Data Science and Management 6 (2023) 239–246

Contents lists available at ScienceDirect

Data Science and Management

journal homepage: www.keaipublishing.com/en/journals/data-science-and-management

Research article

Forecasting stock closing prices with an application to airline company data

Xu Xu [a], Yixiang Zhang [b], Clare Anne McGrory [c], Jinran Wu [d,*], You-Gan Wang [d]

[a] *School of Mathematics and Physics, Wenzhou University, Wenzhou, Zhejiang, 325035, China*
[b] *Olin Business School, Washington University in St. Louis, St. Louis, 63130, Missouri, USA*
[c] *School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, 4001, Australia*
[d] *The Institute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane, Queensland, 4001, Australia*

A R T I C L E I N F O

Keywords:
Chinese airlines
LASSO
Ridge regression
Support vector machine regression
Forecasting

A B S T R A C T

Forecasting stock market movements is a challenging task from the practitioners' point of view. We explore how model selection via the least absolute shrinkage and selection operator (LASSO) approach can be better used to forecast stock closing prices using real-world datasets of daily stock closing prices of three major international airlines. Combining the LASSO method with multiple external data sources in our model leads to a robust and efficient method to predict stock behavior. We also compare our approach with ridge, tree, and support vector machine regressions, as well as neural network approaches to model the data. We include lags of each external variable and response variable in the model, resulting in a total of 870 predictor variables. The empirical results indicate that the LASSO-fitted model is the most effective when compared to other approaches we consider. The results show that the closing price of an airline stock is affected by its closing price for the previous days and those of other types of airlines and is significantly correlated with the Shanghai Composite Index for the previous day and 3 days prior. Other influencing factors include the positive impact of the Shanghai Composite Index daily share volume, the negative impact of loan interest rates, the amount of highway passenger and railway freight turnover, etc.

1. Introduction

This study aims to show that it is possible to find a simple but reliable model to predict movement in stock prices by considering the external factors that influence a company. We illustrate this approach on the stock price data of China's three largest air carriers: Air China, China Southern, and China Eastern Airlines.

The extant literature notes that financial time series are often statistically and economically insignificant, complicating efforts to find an appropriate model to guide stock predictions (Oztekin et al., 2016). Previous research has suggested that, even in challenging cases where statistical significance is not seen in changes in the conditional mean stock values, there may be a predictable pattern indicating the direction of stock price changes (Liu and Kemp, 2019). For this reason, some studies on this topic have focused on predicting the direction of stock price changes rather than the stock prices themselves.

Despite the practical challenges, numerous statistical methods have been proposed for fitting models aimed at predicting stock prices. See Gandhmal and Kumar (2019) for a recent review of this complex task. The importance of finding a simple model is also stressed by Oztekin et al. (2016) because, in practice, the majority of practitioners lack access to vast complex datasets or cannot wait for highly complex analyses to be completed. In light of these challenges, Zhou et al. (2020) notes in a recent extensive literature review that the forecast accuracy levels ranged from 55% to 65% in terms of the proposed models' ability to forecast the direction of a stock. Their method, which necessitates the use of a wide range of data sources, reportedly achieved levels of up to 75% when applied to Chinese stock data.

Airline stock prices are influenced by a myriad of internal and external influencing factors, including consumer confidence (Goh and Rasli, 2014), the fundamentals of enterprise operation (Nowak and Anderson, 2014), the external competitive environment (Gong et al., 2008), organizational alliances (Song et al., 2007), and technical efficiency (Alam and Sickles, 1998). However, the majority of the literature

Peer review under responsibility of Xi'an Jiaotong University.
* Corresponding author.
E-mail address: ryan.wu@acu.edu.au (J. Wu).

https://doi.org/10.1016/j.dsm.2023.09.005
Received 13 July 2023; Received in revised form 26 September 2023; Accepted 30 September 2023
Available online 5 October 2023

discussing airline performance focuses on operations and macro reasons that affect returns, while the external factors influencing performance were largely overlooked (Xu et al., 2021). Accordingly, this study focuses on these largely overlooked external factors. Finding a high-performing prediction model involves not only choosing carefully which external factors to include but also deciding on a trade-off between accuracy and model complexity.

Many studies focus specifically on the impact of exchange rates and oil prices on general stock prices, including airline stock performance. For example, Naser and Alaali (2018) proposed that oil price, macroeconomic, and financial variables were used to predict the profitability of the S&P 500 price index. The empirical results suggest that the prediction performance of the model is better when the oil price is included in the prediction factors. As was described by Arekar and Jain (2017), Peng et al. (2020), and Sonenshine and Cauvel (2017), oil purchases account for approximately one-third of airline companies' operating costs; therefore, we considered oil prices in our modelling. Airline shares are also potentially affected by exchange rates, which was mentioned by Forsyth and Dwyer (2010), Hamrita and Trifi (2011), Mok (1993), and Siami-Namini (2017). Thus, we explored oil prices and exchange rates as factors in our modelling.

We also postulate that daily movements in stock prices of other stocks can be used to improve the model's performance. This includes data on competitor airlines operating in similar market conditions; for instance, two of Air China's main competitors are China Eastern Airlines and China Southern Airlines. Hence, we included competitors' stock data in our model.

As aforementioned, scholars have proposed many alternative approaches for fitting statistical models to stock price data. While their approaches did not include external influencing factors as ours does, the use of the least absolute shrinkage and selection operator (LASSO) method for this task was suggested by Roy et al. (2015) to predict the stock price based on information obtained from the Goldman Sachs Group Inc. from 1999 to 2014, including the low price, high price, opening price, closing price, and trading volume. Their results imply that the LASSO-based model has good potential and outperforms the ridge linear regression model for the problem. LASSO-based approaches are widely regarded as achieving high prediction accuracy in various analytical settings. The accuracy of the approach is boosted by the fact that the method intrinsically performs shrinkage of superfluous coefficients. This feature reduces variance in the fit and minimizes bias. The LASSO also well in situations where the numbers of observations and factors are small and large, respectively, as is often the case in financial modelling. We believe that overall the LASSO method has been greatly underused for stock prediction since the work by Roy et al. (2015), which is why we elect to use it and demonstrate its strengths.

Bao et al. (2004) mentioned that applications of machine learning technology, namely support vector regression (SVR), were explored for predicting stock prices. The daily closing price of Haier, which is a famous Chinese enterprise in Shanghai stock market, was used as the data set to test the effectiveness of SVR. Their experiments suggest that SVR may be a valuable method to predict stock prices. Indeed, such modern techniques are often recommended as the current state-of-the-art for addressing this problem. Therefore, we compare our LASSO approach with this alternative one in our motivating examples. Furthermore, we examine recent research pertaining to stock forecasting, as outlined in Table 1. Notably, we observe a scarcity of studies dedicated to airline stock price prediction. Most contemporary research in this domain predominantly concentrates on time series forecasting, a departure from our approach, which involves identifying significant factors influencing airline stock closing prices amid a multitude of extraneous variables. In this study, we mainly compare the LASSO approach with the closely related approach of support vector regression (SVR), ridge regression (RR), tree regression (TR), and the neural networks (NN) approach.

This study contributes to determining the main factors that affect a stock's price from numerous complicated external factors. In addition to using the traditional stock forecasting model, we utilize technical indicators (e.g., closing price, opening price, sector index, broad market index, and trading volume) to predict airline stock prices. Moreover, we introduce fundamental indicators, including company performance. In particular, we introduce external indicators (e.g., oil price; exchange and interest rates; railway and highway passenger and freight turnover) into the forecast model. In this way, we significantly improve the accuracy of the forecast model to better ascertain the primary and secondary influencing factors that determine stock price. This approach clarifies the rise and fall mechanism of stocks and provides an economic explanation for each influencing factor.

Table 1
Summary of forecasting references in previous literature.

References	Forecasting target	Proposed model
Bao et al. (2004)	Closing prices of Haier	SVM
Roy et al. (2015)	Goldman Sachs Group Inc. stock	LASSO
Naser and Alaali (2018)	SPX price index returns	A dynamic model averaging model
Oztekin et al. (2016)	Borsa Istanbul BIST 100 Index	SVM
Liu and Kemp (2019)	U.S. oil and gas industry stock index	An adaptive probit model
Zhou et al. (2020)	China A-share market index	SVM
Kamara et al. (2022)	Sure-enough stock market datasets	A brand-new end to end algorithm
Kirisci and Cagcag Yolcu (2022)	TAIEX and FTSE	CNN-based forecasting model
Chandola et al. (2022)	Stock price of APPLE	Word2Vec-LSTM
Lin et al. (2022)	RV of CSI300, SPX, and STOXX50 indices	CEEMDAN and LSTM
Abraham et al. (2022)	15 selected stocks	GA-RF
Deng et al. (2022)	SPX, SSE, and HSI	MEMD-LSTM
Md et al. (2023)	Samsung's stock prices	Multi-Layer Sequential LSTM
Staffini (2022)	FTSE MIB	DCGAN
Sadorsky (2022)	SESP, SP, OP	Tree-based machine learning
Kurani et al. (2023)	General stock data	Combination of NNs and statistical method
Liu et al. (2023)	Shanghai, Shenzhen Stock Exchange	EMD-HAM

Note: support vector machine (SVM); Taiwan stock exchange capitalization weighted stock index (TAIEX); financial time stock exchange (FTSE) for London stock market data; Standard & Poor 500 index (SPX); Shanghai stock exchange (SSE); Hang Seng index (HSI); financial times stock exchange Milano indice di Borsa (FTSE MIB); solar energy stock prices (SESP); silver prices (SP); oil price (OP); neural networks (NN); empirical mode decomposition (EMD); heterogeneous autoregressive model (HAM); long short-term memory model (LSTM); multivariate empirical mode decomposition (MEMD); Genetic Algorithm (GA); Random Forest (RF); deep convolutional generative adversarial Network (DCGAN).

The other main contributions of this paper are the construction of a reliable and fast algorithm for stock prediction that makes use of data that are readily accessible to practitioners and can be easily applied and interpreted. We are pleased to report that the mean average percentage error (MAPE), the mean squared error (MSE), and the mean absolute error (MAE) for our models fitted via LASSO suggest a high degree of prediction accuracy. For Air China, these measures were 2.17%, 0.28, and 0.20, respectively; for China Southern Airlines, they were 2.06%, 0.24, and 0.17, respectively; and for China Eastern Airlines, they were 1.94%, 0.16 and 0.12, respectively.

The remainder of this paper is organized as follows. Section 2 formulates the problem of airline stock forecasting with the LASSO method. Section 3 describes the three airline stock datasets and their potential predictor variables and discusses the preprocessing done on the datasets. Section 4 reports the forecasting results, and Section 5 concludes the paper.

2. Problem formulation

The LASSO approach is a type of supervised statistical learning algorithm (Tibshirani, 1996) for performing regression. The name LASSO is an acronym for least absolute shrinkage and selection operator. LASSO algorithms can be applied to forecast time series data, such as stock closing prices; Panagiotidis et al. (2018) and Xu et al. (2021) expressed the time series problem as a regression. LASSO can be thought of as a more modern approach to estimating the parameters in a regression model than the traditional ordinary least squares approach. The method constructs a penalty function to obtain a relatively refined model that shrinks some regression coefficients and removes them from the model. Therefore, it retains the advantage of subset contraction and is a biased estimate for processing data with complex collinearity. Intuitively, we would expect there to be collinearity among the variables in our model because they involve data from companies competing in the same markets under very similar conditions. In this way, LASSO overcomes many of the problems that the least squares approach would encounter in such a situation, which was described by Li and Chen (2014) and Ludwig et al. (2015).

Our response dataset is a time series $y_1, \ldots, y_n$ of the daily closing stock prices for Air China observed over n points in time between 2006 and 2019. If we formulate this time series modelling problem as a regression problem, an appropriate model is a multiple linear regression for a response at a given time point t, and this takes the form:

$$y_i = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_p x_{tp} + \epsilon$$

where the x_t's are the observed values of the predictor variables at time point t, and $\beta = \beta_1, \ldots, \beta_p$ is the slope coefficients for each predictor. The value β_0 is a constant intercept term, and ϵ is the error term (residuals) in the regression model.

It is intuitively logical to select the predictor variables ($\beta_1, \ldots, \beta_p$) and include a lagged value of the response (daily stock closing prices for Air China) and lagged values of other relevant external variables. The resulting model we use can be described more specifically as an autoregressive distributed lag model because we include lag-30 values of all of the predictor variables. Note also that we adjust our predictor variables for seasonality and trend where appropriate before fitting the model by removing the long-term factors.

As mentioned, the LASSO approach (Tibshirani, 1996) can be thought of as an alternative (Birkes and Dodge, 2011) to ordinary least squares estimation of regression parameters $\hat{\beta}$. LASSO is based on a penalized least squares approach to minimizing the residual sum of squares of the problem subject to a bound on its L_1 norm. Given the response vector y (centered around its mean), for values t of the bound less than the L_1 norm of the ordinary least squares estimate of β, LASSO estimates of the predictors are described as solutions to:

$$\widehat{\beta}_L = argmin_{\beta}(y - X\beta)^T(y - X\beta) + C\|\beta\|_1, \tag{1}$$

where $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$.

The parameter $C \geq 0$ determines the amount of shrinkage in the model. The simultaneous model selection and parameter estimation that result is considered an advantageous feature of LASSO, particularly when the number of potential predictors is enormous. Xu et al. (2021) discussed it more in detail. In contrast, there is no comparable parameter in an ordinary least squares approach.

If there is a lot of shrinkage in the LASSO fit, there is a greater tendency for predictors to be excluded from the model and deemed unnecessary. Conversely, in the special case where $C = 0$, i.e., no shrinkage, the LASSO estimator simply equates to the ordinary least squares estimator as no predictors are removed or shrunk in the estimation. In our approach, we employ cross-validation to select the values of C that optimize the predictive performance of the model, or, in other words, values that minimize the mean squared prediction error.

3. The investigated data

3.1. Data description

All 15 predictor variables listed in Table 2 below were explored in the model. However, note that we include 30 lags of each of these variables in the model, as well as lags of the response variable, which is autoregressive in the model. When predicting the stock price of a certain airline company, predictors not only include information such as oil price, exchange rate, and railway carrying capacity but also relevant stock information of the other two airlines, for a total of 870 lagged variables. This means that we use a total of 870 predictor variables in our model, labelled as β_1–β_{870}, respectively. Fig. 1 displays correlations between the non-cumulative variables in the study. Data collected between the years

Table 2
List of predictor variables used in our modelling.

Variable	Description	Data sources	Frequency
AC close	Air China daily stock closing price	Wind data Service	Daily
CS close	China Southern Airlines daily stock closing Price	Wind data Service	Daily
CE close	China Eastern Airlines daily stock closing price	Wind data Service	Daily
S open	Shanghai composite indexdaily opening price	Wind data Service	Daily
S high	Shanghai composite index daily high price	Wind data Service	Daily
S low	Shanghai composite index daily low price	Wind data Service	Daily
S close	Shanghai composite index daily closing price	Wind data Service	Daily
S volume	Shanghai composite index daily share volume	Wind data Service	Daily
An open	Aviation sector index daily opening price	Wind data Service	Daily
A high	Aviation sector index daily high price	Wind data Service	Daily
A low	Aviation sector index daily low price	Wind data Service	Daily
A close	Aviation sector index daily closing price	Wind data Service	Daily
A volume	Aviation sector index daily share volume	Wind data Service	Daily
Crude	Daily crude oil spot price	EIA	Daily
Exchange rate	Daily exchange rate (spot rate: USD to CNY)	CFETS	Daily
Average wages	Average wages of employees	National Bureau of Statistics	Yearly
GDP	Gross domestic product	National Bureau of Statistics	Seasonal
PCDI	Per capita disposable income of national residents	National Bureau of Statistics	Seasonal
AC net profit	The net profit of Air China	Financial announcement	Seasonal
CS net profit	The net profit of China Southern Airlines	Financial announcement	Seasonal
CE net profit	The net profit of China Eastern Airlines	Financial announcement	Seasonal
Import export	Import and export amount	Customs Head Office	Monthly
Railway passenger	Railway passenger turnover	National Bureau of Statistics	Monthly
Railway freight	Railway freight turnover	National Bureau of Statistics	Monthly
Highway passenger	Highway passenger turnover	National Bureau of Statistics	Monthly
Highway freight	Highway freight turnover	National Bureau of Statistics	Monthly
Passenger volume	Civil aviation passenger volume	National Bureau of Statistics	Monthly
Interest rate	Short term loan interest rate per year	People's Bank of China	Monthly
Brent oil	The price of brent oil	EIA	Daily

Note: Energy Information Administration (EIA) of the U.S.

2006 and 2019 were available, and they were split into two sets for training and testing, respectively. The model is first fit to the training data: the training set comprises time series data observed between 2006 to part-way through 2014. We then use this trained model on the test data, which comprised the remaining available data from the latter half of 2014 to the end of 2019 to make our predictions. The data that support the findings of this study are available from the corresponding author upon request.

3.2. Data preprocessing

When seasonality is present in a predictor, it is necessary to preprocess the corresponding data to adjust for this before performing an analysis.

The package "seastests", which is available for download and use with the R language statistical programming suite (Team et al., 2013), can be used to test for seasonality in the data corresponding to each of the predictors. The specific test applied is referred to as the Webel-Ollech overall seasonality test. This test incorporates results from various seasonality tests, which was proposed by Webel and Ollech (2018).

Table 3 indicates whether or not each of our predictors exhibits seasonality according to the Webel-Ollech overall seasonality test. Fig. 2 shows the predictors with seasonality before and after adjustment.

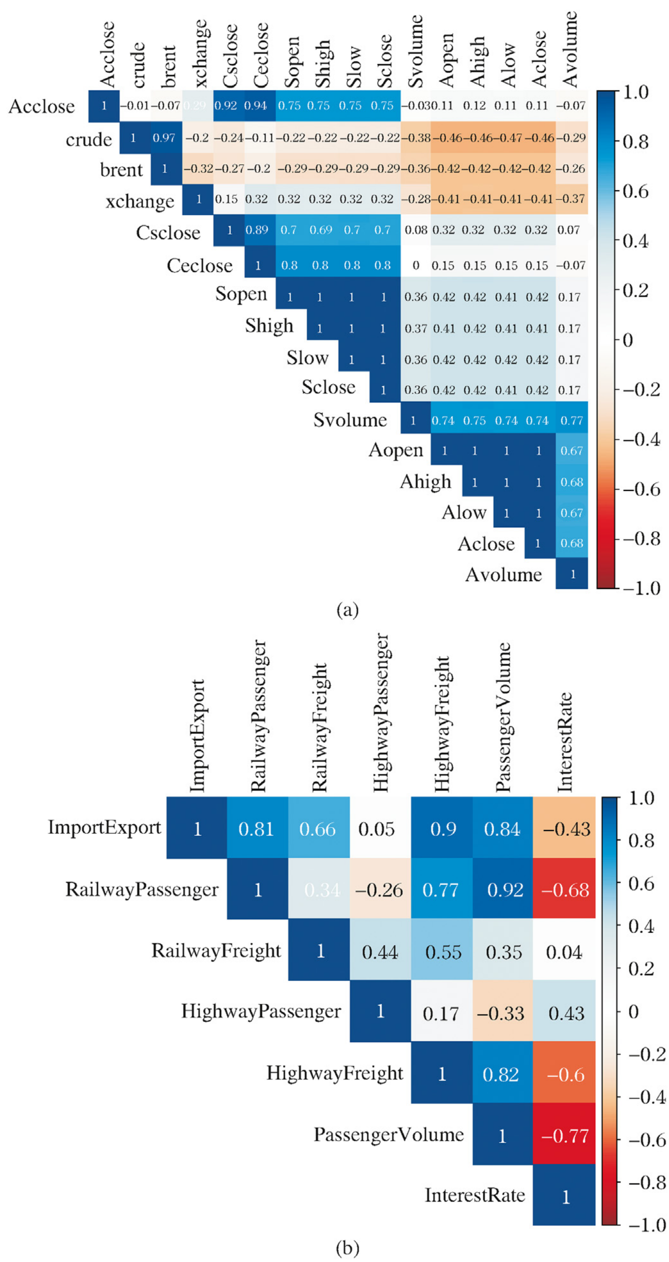

Fig. 1. Correlations between the variables used in the modelling. (a) Daily variables and (b) monthly variables.

4. Results and discussion

This section discusses the forecasting performances of LASSO for the three airlines' stock closing prices. RR, TR, SVR, and NN are selected as benchmark models to illustrate the effectiveness of LASSO. In addition, LASSO techniques and RR are implemented within the package "glmnet" (Friedman et al., 2010) in the statistical programming language R, which we used to produce the main numerical results for this paper. Also available in the statistical programming language R (Team et al., 2013) is the following: (1) TR is available as part of the package "rpart" (Therneau et al., 2015); (2) SVR regression can be implemented using the package "e1071" (Dimitriadou et al., 2008); and (3) NN can be applied using the package "neuralnet" (Günther and Fritsch, 2010).

4.1. Forecasting performances

Table 4 provides measures of the performance of the LASSO-based prediction and compares them with results from alternative analyses of the same data. The table reports the MAPE, RMSE, and MAE for the predicted daily values. Judging by these measures, the LASSO approach outperforms the other two methods. For this reason, the following results we report are from the LASSO approach.

Table 5 displays the estimated values for the selected predictors obtained from the LASSO approach and ranks them in order of influence. In the fitted model, many lagged coefficients have been shrunk to 0 (i.e., they are not estimated to be significant in the model). Note that the shrunk coefficients are not listed in the table. Here, we find the following. (1) Given the body of literature, it is not surprising that the first lag of the Air China closing price variable is ranked as the most influential predictor among those available. The other main influential factors are the closing price of Air China's major competitor, and China Southern Airlines. Perhaps the most surprising outcome in our results is that the effects of crude oil price and exchange rate were completely shrunk from the model. (2) Once again, and unsurprisingly, we find that the first lag of China Southern Airline's closing price variable is ranked as the most influential predictor among those available. In our results, the effects of crude oil price and exchange rate were again completely shrunk from the model. (3) Once more we find that China Southern Airline's closing price variable is ranked as the most influential predictor among those

Table 3
Indication of whether each predictor variable exhibits seasonality.

Variable	Seasonality present?	Variable	Seasonality present?
Average wages	No	Brent	No
GDP	Yes	Exchange rate	No
PCDI	Yes	AC close	No
AC net profit	Yes	CS close	No
CS net profit	Yes	S open	No
CE net profit	Yes	S high	No
Import export	Yes	S low	No
Railway passenger	Yes	S close	No
Railway freight	Yes	S volume	No
Highway passenger	Yes	A open	No
Highway freight	Yes	A high	No
Passenger volume	Yes	A low	No
Interest rate	No	A close	No
Crude	No	A volume	No
CE close	No	-	-

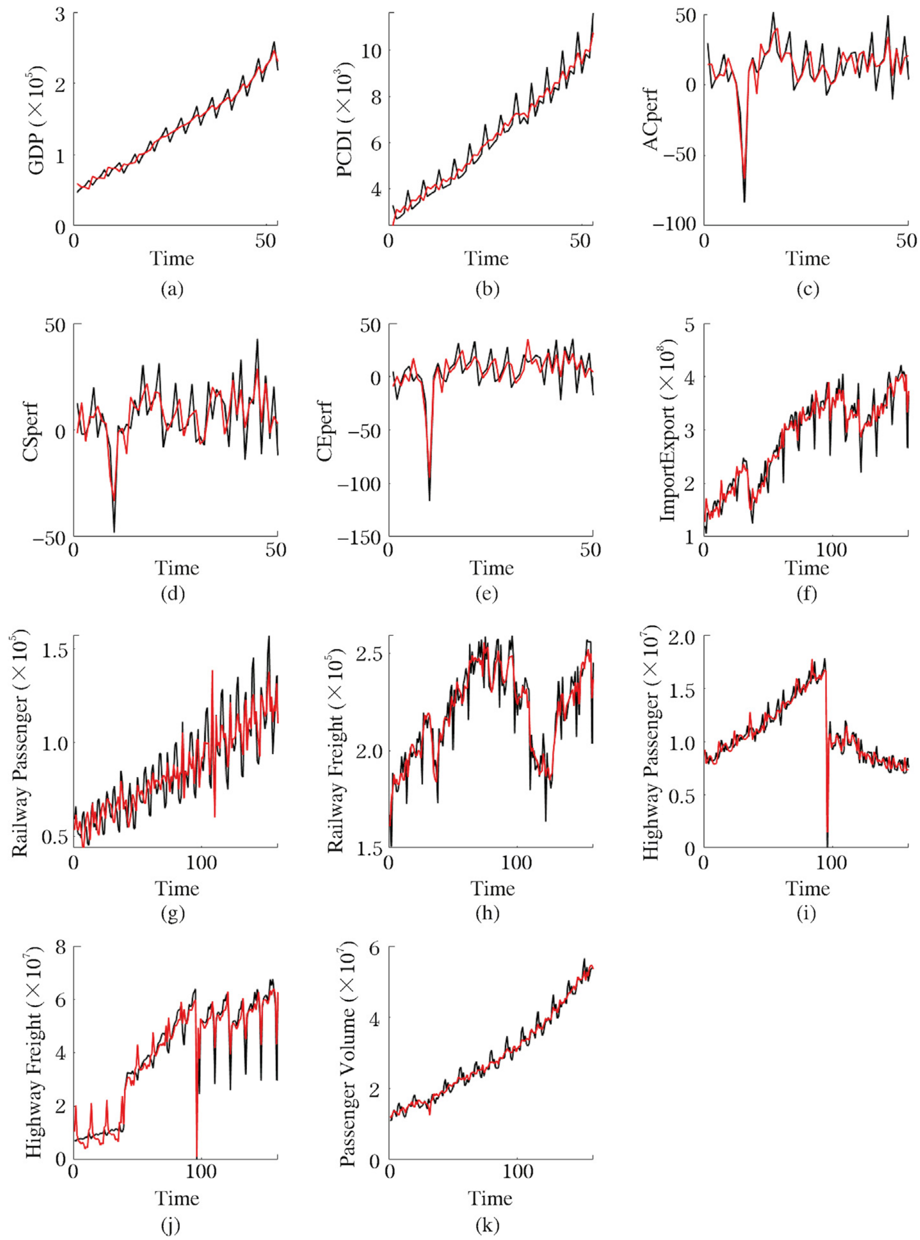

Fig. 2. Plots of predictors that exhibited seasonality before and after adjustment. The black lines correspond to the raw data and the red lines correspond to the data after adjustment for seasonality.

Table 4
Performance measures for the forecast accuracy comparison for three airlines.

Airline	Index	30-day forecast using				
		LASSO	RR	TR	SVR	NN
Air China	MAPE (%)	2.17	13.47	5.61	11.52	21.33
	RMSE	0.28	1.52	0.72	1.15	3.03
	MAE	0.20	1.33	0.55	0.99	2.31
China Eastern Airlines	MAPE (%)	1.94	5.56	5.84	19.84	19.63
	RMSE	0.16	0.50	0.45	1.27	1.77
	MAE	0.12	0.38	0.36	1.09	1.38
China Southern Airlines	MAPE (%)	2.06	5.12	5.77	9.52	26.54
	RMSE	0.24	0.62	0.58	0.86	3.11
	MAE	0.17	0.47	0.47	0.69	2.50

Note: The least absolute shrinkage and selection operator (LASSO); ridge regression (RR); tree regression (TR); support vector regression (SVR); multilayer perception neural networks (NN).

available. The Brent oil price is ranked as fairly influential in this model, but the exchange rate was again completely shrunk from the model.

To test this LASSO-based model's capacity to predict other datasets, we use the test dataset. The coefficients we use in the model are those estimated from the training data, but the autoregressive and lagged response predictor values are replaced with the test set data as appropriate. These coefficients are used to predict the future daily stock close data in the series at each point; the actual data are shown (black lines) with the LASSO predicted values (red lines) superimposed in the plots. Fig. 3 shows the actual data plotted 30 days ahead with the prediction for that same corresponding day superimposed, the agreement between two series, and the predictions achieved via other methods we considered. We can see that LASSO outperforms the alternatives.

Moreover, to illustrate the significance between the LASSO results with considered benchmark models, we employ the *t*-test to identify variations in absolute errors between predictions and actual observations. Table 6 succinctly presents the outcomes, where all *p*-values are approximately 0. This indicates that the predictions generated by both

Table 5
Ranked the least absolute shrinkage and selection operator (LASSO) estimated coefficients of the autoregressive distributed lag model predicting three airlines' stock prices.

Rank	Air China		China Eastern Airlines		China Southern Airlines	
	Variable	Estimates	Variable	Estimates	Variable	Estimates
1	AC close - lag 1	0.98	CE close - lag 1	0.93	CS close -lag 1	0.97
2	CS close - lag 13	0.0007	AC close - lag 12	0.0088	AC close - lag 1	0.0005
3	CS close - lag 1	0.00003	AC close - lag 1	0.0080	CS close - lag 13	0.0003
4	-	-	CE close - lag 25	0.0071	Interest rate - lag 11	0.000021
5	-	-	Brent - lag 20	0.00026	S close - lag 1	2.11E-05
6	-	-	CE close - lag 27	0.001598	S low - lag 3	8.20E-06
7	-	-	CE net profit - lag 6	0.00012	Interest rate - lag 20	−0.00306
8	-	-	S close - lag 1	7.61E-05	Interest rate - lag 3	−0.00606
9	-	-	S open - lag 30	2.31E-06	Interest rate - lag 15	−0.00901
10	-	-	S low - lag 3	3.17E-07	Interest rate - lag 12	−0.00951
11	-	-	Railway freight - lag 1	−4.08E-07	Highway passenger - lag 15	−2.90E-09
12	-	-	Highway passenger - lag 11	−2.44E-09	Highway passenger - lag 18	−4.64E-09
13	-	-	Highway passenger - lag 18	−7.03E-09	S volume - lag 8	1.36E-10
14	-	-	S volume - lag 8	3.17E-11	S volume - lag 3	1.43E-11
15	-	-	Highway freight - lag 18	4.03E-11	S volume - lag 1	1.45E-12
16	-	-	S volume - lag 1	1.89E-12	-	-

Note: Coefficients with the same name correspond to different lags of the same variable.

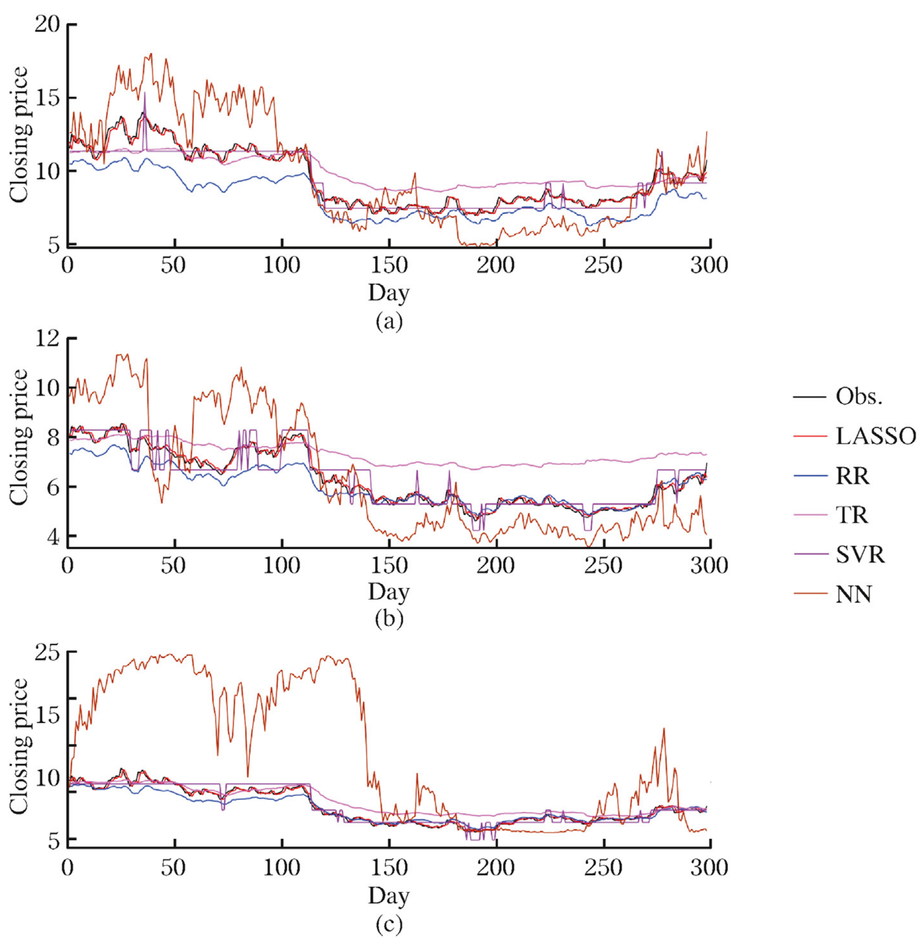

Fig. 3. Plots of the predictions achieved via investigated models. (a) Air China, (b) China Eastern Airlines and (c) China Southern Airlines.
Note: The least absolute shrinkage and selection operator (LASSO); observation (Obs.); Ridge regression (RR); tree regression (TR); support vector regression (SVR); neural networks (NN).

LASSO and the alternative models remain statistically significant across all three airlines.

4.2. Further discussion on economic insights from feature selection using LASSO

Through the analysis of Table 4, the following can be concluded:

(1) The shares of China's three major airlines are closely correlated with the previous trading day's closing price, with correlation coefficients as high as 93%–98%. For Chinese A-shares, the daily closing price fluctuates between ± 10% of the previous day's closing price; therefore, the three airlines' stock prices have the greatest correlation with the closing price of the previous day's stock. In addition, because there is a certain trend in the decline or

rise of stocks, the closing price of the stock has a certain correlation with the historical closing price.

(2) An airline's closing price and other airlines' recent closing prices present a positive correlation. When shares of similar airlines rise rapidly, there is usually hot news or policy support. After hot news forms, the same type of airline will enjoy a money-making effect, thus attracting more investment to the airline stock. Thus, the supply and data type of airline stock influence each other. However, this impact may be delayed for a certain period.

(3) Airline stocks have also been positively affected by the trading prices (opening, closing, and lowest prices) of the Shanghai composite index in recent days, among which the most significant impact is the price of the Shanghai composite index in the previous trading day. In addition, share prices are affected by the volume of recent trading in the Shanghai composite index, with the greater trading volume resulting in higher prices of airline shares. General trading volume continues to grow with the arrival of a bull market. Conversely, ongoing shrinking trading volume is generally considered to be one of the characteristics of a bear market. Therefore, airlines' share prices are closely related to the boom degree of the stock market. It is difficult for airline stock prices to rise independently when the stock market is experiencing a downward trend and shrinking trading volume.

(4) The stock price of an airline is negatively correlated with railway freight and highway passenger turnover. With the construction of increasingly more high-speed railways and expressways in China, high-speed rail and highways have become a substitute for aviation, so many passengers or goods are transported by train or car rather than airplane. The increase in passenger and freight traffic by railway or high-speed railway will reduce airline passenger and freight traffic, which impacts their performance, and in turn, their stock price.

(5) Airline stock closing prices are negatively affected by short-term loan interest rates. The reduction in bank lending rates will indirectly benefit the increase in airline share prices. First of all, after the bank loan interest rate is lowered, the cost of capital use will fall, thus allowing more funds to actively participate in the stock market. Second, lower lending rates can stimulate consumption and investment, which not only promotes business travel and trips but also boosts investor confidence in buying airline stocks. Airlines can also give back to shareholders by increasing their holdings and paying dividends, thus forming a virtuous circle. The same is true of higher interest rates on loans.

(6) Xu et al. (2021) found that oil prices and exchange rates affected airline performance, but the effect on stock prices was not obvious. From the three models explored in this paper, there is no significant evidence that oil prices or exchange rates affect stock prices. China's stock market has long been criticized as an immature market, with a large, well-funded retail base but a lack of performance analysis skills. It can be concluded from the three models that Chinese investors' preference for technical analysis is far greater than that for fundamental analysis, and their preference for short-term analysis is far greater than that for long-term analysis.

(7) The impact of airline net profit on stock prices is not significant. On one hand, although exchange rates and oil prices affect performance, the impact on stock prices is relatively limited because stocks are affected by macro policies, alternative modes of transportation, interest rates, the strength of the stock market, and other factors. On the other hand, Chinese investors pay more attention to short-term stock trends, peer stock trends and capital intervention are not very sensitive to company performance.

5. Conclusion

Although a great deal of literature has focused on finding methods for stock price prediction, it remains a highly challenging task with much scope for further exploration to improve prediction capabilities.

We have presented a simple but effective model fitted via the LASSO approach for improving predictive capabilities around three major Chinese airlines' stock closing prices. The model can be utilized to predict the closing prices using lags of predictor variables as far as 30 days in advance. The empirical results indicated that the LASSO-fitted model outperformed the alternative approaches we considered. The key contributions of this study can be described as the proposal of a flexible methodology for stock prediction and the great potential for the transferability of these ideas to other similarly behaving stocks.

In sum, the LASSO model illustrates that the main factors that affect the closing price of airline stocks are the previous day's closing price of the stock and the previous day's closing price of the other two airlines. The less influential factors include the prices of the previous day and three days prior in the Shanghai composite index, the share volume of the Shanghai composite index, highway passenger turnover, short-term loan interest rates, railway and highway freight turnovers, and airline net profit. The non-significant factors are exchange rate, import and export trade volume, aviation sector composite index, gross domestic product (GDP), etc.

The results of our study might change the approach investors and others take when seeking to predict airline stock prices. While much of the existing literature on stock price prediction agrees that it is worthwhile to take account of well-known influential external factors, we have shown that competitors' performance-related data also have a role to play in improving predictive capabilities for the stock of interest, thus making an additional contribution to the literature.

Table 6
Results of statistical tests between LASSO and other benchmark models for three airlines.

Model	Air China		China Eastern Airlines		China Southern Airlines	
	t-satistics	*p*-value	*t*-satistics	*p*-value	*t*-satistics	*p*-value
RR	−27.48	0	−13.84	0	−13.23	0
TR	−22.45	0	−25.16	0	−16.92	0
SVR	−12.73	0	−14.98	0	−15.75	0
NN	−21.73	0	−25.67	0	−19.11	0

Note: ridge regression (RR); tree regression (TR); support vector regression (SVR); neural networks (NN).

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the Australian Research Council project (Grant No.: DP160104292), the Zhejiang Province Soft Science Project, the Wenzhou Basic Soft Science Research Key Project (First Batch, NO.7), and "Chunhui Program" Collaborative Scientific Research Project (Grant No.: 202202004).

References

Abraham, R., Samad, M.E., Bakhach, A.M., et al., 2022. Forecasting a stock trend using genetic algorithm and random forest. J. Risk Financ. Manag. 15 (5), 188.
Alam, I.M.S., Sickles, R.C., 1998. The relationship between stock market returns and technical efficiency innovations: evidence from the US airline industry. J. Prod. Anal. 9 (1), 35–51.
Arekar, K., Jain, R., 2017. Influence of oil price volatility of developed countries on emerging countries stock market returns by using threshold based approach. Theor. Econ. Lett. 7 (6), 1834–1847.
Bao, Y., Lu, Y., Zhang, J., 2004. Forecasting stock price by SVMs regression. In: International Conference on Artificial Intelligence: Methodology, Systems, and Applications. Springer, pp. 295–303.
Birkes, D., Dodge, Y., 2011. Alternative Methods of Regression. John Wiley & Sons, New York.

Chandola, D., Mehta, A., Singh, S., et al., 2022. Forecasting directional movement of stock prices using deep learning. Ann. Data Sci. 10 (5), 1361–1378.
Deng, C., Huang, Y., Hasan, N., et al., 2022. Multi-step-ahead stock price index forecasting using long short-term memory model with multivariate empirical mode decomposition. Inf. Sci. 607 (Aug.), 297–321.
Dimitriadou, E., Hornik, K., Leisch, F., et al., 2008. Misc functions of the Department of Statistics (e1071), TU Wien. R Package 1 (Jan.), 5–24.
Forsyth, P., Dwyer, L., 2010. Exchange rate changes and the cost competitiveness of international airlines: the aviation trade weighted index. Res. Transport. Econ. 26 (1), 12–17.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Software 33 (1), 1.
Gandhmal, D.P., Kumar, K., 2019. Systematic analysis and review of stock market prediction techniques. Comput. Sci. Rev. 34 (Nov.), 100190.
Goh, C.F., Rasli, A., 2014. Stock investors' confidence on low-cost and traditional airlines in Asia during financial crisis of 2007–2009. Procedia Soc. Behav. Sci. 129 (May), 31–38.
Gong, S.X., Firth, M., Cullinane, K., 2008. International oligopoly and stock market linkages: the case of global airlines. Transp. Res. E Logist. Transp. Rev. 44 (4), 621–636.
Günther, F., Fritsch, S., 2010. Neuralnet: training of neural networks. R J 2 (1), 30.
Hamrita, M.E., Trifi, A., 2011. The relationship between interest rate, exchange rate and stock price: a wavelet analysis. Int. J. Econ. Financ. Issues 1 (4), 220–228.
Kamara, A.F., Chen, E., Pan, Z., 2022. An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices. Inf. Sci. 594 (May), 1–19.
Kirisci, M., Cagcag Yolcu, O., 2022. A new CNN-based model for financial time series: TAIEX and FTSE stocks forecasting. Neural Process. Lett. 54 (4), 3357–3374.
Kurani, A., Doshi, P., Vakharia, A., et al., 2023. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Ann. Data Sci. 10 (1), 183–208.
Li, J., Chen, W., 2014. Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. Int. J. Forecast. 30 (4), 996–1015.
Lin, Y., Lin, Z., Liao, Y., et al., 2022. Forecasting the realized volatility of stock price index: a hybrid model integrating CEEMDAN and LSTM. Expert Syst. Appl. 206 (Nov.), 117736.
Liu, J., Kemp, A., 2019. Forecasting the sign of us oil and gas industry stock index excess returns employing macroeconomic variables. Energy Econ. 81 (Jun.), 672–686.
Liu, M., Choo, W.C., Lee, C.C., et al., 2023. Trading volume and realized volatility forecasting: evidence from the China stock market. J. Forecast. 42 (1), 76–100.
Ludwig, N., Feuerriegel, S., Neumann, D., 2015. Putting big data analytics to work: feature selection for forecasting electricity prices using the lasso and random forests. J. Decis. Syst. 24 (1), 19–36.
Md, A.Q., Kapoor, S., AV, C.J., et al., 2023. Novel optimization approach for stock price forecasting using multi-layered sequential LSTM. Appl. Soft Comput. 134 (Feb.), 109830.
Mok, H.M., 1993. Causality of interest rate, exchange rate and stock prices at stock market open and close in Hong Kong. Asia Pac. J. Manag. 10 (Oct.), 123–143.
Naser, H., Alaali, F., 2018. Can oil prices help predict US stock market returns? Evidence using a dynamic model averaging (DMA) approach. Empir. Econ. 55 (4), 1757–1777.
Nowak, S., Anderson, H.M., 2014. How does public information affect the frequency of trading in airline stocks? J. Bank. Finance 44 (Jul.), 26–38.
Oztekin, A., Kizilaslan, R., Freund, S., et al., 2016. A data analytic approach to forecasting daily stock returns in an emerging market. Eur. J. Oper. Res. 253 (3), 697–710.
Panagiotidis, T., Stengos, T., Vravosinos, O., 2018. On the determinants of bitcoin returns: a lasso approach. Finance Res. Lett. 27 (Dec.), 235–240.
Peng, Y., Chen, W., Wei, P., et al., 2020. Spillover effect and Granger causality investigation between China's stock market and international oil market: a dynamic multiscale approach. J. Comput. Appl. Math. 367 (C), 112460.
Roy, S.S., Mittal, D., Basu, A., et al., 2015. Stock market forecasting using lasso linear regression model. In: Afro-European Conference for Industrial Advancement. Springer, Cham, pp. 371–381.
Sadorsky, P., 2022. Forecasting solar stock prices using tree-based machine learning classification: how important are silver prices? N. Am. J. Econ. Finance 61 (Jul.), 101705.
Siami-Namini, S., 2017. Granger causality between exchange rate and stock price: a Toda Yamamoto approach. Int. J. Econ. Financ. Issues 7 (4), 603–607.
Sonenshine, R., Cauvel, M., 2017. Revisiting the effect of crude oil price movements on US stock market returns and volatility. Mod. Econ. 8 (5), 753.
Song, Y.I., Woo, W., Rao, H.R., 2007. Interorganizational information sharing in the airline industry: an analysis of stock market responses to code-sharing agreements. Inf. Syst. Front 9 (Apr.), 309–324.
Staffini, A., 2022. Stock price forecasting by a deep convolutional generative adversarial network. Front. Artif. Intell. 5 (Feb.), 837596.
Team, R.C., 2013. R: A Language and Environment for Statistical Computing. Available at: https://www.r-project.org/.
Therneau, T., Atkinson, B., Ripley, B., et al., 2015. Package 'rpart'. Available at: https://cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol. 58 (1), 267–288.
Webel, K., Ollech, D., 2018. An overall seasonality test based on recursive feature elimination in conditional random forests. In: Proceedings of the 5th International Conference on Time Series and Forecasting, pp. 20–31.
Xu, X., McGrory, C.A., Wang, Y.G., et al., 2021. Influential factors on Chinese airlines' profitability and forecasting methods. J. Air Transport. Manag. 91 (Mar.), 101969.
Zhou, Z., Gao, M., Liu, Q., et al., 2020. Forecasting stock price movements with multiple data sources: evidence from stock market in China. Physica A Stat. Mech. Appl. 542 (Mar.), 123389.