Data Science in Finance and Economics

http://www.aimspress.com/journal/dsfe

DSFE, 3(1): 14–29
DOI:10.3934/DSFE.2023002
Received: 01 December 2022
Revised: 08 February 2023
Accepted: 14 February 2023
Published: 23 February 2023

Research article

Research of daily stock closing price prediction for new energy companies in China

Qian Shen[1], **Yifan Zhang**[2], **Jiale Xiao**[2], **Xuhua Dong**[2] **and Zifei Lin**[2,*]

[1] School of Statistics, Xi'an University of Finance and Economics, Xi'an 710100, China
[2] School of Mathematics, Xi'an University of Finance and Economics, Xi'an 710100, China

* **Correspondence:** Email: zifeilinxaufe@163.com.

Abstract: Nowadays, China is developing new energy industries to reduce carbon emissions to meet the challenge of world climate change, so investors can consider to invest in stocks of Chinese new energy companies to gain income. In order to study how to forecast stock closing prices of new energy companies in China, we have chosen 12 representative companies, and first used autoregressive univariate time series models to predict the trends of the stock closing prices in the next month. The results show that Seasonal Autoregressive Integrated Moving Average model has the best out-of-sample trend prediction effect. Second, we use multivariate time series forecasting models to predict the stock closing prices of each day through external variables. The results show that Temporal Convolutional Attention Neural Networks has the best effect of out-of-sample prediction. We recommend that investors who are interested in investing in new energy companies in China first use the Seasonal Autoregressive Integrated Moving Average model to predict the short-term stock closing price trend in the future, and then use the Temporal Convolutional Attention Neural Networks model to predict the stock closing price on the next day to decide whether to invest.

Keywords: stock price; Chinese new energy companies; time series prediction; neural network; automated machine learning

JEL Codes: C22, C32, C45, C53

1. Introduction

In the 21st century, the challenge of climate change has always been an important issue in the international community. As major contributors to climate change and environmental pollution, carbon emissions come mainly from unenvironmentally friendly power generation, unscientific waste disposal, and fuel burning from transports and other human activities. In order to cope with the severe challenges in environmental protection, all major economies in the world have taken certain measures. Among the many measures, the development of new energy industry is very important. The development of new energy industry can enable human beings to greatly reduce carbon emissions in production activities, and fundamentally alleviate the problem of environmental pollution. Existing research works show that the development of new energy industry can effectively promote the reduction of carbon emissions. Fan and Lei (2015) believed that promoting the popularization of new energy vehicles in China can effectively promote energy conservation and reduce carbon emissions. Wang et al. (2021) studied the effects of new energy vehicles on air quality by applying a time-varying parameter–stochastic volatility–vector autoregression model. The result showed that the increase in new energy vehicles can reduce PM2.5 emissions. Zhao et al. (2016) think that the development of waste-to-energy has a significant effect on China's energy conservation, emission reduction and environmental protection. Saleh et al. (2021) claimed that using municipal solid waste to generate electricity can effectively reduce carbon dioxide emissions. The study by Li et al. (2020) pointed out that wind power projects have huge emission reduction potential compared with general coal-fired power plants. The research work of Liu et al. (2021) found that the carbon emissions of grassland wind farms in Inner Mongolia in China are significantly less than those of the four non-renewable energy sources of coal, natural gas, oil and nuclear power. In the experiment by Peng et al. (2013), the carbon emission of photovoltaic power generation is much lower than that of traditional fossil energy. Research by Jäger-Waldau et al. (2020) suggests that the development of the photovoltaic industry can make an important contribution to reducing Europe's carbon emissions in the future.

As one of the important economies, China has taken many measures to promote the development of new energy industries. "Technical Roadmap 2.0 for Energy-Saving and New Energy Vehicles" jointly issued by the Ministry of Industry and Information Technology of China and the Society of Automotive Engineers of China claims that by 2035, China's new energy vehicles will account for more than 50% of total vehicle sales, where the number of hydrogen fuel cell vehicles will reach about 1 million. It is estimated that by 2030 and 2050, China's wind power installed capacity will reach 400 million and 1 billion kilowatts, respectively. In 2050, wind power will meet 17% of domestic electricity demand in China. In addition, other new energy industries, such as photovoltaic power generation and waste-to-energy generation, have also received policy support to varying degrees from the government of China.

Considering the above policies and plans, we have reason to believe that the prospects of China's new energy industry will be better in the coming period. As an investor, investing in stocks of companies in promising industries over a period of time means potentially lucrative returns. However, there are only few studies looking at stock price predictions in one industry, let alone new energy industry in China. As a result, we hope to study how to effectively predict future stock prices of new energy companies in China, in order to provide possible assistance for interested investors.

2. Theoretical framework

As a typical time series data, stock price data is usually forecasted by scholars using time series models. With the rapid development of artificial intelligence in recent years, machine learning and deep learning are being widely used in the work of stock price prediction.

Adebiyi et al. (2014) examined the forecasting performance of ARIMA(ompaniesAutoregressive Integrated Moving Average) and artificial neural network models on stock data published by the New

York Stock Exchange. The empirical results showed that the neural network model outperforms the ARIMA model. The study of Weng et al. (2022) showed that ARIMA is superior to Naive methods at predicting changes in Tesla's share price movements. Sharma et al. (2017) believed that traditional methods, such as fundamental and technical analysis in stock price forecasting, may not be able to ensure the reliability of the forecast, so they constructed a regression analysis model for stock price forecasting. Zhang et al. (2018) proposed a novel approach for forecasting stock prices by combining the SVR with the firefly algorithm (FA). The applicability and superiority of the proposed methods were proved by comparative experiments. Patel et al. (2015) used ANN, SVM, random forest and naive Bayes model to forecast two Indian stocks and two stock indexes, and believed that the random forest model had the best forecast effect. Hu et al. (2022) used LSTM neural network and GRU neural network to effectively predict tock price trend of e-commerce platforms. Sethia and Raut (2018) used LSTM, GRU, ANN and SVM models to forecast the JP 500, and found that LSTM outperformed other models.

However, current research often only studies univariate time series, or only considers multivariate time series including auxiliary variables from a single perspective, and most of the strategies they propose are complex and suitable for financial institutions rather than retail investors. Moreover, there is also a lack of research on stock price forecast for new energy companies in China.

time series analysis model from a statistical perspective usually analyzes the univariate time series data itself. These models can learn the law of change in the trend of time series data, and predict the data in next some time steps in the future. As a traditional time series forecasting model, SARIMA (Seasonal Autoregressive Integrated Moving Average) takes into account the effect of seasonal changes. In the syudy of Divisekara et al. (2020), they used SARIMA to predict red lentil prices in Canada, and got good out-sample performance; Prophet is an univariate time series analysis model proposed by Meta. Dutta and Roy (2021) have used multiple time series forecasting models to predict indoor pollutant data. Among them, SARIMA achieved the best prediction effect. In the work of Kumar Jha, B. and Pande, S.M. (2021), they used Prophet, the additive model and ARIMA to predict supermarket data. Their results showed that Prophet is a better prediction model in terms of low error, better prediction, and better fitting. A successor of Prophet model–NeuralProphet improves forecast accuracy by 55 to 92 percent than it did for short to medium-term (Triebe et al. (2021)). AutoTS is a time series package for Python designed for rapidly deploying high-accuracy forecasts at scale (Wang et al. (2022)).

Unlike univariate time series models, multivariate time series models typically consider data that is not limited to the time series data itself. Multivariate time series models typically fit a linear or nonlinear relationship between auxiliary variables and time series data. For example, multivariate time series models are used to fit closing prices of the stock for each day and other data related to stock trading, such as transaction amount, highest price, lowest price, etc. Since these models usually rely on other data that cannot be obtained in advance (we can't know tomorrow's highest price for a stock before tomorrow's trading closes), these models can often only predict data for the next time step. The classic sequence forecast deep learning model LSTM (Long Short-Term Memory) network is developed from the recurrent neural network model (Hochreiter and Schmidhuber (1997)). In the research work of Jiang et al. (2018), they used LSTM and RNN to establish stock price prediction models, and their results proved that LSTM can be well used in stock price forecasting. Temporal Convolutional Neural Networks (TCNN) was proposed by Bai et al. (2018), and proved to be a very effective sequence model. On the basis of TCNN, Temporal Convolutional Attention Neural Networks (TCAN) was proposed by Lin et al. (2021). TCAN uses attention mechanism and time convolution for probabilistic prediction, and demonstrates its performance in empirical studies of solar power generation prediction. Although deep learning is currently the mainstream of sequence forecast, Elsayed et al. (2021) believed that when researchers make time series forecasting, they should not ignore simple machine learning baseline models, and should be configured more carefully to make them more suitable for the tasks. AutoGluon is an open-source AutoML framework developed by Erickson et al. (2020) that can train highly accurate machine learning models on unprocessed tabular datasets (such as CSV files) with just one line of Python, and it was used two popular kaggle competitions, beating 99% of the participants. Thus we use AutoGluon as a baseline model.

3. Data acquisition and experimental design

3.1. Data and variables

We collect the historical stock price data of 12 Chinese new energy companies. Among them, four companies are mainly engaged in wind power generation and optical power generation, four companies are mainly engaged in waste-to-energy generation. The remaining four companies are mainly involved in the new energy vehicle industry, or are in the process of transforming into new energy vehicle companies. More details of the 12 companies are shown in Table 4. The data cover the trading days from the beginning of 2016 through the last trading day of 2021. In order to get auxiliary variables to establish multivariate time series model, we chose daily opening and closing prices, highest and lowest prices, percentage of price change, amount of decrease and increase, and the daily transaction amount as our variables in this analysis.

Considering that financial data often have the characteristics of spikes and thick tails, we calculated the kurtosis of the stock closing price data for 12 Chinese new energy companies. The results show that the values of the kurtosis coefficients of these data are all smaller than the kurtosis of the normal distribution, so we assume that these data can continue to be used.

Figure 1 is a time series plot of stock price data for 12 companies. Observing the graph, it can be found that the stock price of each company presents a certain randomness.

3.2. Design of experiments

3.2.1. Ideas of the research

The purpose of our research is to analyze which model is more suitable for forecasting the future stock prices of Chinese new energy companies. So, we want to make this model more suitable for retail investors in the stock market, while more accurate in forecasting stock prices.

Considering that univariate time series models can predict the closing price of stocks for many days in the future, the multivariate time series model we use can actually only predict the closing price of stocks for the next day, so we consider dividing the work into two parts:

①Trend Forecasting

In financial investment, investors usually need to predict the general trend of financial products in the future. The study of Pongdatu and Putra (2018) argued that the SARIMA model is less accurate in long-term forecasts. Similar to SARIMA, when Satrio et al. (2021) used the Prophet model for time series forecasting, they found that the more days the Prophet model forecasted, the greater the deviation between the forecasting value and the true value became. As a result, we used daily closing

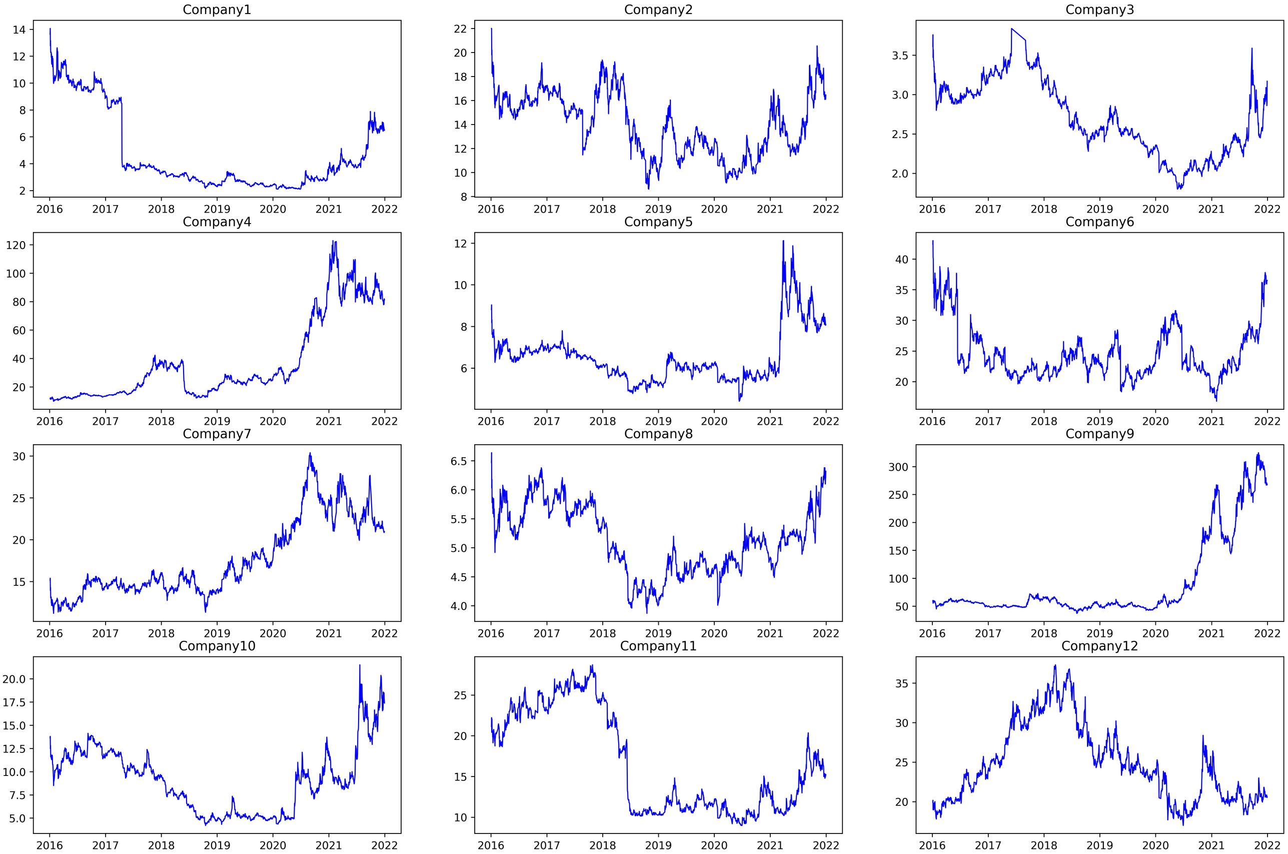

Figure 1. Time series graph of stock price data of 12 companies.

price data from the beginning of 2016 to the end of November 2021 as the training dataset, and stock closing price data for the December 2021 trading day is set as an out-of-sample test dataset. We have used univariate time series models to learn the training dataset, predicted the stock price data of the following trading day, and analyzed the effect of the model prediction according to some commonly used evaluation indicators.

②Closing Price Forecast based on Multivariate Time Series Models

In this part, we have built multi-prediction models that predicts the closing price of the following day based on the opening price, closing price and other data of Chinese new energy companies on the same day. Theoretically, such models can only predict the closing price of stocks on the following trading day at the end of the day's trading. In order to better understand the prediction ability of the models outside the sample, we used the daily closing price data from early 2016 to the end of June 2021 as the training dataset, and the daily stock closing price data from July 2021 to the end of 2022 was set as the out-of-sample test dataset.

3.3. Model evaluation metrics

Considering that time series forecasting and regression models have certain similarities, four commonly used regression model evaluation indicators are selected by us to analyze our prediction results. These model evaluation indicators are all errors in calculating model predictions, so smaller values of these indicators mean better prediction results of the model.

RMSE (Root Mean Square Error) represents the sample standard deviation of the difference between predicted and observed values (called residuals):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y-\hat{y})^2} \tag{1}$$

MAE (Mean Absolute Error) represents the average of the absolute error between the predicted value and the observed value:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i| \tag{2}$$

MAPE (Mean Absolute Percentage Error) is a variant of the MAE that uses a percentage to measure the magnitude of the deviation and is easy to understand and interpret:

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{3}$$

SMAPE (Symmetric Mean Absolute Percentage Error) is a modified indicator based on MAPE, which can improve the problem that MAPE is too large because the true value y_i is small:

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/m} \tag{4}$$

After computing the metrics for each model, we compared these values. By definition, a model with a smaller metric should have better predictive performance, then we take the univariate and multivariate models with minimal values of metrics as the final choice.

Table 1. Parameters of SARIMA models

Company number	1	2	3
parameter	$SARIMA(1,1,1)\times(1,0,1)_{12}$	$SARIMA(1,1,1)\times(1,0,1)_{12}$	$SARIMA(1,1,1)\times(1,0,1)_{12}$
Company number	4	5	6
parameter	$SARIMA(1,1,1)\times(0,1,1)_{12}$	$SARIMA(1,1,1)\times(0,0,1)_{12}$	$SARIMA(1,1,1)\times(0,0,1)_{12}$
Company number	7	8	9
parameter	$SARIMA(0,1,0)\times(1,0,0)_{12}$	$SARIMA(0,1,1)\times(1,0,0)_{12}$	$SARIMA(1,1,1)\times(0,1,1)_{12}$
Company number	10	11	12
parameter	$SARIMA(0,1,0)\times(0,0,1)_{12}$	$SARIMA(0,1,1)\times(1,0,1)_{12}$	$SARIMA(1,1,1)\times(0,0,1)_{12}$

4. Trend forecasting based on univariate time series models

4.1. SARIMA

ARIMA model is a statistical time series model, which makes the time series data stationary by performing d-order difference operations on it, and uses the ARMA model to fit the stationary data. SARIMA model is an improved model of the ARIMA model, which considers the seasonal variation of the time series data. Therefore, based on the ARIMA model, the SARIMA model adds s-step difference operation to eliminate the influence of seasonal effects on data stationarity.

For some non-stationary time series, after d-order difference and D-step seasonal difference, they can become a stationary time series y_t:

$$y_t = \Delta^d \Delta_s^D x_t \tag{5}$$

If x_t satisfies model $ARMA(p,q)\times(P,Q)_s$ with a seasonal period of s, then y_t is called a SARIMA model with a seasonal period of s, a non-seasonal order of p, d, q, and a seasonal order of P, D, and Q, and denoted as $SARIMA(p,d,q)\times(P,D,Q)_s$. That is, x_t satisfies:

$$\phi(B)\Phi(B)\Delta^d\Delta_s^D x_t = c + \theta(B)\Theta(B)\epsilon(t), \epsilon(t) \sim WN(0,\sigma^2), \tag{6}$$

where c is a constant term, $\phi(B)$ and $\theta(B)$ are autoregressive coefficient polynomials and moving average coefficient polynomials, respectively, $\Phi(B)$ and $\Theta(B)$ are seasonal autoregressive coefficient polynomials and seasonal moving average coefficient polynomials, respectively. Further $\Delta^d = (1-B)^d$, $\Delta_s^D = (1-B^s)^D$.

We used SARIMA to fit training data sets of the 12 companies' stock closing price data. After the first SARIMA models fit, we performed white noise tests on the residuals of the models. If the residuals of the models still have a cluster effect, we will transform the original data and then fit the SARIMA model to avoid heteroscedasticity as much as possible. Table 1 shows the details of the models.

4.2. Neural prophet

Prophet model was proposed by Meta (formerly Facebook), which is an open source time series forecasting tool. Its powerful forecasting ability for current variables can be used to forecast univariate time series data in most actual scenarios. Prophet decomposes time series data into an additive model:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{7}$$

In Equation 7 above, $g(t)$ is the trend function, $s(t)$ is the periodic function, $h(t)$ is the holiday function and ϵ_t means the error or stochastic disturbance.

There are two trend functions $g(t)$; one is the logistic growth model in Equation 8, which is a nonlinear model:

$$g(t) = \frac{C}{1 + exp(-k(t-m))} \tag{8}$$

Among them, the maximum value of $g(t)$ is infinitely close to C, k and m represent the slope and intercept of the leading part, respectively. In the actual calculation, C, k and m are all time-varying values, so for k, by detecting the change point and adjusting it on this basis, the piecewise logistic growth model of formula 9 is obtained:

$$g(t) = \frac{C(t)}{1 + exp((-(k + a(t)^t))(t - (m + a(t)^T\gamma)))} \tag{9}$$

Suppose that at time s_j, $j = 1, \ldots, S$ has S change points, the corresponding rate of change is $\theta = (\theta_1, \ldots, \sigma_S)$, Each component of the vector a(t) is defined as, if $t \geq s_j$, $a(j)_t = 1$, otherwise $a(j)_t = 0$.

Each component of the vector γ is defined as follows: calculated by the equal value of the piecewise function at the boundary S_j

$$\gamma_j = (S_j - m - \sum_{lj} \gamma l)(1 - \frac{k + \sum_{lj} \sigma_l}{k + \sum_{l \leq j} \sigma_l}) \tag{10}$$

Another trend function is Linear Trend with Changepoints, which is essentially a piecewise linear function:

$$g(t) = (k + a(t)^t)t + m + a(t)^T\gamma \tag{11}$$

Most of the parameters are the same as the previous model, the only difference is $\gamma_j = -s_j\sigma_j$.

The fit of the periodic function s(t) is approximated by a standard Fourier series, as shown in Equation 12:

$$s(t) = \sum_{n=1}^{N} (a_n cos(\frac{2\pi nt}{P}) + b_n sin(\frac{2\pi nt}{P})) \tag{12}$$

The period of the formula is P, and the parameters to be estimated is $\beta = [a_1, b_1, \ldots, a_N, b_n]^T$, and

$$s(t) = X(t)\beta \tag{13}$$

It is usually assumed that the prior distribution of β is a normal distribution $Normal(0, \sigma^2)$.

According to different holidays in different countries, for each holiday i, let D_i represent the period before and after the holiday. Then, get the following holiday function:

$$h(t) = Z(t)k = [1(t \in D_1), \ldots, 1(t \in D_L)]k, \tag{14}$$

where the prior distribution $k \sim Normal(0, v^2)$

Compared to Prophet, Neural Prophet is optimized using PyTorch's Gradient Descent, making the modeling process much faster than Prophet. Neural Prophet also uses the AR-Net to modeling time series autocorrelation, leverages deep learning ideas, such as configurable nonlinear layers of feedforward neural networks and custom losses and metrics, which makes Neural Prophet have the faster computing speeds and the more accurate forecast results.

Table 2. Different evaluation indexes in Univariate Time Series Models

		1	2	3	4	5	6	7	8	9	10	11	12
	SARIMA	0.173	0.315	0.064	1.237	0.129	0.758	0.183	0.072	1.128	0.645	0.293	0.251
MAE	Neural-Prophet	0.591	0.982	0.077	11.622	0.196	6.443	0.984	0.216	12.281	1.523	0.800	0.579
	AutoTS	0.209	0.969	0.356	11.276	0.715	2.893	1.706	2.907	29.875	1.791	1.059	0.594
	SARIMA	0.233	0.456	0.09	1.607	0.171	1.089	0.278	0.097	1.128	0.803	0.403	0.317
RMSE	Neural-Prophet	0.620	1.233	0.115	12.413	0.239	6.694	1.187	0.282	15.873	1.803	0.913	0.658
	AutoTS	0.278	1.200	0.376	11.645	0.735	3.390	1.754	2.927	36.624	2.074	1.252	0.643
	SARIMA	2.58%	1.79%	2.22%	1.48%	1.55%	2.17%	0.85%	1.16%	0.37%	3.53%	1.83%	1.21%
MAPE	Neural-Prophet	8.73%	5.51%	2.61%	14.17%	2.36%	18.01%	4.61%	3.48%	4.42%	8.14%	5.01%	2.79%
	AutoTS	3.04%	5.75%	12.07%	12.07%	8.57%	7.90%	7.96%	46.92%	10.85%	9.46%	6.89%	2.85%
	SARIMA	2.56%	1.81%	2.19%	1.49%	1.55%	2.15%	0.85%	1.16%	0.37%	3.52%	1.83%	1.21%
SMAPE	Neural-Prophet	9.17%	5.74%	2.65%	13.11%	2.38%	19.96%	4.53%	3.56%	4.31%	8.50%	5.05%	2.75%
	AutoTS	3.13%	5.63%	12.92%	12.92%	8.98%	8.34%	8.31%	37.93%	10.03%	10.08%	6.58%	2.82%

4.3. AutoTS

AutoTS is a framework for automatic time series analysis, which includes ARIMA, exponential smoothing model, Prophet model, etc. AutoTS can fit the input data and get a model that is relatively most suitable for the data. Since AutoTS contains a very rich univariate time series forecasting model, we regard it as the Baseline model in univariate time series forecasting to provide a benchmark for the forecasting effects of other models.

4.4. Analysis of results

Table 2 shows the results of the trend prediction of the closing price of the stocks of Chinese new energy companies by the three models. It isn't hard to find that on out-of-sample data, all four model evaluation metrics values of SARIMA are much smaller than the other two models. However, some of the model evaluation metrics values of Neural-Prophet are even larger than the baseline model AutoTS. As a result, we believe that SARIMA is relatively more suitable for forecasting future short-term trends in the closing price data of Chinese new energy companies' stocks.

Based on the above conclusions and analysis, we finally believe that the SARIMA model is more suitable for forecasting the short-term trend of the stock prices of Chinese new energy companies.

5. Stock price forecast based on multivariate time series model

5.1. LSTM

Traditional recurrent neural network (RNN) is only based on the memory of the previous moment, which is a short memory. In order to make up for this deficiency, the long short-term memory (LSTM) network improves the network structure of RNN. The new structure has the function of filtering information. This function means that LSTM will filter out useless information and retain useful information. When selectively memorizing, the model is often no longer only based on the memory of the previous moment, but can remember the information of the earlier moment. In addition, LSTM pays more attention to the contextual relationship between information. The essence of LSTM is to have a self-measurement mechanism, which measures the weight between the new input information in the current state and the information in the previous memory, and the result of the measurement determines which information is filtered out and which information is retained.

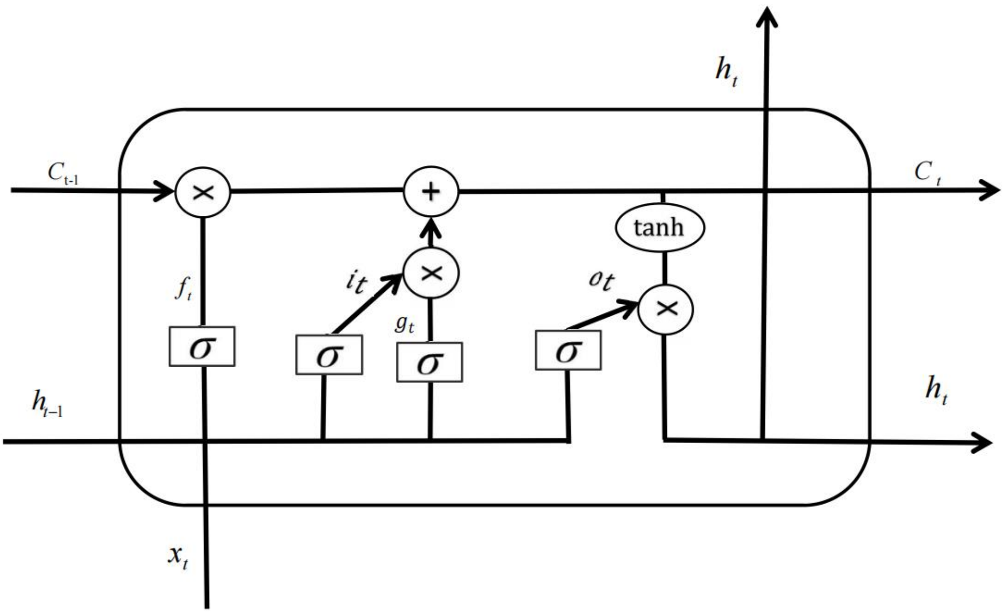

Figure 2. Basic structure of LSTM.

Figure 2 shows the basic structure of LSTM, indicating that the basic structure of LSTM consists of a forget gate f_t, a memory gate i_t, an output gate o_t and a memory unit g_t. The LSTM unit can be calculated by the following formula (Chen and Hu (2022)):

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{15}$$

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right) \tag{16}$$

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + b_o\right) \tag{17}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{18}$$

$$g_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{19}$$

$$h_t = o_t \odot \tanh\left(c_t\right) \tag{20}$$

W is the weight matrix, b is the bias, θ is sigmoid function, c_t is the value of the memory unit, c_{t-1} is the value of the last LSTM unit, x_t is the input of the LSTM unit, h_t is the output of the LSTM Unit, and h_{t-1} is the output of the LSTM unit at the moment before.

Moreover, in order to avoid overfitting as much as possible, we use dropout in regularization. In each layer, we set each neuron to have a 20% probability of being dropped.

5.2. TCAN

TCNN is proposed by Bai et al. (2018), which is a general-purpose convolutional network proposed for sequence modeling tasks with causal constraints. In TCNN, causal constraints on the network mean that the predicted sequence depends only on the input sequence, not future inputs. To impose causal constraints, TCNN consists of causal convolutional layers and output layer. TCNN's causal convolution ensures that information does not leak from the future to the past, and its expanded convolution helps increase TCNN's receptive field. In addition, TCNN consists of residual blocks, so deep networks can be adequately trained using residual learning.

TCAN is proposed by Lin et al. (2021). It consists of three parts: a temporal convolution layers, sparse attention layer and an output layer. In temporal convolution layers, TCAN uses the multiple dilated temporal convolutional layers (TC) to get the temporal latent factors $h_{t-T_{l:t}}$ from the historical data within the input window $y_{t-T_{l:t}}, x_{t-T_{l:t}}$, as:

$$h_{t-T_{l:t}} = TC(y_{t-T_{l:t}}, x_{t-T_{l:t}}) \tag{21}$$

In the sparse attention layer of TCAN, the sparse attention layer takes the temporal latent factors $h_{t-T_{l:t}}$ as input and generates the attention vector $\widetilde{h_t}$ that is used to make the prediction. Then, α-entmax attention (Peters et al. (2019)) is applied in the sparse attention layer. After that, concatenation is employed to combine the information from the attention context vector c_t and target hidden state h_t to produce the attention vector $\widetilde{h_t}$.

The output layer of TCAN uses the attention vectors to predict. The attention vector is transferred as the forecasting results, including the mean and variance of the distribution, as illustrated in Equation 22 and 23. In Equation 23, softplus function can guarantee the variance is always positive.

$$\hat{y}_t = linear\widetilde{h_t} \tag{22}$$

$$\sigma_t^2 = softplus(linear(\widetilde{h_t})) = log(1 + exp(linear(\widetilde{h_t}))) \tag{23}$$

5.3. AutoGluon

Similar to AutoTS used in univariate time series forecasting, AutoGluon is an automated machine learning framework developed by Amazon AI. The framework includes LightGBM, Random Forest, Catboost, XGBoost, NeuralNetMXNet, etc., as well as some combined models that weight and integrate each model. AutoGluon uses these models to fit the multivariate data separately, according to the principle of pursuing the smallest RMSE, to determine the most suitable model for our data set, and save it. Since AutoGluon actually contains a lot of machine learning models, we set it as the baseline model as a standard for evaluating the effect of multivariate time series forecasting models.

5.4. Analysis of results

To facilitate the comparison of model performance, we set the number of training epochs for both LSTM and TCAN in the training dataset to 120 epochs. Then, the results is showed as Table 3. TCAN obtained much lower model evaluation metrics values than the other two models in the off-sample closing price predictions for most corporate stocks. Although LSTM has lower model evaluation metrics

Table 3. different evaluation indexes in Multivariate Time Series Models

		1	2	3	4	5	6	7	8	9	10	11	12
MAE	LSTM	0.316	0.584	0.102	3.318	0.280	0.993	0.719	0.128	10.486	0.867	0.765	0.476
	TCAN	0.206	0.440	0.058	0.755	0.194	0.607	0.342	0.091	8.552	0.639	0.483	0.294
	AutoGluon	0.709	0.544	0.058	0.585	0.214	0.665	0.346	0.086	16.629	2.301	0.510	0.300
RMSE	LSTM	0.412	0.794	0.154	3.951	0.373	1.385	1.090	0.169	14.108	1.120	0.910	0.649
	TCAN	0.287	0.591	0.090	0.966	0.274	0.845	0.527	0.129	10.796	0.841	0.629	0.417
	AutoGluon	0.919	0.727	0.088	0.815	0.286	0.971	0.532	0.125	20.129	2.701	0.681	0.420
MAPE	LSTM	5.50%	3.59%	3.61%	3.84%	3.23%	3.29%	3.14%	2.27%	3.79%	5.26%	4.71%	2.31%
	TCAN	3.38%	2.70%	2.07%	0.87%	2.23%	2.04%	1.44%	1.62%	3.07%	3.89%	2.99%	1.44%
	AutoGluon	11.75%	3.30%	2.04%	0.67%	2.48%	2.19%	1.52%	1.50%	5.70%	13.30%	3.11%	1.46%
SMAPE	LSTM	5.40%	3.60%	3.70%	3.79%	3.25%	3.38%	3.10%	2.28%	3.83%	5.33%	4.70%	2.33%
	TCAN	3.44%	2.72%	2.06%	0.87%	2.24%	2.03%	1.44%	1.60%	3.02%	3.85%	2.96%	1.43%
	AutoGluon	10.77%	3.23%	2.04%	0.66%	2.49%	2.22%	1.53%	1.51%	5.92%	14.54%	3.17%	1.46%

values than the baseline model, its out-of-sample prediction effect is still insufficient compared with TCAN. Considering that most of prediction errors of TCAN are small enough to be accepted, we believe that TCAN is a relatively more suitable multivariate time series forecasting model for predicting the closing price data of the stocks of Chinese new energy companies.

6. Conclusions

6.1. Research findings

Based on the results above, we believe that the effective way to forecast the stock closing price of Chinese new energy companies is: first, use SARIMA models to forecast the short-term trend of the stock closing price, and then judge whether to invest; if the investors decide to invest the stock, they can use TCAN model to fit the stock closing price each day and the stock's opening and closing prices, highest and lowest prices, percentage of price change, amount of decrease and increase, and the daily transaction amount in the last day, and predict the stock's closing price in the next day as a powerful help for their investment.

6.2. Research contributions

Our research focuses on predicting the closing price of shares of new energy companies in China, providing some experience for stock price forecasting in China's new energy industry. Since there are currently few research studies on stock price forecasts for China's new energy industry, the prediction errors of the final models we selected, SARIMA and TCAN, are almost small enough, so we believe that our research has certain advantages for the stock price prediction of new energy companies in China. Because the market prospect of China's new energy companies is good, we believe that such research will help investors in the future. In addition, our research uses the newly proposed Neural Prophet, TCAN and automated machine learning models AutoGluon and AutoTS, and compares them with traditional models with the help of four model evaluation metrics.

6.2.1. Research limitations

Although the results of this study have some academic and industrial significance, there are still several aspects that can be further strengthened. Firstly, our study only uses the existing models. In future studies, we hope to propose innovative models that are more suitable to forecast the closing price of stocks of new energy companies in China. Secondly, we only selected the data of 12 representative companies in the main new energy industry for research. We hope to select as many Chinese new energy companies as possible in further research, so as to get more universal conclusions. Finally, we selected some of the most easily available stock price indicators in the study. We hope to explore how to more accurately predict the stock prices of Chinese new energy companies based on more indicators in the future.

Acknowledgments

The research was supported by the National Natural Science Foundation of China (NNSFC) (Grant Nos. 11902234 and 11972270), Natural Science Basic Research Program of Shaanxi (Program Nos. 2020JQ-853 and 2020JQ-852), Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No.20JG011), and the Young Talents Development Support Program of Xi'an University of Finance and Economics.

Conflict of interest

The authors declare no conflict of interest.

References

Adebiyi AA, Adewumi AO, Ayo CK (2014) Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *J Appl Math* 2014. https://doi.org/10.1155/2014/614342

Bai S, Kolter JZ, Koltun V (2018) An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *ArXiv* https://doi.org/10.48550/arXiv.1803.01271

Chen X, Hu Y (2022) Volatility forecasts of stock index futures in China and the US–A hybrid LSTM approach. *PLoS One* 17. https://doi.org/10.1371/journal.pone.0271595

Divisekara RW, Jayasinghe G, Kumari KW (2020) Forecasting the red lentils commodity market price using SARIMA models. *SN Bus Econ* https://doi.org/10.1007/s43546-020-00020-x

Dutta J, Roy S (2021) IndoorSense: context based indoor pollutant prediction using SARIMAX model. *Multimed Tools Appl* 80: 19989–20018. https://doi.org/10.1007/s11042-021-10666-w

Erickson N, Mueller J, Shirkov A, et al. (2020) AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. *ArXiv* https://doi.org/10.48550/arXiv.2003.06505

Elsayed S, Thyssens D, Rashed A, et al. (2021) Do We Really Need Deep Learning Models for Time Series Forecasting? *ArXiv* https://doi.org/10.48550/arXiv.2101.02118

Fan F, Lei Y (2015) Decomposition analysis of energy-related carbon emissions from the transportation sector in Beijing. *Transport Res D* 42: 135–145 https://doi.org/10.1016/j.trd.2015.11.001

Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Comput* 9: 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu Y, Zhou Z, Wang T (2022) Stock Price Prediction of E-commerce Platforms under COVID-19's Influence Based on Machine Learning. *2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)* 431–436. https://doi.org/10.1109/MLISE57402.2022.00092

Jiang Q, Tang C, Chen C, et al. (2018) Stock Price Forecast Based on LSTM Neural Network. *Proceedings of the Twelfth International Conference on Management Science and Engineering Management.* 393–408. https://doi.org/10.1007/978-3-319-93351-1_32

Jäger-Waldau A, Kougias I, Taylor NA, et al. (2020) How photovoltaics can contribute to GHG emission reductions of 55% in the EU by 2030. *Renew Sust Energ Rev* 126: 109836. https://doi.org/10.1016/j.rser.2020.109836

Kumar Jha B, Pande SM (2020) Time Series Forecasting Model for Supermarket Sales using FB-Prophet. *Renewable & 2021 5th International Conference on Computing Methodologies and Communication (ICCMC).* 547–554. https://doi.org/10.1109/ICCMC51019.2021.9418033

Li J, Li S, Wu F (2020) Research on carbon emission reduction benefit of wind power project based on life cycle assessment theory. *Renew Energ* 155: 456–468. https://doi.org/10.1016/j.renene.2020.03.133

Lin Y, Koprinska I, Rana M (2021) Temporal Convolutional Attention Neural Networks for Time Series Forecasting. *2021 International Joint Conference on Neural Networks (IJCNN)* 1–8. https://doi.org/10.1109/IJCNN52387.2021.9534351

Liu P, Liu L, Xu X, et al. (2021) Carbon footprint and carbon emission intensity of grassland wind farms in Inner Mongolia. *J Clean Prod* 313:127878. https://doi.org/10.1016/J.JCLEPRO.2021.127878

Peng J, Lu L, Yang H (2013) Review on life cycle assessment of energy payback and greenhouse gas emission of solar photovoltaic systems. *Renew Sust Energ Rev* 19: 255–274. https://doi.org/10.1016/J.RSER.2012.11.035

Peters B, Niculae V, Martins AF (2019) Sparse Sequence-to-Sequence Models. *ArXiv* https://doi.org/10.48550/arXiv.1905.05702

Pongdatu GA, Putra YH (2018) Seasonal Time Series Forecasting using SARIMA and Holt Winter's Exponential Smoothing. *IOP Conference Series: Materials Science and Engineering* 407. https://doi.org/10.1088/1757-899X/407/1/012153

Patel J, Shah SR, Thakkar P, et al. (2015) Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Syst Appl.* 42: 259-268. https://doi.org/10.1016/j.eswa.2014.07.040

Shen Q (2022) Stock Price Data of 12 New Energy Companies. *figshare.com* https://doi.org/10.6084/m9.figshare.21383679.v1

Sharma A, Bhuriya D, Singh U (2017) Survey of stock market prediction using machine learning approach. *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)* 2: 506–509. https://doi.org/10.1109/ICECA.2017.8212715

Satrio CB, Darmawan W, Nadia BU, et al. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Comput Sci* 179: 524–532. https://doi.org/10.1016/j.procs.2021.01.036

Saleh SY, El-Ghussain A, Muaamaer AA, et al. (2021) Impact Assessment of Utilizing Municipal Solid Waste for Energy Generation: A case study for MSW management plant at Sofa landfill site. *2021 International Conference on Electric Power Engineering – Palestine (ICEPE- P)* 1–6. https://doi.org/10.1109/ICEPE-P51568.2021.9423474

Sethia A, Raut P (2018) Application of LSTM, GRU and ICA for Stock Price Prediction. *Information and Communication Technology for Intelligent Systems* https://doi.org/10.1007/978-981-13-1747-7_46

Triebe O, Hewamalage H, Pilyugina P, et al. (2021) NeuralProphet: Explainable Forecasting at Scale. *ArXiv* https://doi.org/10.48550/arXiv.2111.15397

Wang C, Chen X, Wu C, et al. (2022) AutoTS: Automatic Time Series Forecasting Model Design Based on Two-Stage Pruning. *ArXiv* https://doi.org/10.48550/arXiv.2203.14169

Weng Q, Liu R, Tao Z (2022) Forecasting Tesla's Stock Price Using the ARIMA Model. *Proc Bus Econ Stud* https://doi.org/10.26689/pbes.v5i5.4331

Wang Q, Su C, Hua Y, et al. (2021) How can new energy vehicles affect air quality in China?— From the perspective of crude oil price. *Energ Environ* 33: 1524–1544. https://doi.org/10.1177/0958305X211044388

Zhang J, Teng Y, Chen W (2018) Support vector regression with modified firefly algorithm for stock price forecasting. *Appl Intell* 49: 1658–1674. https://doi.org/10.1007/s10489-018-1351-7

Zhao XG, Jiang GW, Li A, et al. (2016) Technology, cost, a performance of waste-to-energy incineration industry in China. *Renew Sust Energ Rev* 55: 115–130. https://doi.org/10.1016/J.RSER.2015.10.137

Appendix A. Data Sources

All the data we use comes from the data published by the Shanghai Stock Exchange and Shenzhen Stock Exchange, obtained through the Python package Tushare. Now the data are published on figshare.com (Shen (2022)).

Appendix B. Details of the Companies

Table 4. Details of 12 new energy companies in China

Company's ID	Stock Symbol	Company's Name
Company 1	SH601016	Cecep Wind-power Corporation Co.,Ltd.
Company 2	SZ002202	Xinjiang Goldwind Sci & Tech Co.,Ltd.
Company 3	SH600795	GD Power Development Co.,Ltd.
Company 4	SH601012	Longi Green Energy Technology Co.,Ltd
Company 5	SZ000027	Shenzhen Energy Group Co.,Ltd.
Company 6	SH603568	Zhejiang Weiming Environment Protection Co.,Ltd
Company 7	SH600323	Grandblue Environment Co.,Ltd
Company 8	SZ000598	Chengdu Xingrong Environment Co.,Ltd.
Company 9	SZ002594	BYD Co.,Ltd.
Company 10	SH600418	Anhui Jianghuai Automobile Group Co.,Ltd.
Company 11	SH601238	Guangzhou Automobile Group Co., Ltd.
Company 12	SH600104	SAIC Motor Co., Ltd.