Air Transportation Direct Share Time Series Analysis and Forecast

Xufang Zheng*, Peng Wei†
Department of Aerospace Engineering, Iowa State University, Ames, IA, 50014

Air transportation direct share is the ratio of direct passengers to total passengers on a directional origin and destination (O&D) pair during a period of time. The O&D direct share shows the air passengers' general preference for direct flight services under a certain market state. The O&D direct share time series shows the changing process of a direct flight market. A better understanding and more accurate forecasting of direct share can benefit air traffic planners, airlines, and airports in multiple ways. In previous studies and practice, the direct share is usually assumed as a fixed average, which is not true for air transportation practice. The characteristics of O&D direct share time series are investigated in this research. Time series models and machine learning models are explored for the direct share time series forecasting. 1295 busiest O&D pairs across the U.S. are considered in this research. A framework for direct share time series modeling and forecasting is proposed in this research, which is a promising replacement to the model used for the direct share forecasting by the Federal Aviation Administration Terminal Area Forecast.

I. Introduction

Air transportation origin and destination (O&D) pair is the airport pair on which the airlines provide direct and non-direct flight services. Air transportation direct share is the ratio of direct passengers to total passengers on a certain O&D during a period of time. Fig 1 illustrates the passengers flow distribution on direct and non-direct itineraries from ALB (Albany International Airport) to ATL (Hartsfield-Jackson Atlanta International Airport). In this research, the quarterly direct share is investigated. The passengers flying directly from ALB to ATL (n_1) and the passengers taking one-connect at BWI (Baltimore/Washington International Thurgood Marshall Airport) without flight change (n_2) are direct passengers. The non-direct passengers include the passengers taking one-connect at BWI with flight change (n_3) and the passengers taking more than one connect (n_4). Assuming all the existing itineraries on ALB → ATL are shown in Fig 1, the direct share on ALB → ATL in quarter t ($directShare_{ALB\rightarrow ATL,t}$) can be computed by Eq 1, in which $n_{D,t}$ and $n_{P,t}$ are the quarterly direct passengers and total passengers respectively.

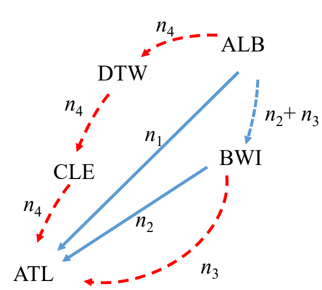

Fig. 1 Illustration of passengers flow distribution on ALB → ATL

$$directShare_{ALB\rightarrow ATL} = \frac{n_{D,t}}{n_{P,t}} = \frac{n_1 + n_2}{n_1 + n_2 + n_3 + n_4} \tag{1}$$

On a certain O&D market, there are commonly multiple flight services provided by the airlines. Direct flight services are generally preferred by air travelers because they are timesaving and more convenient. In some cases, the non-direct

*Ph.D. Candidate, Department of Aerospace Engineering, Iowa State University, IA, 50014, AIAA Student Member.
†Assistant Professor, Department of Aerospace Engineering, Iowa State University, IA, 50014, AIAA Senior Member.

flight services are preferred by the passengers because of their competitive pricing. Direct share shows the passengers' general preference for direct flight services under a certain market state, which changes with passenger demand and the airlines' competition. With increasing traveling demand, airlines tend to provide more direct flight services to compete for the local departing passengers. With more direct flight services provided by airlines with relatively lower pricing, the direct share on the O&D will increase.

O&D direct share shows the air passengers' distribution on direct and non-direct itineraries, which is an important feature for air transportation planners' decision making. In practice, the proportion of direct passengers to non-direct passengers is critical for Federal Aviation Administration (FAA) decision making on airport planning and investment. Under the certain market state, direct share shows passengers' general preference for the direct flight services, which is an essential factor for the study of passengers' traveling and booking preference and behavior [1–4]. The air passengers' preference for direct flight services and the changing process of the direct flight market are critical for airlines' market strategy making. For airport operation, the frequency of direct services to the chosen destination can have an impact on airport utility [4]. The proportion of direct passengers can impact airport capacity planning [5] and airport competition in multiple airport metropolis regions [6]. A better understanding and more accurate forecasting of direct share can benefit the air traffic planners, airlines and airports in various ways.

The O&D direct share time series shows the changing process of the direct flight market over time. Shown in Fig 2 are direct share time series on four different O&Ds. On O&D pair MSY (Louis Armstrong New Orleans International Airport) → HOU (William P. Hobby Airport), there are only direct flight services provided. The direct share time series is relatively high and stable over time. It is reasonable to forecast the direct share as a constant on this type of O&Ds. However, O&D market such as MSY → HOU only take a very small percentage of the O&D markets across the U.S. For O&D pairs such as ORD (O'Hare International Airport) → PSP (Palm Springs International Airport), there is a strong seasonality in the direct flight market. Compared to ORD → PSP, the seasonality in the direct share time series on CLT (Charlotte Douglas International Airport) → PVD (T. F. Green Airport) is relatively weaker. There are also O&D pairs, such as CMH (John Glenn Columbus International Airport) → RDU (Raleigh-Durham International Airport), on which the direct share fluctuates randomly and dramatically over time. It is difficult to identify the seasonal or trend patterns in such direct share time series.

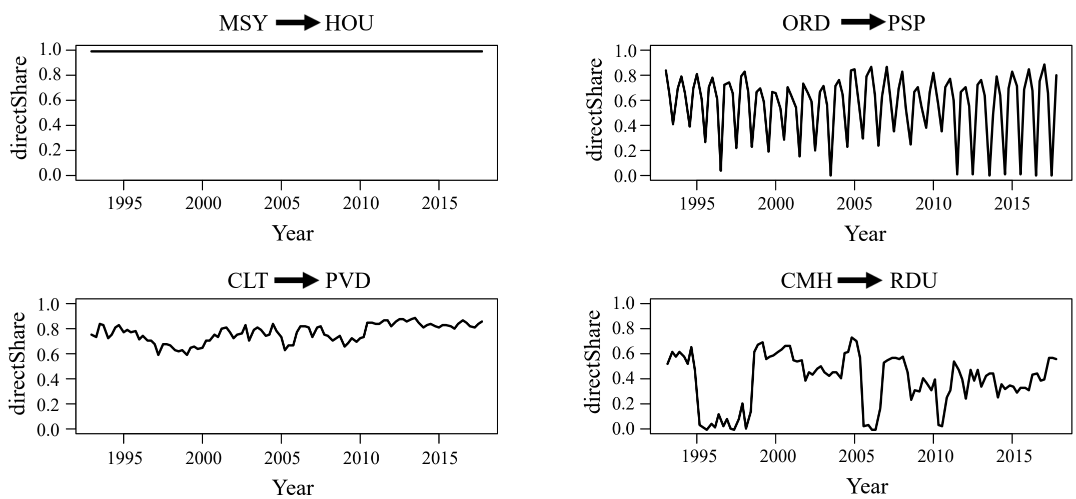

Fig. 2 Direct share time series on different O&Ds

Because the direct flight market states (including passenger demand, competing airlines, type of domain airlines, etc.) vary on different O&D pairs, the direct share time series differ for O&Ds. In most previous studies and practice, the direct share is generally assumed to be a constant proportion during a long period of time. The characteristics of O&D direct share time series are seldom studied previously. In the FAA Terminal Area Forecast (TAF), the model used for direct share forecasting assumes the O&D direct share same as the latest observation, even for long-term forecasting [7]. In this research, both time series models and machine learning models are investigated to develop the direct share time series forecasting model, which can properly capture the characteristics of direct share time series on different O&Ds. To make this study practical and inclusive, 1295 busiest O&D pairs across the U.S. are considered in this research. A framework for direct share time series modeling and forecasting is proposed, which can automatically select the best model for each O&D.

Models developed especially for time sequence data are time series models. To properly model time series with different trends, periodicities and noises, various time series models have been proposed. The signal components model the pattern generated by the intrinsic dynamics of the underlying process [8] and the noise components are used to estimate the noise at the current step. The exponential smoothing models and the autoregressive integrated moving average (ARIMA) models are the most widely used time series models [9]. Exponential smoothing models are based on exponential window functions, which fit the trend in the time series in a polynomial manner [10]. The Holt-Winters filter is one type of exponential smoothing models, which is capable of modeling and forecasting for both seasonal and non-seasonal time series [11]. ARIMA is a generalization of the broad family of ARMA models [10, 12, 13]. There are generally three types of components in an ARIMA model, which are AR(p), I(d) and moving average MA(q). The ARIMA model with seasonal AR(P), I(D) and MA(Q) components is known as the SARIMA model.

Time series models take advantage of the in nature ordering of the observations to forecast the future. Comparing to the time series models, machine learning models make full use of describing the response by the related features, which can automatically extract knowledge about the relation between the response and features [14, 15]. Under the machine learning context, direct share at the current step can be estimated by a feature set, which is a typical supervised learning problem, more specifically, is a regression problem. Based on whether there is a predetermined form of the underlying model, the regression models can be categorized into parametric models and non-parametric models [16, 17]. The most basic and easily interpretable regression model is the multiple linear regression model (MLR). The relation between the response and the features are modeled in an additive linear formulation in the MLR [18]. Non-parametric models provide a more flexible mapping between the response and the features in a nonlinear manner. Gradient boosting machine (GBM) is a tree-based non-parametric regression model with wide applications [19, 20].

The remainder of this paper is organized as follows. The data and data processing are introduced in Section II. Model development is discussed in detail with examples in Section III. The framework for direct share time series modeling and forecasting is proposed in Section IV with model comparison and analysis. In Section IV, we draw the conclusions of this research.

II. Data and Data Processing

To generate the O&D direct share time series, data mining is carried on the Airline Origin and Destination Survey (DB1B) database. The DB1B database is a quarterly 10% sample of airline tickets from carriers' reports, which is a publicly available database from the Bureau of Transportation Statistics (BTS) starting from 1993 [21]. The DB1B Coupon, DB1B Market and DB1B Ticket are the three data tables, which organize the reported ticket data under three levels. The DB1B Market data table is employed in this research, which includes useful data columns of *Passengers* (number of passengers) and *AirportGroup* (ordered airport group for a certain itinerary). The 1295 O&D pairs investigated in this research connect the 223 busiest hubs across the U.S [22]. Because of the data availability, for each O&D direct share time series, there are 100 observations starting from the first quarter in 1993 (1993 Q1) to the fourth quarter in 2017 (2017 Q4).

For machine learning modeling, the feature set is employed to estimate the response. In this research, the feature set is composed of ordered historical observations. Fig 3 illustrates the generation of the data frame for machine learning models. Assuming the width of the time window (lag) is 4, there will be four features in the feature set, which are [$x1$, $x2$, $x3$, $x4$]. Denote the observation at t as y_t. The feature set pairing with y_t are the observations at t-4, t-3, t-2 and t-1, which form the feature set [$x1_t$, $x2_t$, $x3_t$, $x4_t$]. The information contained in the feature set can have a significant impact on modeling performance. In general, more information can benefit the learner, but may cause unnecessary model complexity and poor forecasting performance at the same time. In this research, the lag is a hyperparameter to tune for the machine learning models.

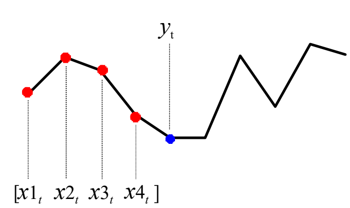

Fig. 3 Data frame generation for machine learning models

III. Model Development

Since direct share is a random variable between 0 and 1, logit and logistic transformations are necessary to guarantee the forecasting boundary for parametric models. For fitting and testing purposes, each direct share time series is split into two segments. The training segment contains the observations from 1993 Q1 to 2011 Q4 (76 observations), while the forecasting segment contains the observations from 2012 Q1 to 2017 Q4. For each model, a 24-step ahead forecasting will be generated for forecasting performance measurement. The testing root of mean square error (RMSE) is employed as the measurement of forecasting accuracy, which is as Eq 2.

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

A. Holt-Winters Filter

The Holt-Winters filter is an exponential smoothing model with additive components. Shown as Eq 3 is the general formulation of a Holt-Winters filter. Denote the tth observation in a time series as y_t. The level, trend, and seasonal components are denoted as L_t, T_t, and S_t respectively. α, β, and γ are the weights of each component, which are all between 0 and 1. The weights determine the variable depending more on the recent observations or observations from further in the past.

$$\begin{aligned} \hat{y}_{t+h} &= L_t + hT_t + S_{t-m+h_m^+} \\ L_t &= \alpha(y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \\ S_t &= \gamma(y_t - L_{t-1} - T_{t-1}) + (1-\gamma)S_{t-m} \end{aligned} \tag{3}$$

The Holt-Winters filter models are developed for the direct share time series on CLT → PVD and MCO (Orlando International Airport) → RDU (Raleigh-Durham International Airport). The details of the developed models and the modeling performance are shown in Table 1. Fitting and testing RMSE are used to measure the training and forecasting performance respectively. To visualize the modeling performance in a more intelligible way, the fitting and forecasting values are depicted together with the historical direct share time series in Fig 4.

Table 1 Holt-Winters modeling result and performance

O&D	α	β	γ	L_0	T_0	Fitting RMSE	Testing RMSE
CLT → PVD	0.7215	0.0001	0.0024	1.2794	-0.0048	0.0442	0.0221
MCO → RDU	0.7972	0.0095	0.0001	0.3375	0.0209	0.0591	0.0259

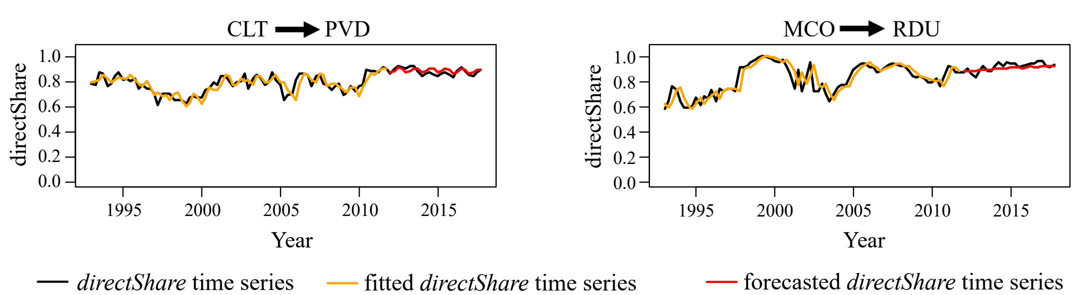

Fig. 4 Holt-Winters filter modeling performance

There is a relatively obvious seasonality in the direct share time series on CLT → PVD compared with the direct share time series on MCO → RDU. The seasonality in the direct share time series on CLT → PVD is modeled properly by the developed Holt-Winters filter model. There is a slightly increasing trend in the forecasting of direct share time series on MCO → RDU, while the forecasting of direct share time series on CLT → PVD is relatively stable, which can also be observed from the estimation of the coefficients. Based on the modeling performance, the Holt-Winters filter models can properly capture the characteristics of the seasonal and non-seasonal time series and provide reasonable forecasting of direct share time series.

B. SARIMA

The autoregressive moving average (ARMA) model was specially developed for stationary time series, which is based on the backward shift operator B [10]. The AR(p) components are related to the one-step or multiple-steps lagged observations in the time series which is formulated as Eq 4. w_t is the white noise at time t. MA(q) components model the noise based on the white noise errors at current and previous steps, which is formulated as Eq 5. The integrated components, I(d), are introduced in the ARMA model to expand its modeling capability for non-stationary time series, which is known as the ARIMA model. I(d) components indicate the differencing process on time series, which is as Eq 6. The general formulation of an ARIMA model, ARIMA(p,d,q), is as Eq 7. ARIMA model with seasonal components is known as the SARIMA model, SARIMA(p, d, q)(P, D, Q)$_m$, which is generally formulated as Eq 8. m is the term of the periodicity.

$$\begin{gathered} y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + w_t \\ \phi(B) y_t = w_t \end{gathered} \tag{4}$$

$$\begin{gathered} y_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_p w_{t-q} \\ y_t = \theta(B) w_t \end{gathered} \tag{5}$$

$$\nabla^d y_t = (1-B)^d y_t \tag{6}$$

$$\phi(B)(1-B)^d y_t = \theta(B) w_t \tag{7}$$

$$\Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d d s_t = \delta + \Theta_Q(B^s)\theta(B) w_t \tag{8}$$

The critical problem for SARIMA model development is to predetermine the term of each component. In most of the previous research, the hyperparameters are determined based on the observation of the time series plot and diagnose graphs. This method is subjective, and not practical when developing SARIMA models for a large number of time series. In this research, the hyperparameters are automatically selected by the step-wise algorithm [23]. To balance the modeling performance and model complexity, Akaike Information Criterion (AIC) is used as the model selection criteria. SARIMA models are developed for the direct share time series on ORD → PSP and BOS (Boston Logan International Airport) → DEN (Denver International Airport). The modeling result and performance are shown in Table 2.

Table 2 SARIMA modeling result and performance

O&D	SARIMA	Fitting RMSE	Testing RMSE
ORD → PSP	$(0,0,0)(0,1,1)_4$	0.1882	0.0858
BOS → DEN	$(2,1,2)(0,0,0)$	0.0639	0.0275

A SARIMA model with seasonal components is developed for the direct share time series on ORD → PSP with periodicity equals to 4. An ARIMA model is developed for the direct share time series on BOS → DEN with four coefficients and first order differencing. The Holt-Winters filter models are developed on the two direct share time series as well, and the modeling performance of the two models are compared in Fig 5.

There is a strong seasonality in the direct share time series on ORD → PSP, which is caused by the seasonal tourism at the destination (Palm Springs, CA). Based on the comparison of Fig 5 (a) and (c), the developed SARIMA model can provide better training and forecasting performance compared to the developed Holt-Winters model. The trend in the time series is modeled in a more dynamic non-parametric manner in the SARIMA models compared to the Holt-Winters filter. The trend in the training set is usually carried on in the long-term forecasting for the Holt-Winters filter modeling, which can be observed from Fig 5 (c). For the Holt-Winters filter model developed for BOS → DEN, $\gamma = 0.2103$, because of which there is a seasonality in the forecasting time series. While, in the developed SARIMA model, there is no seasonal component, the fluctuations in the forecasting of direct share are in a random manner. Based on the comparison, the Holt-Winters filter is more sensitive to the overall trend and possible periodic fluctuations in

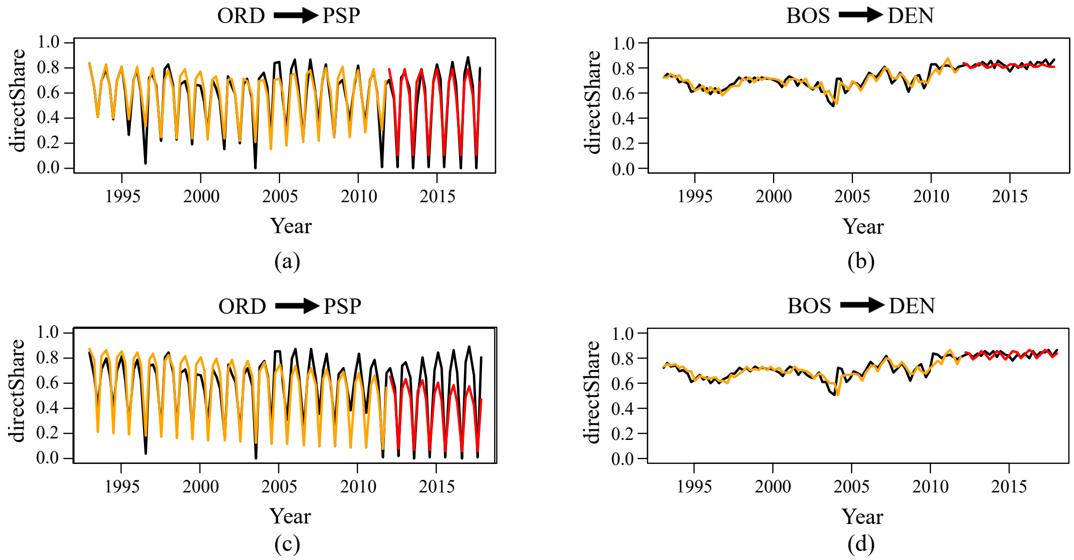

Fig. 5 Modeling performance comparison between SARIMA and Holt-Winters filter. (a) SARIMA modeling. (b) SARIMA modeling. (c) Holt-Winters filter modeling. (d) Holt-Winters filter modeling.

the time series. It may generate forecasting with trend or seasonality for time series, for which there may be no such characteristic. SARIMA model is based on the backward shift operator, which can model the time series in a more dynamic way avoiding the impacts from the observations from further in the past.

C. Multiple Linear Regression

The multiple linear regression (MLR) models the relation between the response and the features in an additive linear formulation, which is the most basic and easily interpretable machine learning model. Shown as Eq 9 is the general formulation of MLR. The coefficients are estimated by minimizing the root sum square (RSS), which is formulated as Eq 10.

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_j x_{ij} \tag{9}$$

$$RSS = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j^2 x_{ij})^2 \tag{10}$$

To tune the hyperparameter lag, a grid searching method is employed in this research, and AIC is used as the tuning criteria. Shown in Fig 6 are the lag tuning processes of the MLR models for direct share time series on ORD → MDT (Harrisburg International Airport) and RSW (Southwest Florida International Airport) → CVG (Cincinnati/Northern Kentucky International Airport).

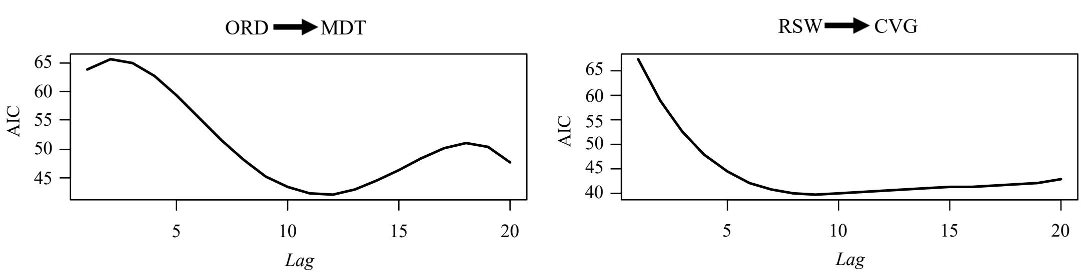

Fig. 6 Tuning hyperparameter lag for multiple linear regression models

As shown in Fig 6, the tuning curves vary for different time series. The MLR models are developed for the two direct share time series with the tuned lag. Shown in Table 3 and Fig 7 are the modeling results and performance of the developed MLR models.

Table 3 Multiple linear regression modeling result and performance

O&D	lag	Fitting RMSE	Testing RMSE
ORD → MDT	12	0.0331	0.0370
RSW → CVG	9	0.0639	0.0275

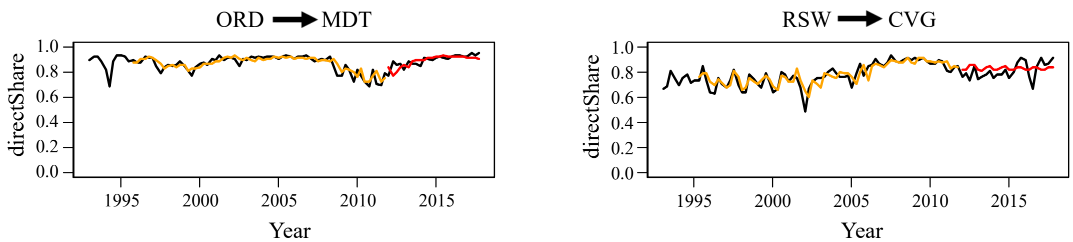

Fig. 7 Multiple linear regression modeling performance

For the two developed MLR models, observations within 12 and 9 previous steps are included in the feature sets respectively. The estimation of direct share at the current step will only depend on a limited number of observations from previous steps. This modeling method can efficiently avoid the impacts from the observations from further in the past, which is greatly useful for modeling time series generated from an unstable process. Based on the modeling performance of the two examples, the MLR model can provide promising fitting and forecasting performance for both seasonal and non-seasonal direct share time series.

D. Gradient Boosting Machine

Compared to the parametric models, there is no predetermined form of the model for the non-parametric models. The gradient boosting machine (GBM) is a non-parametric supervised learning model, which is based on the decision trees. The basic architecture of a GBM is determined by three hyperparameters. *LearningRate* is a value usually between 0 and 0.1 which indicates how much information should be learned from the previous tree when growing a new tree [24, 25]. $MaxDepth$ determines how deep each tree should grow and $Ntrees$ indicates how many trees should be grown in a GBM model. In this research, lag is another hyperparameter to tune. Grid searching method is employed for hyperparameter tuning. Because when developing non-parametric models, the number of coefficients will increase dramatically, the 5-fold cross-validation RMSE is used as the tuning criteria. Shown in Fig 8 is the grid searching result for GBM model for direct share time series on CMH → RDU.

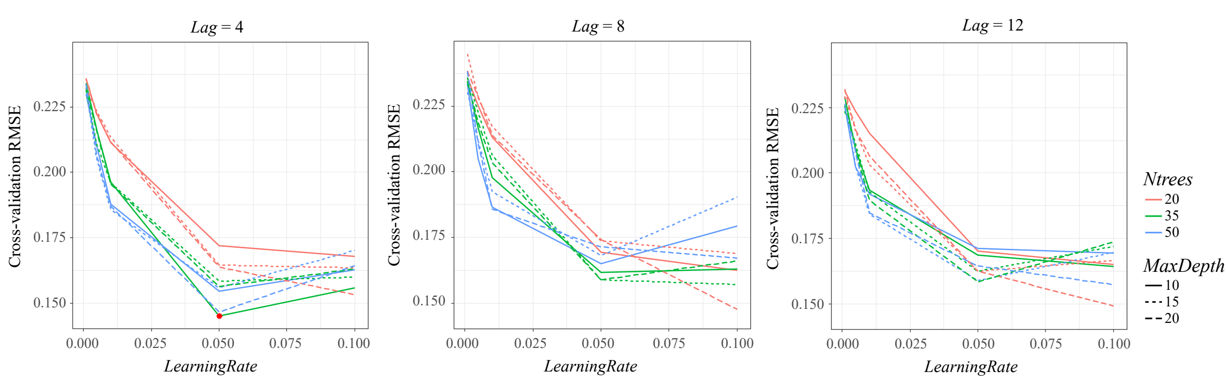

Fig. 8 Hyperparameter tuning for GMB model

The GBM model is developed with the tuned hyperparameters ($LearningRate = 0.05$, $Ntrees = 35$, $MaxDepth = 10$, and $lag = 4$). The fitting and testing RMSE are 0.1187 and 0.0988 respectively. The Holt-Winters filter, SARIMA, and MLR model are developed for the same direct share time series as well. The modeling performance is compared in Fig 9.

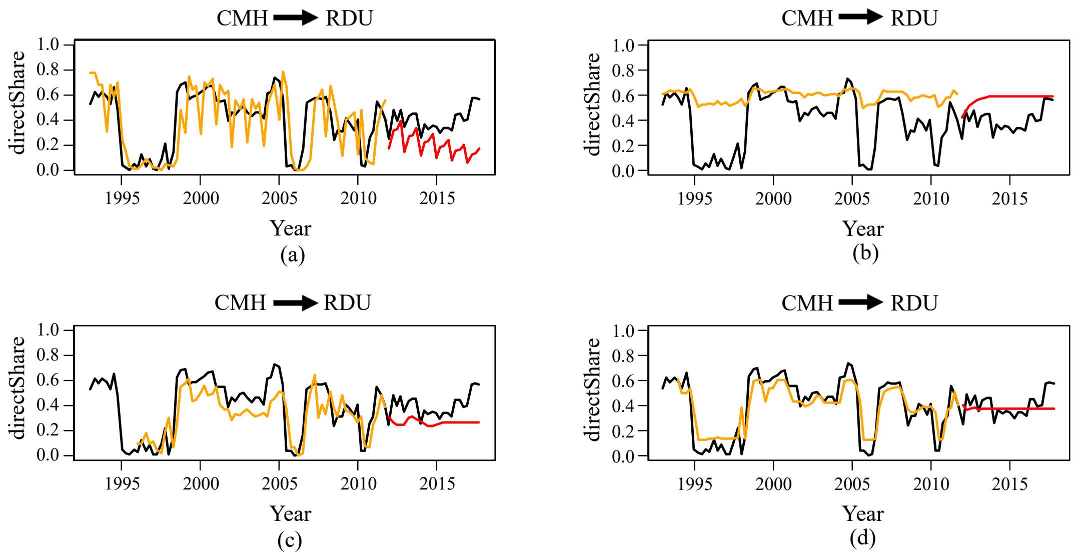

Fig. 9 Modeling performance comparison for direct share time series on CMH → RDU. (a) Holt-Winters filter modeling. (b) SARIMA modeling. (c) MLR modeling. (d) GBM modeling.

For O&D pairs such as CMH → RDU, there are random and dramatic fluctuations in the direct share time series. Based on the comparison in Fig 9, the developed GBM model can outperform all other models in both training and forecasting performance with a small lag value.

IV. Framework for Direct Share Time Series Modeling and Forecasting

Four different models are investigated for direct share time series forecasting previously. Based on the modeling performance analysis and comparison, for different O&D pairs, the models that can provide the best modeling performance are different. To efficiently develop the direct share model for each O&D, a framework for direct share modeling and forecasting is proposed in this research. For each O&D pair, the Holt-Winters filter, SARIMA, MLR, and GBM models are developed firstly, and the best model is selected based on the testing RMSE. Direct share time series forecasting models for the 1295 O&Ds are developed based on the proposed framework. Shown in Fig 10 is the proportion of the selected models. The selection of each model is nearly even, which indicates that considering the four models for each O&D is necessary in this research.

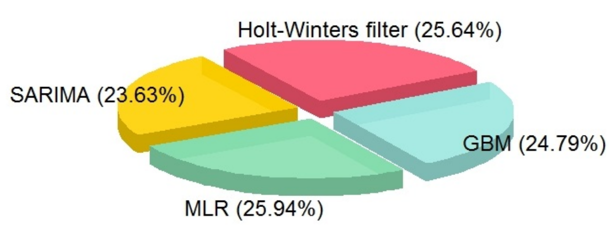

Fig. 10 Proportion of selected models

The model used for direct share forecasting by FAA TAF is introduced in this research as the benchmark, which is denoted as M_{TAF}. For M_{TAF}, the forecasting of direct share on each O&D is a constant as the latest observation. Shown in Fig 11 are the forecasting performance of M_{TAF} on ORD → PSP and ORD → MDT. For O&D direct share time series with seasonality or trend, the M_{TAF} is not capable of providing reasonable and proper forecasting.

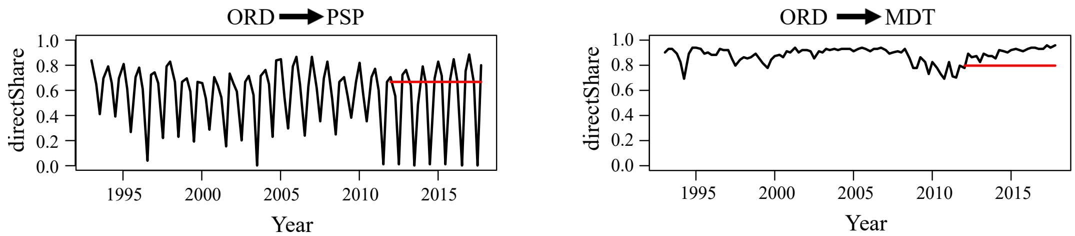

Fig. 11 Forecasting performance of M_{TAF}

The direct share time series forecasting for the 1295 O&Ds based on the M_{TAF} model are generated. The forecasting performance of the proposed framework and the M_{TAF} is compared in Fig 12. Taken the 1295 O&Ds together, the proposed framework can provide better forecasting performance compared to the M_{TAF}. More specifically, the proposed framework can provide a better forecasting performance for 82.46% of the 1295 O&Ds, which shows that the proposed framework is a promising replacement for the M_{TAF}.

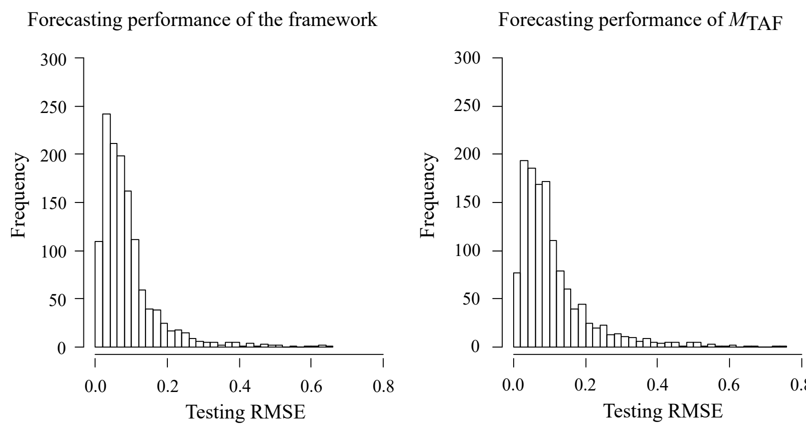

Fig. 12 Forecasting performance comparison

V. Conclusion

Air transportation direct share shows the air passengers' general preference for the direct flight services under a certain market state. The O&D direct share time series can show the change of passengers flow distribution on direct and non-direct itineraries and the changing process of the direct flight market. A better understanding and more accurate forecasting of O&D direct share can benefit air transportation planners, airlines, and airports in multiple ways.

Both time series models and machine learning models are investigated in this research. Based on the modeling performance comparison and analysis, the model that can provide the best forecasting performance varies for different O&Ds. Holt-Winters filter is more capable of modeling and forecasting direct share time series with seasonality and a constant underlying trend. The SARIMA can model the seasonality and trend of direct share time series in a more dynamic and non-parametric way. The MLR and GBM models can extract the knowledge about how to describe the direct share by the feature set automatically from the data. The MLR model is capable of seasonal and non-seasonal direct share time series modeling and forecasting with a limited number of lagged observations. The GBM model shows the advantage in modeling and forecasting direct share time series with random and dramatic fluctuations. A framework is proposed in this research for O&D direct share modeling and forecasting. The direct share forecasting models for 1295 busiest O&Ds across the U.S. are developed automatically by the proposed framework. Based on the forecasting performance comparison, the proposed framework is a promising replacement for the model used for direct share forecasting by the FAA TAF.

Acknowledgments

This research is partially funded and under great support and help from the Office of Aviation Policy and Plans, Federal Aviation Administration. We want to express our appreciation to all the colleagues who provided valuable insights and suggestions, especially David Chien, Roger Schaufele, Chia-Mei Liu and Li Ding.

References

[1] Coldren, G. M., Koppelman, F. S., Kasturirangan, K., and Mukherjee, A., "Modeling Aggregate Air-travel Itinerary Shares: Logit Model Development at A Major US Airline," *Journal of Air Transport Management*, Vol. 9, No. 6, 2003, pp. 361–369.

[2] Warburg, V., Bhat, C., and Adler, T., "Modeling Demographic and Unobserved Heterogeneity in Air Passengers' Sensitivity to Service Attributes in Itinerary Choice," *Transportation Research Record: Journal of the Transportation Research Board*, , No. 1951, 2006, pp. 7–16.

[3] Garrow, L. A., *Discrete Choice Modelling and Air Travel Demand: Theory and Applications*, Routledge, 2010.

[4] Pels, E., Nijkamp, P., and Rietveld, P., "Airport and Airline Choice in A Multiple Airport Region: An Empirical Analysis for the San Francisco Bay Area," *Regional Studies*, Vol. 35, No. 1, 2001, pp. 1–9.

[5] Bradley, M., *Behavioural Models of Airport Choice and Air Route Choice (Chapter 9 of Travel Behaviour Research: Updating the State of Play)*, 1998.

[6] Innes, J. D., and Doucet, D. H., "Effects of Access Distance and Level of Service on Airport Choice," *Journal of Transportation Engineering*, Vol. 116, No. 4, 1990, pp. 507–516.

[7] Office of Aviation Policy and Plans, "Terminal Area Forecast Summary, Fiscal Year 2017-2045," , 2017.

[8] Montgomery, D. C., Jennings, C. L., and Kulahci, M., *Introduction to Time Series Analysis and Forecasting*, John Wiley & Sons, 2015.

[9] De Gooijer, J. G., and Hyndman, R. J., "25 Years of Time Series Forecasting," *International journal of forecasting*, Vol. 22, No. 3, 2006, pp. 443–473.

[10] Brockwell, P. J., Davis, R. A., and Calder, M. V., *Introduction to Time Series and Forecasting*, Vol. 2, Springer, 2002.

[11] Koehler, A. B., Snyder, R. D., and Ord, J. K., "Forecasting Models and Prediction Intervals for the Multiplicative Holt–Winters Method," *International Journal of Forecasting*, Vol. 17, No. 2, 2001, pp. 269–286.

[12] de Oliveira, E. M., and Oliveira, F. L. C., "Forecasting Mid-long Term Electric Energy Consumption through Bagging ARIMA and Exponential Smoothing Methods," *Energy*, Vol. 144, 2018, pp. 776–788.

[13] Alsharif, M. H., Younes, M. K., and Kim, J., "Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea," *Symmetry*, Vol. 11, No. 2, 2019, p. 240.

[14] Alpaydin, E., *Introduction to Machine Learning*, MIT press, 2009.

[15] James, G., Witten, D., Hastie, T., and Tibshirani, R., *An Introduction to Statistical Learning*, Springer, 2013, Vol. 112, Chaps. 3–7.

[16] Hardle, W., Mammen, E., et al., "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, Vol. 21, No. 4, 1993, pp. 1926–1947.

[17] McCune, B., "Non-parametric Habitat Models with Automatic Interactions," *Journal of Vegetation Science*, Vol. 17, No. 6, 2006, pp. 819–830.

[18] Myers, R. H., and Myers, R. H., *Classical and Modern Regression with Applications*, Vol. 2, Duxbury press Belmont, CA, 1990.

[19] Persson, C., Bacher, P., Shiga, T., and Madsen, H., "Multi-site Solar Power Forecasting Using Gradient Boosted Regression Trees," *Solar Energy*, Vol. 150, 2017, pp. 423–436.

[20] Touzani, S., Granderson, J., and Fernandes, S., "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial buildings," *Energy and Buildings*, Vol. 158, 2018, pp. 1533–1543.

[21] Bureau of Transportation Statistics, "Data Profile: Airline Origin and Destination Survey (DB1B)," , 2018. URL https://www.transtats.bts.gov/DatabaseInfo.aspDB_ID=125&DB_Name=Airline%20Origin%20and%20Destination%20Survey%20%28DB1B%29.

[22] Federal Aviation Administration, "Airport Categories Airports," , 2018. URL https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/categories/.

[23] Hyndman, R. J., Khandakar, Y., et al., *Automatic Time Series for Forecasting: the Forecast Package for R*, 6/07, Monash University, Department of Econometrics and Business Statistics . . . , 2007.

[24] Friedman, J., Hastie, T., and Tibshirani, R., *The Elements of Statistical Learning*, Springer series in statistics New York, 2001, Vol. 1, Chaps. 3,7,9.

[25] Friedman, J. H., "Stochastic Gradient Boosting," *Computational Statistics & Data Analysis*, Vol. 38, No. 4, 2002, pp. 367–378.