y_t . Further, the mean absolute error (MAE) is also considered:

$$\text{MAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|. \quad (11)$$

In all scenarios, smaller RMSE and MAE indicate better performance. Apart from the numerical measurement on the methods, we also highlight the pictorial behaviors of the predictions during the *peaks* and *troughs* of the predicted time series.

To further evaluate the significance of difference, we follow the DM test (Diebold & Mariano, 1995; Harvey et al., 1997; Diebold, 2015). Given the MAE losses of two predicted time series (\mathbf{f}, \mathbf{g}) with the true sequence \mathbf{y} , the two-sided DM test follows the following null and alternative hypothesis:

$$\begin{aligned} H_0 : \text{MAE}(\mathbf{y}, \mathbf{f}) - \text{MAE}(\mathbf{y}, \mathbf{g}) &= 0; \\ H_1 : \text{MAE}(\mathbf{y}, \mathbf{f}) - \text{MAE}(\mathbf{y}, \mathbf{g}) &\neq 0. \end{aligned} \quad (12)$$

When the null hypothesis is rejected, there is evidence the two predicted sequences are different. Let the *loss-differential* d_i be defined as the following absolute deviation:

$$d_i = |f_i - y_i| - |g_i - y_i|, \quad (13)$$

where f_i, g_i, y_i are i -th elements of $\mathbf{f}, \mathbf{g}, \mathbf{y}$ respectively. When the time series is of length T , define $N = T^{1/3} + 1$, the DM statistics is calculated as follows:

$$\text{DM} = \frac{\bar{d}}{\sqrt{[\eta_0 + 2 \sum_{k=1}^{N-1} \eta_k] / T}} \sim \mathcal{N}(0, 1), \quad (14)$$

where $\bar{d} = \frac{1}{T} \sum_{i=1}^T d_i$, $\eta_k = \frac{1}{T} \sum_{i=k+1}^T (d_i - \bar{d})(d_{i-k} - \bar{d})$. Thus, there is significant difference between the two sequences if $|\text{DM}|$ is larger than the two-tailed critical value of the standard normal distribution (i.e., 1.96 if the significance value is selected to be 0.05).

2.3 Augmented Blending Volatility Forecasting

Blending-ARCH (BARCH). Applying the various autoregressive conditional heteroscedasticity family models on the real market return data, one should find the models can always overestimate the volatility. In real quantitative trading, this will result in losing trading opportunities. To overcome the drawback, we propose the blending-ARCH model. Suppose $h_t^1, h_t^2, \dots, h_t^N$ are predictions of N parametric volatility models at time t (which can be either ARCH, GARCH, GJR, or EGARCH with different parameters, e.g., different p or q in the models.). Then the *uniform blending* of the results can be obtained by

$$h_t = \frac{1}{N} (h_t^1 + h_t^2 + \dots + h_t^N). \quad (15)$$

However, since most of the parametric models will overestimate the volatility, a simple mean of the predictions can still overestimate it. A better proposal is the *linear blending*:

$$h_t = w_0 + w_1 \cdot h_t^1 + w_2 \cdot h_t^2 + \dots + w_N \cdot h_t^N, \quad (16)$$

where the weight vector $\boldsymbol{w} = [w_1, w_2, \dots, w_N, w_0]^T \in \mathbb{R}^{N+1}$ is set to be the same one for different time t and is learned from ordinary least squares (OLS) in our case. Let $\boldsymbol{h} = [\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_T]^T \in \mathbb{R}^T$ be the vector containing the realized volatility of each time $t \in [1, 2, \dots, T]$, \boldsymbol{X} be the matrix containing the prediction of different parametric models:

$$\boldsymbol{X} = \begin{bmatrix} h_1^1 & h_1^2 & \dots & h_1^N & 1 \\ h_2^1 & h_2^2 & \dots & h_2^N & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h_T^1 & h_T^2 & \dots & h_T^N & 1 \end{bmatrix} \in \mathbb{R}^{T \times (N+1)}. \quad (17)$$

Then OLS solution of \boldsymbol{w} can be obtained by

$$\boldsymbol{w} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{h}, \quad (18)$$

where \boldsymbol{w} is also the projection of \boldsymbol{h} onto the column space of \boldsymbol{X} (Strang, 1993; Lu, 2021; 2022c). For each time $t \in [1, 2, \dots, T]$, we want to predict $h_t = \boldsymbol{x}_t^T \boldsymbol{w}$ as closer to \tilde{h}_t as possible where \boldsymbol{x}_t^T is the t -th row of \boldsymbol{X} . Any other machine learning method can be applied to predict h_t :

$$h_t = h(\boldsymbol{x}_t), \quad \forall t \in [1, 2, \dots, T]. \quad (19)$$

In our implementation, we employ the neural network approach to do the prediction, hence termed *BARCH-NN* in the sequel. Since the number of features obtained from the blending method is not large, we apply a simple neural network structure with only fully connected layer (Appendix A).

Time series correlation. To do the feature selection, we want to compute the correlation of different time series data. The feature with high mean correlation with other features will be deleted. Given two vectors $\boldsymbol{x}_1, \boldsymbol{x}_2$, the cosine correlation is defined as:

$$\text{Cosine}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{\boldsymbol{x}_1^T \boldsymbol{x}_2}{|\boldsymbol{x}_1| \cdot |\boldsymbol{x}_2|} \quad (20)$$

Therefore, for the data matrix \boldsymbol{X} in Eq (17), where each column is given by $\hat{\boldsymbol{x}}_i$ for all i in $[1, 3, \dots, N]$, the (i, j) -th element of the feature correlation matrix is given by:

$$\text{Correlation}(\boldsymbol{X})_{i,j} = \text{Cosine}(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j), \quad (21)$$

where $\text{Correlation}(\boldsymbol{X})$ is a matrix of shape $N \times N$ (the bias feature is deleted for the correlation evaluation). We will see, the feature selection based on this correlation matrix is important to improve the prediction performance and provide significant difference with other models in the sequel. While other correlation algorithms are also explored in our experiments, the difference is not that large, e.g., Pearson correlation, temporal-weighted correlation, and generalized correlation (Tulchinsky, 2019). Though other algorithms exist to select the features, e.g., random forest (Anani & Samarabandu, 2018), mixture model (Rasmussen, 1999; Xuan et al., 2001; Murphy, 2012; Lu, 2017a; Murphy, 2022), MCP algorithm (Zhang, 2010), and SCAD algorithm (Fan & Li, 2001), these complex models seem not to provide better performance in our experiments while the computation may require extra resources.

Effective ratio. Kaufman (2013; 1995) suggested replacing the “weight” variable in the exponential moving average (EMA) formula with a constant based on the *efficiency ratio* (ER). And the ER is shown to provide promising results for financial forecasting via classic quantitative strategies Lu (2022b) where the ER of the closing price are calculated to decide the trend of the asset. This indicator is designed to measure the *strength of a trend*, defined within a range from -1.0 to +1.0 where the larger magnitude indicates a larger upward or downward trend. Instead of calculating the ER of the closing price, we want to calculate the ER of the volatility series. Given the window size M , it is calculated with a simple formula:

$$e_t(M) = \frac{s_t}{n_t} = \frac{h_{t-1} - h_{t-1-M}}{\sum_{i=1}^M |h_{t-i} - h_{t-1-i}|} \tag{22}$$

$$= \frac{\text{Total volatility change for a period}}{\text{Sum of absolute volatility change for each bar}'}$$

where $e_t(M)$ is the ER at time t . We carefully notice that the ER at time t is based on $h_{t-1}, h_{t-2}, \dots, h_{t-1-M}$ to avoid a forward bias since we do not know h_t at time t . At a strong trend (i.e., the input volatility is moving in a certain direction, up or down) the ER will tend to 1 in absolute value; if there is no directed movement, it will be a little more than 0. The effective ratio is a versatile tool, it can also be utilized to find the minimum or maximum point in a time period for the time series, while by only identifying the minimal or maximal point can be noisy in this case. Recently, Lu (2022a) shows the ER can be applied to stochastic optimization to select the smoothing constant in an adaptive fashion making the optimizer converge faster and continue to reduce loss.

Augmented BARCH (aBARCH) and augmented SVR-GARCH (aSVR-GARCH). The aBARCH method simply adds the effective ratio to the predicted results of bARCH:

$$h_t = w_0 + w_1 \cdot h_t^1 + w_2 \cdot h_t^2 + \dots + w_N \cdot h_t^N + \sigma \cdot e_t(M), \tag{23}$$

where the window size M and deviation σ are hyperparameters that can be tuned by cross validation (CV). We carefully notice that, when M is larger, $e_t(M)$ will usually tend to 0, thus $\sigma \cdot e_t(M)$ tend to 0; and when $\sigma \rightarrow 0$, $\sigma \cdot e_t(M)$ will also tend to 0, which reduces to the trivial case. Similarly, the aSVR-GARCH can be obtained based on Eq. (8):

$$\tilde{h}_t = g(\tilde{h}_{t-1}, a_{t-1}^2) + \sigma \cdot e_t(M). \tag{24}$$

In practice, $M = 15$ and $\sigma = 0.1$ can be a good candidate since it only puts a small amount of value to the augmented prediction.

Table 1. Descriptive statistics for daily returns.

Statistics	SH300 return	S&P500 return
Observations	2956	2778
Mean	0.00017	-0.00041
Std.dev	0.01429	0.01086
Median	0.00034	-0.00069
Kurtosis	4.42880	19.72931
Skewness	-0.53949	1.23705
Maximum	0.06715	0.13616
Minimum	-0.08748	-0.08578

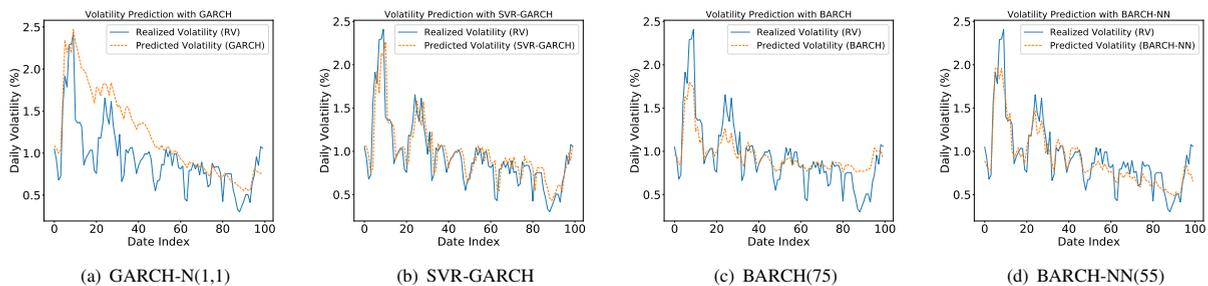


Figure 1. SH300: Pictorial behaviors of GARCH, SVR-GARCH, BARCH models on SH300 data. We can observe the “backward eavesdropping” problem in SVR-GARCH model in Figure 1(b). The BARCH(75) and BARCH-NN(55) models reduce the “backward eavesdropping” problem to some extent. Though the BARCH-NN(55) may look “worse” than SVR-GARCH from the picture at first glance, the RMSE and MAE performances of the former one are better

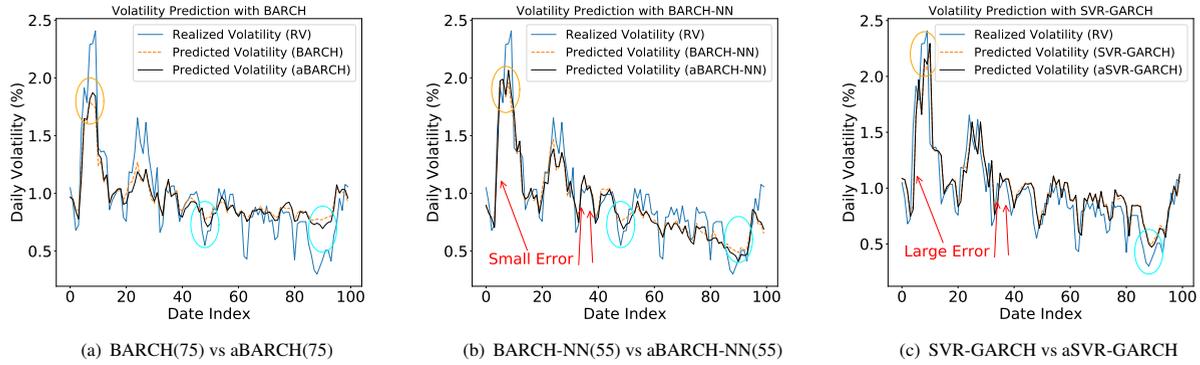


Figure 2. SH300: Demonstration of the augmented method on each model for reducing underestimation (orange cycles) and overestimation (cyan cycles) on SH300 data. Compare Figure 2(b) and Figure 2(c) shows the BARCH-NN(55) and aBARCH-NN(55) models perform better in the rising edge and descending edge of the volatility than the SVR-GARCH model

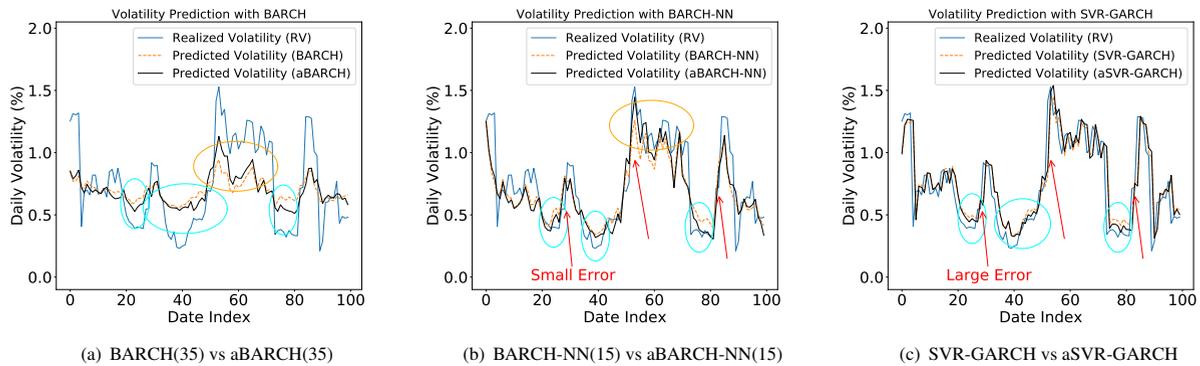


Figure 3. S&P500: Demonstration of the augmented method on each model for reducing underestimation (orange cycles) and overestimation (cyan cycles) on S&P500 data. Compare Figure 3(b) and Figure 3(c) shows the BARCH-NN(15) and aBARCH-NN(15) models perform better in the rising edge and descending edge of the volatility than the SVR-GARCH model

3. Experiments

To evaluate the strategy and demonstrate the main advantages of the proposed BARCH, BARCH-NN, and augmented methods, real market data sets are used. In a wide range of scenarios across various feature settings, BARCH-NN and aBARCH-NN increase the prediction ability in terms of RMSE and MAE measurements on out-of-sample prediction; and augmented method alone can reduce overestimation and underestimation in peak and trough areas.

We obtain the S&P500 index data, which is a market-capitalization-weighted index of 500 leading publicly traded companies in the U.S. with a time period of 11 years (between Jan. 3, 2011 and Jan. 14, 2022). The S&P500 index uses a market-cap weighting method, giving a higher percentage allocation to companies with the larger market-capitalizations. Further, we obtain the SH300 index data, a Chinese alternative for the S&P500 (similar to S&P500, but still has a large difference, which is a market-capitalization weighted index of 300 leading publicly traded companies in China), with a time period of 12 years (between Jan. 4, 2010 and Mar. 7, 2022). Table 1 shows the summary of the descriptive statistics for the SH300 and S&P500 over the whole period we have obtained. For each of ARCH, GARCH, EGARCH, GJR models, four different distributions for innovations are considered, i.e., the normal, the Student's t, the skewed Student's t, and the generalized error distribution (GED). For each innovation, the parameters are tuned by selecting the smallest Bayesian information criterion (BIC) and the best parameters will be represented in the model name (e.g., GARCH-N(1,1) in Table 2 or Table 3). The goodness of fit for SH300 and S&P500 are shown in Table B.6 and Table B.7 respectively in Appendix Appendix B. For the SVR-GARCH model, the parameters are tuned by cross validation.

In all scenarios, we allocate the first 2456 or 2278 observations (for SH300 and S&P500 respectively) for training, the next 252 observations for validation, and the last 252 observations for testing². The training and validation samples constitute the in-sample data, and the test samples constitute the out-of-sample data.

²252 days is usually the total number of trading days per year in the U.S. and is widely used in research.

In all experiments, the features for BARCH or BARCH-NN are obtained with a set of different parameters on ARCH, GARCH, EGARCH, or GJR models (say different p, q , or distributions for innovations). The methods will be denoted as $BARCH(K)$ or $BARCH-NN(K)$ if the number of features is K . In all scenarios, the features are selected randomly from the total feature sets (90 features totally in our experiments, 15 from ARCH, 25 from GARCH, 25 from EGARCH, and 25 from GJR).

Based on Eq. (21), the correlation matrix of the training features is obtained, a matrix of shape $\mathbb{R}^{90 \times 90}$ in our case. Then any feature with a mean correlation smaller than 0.9 is selected afterwards. For both the SH300 and the S&P500 data sets, 10 features are finally selected. The heatmaps of the selected 10 features are shown in Figure 4 for the two data sets. While the correlated based feature selection procedure for the models will be denoted as $BARCH(CO)$ or $BARCH-NN(CO)$ in the experiments.

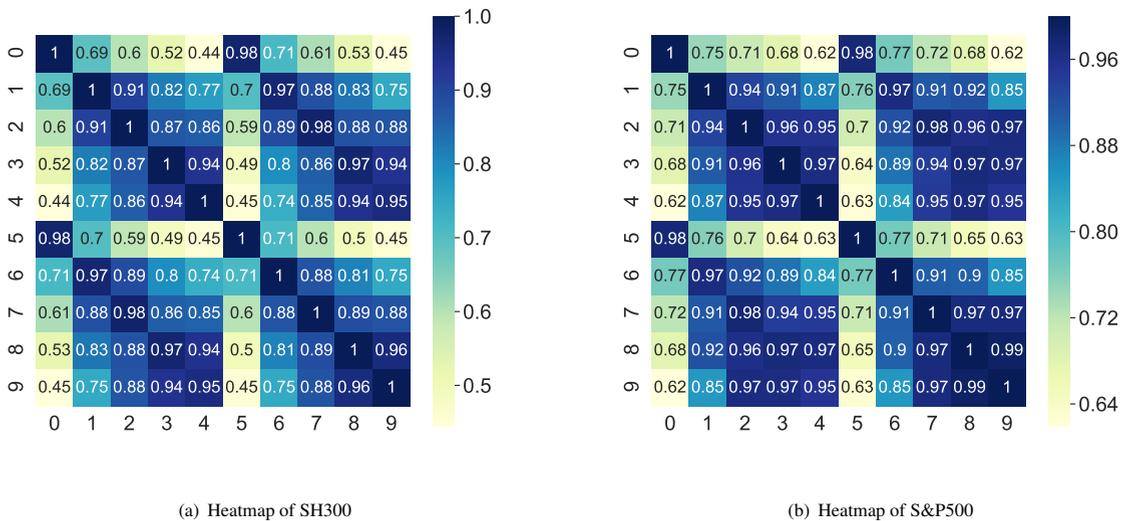


Figure 4. The heatmaps of the correlation among selected features for SH300 and S&P500 data sets. 10 out of 90 features are selected

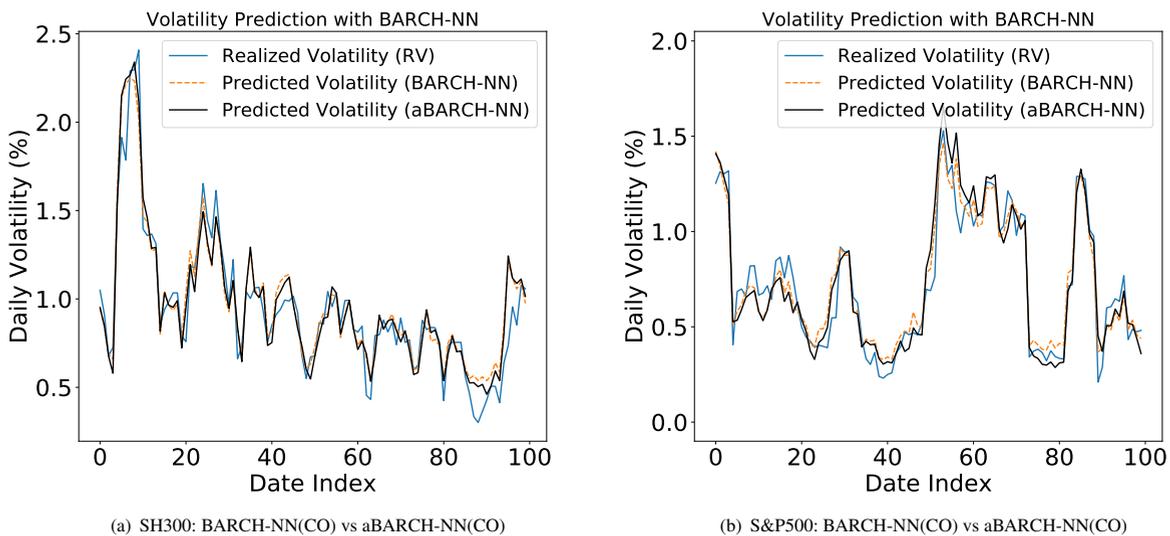


Figure 5. Best out-of-sample performances on SH300 and S&P500 data sets. “Backward eavesdropping” is not observed anymore in this case

3.1 SH300

We first depict typical behaviors on the test data set of SH300³ for different models as shown in Figure 1 where a GARCH model predicts well for the trending of the volatility; however, the performance is low since it cannot reveal the details of the volatility

³Only 100 out of 252 trading days are shown in this and the following figures to save space.

Table 2. SH300: Out-of-sample evaluation on SH300 data. N in BARCH(N), aBARCH(N), BARCH-NN(N), or aBARCH-NN(N) refers to the number of features selected randomly.

Model	RMSE ($\times 10^{-3}$)	MAE ($\times 10^{-3}$)
Eavesdrop	2.279	1.484
BARCH(5)	2.593	1.917
BARCH(15)	2.305	1.753
BARCH(35)	2.287	1.730
BARCH(55)	2.306	1.740
BARCH(75)	2.302	1.737
BARCH(CO)	2.217	1.724
BARCH-NN(5)	2.226	1.703
BARCH-NN(15)	2.045	1.604
BARCH-NN(35)	1.910	1.524
BARCH-NN(55)	1.785	1.419
BARCH-NN(75)	2.045	1.582
BARCH-NN(CO)	1.289	0.959
aBARCH(5)	2.453	1.803
aBARCH(15)	2.170	1.633
aBARCH(35)	2.173	1.641
aBARCH(55)	2.196	1.656
aBARCH(75)	2.192	1.652
aBARCH(CO)	2.121	1.621
aBARCH-NN(5)	2.086	1.600
aBARCH-NN(15)	1.938	1.517
aBARCH-NN(35)	1.848	1.512
aBARCH-NN(55)	1.689	1.359
aBARCH-NN(75)	1.904	1.455
aBARCH-NN(CO)	1.295	0.948
SVR-GARCH	2.223	1.516
aSVR-GARCH	2.421	1.649
ARCH-N(14)	7.261	4.796
ARCH-t(11)	8.094	5.524
ARCH-st(11)	8.106	5.532
ARCH-G(10)	7.251	4.847
GARCH-N(1,1)	5.617	4.041
GARCH-t(1,1)	5.693	4.241
GARCH-st(1,1)	5.699	4.247
GARCH-G(1,1)	5.356	3.936
EGARCH-N(1,1)	6.932	5.167
EGARCH-t(1,1)	7.108	5.441
EGARCH-st(1,1)	7.118	5.453
EGARCH-G(1,1)	6.699	5.058
GJR-N(1,1)	5.773	4.126
GJR-t(1,1)	6.526	4.668
GJR-st(2,1)	6.524	4.671
GJR-G(1,1)	5.702	4.124

Table 3. S&P500: Out-of-sample evaluation on S&P500 data. N in BARCH(N), aBARCH(N), BARCH-NN(N), or aBARCH-NN(N) refers to the number of features selected randomly.

Model	RMSE ($\times 10^{-3}$)	MAE ($\times 10^{-3}$)
Eavesdrop	3.700	2.241
BARCH(5)	5.274	3.967
BARCH(15)	5.080	3.906
BARCH(35)	4.739	3.699
BARCH(55)	4.739	3.715
BARCH(75)	4.753	3.723
BARCH(CO)	5.309	4.161
BARCH-NN(5)	3.728	2.502
BARCH-NN(15)	3.018	2.287
BARCH-NN(35)	3.237	2.138
BARCH-NN(55)	3.242	2.306
BARCH-NN(75)	3.559	2.485
BARCH-NN(CO)	2.262	1.580
aBARCH(5)	5.079	3.792
aBARCH(15)	4.889	3.736
aBARCH(35)	4.560	3.530
aBARCH(55)	4.563	3.556
aBARCH(75)	4.578	3.563
aBARCH(CO)	5.137	3.995
aBARCH-NN(5)	3.617	2.463
aBARCH-NN(15)	2.950	2.189
aBARCH-NN(35)	3.117	2.064
aBARCH-NN(55)	3.121	2.238
aBARCH-NN(75)	3.405	2.345
aBARCH-NN(CO)	2.203	1.568
SVR-GARCH	3.542	2.218
aSVR-GARCH	3.586	2.253
ARCH-N(6)	20.811	8.626
ARCH-t(6)	22.426	9.272
ARCH-st(6)	22.296	9.222
ARCH-G(6)	21.631	8.931
GARCH-N(2,1)	20.484	8.913
GARCH-t(2,1)	21.872	9.468
GARCH-st(2,1)	21.717	9.417
GARCH-G(2,1)	21.181	9.179
EGARCH-N(2,1)	11.691	5.833
EGARCH-t(2,1)	12.805	6.387
EGARCH-st(2,1)	12.584	6.890
EGARCH-G(2,1)	12.234	6.080
GJR-N(1,1)	12.659	7.684
GJR-t(1,1)	13.795	8.241
GJR-st(1,1)	13.777	8.239
GJR-G(1,1)	13.344	8.009

Table 4. SH300: DM test on SH300 data (benchmark is the SVR-GARCH result). N in BARCH(N), aBARCH(N), BARCH-NN(N), or aBARCH-NN(N) refers to the number of features selected randomly.

Model	DM Statistics	P-value
BARCH-NN(15)	-0.061	0.9516
BARCH-NN(35)	0.822	0.4119
BARCH-NN(55)	1.971	0.0499
BARCH-NN(75)	0.175	0.8610
BARCH-NN(CO)	5.916	0.0000
aBARCH-NN(15)	2.464	0.0144
aBARCH-NN(35)	0.991	0.3226
aBARCH-NN(55)	2.762	0.0062
aBARCH-NN(75)	1.571	0.1174
aBARCH-NN(CO)	6.249	0.0000
aSVR-GARCH	-0.570	0.5690

Table 5. S&P500: DM test on S&P500 data (benchmark is the SVR-GARCH result). N in BARCH(N), aBARCH(N), BARCH-NN(N), or aBARCH-NN(N) refers to the number of features selected randomly.

Model	DM Statistics	P-value
BARCH-NN(15)	-0.216	0.8290
BARCH-NN(35)	0.802	0.4234
BARCH-NN(55)	-3.706	0.0003
BARCH-NN(75)	-1.915	0.0567
BARCH-NN(CO)	5.808	0.0000
aBARCH-NN(15)	0.508	0.6119
aBARCH-NN(35)	1.399	0.1629
aBARCH-NN(55)	-2.607	0.0097
aBARCH-NN(75)	-0.812	0.4174
aBARCH-NN(CO)	5.706	0.0000
aSVR-GARCH	0.431	0.6666

trending as shown in Figure 1(a). Figure 1(b) shows a typical graphical representation of the SVR-GARCH model where we may think it predicts well at first glance. However, the disadvantage of “backward eavesdropping” can be easily observed, i.e., predict volatility at time t by deviating the realized volatility at time $t - 1$ to some extent. And the error is accumulated during the time frame. The RMSE for the SVR-GARCH model is 2.223×10^{-3} ; while the error reduces to 1.310×10^{-3} if we shift the prediction one-day-ahead quantifying the extent of the “backward eavesdropping”. The BARCH result shown in Figure 1(c) partly solves the “backward eavesdropping” problem and the RMSE of the BARCH-NN(55) model is the smallest among the four. The RMSE of the BARCH-NN(55) model is 1.785×10^{-3} ; while again if we shift the prediction one-day-ahead, the error increases to 2.585×10^{-3} which again states that the SVR-GARCH model can “backward eavesdrop”. Meanwhile, similar to the GARCH model, BARCH can also lose some detailed information, e.g., the prediction between 80-th and 100-th day in Figure 1(c). This problem is less severe in the BARCH-NN model which is because the neural network is a more powerful machine learning tool that can approximate universal functions (Gelenbe et al., 1999).

Figure 2 compares the augmented version of each model on SH300 data. In all comparisons, we find the augmented method can reduce underestimation during the peak areas (orange cycles in the figures) and reduce overestimation during the trough areas (cyan cycles in the figures) to some extent. As discussed in the introduction section, in real quantitative strategies, trading opportunities always happen in these areas. For example, in option trading, traders tend to sell option if the implied volatility is higher than the predicted volatility (i.e., the proxy of realized volatility (RV)) during the peak area. If the predicted volatility is underestimated, the strategy can lose money.

Though the BARCH-NN(55) and aBARCH-NN(55) shown in Figure 2(b) are not the best ones we have obtained for the BARCH-NN models (we shall come back to the best one in Figure 5(a) in next paragraph), a comparison with the SVR-GARCH result shows the BARCH-NN performs better when the volatility is in the rising edge or descending edge (red arrows in Figure 2(b) and Figure 2(c)).

As discussed previously, a correlation-based procedure is used to select the features for BARCH and BARCH-NN models where 10 out of 90 features are selected. The BARCH-NN(CO) and aBARCH-NN(CO) perform best for the SH300 data set. Figure 5(a) shows the difference between the predicted volatility and the realized volatility is small, and the “backward eavesdropping” problem is solved as well.

Table 2 reports the detailed forecasting volatility performance of different models for SH300 data in terms of RMSE and MAE based on Eq. (10) and (11). We observe the BARCH-NN(CO) is close to the realized volatility, the RMSE and MAE performances also show promising results for the models. Since we mentioned the SVR-GARCH has the “backward eavesdropping” problem, we also report the RMSE and MAE by simply predicting the volatility at time t by that at time $t - 1$. The measure is termed *Eavesdrop* in Table 2. We notice the SVR-GARCH is only a litter better than the *Eavesdrop* result in terms of RMSE, while even worse than the latter in the sense of MAE. However, the proposed BARCH-NN(CO) or aBARCH-NN(CO) models are better in both of the two measurements. Table 4 provides the DM test of the BARCH models (benchmark is the SVR-GARCH result) where the feature selected BARCH-NN(CO) and aBARCH-NN(CO) models indicate significant differences to the SVR-GARCH model.

3.2 S&P500

Similarly, Figure 3 compares the augmented version of each model on S&P500. Again, in all comparisons, we find the augmented method can reduce underestimation during the peak areas (orange cycles in the figures) and reduce overestimation during the trough areas (cyan cycles in the figures) to some extent. Though the BARCH-NN(15) and aBARCH-NN(15) shown in Figure 3(b) are not the best ones we have obtained for the BARCH-NN models, a comparison with the SVR-GARCH shows the BARCH-NN performs better when the volatility is in the rising edge or descending edge (red arrows in Figure 3(b) and Figure 3(c)).

Same as the SH300 case, a correlation-based procedure is used to select the features for BARCH and BARCH-NN models where the same 10 out of 90 features are selected. This is partly because the ARCH, GARCH, EGARCH, or GJR models are consistent over different data sets. The BARCH-NN(CO) and aBARCH-NN(CO) perform best for the S&P500 data set. Figure 5(b) shows the difference between the predicted volatility and the realized volatility is small for our best model, and again the “backward eavesdropping” problem is also not observed on the S&P500 data.

Table 3 reports the detailed forecasting volatility performance of different models for S&P500 data in terms of RMSE and MAE. Again, we observe that the SVR-GARCH model only performs a little better than the *Eavesdrop* result. While the proposed BARCH-NN(CO) or aBARCH-NN(CO) models are much better. Table 5 provides the DM test of the BARCH models (benchmark is the SVR-GARCH result). The feature selected BARCH-NN(CO) and aBARCH-NN(CO) models on S&P500 data still provide evidence of significant difference to the SVR-GARCH model.

4. Conclusion

The aim of this paper is to solve the “backward eavesdropping” problem in the SVR-GARCH model and overcome the overestimation and underestimation in various volatility prediction models. To check the mentioned problems and the performance of the proposed models, real market data sets including SH300 and S&P500 are used in empirical analysis. The article proposes a blending-ARCH model based on correlation feature selection that improves the performance in terms of RMSE, MAE, DM test, and pictorial behavior. Furthermore, the existing ARCH family or SVR-GARCH models tend to overestimate or underestimate in the peak and trough areas of the volatility sequence. The overestimation and underestimation problem is severe in real quantitative trades since many trading strategies tend to bid or ask in the peak and trough areas of the volatility sequence. An augmented method based on the effective ratio can be applied to any existing volatility models and shows it can reduce overestimating and underestimating.

References

- Anani, W., & Samarabandu, J. (2018). Comparison of recurrent neural network algorithms for intrusion detection based on predicting packet sequences. In *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, pages 1–4. IEEE. <https://doi.org/10.1109/CCECE.2018.8447793>.
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905. <https://doi.org/10.2307/2527343>.
- Bezerra, P. C. S., & Albuquerque, P. H. M. (2017). Volatility forecasting via SVR-GARCH with mixture of Gaussian kernels. *Computational Management Science*, *14*(2), 179–196. <https://doi.org/10.1007/s10287-016-0267-0>.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.
- Chen, S., Härdle, W. K., & Jeong, K. (2010). Forecasting volatility with support vector machine-based GARCH model. *Journal of Forecasting*, *29*(4), 406–433. <https://doi.org/10.1002/for.1134>.
- Choi, P., & Nam, K. (2008). Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the SU-normal distribution. *Journal of Empirical finance*, *15*(1), 41–63. <https://doi.org/10.1016/j.jempfin.2006.06.009>.
- Choudhry, T., & Wu, H. (2008). Forecasting ability of GARCH vs Kalman filter method: evidence from daily UK time-varying beta. *Journal of Forecasting*, *27*(8), 670–689. <https://doi.org/10.1002/for.1096>.
- Costello, Z., & Martin, H. G. (2018). A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ systems biology and applications*, *4*(1), 1–14.
- Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, *33*(1), 1–1. <https://doi.org/10.1080/07350015.2014.983236>.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3). <https://doi.org/10.2307/1392185>.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007. <https://doi.org/10.2307/1912773>.
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

- Gelenbe, E., Mao, Z.-H., & Li, Y.-D. (1999). Function approximation with spiked random networks. *IEEE Transactions on Neural Networks*, 10(1), 3–9. <https://doi.org/10.1109/72.737488>.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779–1801. <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x>.
- Goodwin, T. H. (1998). The information ratio. *Financial Analysts Journal*, 54(4), 34–43. <https://doi.org/10.2469/faj.v54.n4.2196>.
- Guidolin, M. (2011). Markov switching models in empirical finance. In *Missing data methods: Time-series methods and applications*. Emerald Group Publishing Limited. [https://doi.org/10.1108/S0731-9053\(2011\)000027B004](https://doi.org/10.1108/S0731-9053(2011)000027B004).
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4).
- Haug, E. G. (2007). *The complete guide to option pricing formulas*. McGraw-Hill Education.
- Jorion, P. (1995). Predicting volatility in the foreign exchange market. *The Journal of Finance*, 50(2), 507–528. <https://doi.org/10.1111/j.1540-6261.1995.tb04793.x>.
- Kaufman, P. J. (1995). *Smarter trading*. McGraw-Hill, New York.
- Kaufman, P. J. (2013). *Trading Systems and Methods*, + Website, volume 591. John Wiley & Sons. <https://doi.org/10.1002/9781119202561>.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- León, Á., & Níguez, T.-M. (2021). Polynomial adjusted Student-t densities for modeling asset returns. *The European Journal of Finance*, pages 1–23. <https://doi.org/10.1080/1351847X.2021.1985561>.
- Levy, H., & Duchin, R. (2004). Asset return distributions and the investment horizon. *The Journal of portfolio management*, 30(3), 47–62. <https://doi.org/10.3905/jpm.2004.412319>.
- Levy, M., & Kaplanski, G. (2015). Portfolio selection in a two-regime world. *European Journal of Operational Research*, 242(2), 514–524. <https://doi.org/10.1016/j.ejor.2014.10.012>.
- Li, Y. (2014). Estimating and forecasting APARCH-skew-t model by wavelet support vector machines. *Journal of Forecasting*, 33(4), 259–269. <https://doi.org/10.1002/for.2275>.
- Lu, J. (2017)a. Hyperprior on symmetric Dirichlet distribution. *arXiv preprint arXiv:1708.08177*.
- Lu, J. (2017)b. Machine learning modeling for time series problem: Predicting flight ticket prices. *arXiv preprint arXiv:1705.07205*.
- Lu, J. (2021). A rigorous introduction for linear models. *arXiv preprint arXiv:2105.04240*.
- Lu, J. (2022)a. Adasmooth: An adaptive learning rate method based on effective ratio. *arXiv preprint arXiv:2204.00825*.
- Lu, J. (2022)b. Exploring classic quantitative strategies. *arXiv preprint arXiv:2202.11309*.
- Lu, J. (2022)c. Matrix decomposition and applications. *arXiv preprint arXiv:2201.00145*.
- Merton, R. C., Scholes, M. S., & Gladstein, M. L. (1978). The returns and risk of alternative call option portfolio investment strategies. *Journal of Business*, pages 183–242.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Murphy, K. P. (2022). *Probabilistic machine learning: an introduction*. MIT press.
- Natenberg, S. (2014). *Option volatility and pricing: Advanced trading strategies and techniques*. McGraw-Hill Education.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370. <https://doi.org/10.2307/2938260>.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256. <https://doi.org/10.1016/j.jeconom.2010.03.034>.
- Pérez-Cruz, F., Afonso-Rodríguez, J. A., & Giner, J. (2003). Estimating GARCH models using support vector machines. *Quantitative Finance*, 3(3), 163.
- Rajankar, S., & Sakharkar, N. (2019). A survey on flight pricing prediction using machine learning. *International Journal Of Engineering Research & Technology (Ijert)*, 8(6), 1281–1284.
- Rasmussen, C. (1999). The infinite Gaussian mixture model. *Advances in neural information processing systems*, 12.
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, 39(1), 119–138. <https://doi.org/10.1086/294846>.
- Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management*, 21(1), 49–58. <https://doi.org/10.3905/jpm.1994.409501>.
- Strang, G. (1993). *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA.
- Tulchinsky, I. (2019). *Finding Alphas: A quantitative approach to building trading strategies*. John Wiley & Sons. <https://doi.org/10.1002/9781119571278>.
- Wirjanto, T. S., & Xu, D. (2009). *The applications of mixtures of normal distributions in empirical finance: A selected survey*. Citeseer.
- Xuan, G., Zhang, W., & Chai, P. (2001). EM algorithms of Gaussian mixture model and hidden Markov model. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 1, pages 145–148. IEEE. <https://doi.org/10.1109/ICIP.2001.958974>.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, 38(2), 894–942. <https://doi.org/10.1214/09-AOS729>.

Appendix A. Neural network structure

For the network structure, we only use a structure with fully connected layers. For each fully connected layer, we denote it by $F(\langle \text{num outputs} \rangle; \langle \text{activation function} \rangle)$. Then the network structure we use can be described by:

$$F(100:\text{Relu}) - F(50:\text{Relu}) - F(50:\text{Relu}) - F(1:\text{MSE}). \tag{A.1}$$

And the parameter to optimize the network is given as follows:

- L2 penalty (regularization term): $\alpha=0.0001$;
- Learning rate: 0.001;
- Batch size: 200;
- Optimizer: Adam($\beta_1 = 0.9, \beta_2 = 0.999$) (Kingma & Ba, 2014);

Appendix B. Goodness of fit for SH300 and S&P500

In this section, we report the goodness of fit for SH300 and S&P500 data sets under ARCH, GARCH, EGARCH, and GJR respectively with different innovations (the normal, the Student’s t, the skewed Student’s t, and the generalized error distribution) in Table B.6 and Table B.7 respectively.

Table B.6. Out-of-sample goodness of fit on SH300 data. LL is the log likelihood, AIC is the Akaike information criterion, and BIC is the Bayesian information criterion.

Model	LL	AIC	BIC
ARCH-N(14)	4937.811	-9.906	-9.995
ARCH-t(11)	4837.193	-9.700	-9.778
ARCH-st(11)	4836.465	-9.701	-9.785
ARCH-G(10)	4835.353	-9.695	-9.767
GARCH-N(1,1)	4925.859	-9.858	-9.876
GARCH-t(1,1)	4824.549	-9.657	-9.681
GARCH-st(1,1)	4823.814	-9.658	-9.688
GARCH-G(1,1)	4817.683	-9.643	-9.667
EGARCH-N(1,1)	4925.239	-9.856	-9.874
EGARCH-t(1,1)	4821.969	-9.652	-9.676
EGARCH-st(1,1)	4821.278	-9.653	-9.683
EGARCH-G(1,1)	4815.950	-9.640	-9.664
GJR-N(1,1)	4925.644	-9.859	-9.883
GJR-t(1,1)	4819.234	-9.650	-9.686
GJR-st(2,1)	4818.685	-9.651	-9.693
GJR-G(1,1)	4817.034	-9.644	-9.674

Table B.7. Out-of-sample goodness of fit on S&P500 data. LL is the log likelihood, AIC is the Akaike information criterion, and BIC is the Bayesian information criterion.

Model	LL	AIC	BIC
ARCH-N(6)	3395.889	-6.806	-6.847
ARCH-t(6)	3350.526	-6.717	-6.764
ARCH-st(6)	3349.848	-6.718	-6.771
ARCH-G(6)	3337.130	-6.690	-6.738
GARCH-N(2,1)	3395.348	-6.799	-6.822
GARCH-t(2,1)	3349.800	-6.710	-6.739
GARCH-st(2,1)	3349.082	-6.710	-6.746
GARCH-G(2,1)	3336.435	-6.683	-6.713
EGARCH-N(2,1)	3398.249	-6.804	-6.828
EGARCH-t(2,1)	3351.507	-6.713	-6.743
EGARCH-st(2,1)	3350.735	-6.713	-6.749
EGARCH-G(2,1)	3337.884	-6.686	-6.715
GJR-N(1,1)	3331.354	-6.671	-6.694
GJR-t(1,1)	3299.961	-6.610	-6.640
GJR-st(1,1)	3299.785	-6.612	-6.647
GJR-G(1,1)	3289.422	-6.589	-6.618

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.