Xu, Xu, McGrory, Clare, Wang, You-Gan, & Wu, Jinran
(2021)
Influential factors on Chinese airlines' profitability and forecasting methods.
Journal of Air Transport Management, *91*, Article number: 101969.

https://doi.org/10.1016/j.jairtraman.2020.101969

Highlights

Influential factors on Chinese airlines profitability and forecasting methods

Xu Xu,Clare Anne McGrory,You-Gan Wang,Jinran Wu

- Effects of a number of revenue and cost factors on airlines profitability are explored
- A number of modern approaches including LASSO and machine learning are investigated
- LASSO effectively removes redundant variables generally outperforming other approaches
- Oil price negatively impacts short term profit of the Chinese airlines

Influential factors on Chinese airlines profitability and forecasting methods

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

[a]College of Mathematics and Physics, Wenzhou University, P.R. China
[b]School of Mathematics and Information Engineering, Oujiang College, Wenzhou University, P.R. China
[c]School of Mathematical Sciences, Queensland University of Technology, Australia
[d]ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, Australia

ARTICLE INFO

Keywords:
Airline
Fuel Costs
Profitability
LASSO
Forecasting

ABSTRACT

We establish profit models to predict the performance of airlines in the short term using the quarterly profit data collected on the three largest airlines in China together with additional recent historical data on external influencing factors. In particular, we propose the application of the LASSO estimation method to this problem and we compare its performance with a suite of other more modern state-of-the-art approaches including ridge regression, support vector regression, tree regression and neural networks. It is shown that LASSO generally outperforms the other approaches in this study. We concluded a number of findings on the oil price and other influential factors on Chinese airline profitability.

1. Introduction

Oil prices and exchange rate fluctuations can be very large, and the performance of the air transport industry is highly sensitive to both of these. So, what is the impact of oil prices and exchange rate on Chinese airlines' profitability? Can the characteristics of fluctuations in oil prices and exchange rate over time be combined with other influencing factors to predict the short, and medium-term, profits of airlines? Can a forecasting method be found for airline managers and institutional investors such as insurance or social security funds? In this article we use and compare some very modern statistical computational methodologies to solve the problems we describe here.

As an important branch of the transportation industry, the air transportation industry has developed rapidly in recent years. Its development reflects the level of national economic and social modernization around the world. Airline companies in China mainly provide passenger, freight and postal services to China, Hong Kong, Macao and Taiwan as well as services with partner international airlines. By the end of 2018, China's civil air transport volume had reached 612 million passengers, up 10.9% over the previous year. The total transport turnover reached 120.653 billion ton-km, up 11.4% over the previous year (data source: 2018 civil aviation industry development statistical bulletin).

With the gradual globalization of the world economy, on one hand, the improvement of the global supply chain will certainly expand the demand for air transport and drive the sustained and rapid development of the air transport industry. On the other hand, with the gradual widening of China's civil aviation transportation market, the competition among airlines in various aspects such as scale, price and service is strong, making it harder for airlines to remain profitable and survive in the market.

In recent years, the popularity of China's high-speed rail network has grown, and there have been improvements in the highway network between cities. In this way, railway and road transportation, which has a relatively lower cost advantage, has also formed a competitive pressure on the aviation transportation industry. The air transport industry itself is also greatly affected by other factors in the external environment: terrorist attacks, natural disasters, strikes and unexpected public health events will affect the normal operation of airlines, thus affecting the companies' performance and long-term development.

More accurate prediction of the airlines performance, based on easily recorded known influential factors, is therefore potentially very important as a means to mitigate impacts from those external factors that can be predicted. Based on reliable predictions, the airlines might more effectively adjust their company's development strategy, use pricing contracts such as fuel hedging as part of their risk management plans, and adjust fares and purchasing plans, among other things. For investors, more insights into the airlines' performance might impact on investment choices in that type of business.

This paper will consider quarterly profit data collected on the three largest airlines in China as research samples, along with other external data sources, to train and test profit forecast models. These can provide practical guidance to decision-makers and investors. China's three largest carriers, usually referred to as the "big three", are Air China, China Southern Airlines and China Eastern Airlines.

Air China was formally established in Beijing in 1988 as the main air transport company controlled by the China national aviation corporation. Air China has the most developed route network and the largest annual passenger volume.

Founded in 1995, China Southern Airlines, headquar-

*Corresponding author
xuxu@wzu.edu.cn (X. Xu); c.mcgrory@qut.edu.au (C.A. McGrory); you-gan.wang@qut.edu.au (Y. Wang); jinran.wu@hdr.qut.edu.au (J. Wu)
ORCID(s): 0000-0001-7545-3491 (X. Xu); 0000-0002-7364-2149 (C.A. McGrory); 0000-0003-0901-4671 (Y. Wang); 0000-0002-2388-3614 (J. Wu)

Influential factors on Chinese airlines profitability and forecasting methods

Anonymous Author[a]

ARTICLE INFO

Keywords:
Airline
Fuel Costs
Profitability
LASSO
Forecasting

ABSTRACT

We establish profit models to predict the performance of airlines in the short term using the quarterly profit data collected on the three largest airlines in China together with additional recent historical data on external influencing factors. In particular, we propose the application of the LASSO estimation method to this problem and we compare its performance with a suite of other more modern state-of-the-art approaches including ridge regression, support vector regression, tree regression and neural networks. It is shown that LASSO generally outperforms the other approaches in this study. We concluded a number of findings on the oil price and other influential factors on Chinese airline profitability.

1. Introduction

Oil prices and exchange rate fluctuations can be very large, and the performance of the air transport industry is highly sensitive to both of these. So, what is the impact of oil prices and exchange rate on Chinese airlines' profitability? Can the characteristics of fluctuations in oil prices and exchange rate over time be combined with other influencing factors to predict the short, and medium-term, profits of airlines? Can a forecasting method be found for airline managers and institutional investors such as insurance or social security funds? In this article we use and compare some very modern statistical computational methodologies to solve the problems we describe here.

As an important branch of the transportation industry, the air transportation industry has developed rapidly in recent years. Its development reflects the level of national economic and social modernization around the world. Airline companies in China mainly provide passenger, freight and postal services to China, Hong Kong, Macao and Taiwan as well as services with partner international airlines. By the end of 2018, China's civil air transport volume had reached 612 million passengers, up 10.9% over the previous year. The total transport turnover reached 120.653 billion ton-km, up 11.4% over the previous year (data source: 2018 civil aviation industry development statistical bulletin).

With the gradual globalization of the world economy, on one hand, the improvement of the global supply chain will certainly expand the demand for air transport and drive the sustained and rapid development of the air transport industry. On the other hand, with the gradual widening of China's civil aviation transportation market, the competition among airlines in various aspects such as scale, price and service is strong, making it harder for airlines to remain profitable and survive in the market.

In recent years, the popularity of China's high-speed rail network has grown, and there have been improvements in the highway network between cities. In this way, railway and road transportation, which has a relatively lower cost advantage, has also formed a competitive pressure on the aviation transportation industry. The air transport industry itself is also greatly affected by other factors in the external environment: terrorist attacks, natural disasters, strikes and unexpected public health events will affect the normal operation of airlines, thus affecting the companies' performance and long-term development.

More accurate prediction of the airlines performance, based on easily recorded known influential factors, is therefore potentially very important as a means to mitigate impacts from those external factors that can be predicted. Based on reliable predictions, the airlines might more effectively adjust their company's development strategy, use pricing contracts such as fuel hedging as part of their risk management plans, and adjust fares and purchasing plans, among other things. For investors, more insights into the airlines' performance might impact on investment choices in that type of business.

This paper will consider quarterly profit data collected on the three largest airlines in China as research samples, along with other external data sources, to train and test profit forecast models. These can provide practical guidance to decision-makers and investors. China's three largest carriers, usually referred to as the "big three", are Air China, China Southern Airlines and China Eastern Airlines.

Air China was formally established in Beijing in 1988 as the main air transport company controlled by the China national aviation corporation. Air China has the most developed route network and the largest annual passenger volume.

Founded in 1995, China Southern Airlines, headquartered in Guangzhou, is the largest carrier in China with the largest number of transport aircraft.

China Eastern Airlines is a state-owned holding airline headquartered in Shanghai, China. Based on the former China Eastern Airlines group, China Eastern Airlines merged China northwest Airlines with China Yunnan Airlines. It is the first Chinese civil aviation airline listed in Hong Kong, New York and Shanghai.

Since their inception, the "big three" airlines have con-

*Corresponding author
anonymous@com (A. Author)
ORCID(s):

tinued to grow in financial and operational scale, showing great momentum to become the leading international aviation market. As of 2018, China's "big three" aircraft fleet is only second to the five big airlines in the United States in size. In recent years, it has been rising in terms of revenue, mileage and passenger volume. From the perspective of scale and influence, the three major airlines are taken as samples, which we expect provide a good indicator of the current trends in the overall air transportation market in China.

The existing research on airline performance is generally limited to looking for the operating and macro reasons that affect performance; this is also the case in the mainstream forecasting literature in general, see e.g. Li (2014) and Verstraete and Aghezzaf (2020). This article differs from the majority of the mainstream literature in that we establish a multi-factor prediction model based on multiple external influencing factors.

Because we can form the airline profit modeling problem as the solution to a statistical model that involves the traditional time series autoregressive distributed lag model with a large number of predictors, we propose that the LASSO is one of the most suitable approaches for this problem. The LASSO of Tibshirani (1996) is a modern alternative to the ordinary least squares regression approach (Birkes, 1993). The LASSO approach is preferable when there are a large number of predictors which also might be highly correlated with one another; ordinary least squares approaches can be unreliable in such settings.

The remainder of this article is organised as follows. In Section 2 we review the relevant literature. Section 3 outlines our proposed modelling framework and gives the results of our analysis. Section 4 concludes the article.

2. Literature Review

Data on 35 large scheduled passenger airlines from 1957 to 1986 are used by Spiller (2006) to estimate the effect of profitability and other aspects of financial health on accident and incident rates. The results indicated that lower profitability is correlated with higher accident and incident rates, particularly for smaller carriers. Therefore improved predictions of airline profitability, which is the aim of this article, is not only motivated by the desire to help investors and owners of air transportation businesses with their income, but by the socially responsible goal of improving air passenger safety.

There is extensive literature in a number of fields related to the profitability of airline companies from the perspectives of airline internal governance and cost control. Two key operational factors that critically impact airline profitability are the design of a network of daily flights (Nero, 1999) and the operational capability to turn around aircrafts quickly (Van Landeghem, 2002). Most articles analyze the causes of their fluctuations of profitability based on historical data or economic theories.

Besides the internal factors of airline companies, there are also quite a few studies on the impact of external environmental changes on the profitability of airlines. Airline profits are well known to be affected by the changes in the market price of crude oil. For instance, Narayan and Sharma (2011) found a significant impact of oil price movements on individual firm returns. In 2015, Miranda (2015) aimed to determine if a combination of terrorist activity and the price of petroleum more accurately predicted airline profitability, and which variable was the most significant. The results suggested that both terrorism and the price of fuel were statistically significant, with the cost of fuel indicating a higher contribution to the model than terrorist activity. The results of the multiple linear regression analysis indicated the model was able to significantly predict airline profitability.

Another major influencer of airline profitability is the exchange rate. However, many published discussions of how changes in exchange rates impact on international airlines tend to be ambivalent. It has been shown that the relative cost competitiveness of an international airline will decline when there is an exchange rate appreciation in its home country as will be the case for other tradable goods and services (Forsyth, 2010).

It is well known that airline profit may be affected by both seasonal and economic cycles. Seasonality and its impact on forecasting have been long established for the airline industry. Airline companies' performance is affected by seasonality because the tourism industry presents obvious seasonality, and the number of tourists directly affects the seat vacancy rate of airline companies. It is clear to see that profits are significantly affected by the peak tourism season when the net profits in the financial statements from 2006 to 2019 of the three major airlines we study are considered. Due to the influence of the summer vacation season, except for 2008, 2009 and 2015, the three airlines had the highest profits in the third quarter, with an average profit of 2.225 billion CNY. See Figure 1. The second most profitable quarter, boosted by the winter holiday and spring break travel peaks, was the first quarter, with an average profit of 932 million CNY. Most of the profits in the second quarter ranked third, with an average profit of 785 million CNY. Most of the lowest profit quarters fell in the fourth quarter, with an average profit of only 947 million CNY, and with the airlines showing a net profit loss in the fourth quarter of most years. The economic cycle's significant impact on the performance of airlines is discussed in, for example, Goh and Rasli (2014).

Other external events affecting the whole market such as economic crises, also naturally have an impact on the air transportation industry (Navarro and Martínez, 2015). Whenever there is a financial crisis in the international economy or changes in the foreign economic policies of major countries, China's import and export may be blocked, leading to a decline in China's economic growth.

Due to various internal airline governance modes, and different means of cost control, internal factors can affect the performance of specific airline companies. However, this paper focuses on how the profit of airline companies will be affected when the external environment changes. The main external factors of airline performance are oil price and ex-

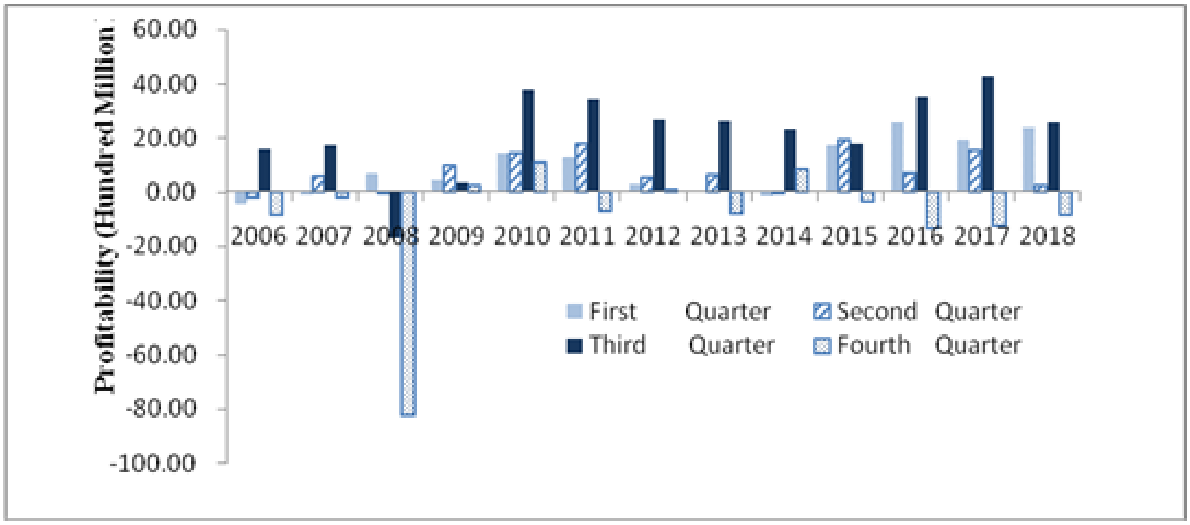

Figure 1: Average quarterly net profit of China's three major airlines from 2006 to 2019.

change rate according to the literature, therefore we explore whether it is possible to use these external factors to find an appropriate forecasting model.

In addition to oil price, exchange rate and seasonal factors, the following impacts should also be considered.

We consider the performance of other comparable businesses in predicting the business of interest. It makes intuitive sense that if the businesses are competing with one another and operating in similar market environments that they might give us useful information.

The prosperity of the civil aviation transportation industry is closely related to domestic and international macroeconomic development. Macroeconomic prosperity directly affects the development of economic activities, residents' disposable income and the increase and decrease of import and export trade volume, which affects the demand for air passenger transport and air freight. This in turn affects the business and operating performance of the group.

The macroeconomic policies formulated by the government, especially the cyclical macro policy adjustments such as credit, interest rate and fiscal expenditure, will also directly or indirectly affect the air transport industry.

Air transportation, railway transportation and highway transportation have certain substitutability in short and medium distance transportation. With the continuous improvement of the high-speed rail network, if an airline company fails to formulate effective marketing strategies to cope with the high-speed rail competition, it may affect the company's operating efficiency.

Finally we note that the aviation industry is greatly affected by the external environment, such as earthquakes, typhoons, tsunamis and other natural disasters, sudden public health events (COVID 19) (**?**) and terrorist attacks. All of these have a negative impact on the normal development of the airline. However, such force majeure and unforeseen risks need to be modeled separately to forecast airline profits; this is not within the scope of our paper.

3. Proposed Framework

The variables available for use in the model are described in Table 1. The chosen set of potential predictor variables aims to encompass all of the potentially influential external factors described above. Descriptive statistics for each of the variables are given in Table 2, and a pairwise scatter plot of the predictor variables against one another is shown in Figure 2.

3.1. Model Settings and Predictors

We model time series of Air China profit data, these are our responses $y_1, ..., y_T$ observed over T points in time. A multiple linear regression for a reponse at a given time point t takes the form:

$$y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_p x_{tp} + \epsilon,$$

where the $x_{t.}$'s are the observed values of the predictor variables at time point t, and the β_1 ... β_p are the slope co-efficients for each predictor. The value β_0 is a constant intercept term, and ϵ is the error term (residuals). In the special case where we are interested in incorporating lagged values of the response and predictor variables in the model as the β_1 ... β_p, the model we use is the autoregressive distributed lag model. We do this because we believe that effects in the response do not occur instantaneously, but instead in a dynamic fashion distributed over time. The slope co-efficients we estimate are referred to as the distributed lag weights and with these we aim to capture a distribution over the predictors that can more accurately anticipate profit changes.

More specifically, the responses we model here and some of the lagged predictors are the quarterly profits of the big three Chinese airlines. We have profit data for Air China, China Southern Airlines and China Eastern Airlines from the first quarter of 2006 through to the first quarter of 2019. We also plotted the autocorrelation functions for each of the "big three" profits timeseries. Inspecting Figure 3 suggests

Table 1
Data sources for the predictor variables.

Variable	Description
Oil Price	Brent crude oil spot price: Energy Information Administration(EIA)
Exchange Rate	Exchange rate (Spot rate: USD to CNY): China Foreign Exchange Trade System (CFETS)
Air China Profit	Profitability (net profit) from the published financial statements of Air China
China Southern Airlines Profit	Profitability (net profit) from the published financial statements of China Southern Airlines
China Eastern Airlines Profit	Profitability (net profit) from the published financial statements of China Eastern Airlines
GDP	Gross domestic profit in China
Railway Passengers	Railway passenger turnover in China
Railway Freight	Railway freight turnover in China
Highway Passengers	Highway passenger turnover in China
Highway Freight	Highway freight turnover in China
Passenger Volume	Civil aviation passenger volume in China
Interest Rate	Short term loan interest rate per year in China 6 months to 1 year inclusive
Labour	Average wages of employees: transportation, warehousing and postal industry in China
Import Export	Import and export amount in China
PCDI	Per capita disposable income of Chinese national residents

Table 2
Summary statistics for the predictor variables from the first quarter of 2006 to the first quarter of 2019

Variable	Mean	Standard Deviation	Range	Units
Oil Price	78.66	25.58	26.01 - 143.95	USD
Exchange Rate	6.73	0.53	6.04 - 8.07	USD to CNY
Air China Profit	12.32	19.76	-83.43 - 51.67	Billion CNY
China Southern Airlines Profit	6.64	15.19	-48.08 - 42.82	Billion CNY
China Eastern Airlines Profit	4.46	22.48	-116.35 - 35.68	Billion CNY
GDP	135775.10	54991.42	47078.90 - 258808.90	100 Million CNY
Railway Passengers	85090.13	24091.61	45168.00 - 157377.80	Million Persons
Railway Freight	218946.40	23947.69	150538.00 - 259215.60	Million Tons
Highway Passengers	10813475.00	2876492.00	1736.85 - 17808882.00	10,000 Persons
Highway Freight	38866779.00	19736107.00	6404.75 - 67621700.00	10,000 Tons
Passenger Volume	29020880.00	12081998.00	11000000.00 - 56567803.00	10,000 Persons
Interest Rate	5.59	0.94	4.35 - 7.47	%
Labour	53747.11	20360.67	24111.00 - 88508.00	CNY
Import Export	283229055.00	81471740.00	105792012.00 - 421683800.00	Thousand USD
PCDI	6240.31	2288.00	2704.00 - 11633.00	CNY

that correlation between each series and the same time one year previously is significant.

The Webel-Ollech overall seasonality test, which combines results from different seasonality tests, was used to test for seasonality in the data. As expected, the data relating to all of the "big three" airlines had statistically significant seasonility present, this was adjusted for before the modelling was carried out; this allows the dataset to achieve stationarity which is a condition we must satisfy in order to analyse the data in an autoregressive distributed lag regression model framework. By removing the long term seasonal factors in this way, our model should capture the shorter term relationships between our predictors and response that are of interest to us here, rather than capturing well-known long term behaviours. Timeseries plots of the profits for each of the airlines with their seasonally adjusted timeseries super-

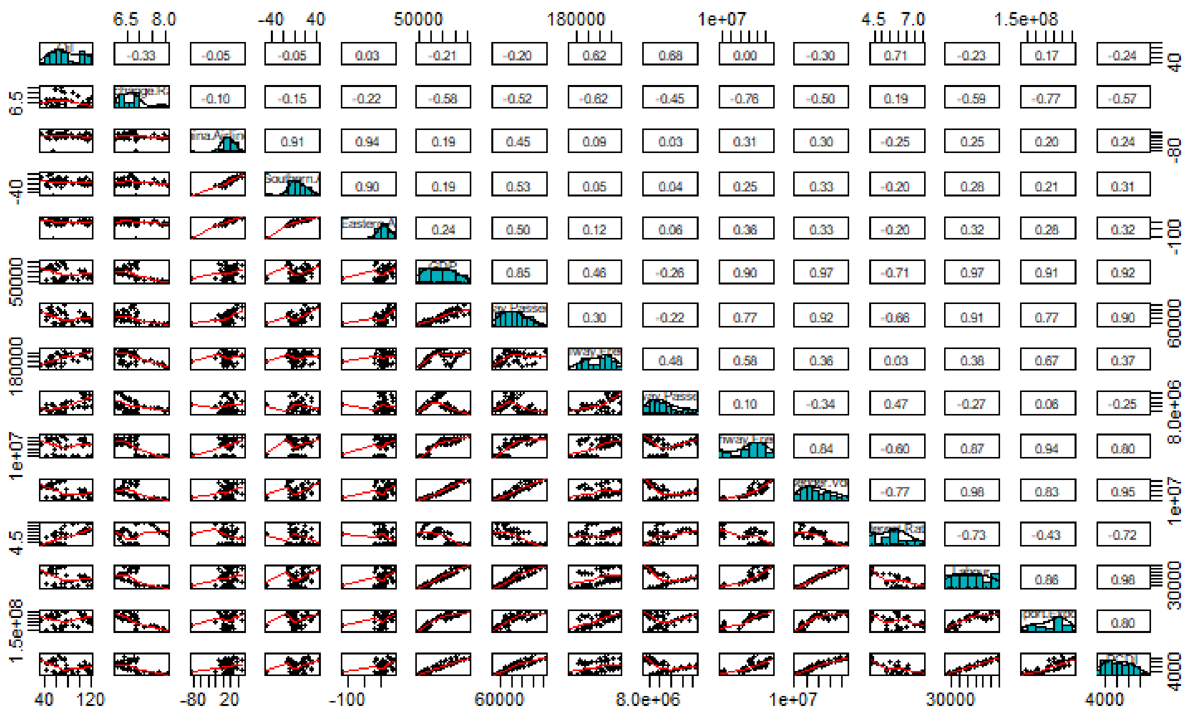

Figure 2: Pairwise scatter plots for each of the independent variables. The numbers in the upper panel are the Pearson correlation coefficients for each pair.

imposed are shown in Figures 4, 5 and 6. There is some similarity between the profits timeseries for each of the big three airlines; this is not surprising given that they are comparable businesses operating within the same economic zone.

Another two of the lagged predictors are the quarterly exchange rate for USD to CNY and the quarterly Brent crude oil price obtained by aggregating from the daily data. The effect of seasonality in the oil prices and in the exchange rate data was not statistically significant according to the test therefore only removing the trend from the exhange rate data was necessary in the data pre-processing stage. The timeseries plots of the predictors, also between the first quarter of 2006 through to the first quarter of 2019, are shown in Figures 7 and 8.

For the remaining lagged predictor variables, as listed in Table 1, we completed the same procedure of checking for seasonality and trend and adjusting for it where appropriate.

All of the statistical analyses performed in this article were carried out using the statistical programming language R (R Core Team, 2017).

3.2. Approaches for Modelling Airlines' Profits in Our Comparison Study

Because it is of interest to explore a large number of lagged covariates relative to the sample size, and there is a high likelihood of multicollinearity in the set of predictor variables, a standard regression might give unreliable parameter in this setting. We compare the out-of-sample predictive performance of several very modern alternatives to standard regression to our proposed approach.

3.2.1. LASSO Approach

As outlined by Tibshirani (1996) the least absolute shrinkage and selection operator, referred to as the LASSO, is a usfeul alternative to the standard linear regression model in this scenario. LASSO has been applied to many prediction fields, and has obtained a lot of success in research results, see e.g. Li (2014) and Verstraete and Aghezzaf (2020). For forecasting macroeconomic time series, Li (2014) shows that all of the LASSO-based models they considered outperformed dynamic factor models in out-of-sample forecast evaluations. Verstraete and Aghezzaf (2020) compares the LASSO with the standard regression approach in the context of tactical sales data forecasting demonstrating the value in using the LASSO instead.

LASSO is a form of penalized least squares estimation. The model estimates the set of weighted lag coefficients just as an ordinary least squares approach would, but the LASSO approach involves shrinkage of the estimated vector of regression coefficients towards zero, and in some cases some coefficients can be removed from the model. When the LASSO is fitted, it estimates values across a set of shrinkage parameters; the smallest estimated value of the shrinkage parameter being chosen as the critical and most useful value to use in

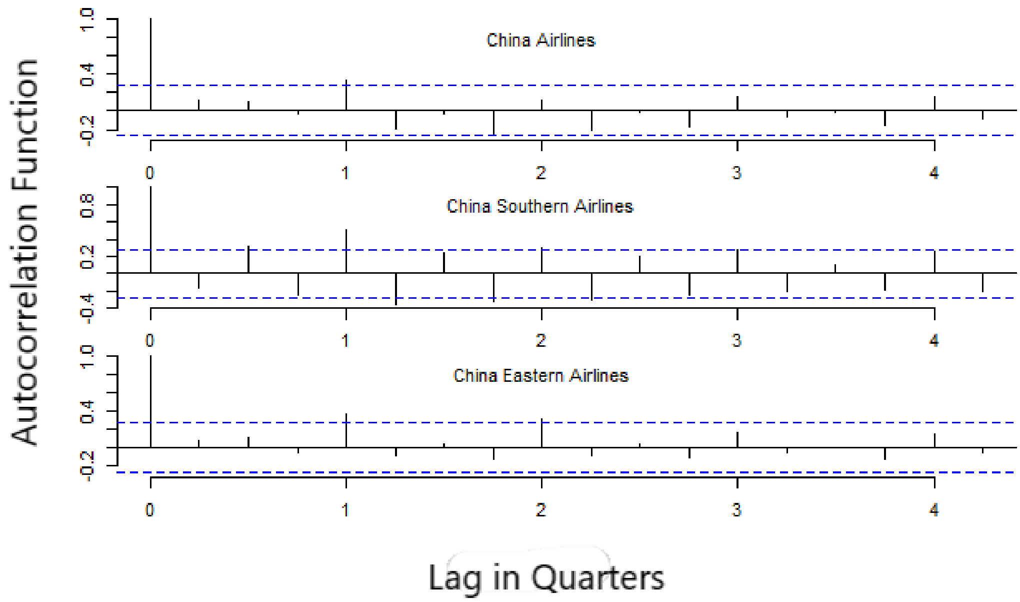

Figure 3: Autocorrelation function for each airline profits time series

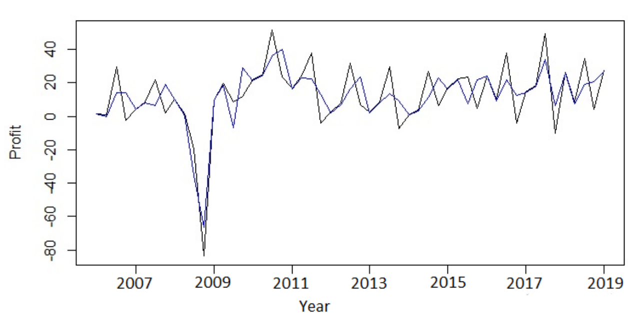

Figure 4: Quarterly profits (billion CNY) for China Airlines between the years 2006 to 2019. The black line is the raw data, and the blue line is the seasonally adjusted data.

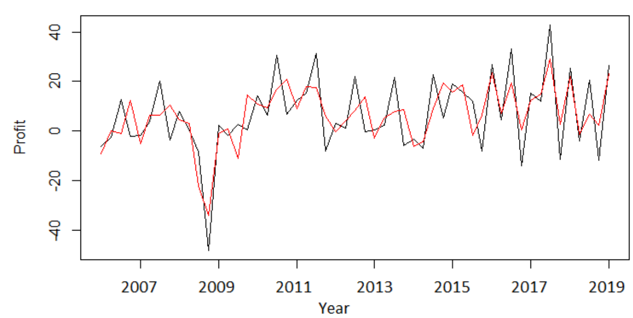

Figure 5: Quarterly profits (billion CNY) for China Southern Airlines between the years 2006 to 2019. The black line is the raw data, and the red line is the seasonally adjusted data.

co-efficient estimation from the resulting model fit.

Algebraically this can be expressed as minimizing:

$$\sum_t (y_t - \beta_0 - \sum_p \beta_p x_{tp})^2 + \lambda \sum_p |\beta_p|,$$

where $t > 0$ and $\lambda \geq 0$ is the shrinkage parameter that is chosen by performing cross-validation over a training set. The penalty term used is referred to as L1 regularisation. The LASSO approach can handle hundreds of data series simultaneously if necessary, and extract useful information for forecasting. With the gradual use of more modern statistical computational learning algorithms and artificial intelligence prediction methods, the LASSO model has been widely used in power, energy, company bankruptcy, stock price, macroeconomic indicators and other forecasting fields. German electricity spot prices are predicted in Ludwig and Feuerriegel (2015) with reference to historical prices and a set of weather variables using the LASSO; it manages to improve forecasting accuracy by up to 16.9% in terms of mean average error.

The LASSO has hardly been used in the prediction of the

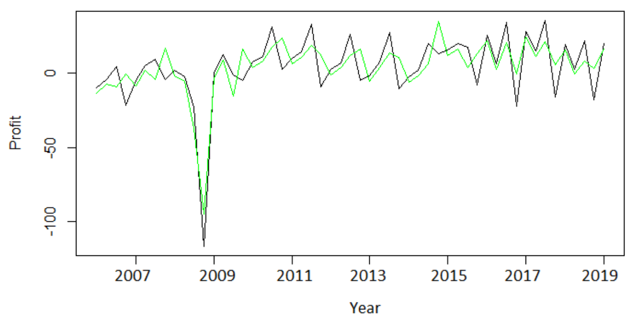

Figure 6: Quarterly profits (billion CNY) for China Eastern Airlines between the years 2006 to 2019. The black line is the raw data, and the green line is the seasonally adjusted data.

Figure 7: Monthly BRT oil prices

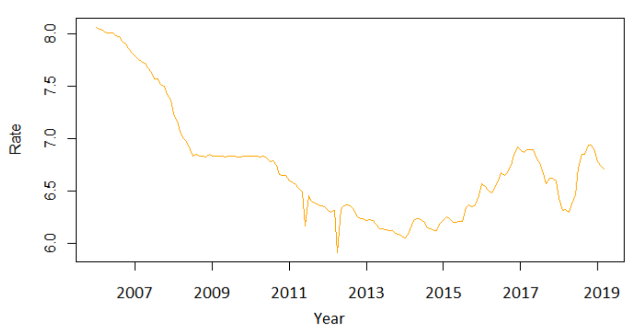

Figure 8: Monthly USD to CNY exchange rate

profits of companies with cyclical business activities. This paper shows how the LASSO model can be used to study the impact of crude oil price, exchange rate and other factors on airline profits. In particular, the LASSO estimation was done by making use of the package 'glmnet' (Friedman and Hastie, 2013) within the R software suite.

3.2.2. Ridge regression

The difference between the LASSO approach and ridge regression, is that where the LASSO uses the L1 regularizsation penalty, an L2 regularisation penalty term (the sum of the squares of the coefficients) is used instead in ridge regression. This approach is also implemented using the R package glment (Friedman and Hastie, 2013). Again, cross validation is used to select the best regularization parameter λ for our ridge regression training. See for example, Hoerl (1970).

3.2.3. Support vector regression

Support vector regression (SVR) is an efficient statistical learning method, it combines L2 regularisation with an insensitive Laplace distribution as a robust optimized objective for regression model training. For our SVR training, the R package e1071 is employed (Meyer et al. (2019)). The parameters in SVR are set as: the regularization coefficient 1, the insensitive parameter 0.1, the kernel radius basic function (rbf). See for example, Drucker et al. (1997).

3.2.4. Multilayer perceptron

A multilayer perceptron (MLP) is a type of feedforward artificial neural network. A MLP model is trained 100 times to the data here to obtain average performance using the package neuralnet (Gunther, 2010). More specifically, the MLP structure is set with three layers (20 input nodes, 10 hidden nodes, and 1 output node). Moreover, the logistic function is selected as the activation function in the hidden nodes. See for example, Pal (1992).

3.2.5. Bagging regression tree

As one of the ensemble learning methods, the bagging regression tree, is chosen as the benchmark model for these in our profitability modelling comparison. In this method, each regression tree is regarded as a voter to make a more intelligent group decision. Our experiment is carried out 100 times to achieve the average performance in R with the package caret (Kuhn, 2008). The default settings are employed in the model training. See for example, Sutton (2005).

3.3. Results

3.4. Training the Model

There are a total of 60 predictor variables used in the model. These are lags over the 4 preceding quarters for: the oil prices and exchange rate predictors, β_1 - β_4 and β_5 - β_8, respectively; the Air China profit β_9 - β_{12}; the China Southern Airlines profit β_{13} - β_{16}; the China Eastern Airlines profit β_{17} - β_{20}; GDP in China β_{21} - β_{24}; railway passengers in China β_{25} - β_{28}; railway freight in China β_{29} - β_{32}; highway passengers in China β_{33} - β_{36}; highway freight in China β_{37} - β_{40}; passenger volume in China β_{41} - β_{44}; the interest rate in China β_{45} - β_{48}; labour costs in China β_{49} - β_{52}; import and export amount in China β_{53} - β_{56}; and per capita disposible income of Chinese national residents β_{57} - β_{60}. This corresponds to the same 1-year period of historical predictive data being used for all predictors considered.

The LASSO is applied three times using each of the "big three" airlines in turn as a response. The R function cv.glmnet helps the user to select the most appropriate value for λ by performing n-fold cross-validation. In order to choose the most appropriate value for λ, the LASSO method extracts different values for it, such as λ_{min} that gives minimum mean cross-validated error. In each fit, we chose the values for the predictors $\beta_1, ..., \beta_{60}$ as those fitted when λ equals λ_{min}.

In each case, many of the weighted lag co-efficients have been shrunk to 0, that is, they are not all significant in the model in other words. The results obtained here and shown in Tables 3, 4 and 5 give us a valuable insight into which predictors are statistically significant as useful predictors of profit for the big three airlines.

The previous quarter's oil price has a significant effect on the net profits of Air China in the current quarter. Generally, the higher the oil price, the lower the profitability of the airline. The rise or fall of oil price itself has a certain degree of autocorrelation and periodicity, and the impact on profits will last for about 3 months. There is a lag in the impact of oil price fluctuations on profits because the rise and fall of oil has a more obvious trend in the short term, and it is also related to the lag of China Aviation Gasoline Price Adjustment for Crude Oil Price for a period.

Perhaps somewhat surprisingly, oil price was not found to be a significant predictor for the profits of the other two of the big three airlines. However, it is important to note the relationship between the oil and exchange rate variables because oil is traded in US dollars. The appreciation of CNY compared to the US dollar lowers the fuel cost for the airlines and should encourage Chinese passengers to travel abroad leading to higher demand and revenues. Conversely it becomes more expensive for American passengers to travel to China leading to lower demand and revenues. This means that the impact of an increasing oil price can be offset by an appreciation of the CNY and vice versa. As expected, the exchange rate was found to be a significant predictive factor for all three of the big airlines.

The model suggests that the performance of each airline company in the current quarter is significantly affected by the performance of itself and/or its peers in the previous quarter.

Using the LASSO model, it can be concluded that the civil aviation passenger volume in China is a direct influencing factor of airline's profit. Using the civil aviation passenger volume of the past quarter and the same period of last year can significantly help to predict the airline's profit. The larger the civil aviation passenger volume is, the better the future profit will be. In contrast, the impact of competitors such as railway and high-speed is relatively small, and after coefficient compression, they are all classified as non-significant factors and excluded from the model.

Interest rate has only become a significant factor in the prediction model for China Eastern Airlines and not the other two of the "big three". The lack of significance for the other two may be related to the relatively stable level of interest rate in China.

Average waves of employees have not become a significant factor in the prediction model for any of the airlines considered.

Table 3
Co-efficients Estimated from LASSO to Predict Air China's Profits

Variable	Co-Efficient
Intercept	-834.88
Oil Prices - lag 1	-0.09
Exchange Rate - lag 4	-13.74
Exchange Rate - lag 1	509.37
Air China Profit - lag 3	0.03
Air China Profit - lag 1	0.27
GDP - lag 2	-10.41
Passenger Volume - lag 4	81.86
Passenger Volume - lag 1	241.71
PCDI - lag 2	41.49

Table 4
Co-efficients Estimated from LASSO to Predict China Southern Airline's Profits

Variable	Co-Efficient
Intercept	-401.84
Exchange Rate - lag 2	13.79
Exchange Rate - lag 1	313.94
Air China Profit - lag 1	0.18
Passenger Volume - lag 4	27.73
Passenger Volume - lag 1	59.63
PCDI - lag 1	-9.26

Per capita disposable income increases are directly related to the increase of deposits, purchasing power and long-distance travel ability of Chinese residents. According to our model, increases in per capita disposable income in the short to medium term lead to increased profits for Air China, but a slight decrease in profits for the other two "big three" airlines.

Per capita disposable income of Chinese national residents and gross domestic profit in China are positively correlated. However, increased gross domestic profit demand surges drive oil price increases which in turn decrease airline profits. This is why there is a negative correlation between gross domestic product and profitability for the Air China and China Eastern Airlines models.

3.5. Forecasting Results

In this section we explore the performance of our model for forecasting future profits, comparing it with other very modern approaches. This means that the fitted model that was trained on part of the Air China data in Section 3.4 is used in a predictive capacity to estimate the timeseries for the remainder of the data. The weighted co-efficents we use in the model are those estimated from the training data, but the autoregressed response predictor values are replaced by

Table 5
Co-efficients Estimated from LASSO to Predict China Eastern Airline's Profit

Variable	Co-Efficient
Intercept	-429.09
Exchange Rate - lag 4	-369.35
Exchange Rate - lag 2	127.27
Exchange Rate - lag 1	375.42
Air China Profit - lag 3	0.03
China Southern Airlines Profit - lag 1	0.42
GDP - lag 2	-60.84
Passenger Volume - lag 4	317.16
Passenger Volume - lag 1	3.98
Interest Rate - lag 4	3.99
PCDI - lag 1	-15.82

the test set airline data as appropriate. Results are plotted in Figures 9, 10 and 11, and we can see that the model predicts profits well in the short term.

We also provide a comparison of our LASSO approach with four alternative forecasting approaches. The results are displayed in Table 8. The LASSO approach outperforms the others in terms of performance as measured by RMSE and MAE.

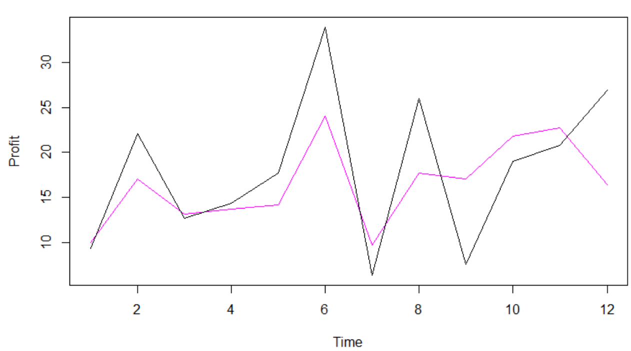

Figure 9: Predicted profits (billion CNY) for China Airlines superimposed upon the actual Profits (billion CNY) for the airline. The black line corresponds to the actual profits four quarters ahead, while the magenta line represents our model's predicted profits for the same period.

4. Conclusions

We have demonstrated a useful application of the LASSO method in areas where it has not been extensively used before and shown in our experiment that it generally out-performed four alternative state-of-the-art forecasting approaches. We also note that this type of model can also be extended to other industries with high external dependence, such as the chemical industry, petrochemical industry and steel industry, which have obvious periodicity and whose performance is affected by external factors such as oil prices and exchange rate.

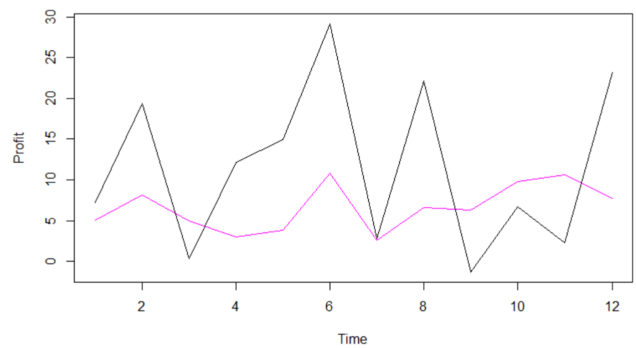

Figure 10: Predicted profits (billion CNY) for China Southern Airlines superimposed upon the actual Profits (billion CNY) for the airline. The black line corresponds to the actual profits four quarters ahead, while the magenta line represents our model's predicted profits for the same period.

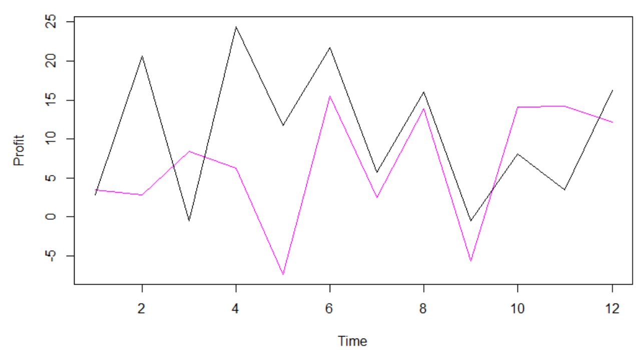

Figure 11: Predicted profits (billion CNY) for China Eastern Airlines superimposed upon the actual Profits (billion CNY) for the airline. The black line corresponds to the actual profits four quarters ahead, while the magenta line represents our model's predicted profits for the same period.

From our modelling we can draw the following key conclusions about Chinese airline profitability:

For Air China, the higher the oil price, the less short-term profit of the airline company. Following successful examples outside of China (American Southwest Airlines Corporation, Mexico Airlines), we would suggest that Air China should make use of crude oil futures, crude oil options, refined oil futures and options, which can effectively manage the sharp price risk of oil and ensure the smooth growth of profits.

The performance of each airline and its competitors is a good predictor of performance in the following quarter. When competitors or themselves make good profits in the past quarter, it means that the aviation industry is in a boom cycle, which is expected to continue for a period.

Our model suggests that the short-term adjustment range

Table 6
Performance measures for the forecast accuracy comparison when predicting China Airline's profits

Method	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)
LASSO	**5.97**	**4.71**
Ridge Regression	7.38	5.78
Support Vector Regression	9.84	8.20
Multilayer Perceptron	14.14	11.13
Bagging Tree Regression	9.94	7.63

Table 7
Performance measures for the forecast accuracy comparison when predicting China Southern Airline's profits

Method	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)
LASSO	10.45	**8.89**
Ridge Regression	**10.24**	9.21
Support Vector Regression	11.43	9.96
Multilayer Perceptron	12.78	11.56
Bagging Tree Regression	11.41	9.22

of employees' salary would have little impact on cost and profit.

China's GDP and per capita disposable income have been in a stable period showing consistent rises. Increases in GDP and per capita disposable income do not necessarily lead to increases in Chinese airlines' profits. This is most likely due to the fact that they drive up the price of oil and this is something that airlines are very exposed to.

Exchange rate is a very important predictor of Chinese airline profits for all of the "big three". The exchange rate has always been considered as the main factor affecting the performance of airlines so this is not a surprising outcome. This paper concludes that if the exchange rate of USD to CNY rises in the last quarter, meaning that overseas travel is more affordable for Chinese nationals, the profit in the subsequent quarter will increase. Conversely, if it has risen in the same period of last year, the profit will decrease. This may be related to the fact that the airline company sets the current fares according to the most recent quarters as well as the popularity of the same period last year. This means that airlines might set higher fares compared with the same period last year.

Another very important factor affecting the performance of airlines in the future is the passenger transportation volume of airlines in the past quarter, and that in the same period of last year. The transport volume plays an important role in the prediction model. The positive effect of the recent transport volume is obvious, followed by the transport volume of the same period last year.

Conflicts of interest

The authors state that there is no conflict of interests regarding the publication of this paper.

Table 8
Performance measures for the forecast accuracy comparison when predicting China Eastern Airline's profits

Method	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)
LASSO	**10.57**	**8.52**
Ridge Regression	11.65	9.52
Support Vector Regression	11.60	9.97
Multilayer Perceptron	14.30	11.40
Bagging Tree Regression	11.82	9.60

References

Birkes, D. & Dodge, Y., 1993. Alternative Methods of Regression. John Wiley and Sons, New York, New York, USA.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A. J. & Vapnik, V., 1997. Support vector regression machines. In Advances in neural information processing systems , 155–161.

Forsyth, P. & Dwyer, L., 2010. Exchange rate changes and the cost competitiveness of international airlines: The aviation trade weighted index. Research in Transportation Economics 26, 12–17.

Friedman, J., Hastie, T. & Tibshirani, R., 2013. glmnet: Lasso and elastic-net regularized generalized linear models. R package version .

Goh, C.F., Rasli, A. & Khan, S., 2014. Stock investors confidence on low-cost and traditional airlines in asia during financial crisis of 20072009. Procedia - Social and Behavioral Sciences 129, 31–38.

Gunther, F. & Fritsch, S., 2010. Neuralnet: Training of neural networks. The R journal 2, 30–38.

Hoerl, A. E. & Kennard, R.W., 1970. Ridge regression: applications to nonorthogonal problems. Technometrics 12, 69–82.

Kuhn, M., 2008. Building predictive models in r using the caret package. Journal of statistical software 28, 1–26.

Li, J. & Chen, W., 2014. Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models. International Journal of Forecasting 30, 996–1015.

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. Journal of Decision Systems 24, 19–36.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.C., Lin, C.C. & Meyer, M.D., 2019. Package e1071. R package version .

Miranda, U., 2015. The relationship between terrorism, oil prices, and airline profitability (doctoral study). Walden Disserations and Doctoral Studies Collection .

Narayan, P.K., Sharma, S.S., 2011. New evidence on oil price and firm returns. Journal of Banking & Finance 35, 3253–3262.

Navarro, A., Martínez, A..T.J.F., 2015. The effect of the economic crisis on the behaviour of airline ticket prices. a case-study analysis of the new yorkmadrid route. Journal of Air Transport Management 47, 48–53.

Nero, G., 1999. A note on the competitive advantage of large hub-and-spoke networks. Transportation Research Part E, Logistics and Transportation Review 34, 225–239.

Pal, S. K. & Mitra, S., 1992. Multilayer perceptron, fuzzy sets, classification. In Proc. IEEE Transactions on Neural Networks , 683–697.

R Core Team, R.F.f.S.C., 2017. R: Statistical computing environment. R: A Language and Environment for Statistical Computing.

Spiller, P., 2006. Assessing the profitability of effects of airline deregulation. Economics Letters 14, 263–267.

Sutton, C.D., 2005. Classification and regression trees, bagging, and boosting, Handbook of statistics Volume 24. Elsevier.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267288.

Van Landeghem, H. & Beuselinck, A., 2002. Reducing passenger boarding time in airplanes: A simulation based approach. European Journal of Operational Research 142, 294–308.

Verstraete, G., Aghezzaf, E-H. & Desmet, B., 2020. A leading macroeconomic indicators based framework to automatically generate tactical sales forecasts. Computers & Industrial Engineering 139, 106–169.