High-frequency CSI300 futures trading volume predicting through the neural network

Xiaojie Xu and Yun Zhang
North Carolina State University at Raleigh, Raleigh, North Carolina, USA

Abstract

Purpose – For policymakers and participants of financial markets, predictions of trading volumes of financial indices are important issues. This study aims to address such a prediction problem based on the CSI300 nearby futures by using high-frequency data recorded each minute from the launch date of the futures to roughly two years after constituent stocks of the futures all becoming shortable, a time period witnessing significantly increased trading activities.

Design/methodology/approach – In order to answer questions as follows, this study adopts the neural network for modeling the irregular trading volume series of the CSI300 nearby futures: are the research able to utilize the lags of the trading volume series to make predictions; if this is the case, how far can the predictions go and how accurate can the predictions be; can this research use predictive information from trading volumes of the CSI300 spot and first distant futures for improving prediction accuracy and what is the corresponding magnitude; how sophisticated is the model; and how robust are its predictions?

Findings – The results of this study show that a simple neural network model could be constructed with 10 hidden neurons to robustly predict the trading volume of the CSI300 nearby futures using 1–20 min ahead trading volume data. The model leads to the root mean square error of about 955 contracts. Utilizing additional predictive information from trading volumes of the CSI300 spot and first distant futures could further benefit prediction accuracy and the magnitude of improvements is about 1–2%. This benefit is particularly significant when the trading volume of the CSI300 nearby futures is close to be zero. Another benefit, at the cost of the model becoming slightly more sophisticated with more hidden neurons, is that predictions could be generated through 1–30 min ahead trading volume data.

Originality/value – The results of this study could be used for multiple purposes, including designing financial index trading systems and platforms, monitoring systematic financial risks and building financial index price forecasting.

Keywords CSI300 futures, Trading volume, Forecasting, Neural network

Paper type Research paper

1. Introduction

For policymakers and participants of financial markets, predictions of trading volumes of financial indices are important issues. This is because such predictions carry significant market implications for financial index prices and their movements (Wang *et al.*, 2013, 2019; Hou and Li, 2014; Sohn and Zhang, 2017; Susheng and Zhen, 2014; Yan and Hongbing, 2018; Ausloos *et al.*, 2020), which are an essential part of various decisioning processes with

purposes of generating trading profits and preventing trading losses. Considering the tremendous importance and need to monitor ever-changing financial markets that ultimately determine the safety and soundness of the financial and economic environment of different countries and regions, it is of particular interest to regulators and traders to understand the issue of financial trading in the high-frequency domain (Xu and Zhang, 2023a). To fulfill this mission, the literature has witnessed a great amount of effort in constructing different types of models for making predictions. These modeling techniques have included traditional regression types of (time-series) econometric models and machine learning models.

1.1 Traditional regression and time series models

Chen *et al.* (2011) have proposed a hierarchical model that has two different components, which combine an intraday approach and a daily approach, to predict the trading volumes of 30 DJIA stocks. Chen *et al.* (2011) have found that their proposed hierarchical method leads to higher prediction accuracy than any of the two individual approaches. Joseph *et al.* (2011) have used a simple linear regression model for predictions of abnormal activities of trading for 470 S&P 500 companies. Joseph *et al.* (2011) have determined that online search activities offer useful predictive information for the prediction horizon of one week. Brownlees *et al.* (2011) have compared a rolling average method and a multiplicative error model for prediction purposes of different exchange-traded fund volumes. Brownlees *et al.* (2011) have found that the multiplicative error model results in higher accuracy. Gharehchopogh *et al.* (2013) have applied a simple linear regression model for predicting the trading volume of the S&P 500 index by using predictive information from the price of the index. Ye *et al.* (2014) have compared static and dynamic versions of a volume-weighted approach for predicting SSE 50 stocks' intra-daily volumes. Ye *et al.* (2014) have suggested that the dynamic version leads to better predictions. Satish *et al.* (2014) have proposed combining an ARIMA model and a rolling average approach for predictions of trading volumes of 30 DJIA stocks. Bordino *et al.* (2014) have found that predictive information from Yahoo Finance could benefit predictions of trading volumes of NYSE and Nasdaq stocks on both daily and hourly frequency. Nasir *et al.* (2019) have used a hybrid framework of vector auto-regressive models and non-parametric methods for predictions of trading volumes of Bitcoin on a weekly basis through predictive information from Google searching activities. Nasir *et al.* (2019) have found that more searching activities are associated with higher trading volumes of Bitcoin. Kao *et al.* (2020) have combined a vector auto-regressive model with a smoothing approach for the purpose of assessing causality between trading volumes and financial returns.

1.2 Modern methods

Chen *et al.* (2016) have compared a state-space model based upon Kalman filtering with a rolling average approach and a multiplicative error method for predictions of trading volumes of many stocks from different exchanges. Chen *et al.* (2016) have determined that the state-space technique leads to higher accuracy for intraday predictions. Ma and Li (2021) have proposed multi-state Kalman filtering for the purpose of generating higher prediction accuracy as compared to two-state filtering for trading volumes of nearly a thousand stocks in the USA.

1.3 Machine learning and deep learning techniques

Kaastra and Boyd (1995) have explored the comparison between a neural network and an ARIMA model for predictions of trading volumes of different agricultural commodities' futures contracts. Kaastra and Boyd (1995) have determined that the neural network generates higher prediction accuracy. Alvim *et al.* (2010) have examined comparisons among

a partial least squares, a support vector machine and a naive no-change model for predicting the trading volumes of nine Bovespa stocks. Alvim *et al.* (2010) have found that the naive model leads to the lowest prediction accuracy. Oliveira *et al.* (2017) have evaluated the usefulness of a support vector machine for predicting trading volumes of Dow Jones and S&P 500 on a daily basis by using predictive information from micro-blogging. Lu *et al.* (2020b) have combined feature extracting functions of a CNN model and predicting functions of an LSTM model for predicting prices of stocks by using predictive information from trading volumes and historical prices. Yan and Yang (2021) have utilized the same predictive information set as that of Lu *et al.* (2020b) in predicting prices of stocks through encoder/decoder LSTM models. Zhao *et al.* (2021) have considered different machine learning models that include a support vector machine, a random forest and an LSTM, and a graph-based method for predicting trading volumes' movement patterns by using predictive information from prices of stocks. Zhao *et al.* (2021) have determined that the graph-based method leads to higher prediction accuracy. Separate from stock markets, Shen *et al.* (2021) have illustrated that an LSTM could be useful for predicting trading volumes of foreign business in different countries and regions. Zhang (2020) has demonstrated that a Levenberg–Marquardt trained neural network could be effectively utilized for predicting trading volumes of exports and imports.

1.4 Time series decomposition approaches
Lu *et al.* (2020a) have explored various machine-learning techniques to predict trading volumes and prices of carbon emission rights. Lu *et al.* (2020a) have paid special attention to the use of ensemble mode decomposition and data smoothing methods. Xie *et al.* (2020) have investigated the usefulness of decomposition approaches for predictions of trading volumes of electricity. For financial indices, Liu *et al.* (2022) have shed light on how to decompose trading activities of stocks into short- and long-run components in assessing potential extreme trading information. Chacón *et al.* (2020) have incorporated ensemble mode decomposition and data smoothing techniques into an LSTM for predictions of prices of stocks. Chacón *et al.* (2020) have determined that such a framework could benefit from improving prediction accuracy.

Regarding the case of the CSI300, recent work has generally focused on predictions of prices through the use of different time-series techniques (e.g. Wang and Chen, 2013; Xu, 2017, Xu, 2018, 2019b; Zhang and Sun, 2017; Huang *et al.*, 2018; Zhou *et al.*, 2019a) and machine learning approaches (e.g. Sun *et al.*, 2015; Yang and Cheng, 2015; Wang *et al.*, 2016; Lu and Li, 2017; Yao *et al.*, 2018; Ning, 2020; Long *et al.*, 2019; Zhou *et al.*, 2019b). Therefore, our present work targets at filling the research gap of trading volume predictions for the CSI300. Specifically, we address such a prediction problem based upon the CSI300 nearby futures by using high-frequency data recorded each minute from the launch date of the futures to roughly two years after constituent stocks of the futures all becoming shortable, a time period witnessing significantly increased trading activities.

The stock market was established in China in the early 1990s. Since then, it has undergone through dramatic developments with economic growth. However, until March 2005, there was no financial index designed for the purpose of reflecting the overall market status. This situation was resolved on 04/08/2005 when the CSI300 was launched. This financial index includes 300 stocks traded in Shanghai and Shenzhen exchanges and reflects 70% of the total market capitalization. For further financial developments, the futures of the CSI300 index was launched on 04/16/2010. Since then, it has turned out to be the most actively traded financial contract in China. Two pilot programs by the China Securities Regulatory Commission are worth noting. First, on 12/05/2011, the CSI300's underlying stocks available for shortable trading increased from 90 to 260. Second, on 01/31/2013, the CSI300's underlying stocks

available for shortable trading increased from 260 to 300. Such programs have contributed to elevated trading volumes of the CSI300 and the nearby futures have attracted the most trading activities (Xu, 2019b). For understanding more institutional backgrounds of the CSI300, one could refer to Yang *et al.* (2012), Hou and Li (2013), Xu (2017, 2018, 2019b) and Xu and Zhang (2021c, 2022b).

To perform our prediction exercise, we adopt the neural network (denoted as NN) for modeling the irregular trading volume series of the CSI300 nearby futures. The NN has been found in the literature to have great prediction potential for financial and economic applications in terms of time-series data (e.g. Yang *et al.*, 2008, 2010; Wang and Yang, 2010; Cabrera *et al.*, 2011; Zhang and Pan, 2014; Yang and Cheng, 2015; Kong and Zhu, 2018; Xu and Zhang, 2021b, 2022d). We concentrated on answering the research questions as follows: are we able to utilize the lags of the trading volume series to make predictions; if this is the case, how far can the predictions go and how accurate can the predictions be; can we use predictive information from trading volumes of the CSI300 spot and first distant futures for improving prediction accuracy and what is the corresponding magnitude; how sophisticated is the model; and how robust are its predictions? Our results show that we could construct a rather simple neural network model with 10 hidden neurons to robustly predict the trading volume of the CSI300 nearby futures using 1–20 min ahead trading volume data. The model leads to the root mean square error of about 955 contracts. Utilizing additional predictive information from trading volumes of the CSI300 spot and first distant futures could further benefit prediction accuracy and the magnitude of improvements is about 1–2%. This benefit is particularly significant when the trading volume of the CSI300 nearby futures is close to be zero. Another benefit, at the cost of the model becoming slightly more sophisticated with more hidden neurons, is that predictions could be generated through 1–30 min ahead trading volume data.

Our contributions to the literature are as follows. First, to our knowledge, the present work is the first one on predictions of trading volumes of the CSI300 nearby futures. Our results here would fill the research gap in terms of gaining understanding of the problem of trading volume predictions based on an important financial index. These would have implications from a practical standpoint for many economic agents, including policymakers, traders and investors and regulatory agencies. Specifically, the results would benefit from monitoring ever-changing trading activities for ensuring the safety and soundness of financial systems. Second, the current study is the first one that employs a powerful machine learning approach for the prediction purpose of the trading volume of the CSI300 nearby futures. Under the special and unique market structure, including a domestic individual investor having a high barrier to enter trading of the futures, as well as a foreign institutional investor with qualifications and relatively low participation ratios of institutional traders as compared to individual investors (Ng and Wu, 2007; Xu, 2017), we successfully illustrate that the NN could effectively make rather accurate and robust predictions of the trading volume, which shows the great potential of the NN under different market structures. Such results should be of practical use to financial indices traded in different countries that share a similar market structure with the CSI300, probably for a particular time period. Third, our analysis is the first one that adopts the high-frequency data recorded on a minute basis for financial trading volume predictions, although previous studies have investigated the issue for different financial indices using intraday trading data (Xu, 2018). A good understanding of high-frequency trading volume predictions could greatly benefit investors and policymakers in risk management and financial price index predictions as part of market evaluations.

The remainder of this study is organized as follows. Section 2 describes data used. Section 3 discusses models. Section 4 presents results. Section 5 provides conclusions.

2. Data
Our data are sourced from Wind Information Co., Ltd, which include trading volumes of the CSI300 spot and CSI300 futures [1]. The trading volumes are recorded on a minute basis, and for each trading day, the data span 9:16 a.m.–11:30 a.m. and 1:01 p.m.–3:15 p.m. The time period analyzed here ranges from 04/16/2010 (the date on which the futures was launched) to 11/14/2014. Thus, there are 299,970 observed trading volumes for each series. The data recorded on the 1-min basis not only reflect more trading activities than the data recorded on the 5- or 10-min basis but also maintain sufficient economic significance so that investigations of the one-minute data carry necessary importance to traders and policymakers (Xu, 2018). Visualization of the trading volumes of the CSI300 spot, CSI300 nearby futures and CSI300 first distant futures [2] is provided in Figure 1. We could see from Figure 1 that these three trading volume series show obvious chaotic and noised patterns. Figure 1 also reveals that the nearby futures contract is most actively traded and its trading volumes have been expanding during the time period considered here. Summary statistics of the three trading volumes series are provided in Table 1. We could observe that the trading volumes are leptokurtic and skewed positively.

3. Models
Two different types of nonlinear autoregressive neural network (denoted as ANN) models have been considered in this work. The first model is named a pure ANN or the CSI300 nearby futures own-lag only model. The model is denoted as follows:

$$y(t) = f(y(t-1), \ldots, y(t-d)), \tag{1}$$

where y is employed to reflect the trading volume of the CSI300 nearby futures, t is employed to reflect the time, d is employed to reflect the number of delays used by the model and f is employed to reflect the function form of the model. It is worth noting that the function form f is yet unknown in advance and the model estimated could be denoted as follows:

$$y(t) = \alpha_0 + \sum_{j=1}^{k} \alpha_j \phi \left(\sum_{i=1}^{d} \beta_{ij} y(t-i) + \beta_{0j} \right) + \varepsilon(t), \tag{2}$$

where k is employed to reflect the number of hidden layers used by the model with the transfer function being ϕ, β_{ij} is employed to reflect the parameter that is associated with the connection's weight between the input unit i and the hidden unit j, α_j is employed to reflect the connection's weight between the hidden unit j and the output unit, β_{0j} and α_0 are employed to reflect the constants that are associated with, respectively, the hidden unit j and the output unit and ε is employed to reflect the error item. The second model is an ANN that has exogenous inputs (ANN–X). The model is denoted as follows:

$$\begin{aligned} y(t) &= f(y(t-1), \ldots, y(t-d), x(t-1), \ldots, x(t-d)) \\ &= \alpha_0 + \sum_{j=1}^{k} \alpha_j \phi \left(\sum_{i=1}^{d} \left[\beta_{ij} y(t-i) + \beta'_{ij} x(t-i) \right] + \beta_{0j} \right) + \varepsilon(t), \end{aligned} \tag{3}$$

where x is employed to reflect the trading volume of the CSI300 spot alone or x is employed to reflect the trading volumes of the CSI300 spot and CSI300 first distant futures together, and β'_{ij} is employed to reflect the parameter that is associated with the connection's weight between the exogenous input unit i and the hidden unit j. The ANN–X model has included more predictive information as compared to the ANN model and thus could explore the potential
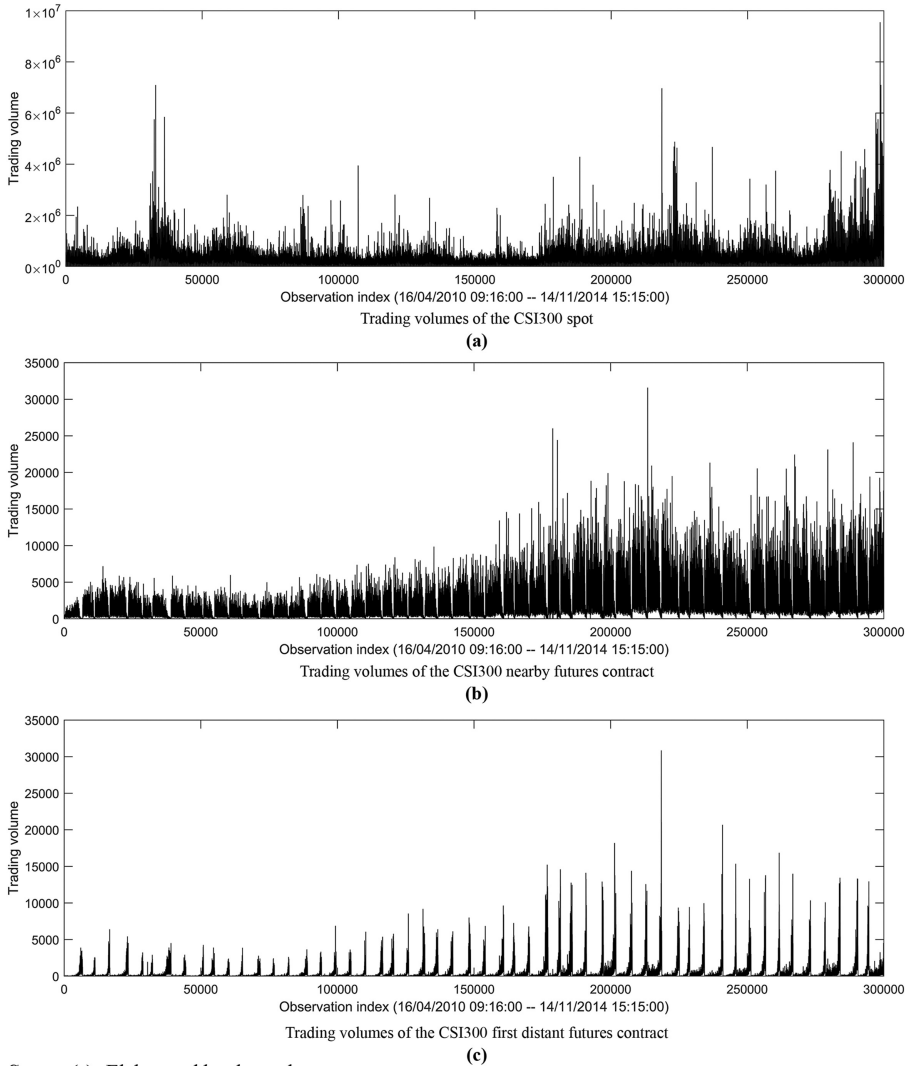
(a)

(b)

(c)

Source(s): Elaborated by the authors

	CSI300 spot	CSI300 nearby futures contract	CSI300 first distant futures contract
Mean	275,070	1,514	292
Median	222,010	1,101	54
Maximum	9,550,879	31,586	30,849
Minimum	0	0	0
Standard deviation	230,860	1,473	718
Skewness	4.945	2.691	6.327
Kurtosis	79.902	16.450	77.031
Jarque–Bera p value	<0.001	<0.001	<0.001

Source(s): Elaborated by the authors

usefulness of the additional predictive information for improving prediction accuracy (Xu, 2019a, 2020).

We make use of the ANN models based upon a two-layer feed-forward network whose hidden layer adopts a transfer function in the form of a logistic sigmoid function as follows:

$$\phi(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

and whose output layer adopts a linear function. It should be noted that the output $y(t)$ would be fed back via the delays to inputs of the neural network and the training process would take on the form of open loops in order to achieve efficiency purposes. In open loops, the true output will be utilized instead of the estimated output. To be more specific, adopting the open loop would help ensure inputs to neural networks being more accurate and that resultant neural networks have pure feed-forward architectures.

In terms of numbers of hidden neurons, we have tested 5, 10, 15, 25, 35 and 50. Regarding the delays, we have examined 1, 2, 5, 10, 20 and 30. As a result, 36 testing pairs are considered. For model estimations, we have segmented the trading volume data by using 70% for training, 15% for validation and 15% for testing. For the consideration of robustness analysis, we have also examined the following alternative data segmentation ratios by reserving 15% of the trading volume series for testing: 60% for training and 25% for validation, 65% for training and 20% for validation, 75% for training and 10% for validation and 80% for training and 5% for validation.

One could employ different algorithms for training a machine learning model. For our case, we explored the following three algorithms: the Levenberg–Marquardt (Levenberg, 1944; Marquardt, 1963) algorithm, the scaled conjugate gradient (Møller, 1993) algorithm and the Bayesian regularization (MacKay, 1992; Foresee and Hagan, 1997) algorithm. These three algorithms have been demonstrated by previous studies in terms of their success in achieving relatively good accuracy under various circumstances (e.g. Doan and Liong, 2004; Kayri, 2016; Khan *et al.*, 2019; Selvamuthu *et al.*, 2019; Xu and Zhang, 2021a, d, 2022a, c, 2023b). Baghirli (2015) and Al Bataineh and Kaur (2018) have carried out targeted studies comparing these three algorithms. Table 2 contains the specification of each ANN and ANN–X model setting examined in the present work. Figure 2 visualizes the architecture of the final neural networks constructed in this study.

Algorithm	Levenberg–Marquardt scaled conjugate gradient Bayesian regularization
Delay	1
	2
	5
	10
	20
	30
Hidden neuron	5
	10
	15
	25
	35
	50
Training vs validation vs testing	60% vs 25% vs 15%
	65% vs 20% vs 15%
	70% vs 15% vs 15%
	75% vs 10% vs 15%
	80% vs 5% vs 15%

Table 2.
Explored ANN and ANN–X model settings for the trading volume prediction of the CSI300 nearby futures

Source(s): Elaborated by the authors
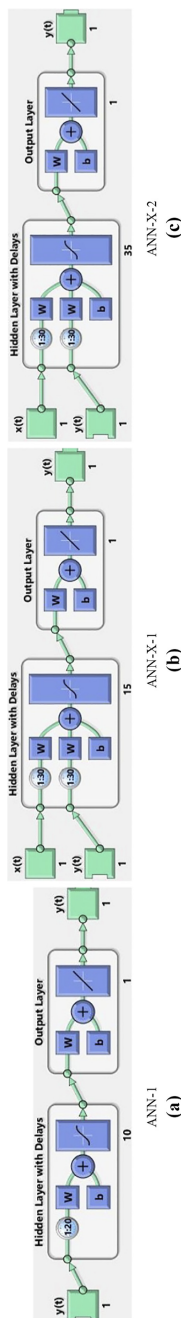
Note(s): (a) The ANN-1 model adopts 20 delays and 10 hidden neurons. (b) The ANN–X-1 model adopts 30 delays and 15 hidden neurons, where X stands for the trading volume of the CSI300 spot. (c) The ANN–X-2 model adopts 30 delays and 35 hidden neurons, where X stands for the trading volumes of the CSI300 spot and CSI300 first distant futures. Mode details of the ANN-1, ANN–X-1, and ANN–X-2 models have been discussed in Section 4

Source(s): Elaborated by the authors

3.1 Levenberg–Marquardt algorithm
The Levenberg–Marquardt algorithm targets the realization of approximating the second-order training speed. By doing this, the algorithm could avoid computing H – the Hessian matrix and thus accelerate training speed (Paluszek and Thomas, 2020). For this algorithm, using a system with weights denoted as w_1 and w_2 as an example, the approximation is as follows:

$$H \approx J^T J, \tag{5}$$

where

$$J = \left[\dfrac{\partial E}{\partial w_1} \quad \dfrac{\partial E}{\partial w_2} \right] \tag{6}$$

for a non-linear function denoted as $E(\cdot)$, which includes information on the sum square error and whose

$$H = \left[\begin{array}{cc} \dfrac{\partial^2 E}{\partial w_1^2} & \dfrac{\partial^2 E}{\partial w_1 \partial w_2} \\[2mm] \dfrac{\partial^2 E}{\partial w_2 \partial w_1} & \dfrac{\partial^2 E}{\partial w_2^2} \end{array} \right]. \tag{7}$$

The gradient would be reflected as follows:

$$g = J^T e, \tag{8}$$

where e is employed to denote a vector of errors. In order to make updates to the weights and the biases, the rule as follows is to be adopted:

$$w_{k+1} = w_k - \left[J^T J + \mu I \right]^{-1} J^T e, \tag{9}$$

where w is employed to denote the weight vector, k is employed to denote the index of the iteration during the model training process, I is employed to denote the identity matrix and μ is employed to denote the combination coefficient (noting that μ is always positive). For the case that $\mu = 0$, this algorithm would turn out to be similar to the Newton approach. When μ is large, this algorithm would turn to be gradient descent based on small step sizes. When a successful step is reached, μ would be decreased due to reduced requirements for the fast gradient descent. The Levenberg–Marquardt algorithm inhabits good properties of steepest-descent types of algorithms and Gauss-Newton types of techniques while avoiding some of their limitations. To be more specific, it would efficiently address the concern of slow convergence (Hagan and Menhaj, 1994).

3.2 Scaled conjugate gradient algorithm
Adjustments of weights are conducted along the steepest descent in a backpropagation algorithm due to the need to realize fast decreases of a given performance function in that direction. But this does not always lead to the fastest convergence and model training. A conjugate gradient algorithm would carry out searching along the conjugate direction, which would generally result in quicker convergence as compared to the steepest descent method. A learning rate is often adopted by different algorithms for deciding the length of an updated weight step size. When it comes to the case of a conjugate gradient algorithm, the step size would be modified in the processes of iterations. Therefore, searching is carried out

along the conjugate gradient direction with the purpose of determining a step size to realize
the reduction of a given performance function. In addition, as line searching by a conjugate
gradient algorithm could be time-consuming, the scaled conjugate gradient algorithm would
be adopted to improve the training speed. The scaled conjugate gradient algorithm is
generally faster as compared to a Levenberg–Marquardt backpropagation-based algorithm.

3.3 Bayesian regularization algorithm

Cross-validation could be lengthy. But it is not required by a Bayesian regularized NN.
Essentially, the Bayesian regularization would convert nonlinear types of regressions to
statistical issues in the matter of ridge types of regressions and the algorithm would explore
the weights' probabilistic nature related to the underlying data under investigation. The
chance of overfitting would, however, increase dramatically when additional hidden layers of
neurons are used. Thus, in a Bayesian regularization algorithm, unreasonable sophisticated
models would be penalized with the corresponding linkage weights pushed to 0. As a result,
the NN would concentrate on non-trivial weights. Naturally, some parameters would
converge to constant values as the NN grows. A Bayesian regularized NN would generally be
more parsimonious as compared to a basic backpropagation network. It would also help
reduce the probability of model overfitting due to the underlying data noises.

4. Results

4.1 ANN

We show, in Figure 3, root mean square errors (RMSEs) generated by the ANN models that
are trained through the Levenberg–Marquardt algorithm and based upon the trading
volume's own lags of the CSI300 nearby futures contract. With the consideration of the need
to balance model prediction accuracy and the stabilities of model performance across the
three phases of training, validation and testing, we make the selection of the ANN model that
uses 10 hidden neurons and 20 delays. We denote this model as ANN-1 and the corresponding
performance in terms of RMSEs is 955.67 for training, 957.94 for validation and 955.32 for
testing. The summary of the ANN-1 model is included in Table 3. A similar analysis has been
conducted for ANN models trained via the other two algorithms (results available upon
request). Combining all analysis results, we present, in Figure 4, the prediction performance
of the top three candidate models and determine that the ANN-1 model is still the optimal
choice for balancing model prediction accuracy and stabilities of model performance. Please
note that in Figure 4, RMSEs are 956.17 for training, 975.75 for validation and 922.75 for
testing for the ANN model trained via the Levenberg–Marquardt algorithm, which is based
upon 15 hidden neurons and 30 delays. And, RMSEs are 951.77 for training and 949.91 for
testing for the ANN model trained via the Bayesian regularization algorithm, which
consumes much more time than the ANN-1 model. For the remaining of this work, our focus
would be the Levenberg–Marquardt algorithm for model training.

We present, in Figure 5, four variations of the ANN-1 model associated with different data
segmentation ratios utilized during model building. Specifically, the ANN-1 model uses 70% of
the data for training, 15% of the data for validation and 15% of the data for testing. The four
variations maintain 15% of the data for testing but consider the following ratios for training and
validation purposes: 60% for training and 25% for validation, 65% for training and 20% for
validation, 75% for training and 10% for validation and 80% for training and 5% for validation.
According to the results shown in Figure 5, one would be able to see that the ratio of 70% for
training, 15% for validation and 15% for testing results in the most stable model performance
across the three phases. Thus, for the remaining of this work, our focus would be this data
segmentation ratio. Through reserving 15% of the trading volume data for the testing purpose,
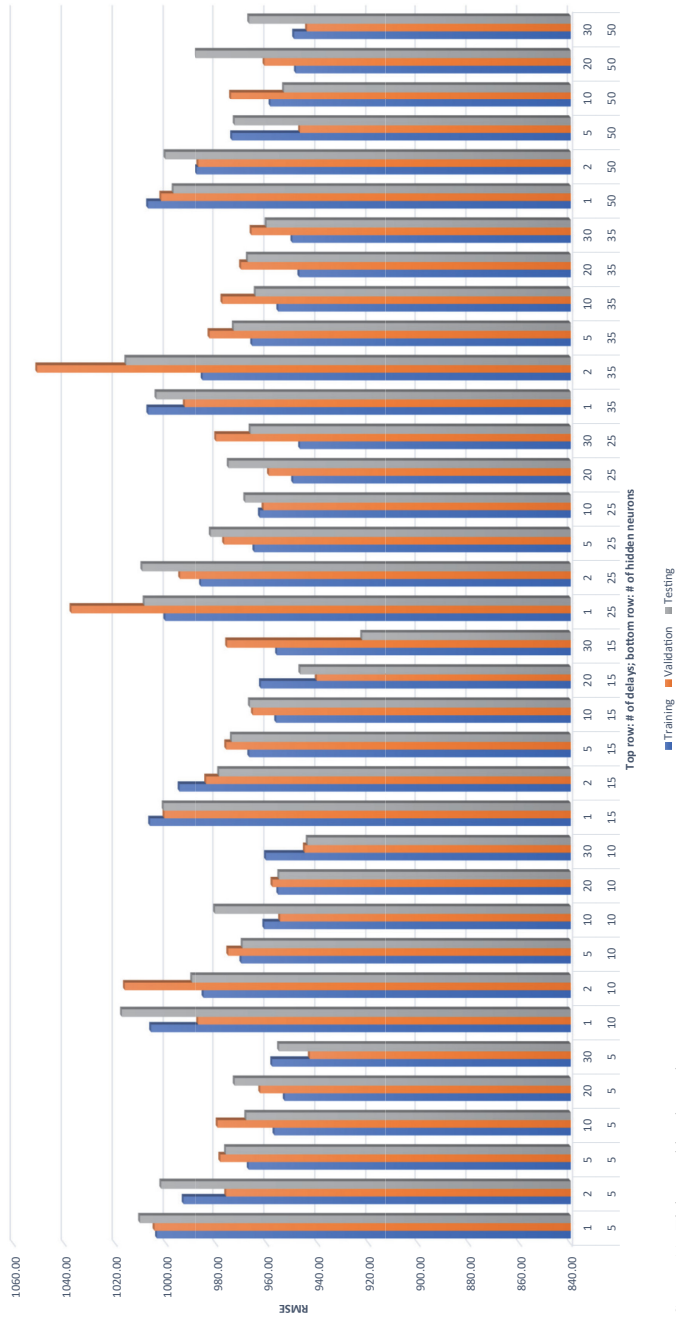
Figure 3.
Performance of ANN
models based upon the
Levenberg–Marquardt
algorithm in terms
of RMSEs

Source(s): Elaborated by the authors

	ANN-1	ANN-X-1	ANN-X-2	AR	ARIMA	AR-X-1	ARIMA-X-1	AR-X-2	ARIMA-X-2
Algorithm		Levenberg–Marquardt							
Delay	20	30	30						
Hidden neuron	10	15	35						
RMSE Training	955.67	949.37	942.44	1155.54	1149.02	1150.67	1136.52	1141.96	1132.73
Validation	957.94	946.50	934.04	2158.31	2143.72	2146.59	2132.80	2137.05	2126.07
Testing	955.32	935.85	940.20	2133.02	2115.28	2118.96	2101.95	2109.03	2099.15

Source(s): Elaborated by the authors

Table 3.
Prediction
performance
comparisons

we have also conducted additional analysis by executing the model based upon each of the five different segmentation ratios for training and validation for one hundred times and comparing means of the validation results. It turns out that the ratios of 60% for training and 25% for validation, 65% for training and 20% validation, 70% for training and 15% for validation, 75% for training and 10% for validation and 80% for training and 5% for validation lead to means of validation RMSEs of 958.59, 953.67, 952.82, 970.23 and 998.45, respectively, which support the choice of the ratio of 70% for training, 15% for validation and 15% for testing.

Visualization of ANN-1's predictions is shown in Figure 6 and visualization of ANN-1's prediction errors is shown in Figure 7. In Figure 6, the dark solid 45-degree line associated with "perfect prediction" stands for the situation for which a point on this line indicates no prediction error. Considering that ANN-1's prediction performance is rather stable, Figures 6 and 7 have combined visualization results for the training, validation and testing data because separating the plots for different phases almost does not create different visualization for the three sub-samples. For the remaining of this work, we would follow this practice for ANN–X models as well.
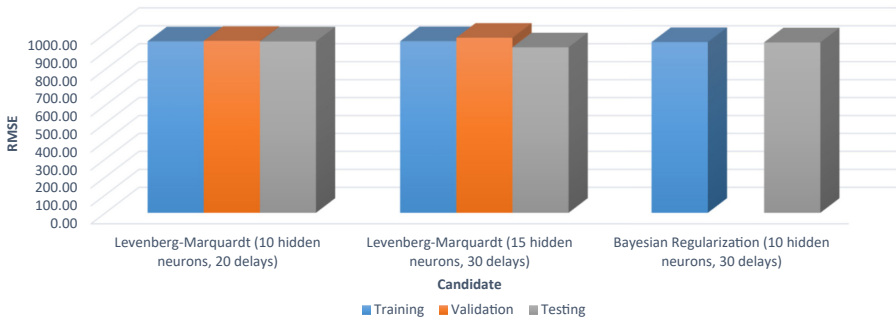
Figure 4.
Performance of ANN models in terms of RMSEs: the top three candidates

Source(s): Elaborated by the authors
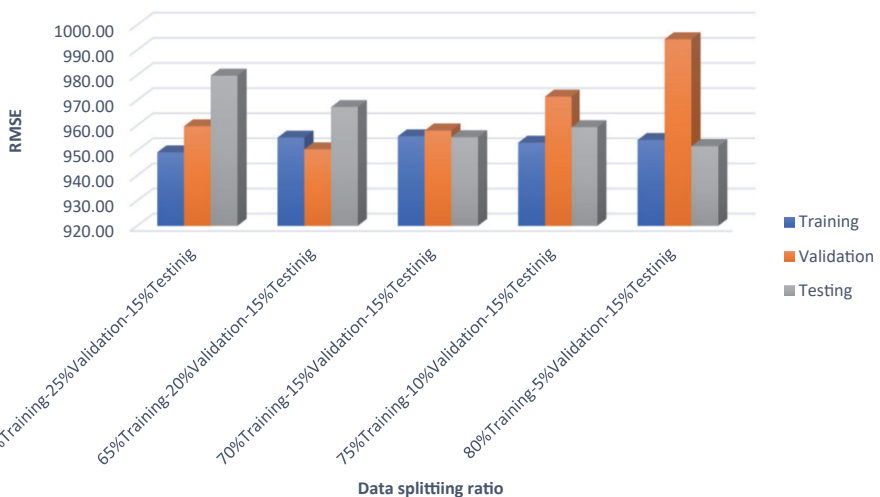
Figure 5.
Performance of ANN-1 variations in terms of RMSEs: different data segmentation ratios

Source(s): Elaborated by the authors
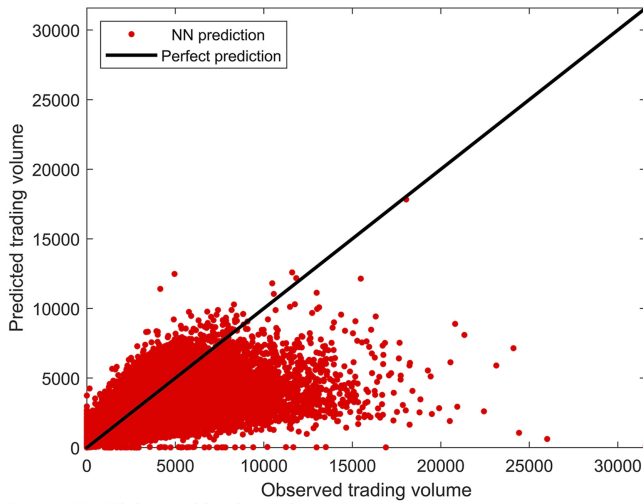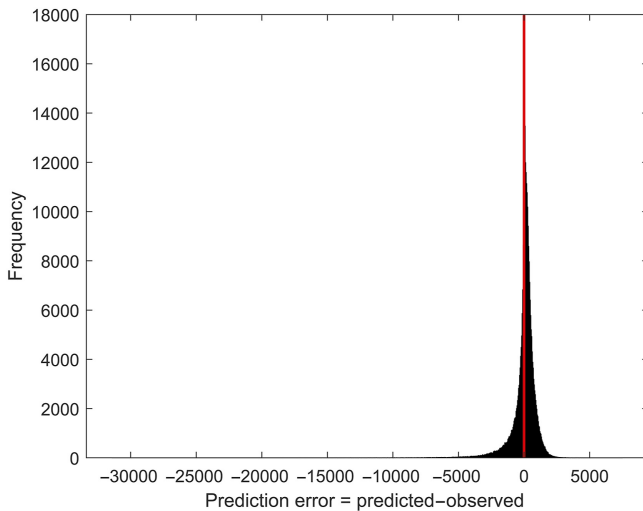
Source(s): Elaborated by the authors

Source(s): Elaborated by the authors

Based upon the analysis of the ANN-1 model, we could observe that the trading volume of the CSI300 nearby futures contract would be able to be predicted with 1–20 min ahead trading volume data, considering that the ANN-1 model is based upon 20 delays and the 1-min data are utilized for model building, through a relatively low complex model that has 10 hidden neurons [3] based on its own lags, leading to rather robust prediction performance with the RMSE being about 955 contracts across the three phases. Prediction errors are concentrated around 0 with a high peak in frequency as visualized in Figure 7. Several trading volumes of the CSI300 nearby futures, including the largest trading volume that is observed, however,

show prediction results of 0 as visualized in Figure 6. This problem is somewhat remediated by further including the trading volumes of the CSI300 spot and CSI300 first distant futures in the ANN–X models that we turn to next.

4.2 ANN–X

Similar to the analysis based on the ANN models, Figure 8 reports RMSEs for the ANN–X models when the trading volume of the CSI300 spot is further included as part of model training. We test the trading volume of the CSI300 spot prior to the trading volume of the CSI300 first distant futures due to the closer relation between the CSI300 nearby futures and CSI300 spot as compared to that between the CSI300 nearby futures and CSI300 first distant futures (Xu, 2019b). Balancing model prediction accuracy and model performance stabilities across different phases, we make the selection of the ANN–X model that has 15 hidden neurons and 30 delays, denoted as ANN–X-1. This model results in RMSEs of 949.37 for training, 946.50 for validation and 935.85 for testing. The summary of the ANN–X-1 model is included in Table 3. Predictions from the ANN–X-1 model are reported in Figure 9 and corresponding prediction errors are reported in Figure 10.

We finally include the trading volumes of both the CSI300 spot and CSI300 first distant futures for model training and report RMSEs for ANN–X models in Figure 11. Again, balancing model prediction accuracy and model performance stabilities across the three phases, we make the selection of the ANN–X model that has 35 hidden neurons and 30 delays, denoted as ANN–X-2. It results in RMSEs of 942.44 for training, 934.04 for validation and 940.20 for testing. The summary of the ANN–X-2 model is included in Table 3. Predictions from the ANN–X-2 model are visualized in Figure 12 and corresponding prediction errors are visualized in Figure 13.

In Figure 14, we make comparisons of model performance based on the ANN-1, ANN–X-1 and ANN–X-2 models. It could be observed that including the trading volumes of the CSI300 spot and CSI300 first distant futures could help improve prediction accuracy by a robust modest magnitude of about 1–2%. This, however, significantly helps some near-zero predictions of the trading volumes via the own-lag-only model, i.e. the ANN-1 model, as can be observed by making comparisons of prediction results shown in Figures 6, 9, and 12. To be more specific, one should be able to observe, in Figure 6, that there exist certain amounts of data points that show predicted trading volumes of zero or near zero while their associated observed trading volumes are not zero. These data points could be visually located in Figure 6, whose associated vertical axis (i.e. the predicted trading volume of the CSI300 nearby futures contract) values are zero or near zero while the corresponding horizontal axis (i.e. the observed trading volume of the CSI300 nearby futures contract) values are not zero. When we turn attention to results in Figures 9 and 12, we would observe that such data points have been largely eliminated, particularly, for the results shown in Figure 12. Another benefit is that the prediction of the trading volume of the CSI300 nearby futures would be generated via 1–30 min ahead trading volume data with the incorporation of the additional series, considering that the ANN–X-1 and ANN–X-2 models are based upon 30 delays and the 1-min data are employed for model building, although the complexity of the models would slightly increase as additional hidden neurons would be required.

4.3 Subperiod analysis

To test whether prediction accuracy would be affected by the number of shortable stocks in the CSI300, we run the models, i.e. ANN-1, ANN–X-1 and ANN–X-2, on three subperiods with different numbers of shortable stocks and compare the results. The first subperiod is April 16, 2010–December 4, 2011, during which 90 stocks in the CSI300 are shortable. The second subperiod is December 5, 2011–January 30, 2013, during which 260 stocks are shortable. The third subperiod is January 31, 2013–November 14, 2014, during which all 300 stocks are shortable. The results are presented in Figure 15, together with those for the whole sample
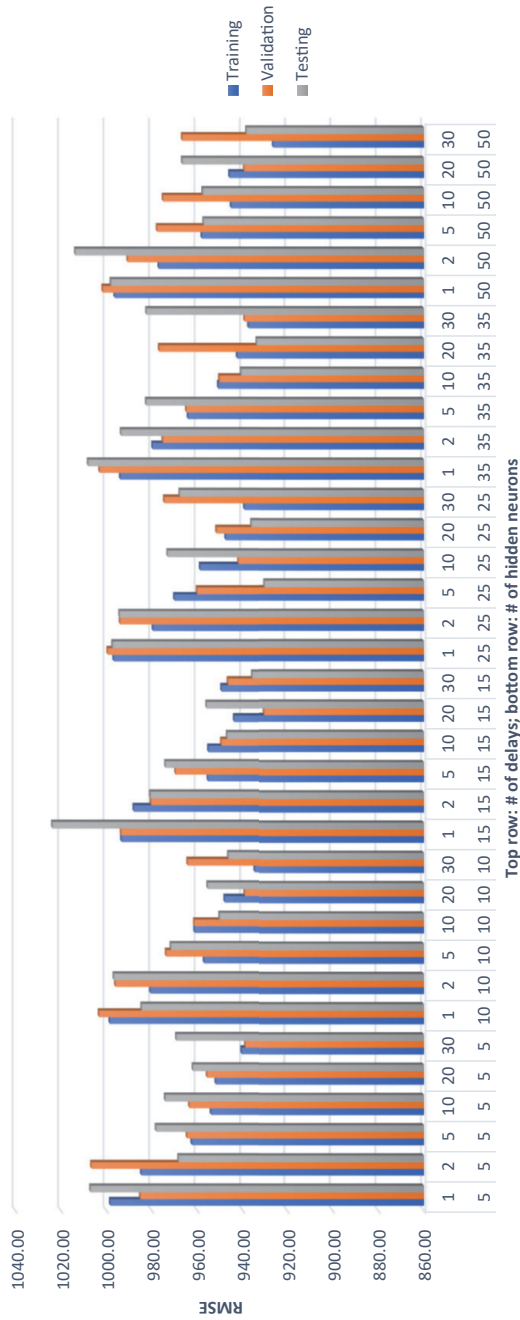
Source(s): Elaborated by the authors

Figure 8.
ANN–X models
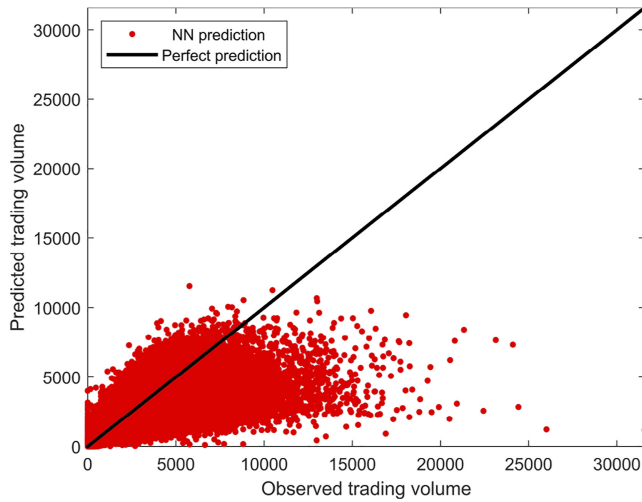(X = the trading
volume of the CSI300
spot): RMSEs

Figure 9.
ANN–X-1 predictions

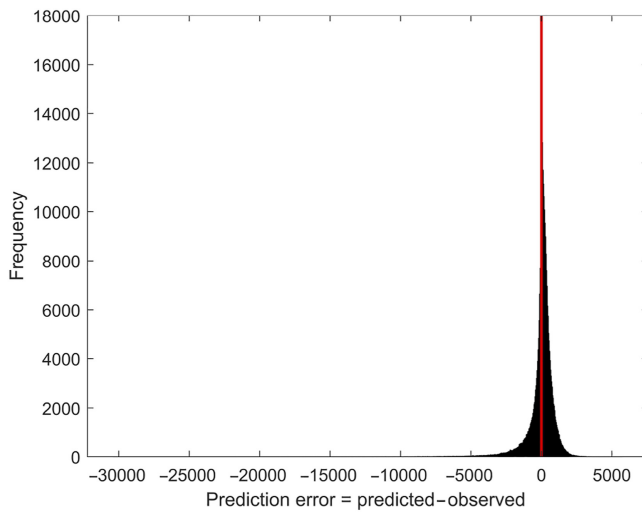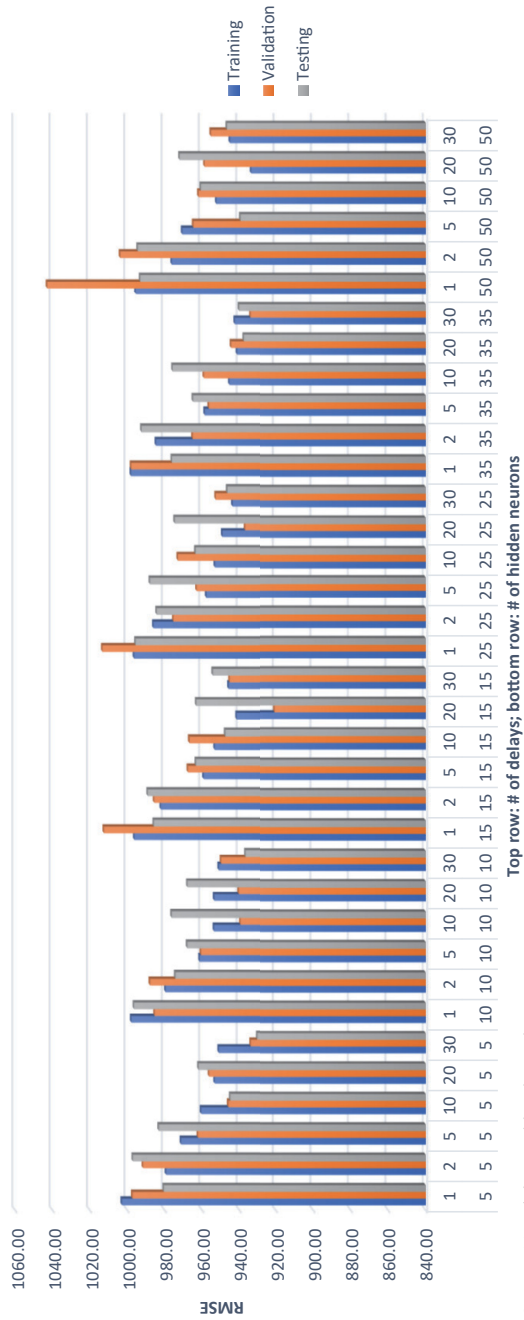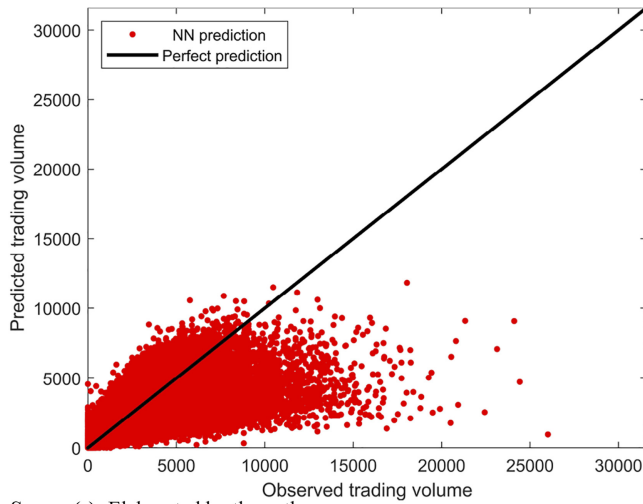**Source(s):** Elaborated by the authors

Figure 10.
ANN–X-1 prediction
errors

Source(s): Elaborated by the authors

from April 16, 2010, to November 14, 2014. We could observe from Figure 15 that the trading volume of the CSI300 nearby futures of the first subperiod is most accurately predicted, followed by the second subperiod and then the third subperiod. This result is intuitive because with more stocks becoming shortable, the trading becomes more volatile and harder to predict. From Figure 15, we still observe that incorporating the spot and first distant futures improves predictions for the three subperiods. Specifically, this can be seen when comparing the result of ANN-1 with those of ANN–X-1 and ANN–X-2 for a given subperiod and subsample.

Source(s): Elaborated by the authors

Figure 11.
ANN–X models
(X = trading volumes
of the CSI300 spot and
CSI300 first distant
futures): RMSEs

Figure 12.
ANN–X-2 predictions

Source(s): Elaborated by the authors
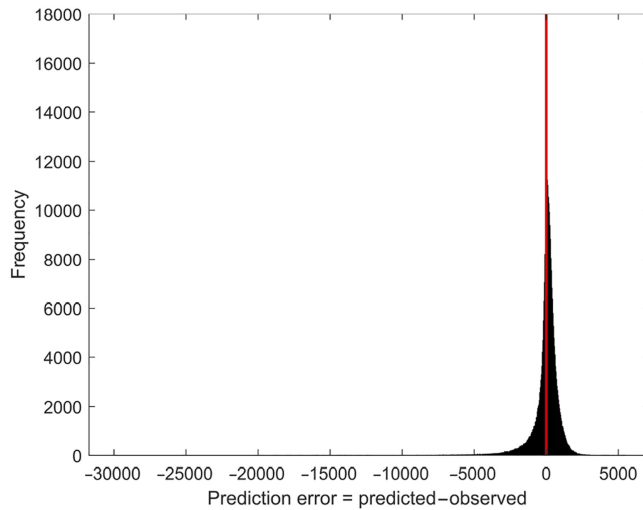
Figure 13.
ANN–X-2 prediction
errors

Source(s): Elaborated by the authors

4.4 Benchmark analysis

We have performed benchmark analysis through comparisons of the ANN-1 model with the linear autoregressive (denoted as AR) model and the linear autoregressive integrated moving average (denoted as ARIMA) model. Considering that the ANN–X-1 and ANN–X-2 models include additional predictive information as compared to the AR and ARIMA models and the performance of ANN–X-1 and ANN–X-2 has been shown to be better than that of the ANN-1 model, we have not benchmarked them against the AR or ARIMA model. In determining the lag of the AR model, the Bayesian information criterion (Schwarz, 1978) has been employed.

The structure of the ARIMA model has also been determined through the Bayesian information criterion (Schwarz, 1978). We have applied the modified Diebold-Mariano (Diebold and Mariano, 2002) test (Harvey *et al.*, 1997) for the purpose of making comparisons of model performance. The modified test mitigates some shortcomings of the original test, particularly the potential over-sized issues. The modified test is based upon d_t as follows:

$$d_t = \left(\text{error}_t^{M_1} \right)^2 - \left(\text{error}_t^{M_2} \right)^2, \tag{10}$$

where $\text{error}_t^{M_1}$ and $\text{error}_t^{M_2}$ are employed to denote two error terms at time t that are generated based upon model M_1 and model M_2, respectively. Here, we would denote AR or ARIMA as model M_1 and ANN-1 as model M_2. The test statistic for comparing model performance is denoted as MDM as follows:

$$MDM = \left[\frac{T + 1 - 2h + T^{-1}h(h-1)}{T} \right]^{1/2} \left[T^{-1} \left(\gamma_0 + 2 \sum_{k=1}^{h-1} \gamma_k \right) \right]^{-1/2} \bar{d}, \tag{11}$$

where T is employed to denote the length of the time period of the testing phase, h is employed to denote the prediction horizon ($h = 1$ for our application), $\bar{d}$ is employed to denote the sample average of d_t,

$$\gamma_0 = T^{-1} \sum_{t=1}^{T} \left(d_t - \bar{d} \right)^2 \tag{12}$$

is employed to denote the variance of d_t, and

$$\gamma_k = T^{-1} \sum_{t=k+1}^{T} \left(d_t - \bar{d} \right) \left(d_{t-k} - \bar{d} \right) \tag{13}$$

is employed to denote the kth auto-covariance of d_t for $k = 1, \ldots, h - 1$ and $h \geq 2$. Under the null that two models being compared result in equal mean squared errors, the MDM test would follow the t – distribution whose degrees of freedom is $T - 1$. We report RMSEs stemming from the AR model and the ARIMA model in Table 3. We have found that the p values of the MDM tests are below 0.001. This result suggests that prediction accuracy stemming from the ANN-1 model is statistically significantly better than prediction accuracy
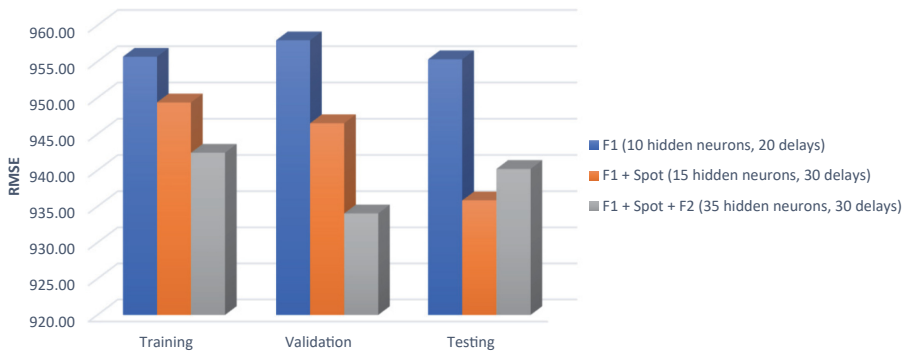
Source(s): Elaborated by the authors
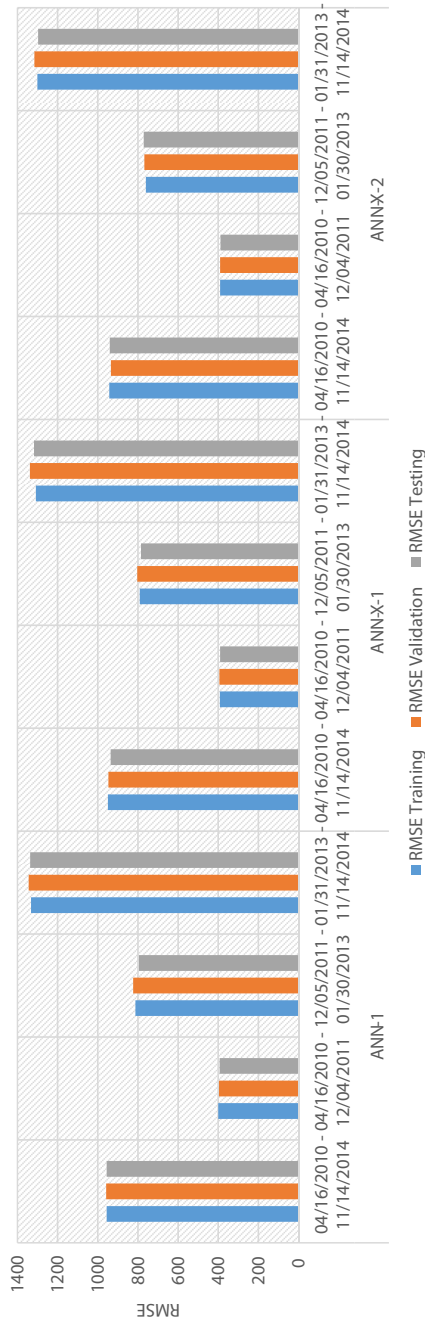
Source(s): Elaborated by the authors

Figure 15.
Subperiod analysis

stemming from the AR model and the ARIMA model. We have also conducted comparisons of performance based on the superior predictive ability (SPA) test (Hansen, 2005) as a robustness check and found that this test determines that the performance of the ANN-1 model is statistically significantly better than that of the AR model and the ARIMA model as well. We note that using the Akaike information criterion (Akaike, 1974) for determining the lag of the AR model and the structure of the ARIMA model does not affect this conclusion. It should be mentioned here that a certain model that is not performing as well as compared to another model would not necessarily mean that the particular model would not be able to contribute to prediction results. Many previous studies on prediction combinations actually have targeted at constructing different weights for different models' predictions with the purpose of potentially improving prediction accuracy. One interesting research field of prediction combinations is combining linear models and nonlinear models. Previous research, such as Stock and Watson (1998) and Blake and Kapetanios (1999), would have offered good examples in this research area. Hansen *et al.* (2011) have introduced the concept of the model confidence set (MCS), which is a useful technique to select optimal models with a given level of confidence.

Following the same idea of comparing the ANN-1 model with the AR and ARIMA models, we have also compared the performance of the ANN–X-1 model with that of the AR–X-1 and ARIMA–X-1 models, where X refers to the trading volume of the CSI300 spot, and performance of the ANN–X-2 model with that of the AR–X-2 and ARIMA–X-2 models, where X refers to the trading volumes of the CSI300 spot and first distance futures. The RMSEs based upon the AR–X-1, ARIMA–X-1, AR–X-2 and ARIMA–X-2 models are reported in Table 3. We have found that the p values of the *MDM* tests are below 0.001 for comparisons between the ANN–X-1 model and the AR–X-1 and ARIMA–X-1 models. This result suggests that the performance of ANN–X-1 is statistically significantly better than that of AR–X-1 and ARIMA–X-1. Similarly, we have found that the p values of the *MDM* tests are below 0.001 for comparisons between the ANN–X-2 model and the AR–X-2 and ARIMA–X-2 models. This result suggests that the performance of ANN–X-2 is statistically significantly better than that of AR–X-2 and ARIMA–X-2. As a robustness check, we have also conducted comparisons of performance based on the SPA test (Hansen, 2005) and found that it still holds that performance of ANN–X-1 is statistically significantly better than that of AR–X-1 and ARIMA–X-1 and performance of ANN–X-2 is statistically significantly better than that of AR–X-2 and ARIMA–X-2.

5. Conclusion
For policymakers and participants of financial markets, predictions of trading volumes of financial indices are important issues. In this present work, we address such a prediction problem based on the CSI300 nearby futures by using high-frequency data recorded on a minute basis, which has never been explored in previous studies. We adopt the neural network for modeling the irregular trading volume series and have key empirical findings as follows. Our results show that we could construct a rather simple neural network model, trained via the Levenberg–Marquardt (Levenberg, 1944; Marquardt, 1963) algorithm, with 10 hidden neurons to robustly predict the trading volume of the CSI300 nearby futures using one to twenty minutes ahead trading volume data. The model robustly leads to the root mean square error of about 955 contracts across the three phases of training, validation and testing. Utilizing additional predictive information from trading volumes of the CSI300 spot and first distant futures could further benefit prediction accuracy and the magnitude of improvements is about 1–2%. This benefit is particularly significant when the trading volume of the CSI300 nearby futures is close to be zero. Another benefit, at the cost of the model becoming slightly more sophisticated with more hidden neurons, is that predictions could be generated through 1–30 min ahead trading volume data as the corresponding neural networks would use 30

delays and 15 or 35 neurons, which also are trained via the Levenberg–Marquardt (Levenberg, 1944; Marquardt, 1963) algorithm. Our results could be used for multiple purposes, including designing financial index trading systems and platforms in terms of ongoing evaluating system/platform limits for processing trading activities, monitoring systematic financial risks in terms of ongoing detecting possible abnormal trading activities and building financial index price forecasting as suggested in the literature that one might make use of predictive information from the trading volume for helping improve the prediction accuracy of financial index prices. Our results here would be useful to policymakers from different countries for the purpose of designing another financial index or reforming an existing financial index. To be more specific, gaining good understanding of the trends of financial trading volumes would help the planning of a relatively new financial index from a thin market to a liquid and mature market. Although our present work focuses on relatively fundamental neural network models, developments in the machine learning field suggest that there exist more advanced models, such as the convolutional neural network and long short-term memory neural network, which have been seen in the literature for financial predictions. Explorations of more advanced models should be a worthwhile avenue for future studies on predicting financial trading volumes, including that of the CSI300 futures.

Notes

1. It is possible that different platforms could generate slightly different trading volume data for each minute. It is worth noting that trading volumes of the CSI300 spot are always 0 from 9:16 a.m. to 9:29 a.m. on a trading day.

2. When different futures contracts are being investigated, the contract that has the closest settlement date is named the nearby contract. The first distant contract is the contract which settles right after the nearby contract.

3. We state "a relatively low complex model" from our empirical judgment that the ANN-1 model with 10 hidden neurons is not so complex for our case.

References

Akaike, H. (1974), "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, Vol. 19 No. 6, pp. 716-723.

Al Bataineh, A. and Kaur, D. (2018), "A comparative study of different curve fitting algorithms in artificial neural network using housing dataset", *NAECON 2018-IEEE National Aerospace and Electronics Conference*, pp. 174-178, IEEE, doi: 10.1109/NAECON.2018.8556738.

Alvim, L., dos Santos, C.N. and Milidiu, R.L. (2010), "Daily volume forecasting using high frequency predictors", *Proceedings of the 10th IASTED International Conference*, p. 248.

Ausloos, M., Zhang, Y. and Dhesi, G. (2020), "Stock index futures trading impact on spot price volatility. the CSI 300 studied with a Tgarch model", *Expert Systems with Applications*, Vol. 160 No. 1, 113688, doi: 10.1016/j.eswa.2020.113688.

Baghirli, O. (2015), "Comparison of Lavenberg-Marquardt, scaled conjugate gradient and Bayesian regularization backpropagation algorithms for multistep ahead wind speed forecasting using multilayer perceptron feedforward neural network".

Blake, A. and Kapetanios, G. (1999), "Forecast combination and leading indicators: combining artificial neural network and autoregressive forecasts", Manuscript, National Institute of Economic and Social Research.

Bordino, I, Kourtellis, N., Laptev, N. and Billawala, Y. (2014), "Stock trade volume prediction with yahoo finance user browsing behavior", *2014 IEEE 30th International Conference on Data Engineering*, pp. 1168-1173, IEEE, doi: 10.1109/ICDE.2014.6816733.

Brownlees, C.T., Cipollini, F. and Gallo, G.M. (2011), "Intra-daily volume modeling and prediction for algorithmic trading", *Journal of Financial Econometrics*, Vol. 9 No. 3, pp. 489-518, doi: 10.1093/jjfinec/nbq024.

Cabrera, J., Wang, T. and Yang, J. (2011), "Linear and nonlinear predictablity of international securitized real estate returns: a reality check", *Journal of Real Estate Research*, Vol. 33 No. 4, pp. 565-594, doi: 10.1080/10835547.2011.12091317.

Chacón, H.D., Kesici, E. and Najafirad, P. (2020), "Improving financial time series prediction accuracy using ensemble empirical mode decomposition and recurrent neural networks", *IEEE Access*, Vol. 8, pp. 117133-117145, doi: 10.1109/ACCESS.2020.2996981.

Chen, S., Chen, R., Ardell, G. and Lin, B. (2011), "End-of-day stock trading volume prediction with a two-component hierarchical model", *The Journal of Trading*, Vol. 6 No. 3, pp. 61-68, doi: 10.3905/jot.2011.6.3.061.

Chen, R., Feng, Y. and Palomar, D. (2016), "Forecasting intraday trading volume: a Kalman filter approach", available at: SSRN 3101695.

Diebold, F.X. and Mariano, R.S. (2002), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, Vol. 20 No. 3, pp. 134-144, doi: 10.2307/1392185.

Doan, C.D. and Liong, S.y. (2004), "Generalization for multilayer neural network bayesian regularization or early stopping", *Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference*, pp. 5-8.

Foresee, F.D. and Hagan, M.T. (1997), "Gauss-newton approximation to bayesian learning", *Proceedings of International Conference on Neural Networks (ICNN'97)*, pp. 1930-1935, IEEE, doi: 10.1109/ICNN.1997.614194.

Gharehchopogh, F.S., Bonab, T.H. and Khaze, S.R. (2013), "A linear regression approach to prediction of stock market trading volume: a case study", *International Journal of Managing Value and Supply Chains*, Vol. 4 No. 3, p. 25, doi: 10.5121/ijmvsc.2013.4303.

Hagan, M.T. and Menhaj, M.B. (1994), "Training feedforward networks with the marquardt algorithm", *IEEE Transactions on Neural Networks*, Vol. 5 No. 6, pp. 989-993, doi: 10.1109/72.329697.

Hansen, P.R. (2005), "A test for superior predictive ability", *Journal of Business and Economic Statistics*, Vol. 23 No. 4, pp. 365-380, doi: 10.1198/073500105000000063.

Hansen, P.R., Lunde, A. and Nason, J.M. (2011), "The model confidence set", *Econometrica*, Vol. 79 No. 2, pp. 453-497, doi: 10.3982/ECTA5771.

Harvey, D., Leybourne, S. and Newbold, P. (1997), "Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, Vol. 13 No. 2, pp. 281-291 No. 2, doi: 10.1016/S0169-2070(96)00719-4.

Hou, Y. and Li, S. (2013), "Price discovery in Chinese stock index futures market: new evidence based on intraday data", *Asia-Pacific Financial Markets*, Vol. 20 No. 1, pp. 49-70, doi: 10.1007/s10690-012-9158-8.

Hou, Y. and Li, S. (2014), "The impact of the csi 300 stock index futures: positive feedback trading and autocorrelation of stock returns", *International Review of Economics and Finance*, Vol. 33, September 2014, pp. 319-337, doi: 10.1016/j.iref.2014.03.001.

Huang, W., Lai, P.C. and Bessler, D.A. (2018), "On the changing structure among Chinese equity markets: Hong Kong, Shanghai, and Shenzhen", *European Journal of Operational Research*, Vol. 264 No. 3, pp. 1020-1032, doi: 10.1016/j.ejor.2017.01.019.

Joseph, K., Wintoki, M.B. and Zhang, Z. (2011), "Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search", *International Journal of Forecasting*, Vol. 27 No. 4, pp. 1116-1127, doi: 10.1016/j.ijforecast.2010.11.001.

Kaastra, I. and Boyd, M.S. (1995), "Forecasting futures trading volume using neural networks", *The Journal of Futures Markets*, Vol. 15 No. 8, p. 953, doi: 10.1002/fut.3990150806.

Kao, Y.S., Chuang, H.L. and Ku, Y.C. (2020), "The empirical linkages among market returns, return volatility, and trading volume: evidence from the S&P 500 VIX futures", *The North American Journal of Economics and Finance*, Vol. 54, November 2020, 100871, doi: 10.1016/j.najef.2018. 10.019.

Kayri, M. (2016), "Predictive abilities of Bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: a comparative empirical study on social data", *Mathematical and Computational Applications*, Vol. 21 No. 2, p. 20, doi: 10.3390/mca21020020.

Khan, T.A., Alam, M., Shahid, Z. and Mazliham, M. (2019), "Comparative performance analysis of Levenberg-Marquardt, Bayesian regularization and scaled conjugate gradient for the prediction of flash floods", *Journal of Information Communication Technologies and Robotic Applications*, Vol. 10 No. 2, pp. 52-58.

Kong, A. and Zhu, H. (2018), "Predicting trend of high frequency csi 300 index using adaptive input selection and machine learning techniques", *Journal of Systems Science and Information*, Vol. 6 No. 2, pp. 120-133, doi: 10.21078/JSSI-2018-120-14.

Levenberg, K. (1944), "A method for the solution of certain non-linear problems in least squares", *Quarterly of Applied Mathematics*, Vol. 2 No. 2, pp. 164-168.

Liu, M., Choo, W.C., Lee, C.C. and Lee, C.C. (2022), "Trading volume and realized volatility forecasting: evidence from the China stock market", *Journal of Forecasting*, Vol. 42 No. 1, doi: 10.1002/ for.2897.

Long, W., Lu, Z. and Cui, L. (2019), "Deep learning-based feature engineering for stock price movement prediction", *Knowledge-Based Systems*, Vol. 164 No. 15, pp. 163-173, doi: 10.1016/j.knosys.2018. 10.034.

Lu, T. and Li, Z. (2017), "Forecasting csi 300 index using a hybrid functional link artificial neural network and particle swarm optimization with improved wavelet mutation", *2017 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pp. 241-246, IEEE, doi: 10.1109/ICCNEA.2017.55.

Lu, H., Ma, X., Huang, K. and Azimi, M. (2020a), "Carbon trading volume and price forecasting in China using multiple machine learning models", *Journal of Cleaner Production*, Vol. 249 No. 10, 119386, doi: 10.1016/j.jclepro.2019.119386.

Lu, W., Li, J., Li, Y., Sun, A. and Wang, J. (2020b), "A CNN-LSTM-based model to forecast stock prices", *Complexity*, Vol. 2020, 6622927, doi: 10.1155/2020/6622927.

Ma, S. and Li, P. (2021), "Predicting daily trading volume via various hidden states", arXiv preprint arXiv:2107.07678.

MacKay, D.J. (1992), "Bayesian interpolation", *Neural Computation*, Vol. 4 No. 3, pp. 415-447, doi: 10. 1162/neco.1992.4.3.415.

Marquardt, D.W. (1963), "An algorithm for least-squares estimation of nonlinear parameters", *Journal of the Society for Industrial and Applied Mathematics*, Vol. 11 No. 2, pp. 431-441, doi: 10.1137/ 0111030.

Møller, M.F. (1993), "A scaled conjugate gradient algorithm for fast supervised learning", *Neural Networks*, Vol. 6 No. 4, pp. 525-533, doi: 10.1016/S0893-6080(05)80056-5.

Nasir, M.A., Huynh, T.L.D., Nguyen, S.P. and Duong, D. (2019), "Forecasting cryptocurrency returns and volume using search engines", *Financial Innovation*, Vol. 5 No. 2, pp. 1-13, doi: 10.1186/ s40854-018-0119-8.

Ng, L. and Wu, F. (2007), "The trading behavior of institutions and individuals in Chinese equity markets", *Journal of Banking and Finance*, Vol. 31 No. 9, pp. 2695-2710, doi: 10.1016/j.jbankfin.2006.10.029.

Ning, S. (2020), "Short-term prediction of the csi 300 based on the bp neural network model", *Journal of Physics: Conference Series*, 012054, IOP Publishing, doi: 10.1088/1742-6596/1437/1/012054.

Oliveira, N., Cortez, P. and Areal, N. (2017), "The impact of microblogging data for stock market prediction: using twitter to predict returns, volatility, trading volume and survey sentiment

indices", *Expert Systems with Applications*, Vol. 73 No. 1, pp. 125-144, doi: 10.1016/j.eswa.2016.
12.036.

Paluszek, M. and Thomas, S. (2020), *Practical MATLAB Deep Learning: A Project-Based Approach*,
Apress, New York.

Satish, V., Saxena, A. and Palmer, M. (2014), "Predicting intraday trading volume and volume
percentages", *The Journal of Trading*, Vol. 9 No. 3, pp. 15-25, doi: 10.3905/jot.2014.9.3.015.

Schwarz, G. (1978), "Estimating the dimension of a model", *The Annals of Statistics*, Vol. 6 No. 2,
pp. 461-464, doi: 10.1214/aos/1176344136.

Selvamuthu, D., Kumar, V. and Mishra, A. (2019), "Indian stock market prediction using artificial
neural networks on tick data", *Financial Innovation*, Vol. 5 No. 16, p. 16, doi: 10.1186/s40854-019-
0131-7.

Shen, M.L., Lee, C.F., Liu, H.H., Chang, P.Y. and Yang, C.H. (2021), "Effective multinational trade
forecasting using lstm recurrent neural network", *Expert Systems with Applications*, Vol. 182
No. 15, 115199, doi: 10.1016/j.eswa.2021.115199.

Sohn, S. and Zhang, X. (2017), "Could the extended trading of csi 300 index futures facilitate its role of
price discovery?", *Journal of Futures Markets*, Vol. 37 No. 7, pp. 717-740, doi: 10.1002/fut.21804.

Stock, J.H. and Watson, M.W. (1998), "A comparison of linear and nonlinear univariate models for
forecasting macroeconomic time series". doi: 10.3386/w6607.

Sun, B., Guo, H., Karimi, H.R., Ge, Y. and Xiong, S. (2015), "Prediction of stock index futures prices
based on fuzzy sets and multivariate fuzzy time series", *Neurocomputing*, Vol. 151, Part 3,
pp. 1528-1536, doi: 10.1016/j.neucom.2014.09.018.

Susheng, W. and Zhen, Y. (2014), "The dynamic relationship between volatility, volume and open
interest in csi 300 futures market", *WSEAS Transactions on Systems*, Vol. 13, pp. 1-11.

Wang, C. and Chen, R. (2013), "Forecasting csi 300 volatility: the role of persistence, asymmetry, and
distributional assumption in Garch models", *2013 Sixth International Conference on Business
Intelligence and Financial Engineering*, IEEE, pp. 355-358, doi: 10.1109/BIFE.2013.74.

Wang, D.H., Suo, Y.Y., Yu, X.W. and Lei, M. (2013), "Price–volume cross-correlation analysis of
csi300 index futures", *Physica A: Statistical Mechanics and Its Applications*, Vol. 392 No. 5,
pp. 1172-1179, doi: 10.1016/j.physa.2012.11.031.

Wang, S., Li, G. and Wang, J. (2019), "Dynamic interactions between intraday returns and trading
volume on the CSI 300 index futures: an application of an SVAR model", *Mathematical
Problems in Engineering*, Vol. 2019, 8676531, doi: 10.1155/2019/8676531.

Wang, T. and Yang, J. (2010), "Nonlinearity and intraday efficiency tests on energy futures markets",
Energy Economics, Vol. 32 No. 2, pp. 496-503, doi: 10.1016/j.eneco.2009.08.001.

Wang, J., Hou, R., Wang, C. and Shen, L. (2016), "Improved v-support vector regression model based
on variable selection and brain storm optimization for stock price forecasting", *Applied Soft
Computing*, Vol. 49, December 2016, pp. 164-178, doi: 10.1016/j.asoc.2016.07.024.

Xie, M., Zhang, M., Liu, X., Ma, G. and He, P. (2020), "Decomposition model framework of trading
volume of cascade hydropower stations under the linking mode of medium-long term and
spot market", *Proceedings of PURPLE MOUNTAIN FORUM 2019-International Forum on
Smart Grid Protection and Control*, pp. 897-905, Springer, doi: 10.1007/978-981-13-9779-
0_73.

Xu, X. (2017), "The rolling causal structure between the Chinese stock index and futures",
Financial Markets and Portfolio Management, Vol. 31 No. 4, pp. 491-509, doi: 10.1007/s11408-
017-0299-7.

Xu, X. (2018), "Intraday price information flows between the csi300 and futures market: an application
of wavelet analysis", *Empirical Economics*, Vol. 54 No. 3, pp. 1267-1295, doi: 10.1007/s00181-017-
1245-2.

Xu, X. (2019a), "Contemporaneous and granger causality among us corn cash and futures prices", *European Review of Agricultural Economics*, Vol. 46 No. 4, pp. 663-695, doi: 10.1093/erae/jby036.

Xu, X. (2019b), "Contemporaneous causal orderings of csi300 and futures prices through directed acyclic graphs", *Economics Bulletin*, Vol. 39 No. 3, pp. 2052-2077.

Xu, X. (2020), "Corn cash price forecasting", *American Journal of Agricultural Economics*, Vol. 102 No. 4, pp. 1297-1320, doi: 10.1002/ajae.12041.

Xu, X. and Zhang, Y. (2021a), "Corn cash price forecasting with neural networks", *Computers and Electronics in Agriculture*, Vol. 184, May 2021, 106120, doi: 10.1016/j.compag.2021.106120.

Xu, X. and Zhang, Y. (2021b), "House price forecasting with neural networks", *Intelligent Systems with Applications*, Vol. 12, November 2021, 200052, doi: 10.1016/j.iswa.2021.200052.

Xu, X. and Zhang, Y. (2021c), "Individual time series and composite forecasting of the Chinese stock index", *Machine Learning with Applications*, Vol. 5 No. 15, 100035, doi: 10.1016/j.mlwa.2021.100035.

Xu, X. and Zhang, Y. (2021d), "Network analysis of corn cash price comovements", *Machine Learning with Applications*, Vol. 6 No. 15, 100140, doi: 10.1016/j.mlwa.2021.100140.

Xu, X. and Zhang, Y. (2022a), "Commodity price forecasting via neural networks for coffee, corn, cotton, oats, soybeans, soybean oil, sugar, and wheat", *Intelligent Systems in Accounting, Finance and Management*, Vol. 29 No. 3, pp. 169-181, doi: 10.1002/isaf.1519.

Xu, X. and Zhang, Y. (2022b), "Neural network predictions of the high-frequency CSI300 first distant futures trading volume", *Financial Markets and Portfolio Management*. doi: 10.1007/s11408-022-00421-y.

Xu, X. and Zhang, Y. (2022c), "Residential housing price index forecasting via neural networks", *Neural Computing and Applications*, Vol. 34 No. 17, pp. 14763-14776, doi: 10.1007/s00521-022-07309-y.

Xu, X. and Zhang, Y. (2022d), "Steel price index forecasting through neural networks: the composite index, long products, flat products, and rolled products", *Mineral Economics*. doi: 10.1007/s13563-022-00357-9.

Xu, X. and Zhang, Y. (2023a), "A high-frequency trading volume prediction model using neural networks", *Decision Analytics Journal*, Vol. 7, June 2023, 100235, doi: 10.1016/j.dajour.2023.100235.

Xu, X. and Zhang, Y. (2023b), "Regional steel price index forecasts with neural networks: evidence from East, South, North, Central South, Northeast, Southwest, and Northwest China", *The Journal of Supercomputing*. doi: 10.1007/s11227-023-05207-1.

Yan, Y. and Hongbing, O. (2018), "Dynamic probability of informed trading and price movements: evidence from the csi300 index futures market", *Applied Economics Letters*, Vol. 25 No. 14, pp. 998-1003, doi: 10.1080/13504851.2017.1391990.

Yan, Y. and Yang, D. (2021), "A stock trend forecast algorithm based on deep neural networks", *Scientific Programming*, Vol. 2021 No. 4, 7510641, doi: 10.1155/2021/7510641.

Yang, L. and Cheng, X. (2015), "Predictive analytics on CSI 300 index based on ARIMA and RBF-ANN combined model", *Journal of Mathematical Finance*, Vol. 5 No. 4, p. 393, doi: 10.4236/jmf.2015.54033.

Yang, J., Cabrera, J. and Wang, T. (2010), "Nonlinearity, data-snooping, and stock index ETF return predictability", *European Journal of Operational Research*, Vol. 200 No. 2, pp. 498-507, doi: 10.1016/j.ejor.2009.01.009.

Yang, J., Su, X. and Kolari, J.W. (2008), "Do euro exchange rates follow a martingale? Some out-of-sample evidence", *Journal of Banking and Finance*, Vol. 32 No. 5, pp. 729-740, doi: 10.1016/j.jbankfin.2007.05.009.

Yang, J., Yang, Z. and Zhou, Y. (2012), "Intraday price discovery and volatility transmission in stock index and stock index futures markets: evidence from China", *Journal of Futures Markets*, Vol. 32 No. 2, pp. 99-121, doi: 10.1002/fut.20514.

Yao, S., Luo, L. and Peng, H. (2018), "High-frequency stock trend forecast using LSTM model", *2018 13th International Conference on Computer Science and Education (ICCSE)*, pp. 1-4, IEEE, doi: 10.1109/ICCSE.2018.8468703.

Ye, X., Yan, R. and Li, H. (2014), "Forecasting trading volume in the Chinese stock market based on the dynamic VWAP", *Studies in Nonlinear Dynamics and Econometrics*, Vol. 18 No. 2, pp. 125-144, doi: 10.1515/snde-2013-0023.

Zhang, Z. (2020), "Bp neural network trade volume prediction and enterprises HRM optimization model based on ES-LM training", *Journal of Intelligent and Fuzzy Systems*, Vol. 39 No. 5, pp. 5883-5894, doi: 10.3233/JIFS-219218.

Zhang, C. and Pan, H. (2014), "Experimenting with 3 different input-output mapping structures of ANN models for predicting CSI 300 index", *Management Science and Engineering*, Vol. 8 No. 1, pp. 22-34, doi: 10.3968/j.mse.1913035X20140801.4274.

Zhang, Y.T. and Sun, B. (2017), "Analysis of CSI 300 stock index futures price trend based on ARIMA model", *DEStech Transactions on Social Science, Education and Human Science*. doi: 10.12783/dtssehs/seme2017/18022.

Zhao, L., Li, W., Bao, R., Harimoto, K. and Sun, X. (2021), "Long-term, short-term and sudden event: trading volume movement prediction with graph-based multi-view modeling", arXiv preprint arXiv:2108.11318.

Zhou, W., Pan, J. and Wu, X. (2019a), "Forecasting the realized volatility of csi 300", *Physica A: Statistical Mechanics and Its Applications*, Vol. 531 No. 1, 121799, doi: 10.1016/j.physa.2019.121799.

Zhou, Y.L., Han, R.J., Xu, Q., Jiang, Q.J. and Zhang, W.K. (2019b), "Long short-term memory networks for CSI300 volatility prediction with baidu search volume", *Concurrency and Computation: Practice and Experience*, Vol. 31 No. 10, e4721, doi: 10.1002/cpe.4721.

Corresponding author
Xiaojie Xu can be contacted at: xxu6@alumni.ncsu.edu