The chaos on US domestic airline passenger demand forecasting caused by COVID-19

Nahid Jafari

School of Business,
State University of New York-Farmingdale,
NY, USA
Email: nahid.jafari@hotmail.com

Abstract: Commercial aviation is a major contributor to the US economy, directly or indirectly generating approximately US$680 billion, or 4% of GDP, and supporting millions of jobs. Approximately 965 million passengers flew to US destinations in 2017 (https://rosap.ntl.bts.gov/view/dot/37861). Given the importance of the industry, accurate forecasting of air passenger demand is valuable, and the most sophisticated forecasting technologies can be applied to this endeavour. The ongoing COVID-19 crisis has had an unprecedented impact on air traffic. Effective forecast of passenger demand would benefit airlines to develop adequate recovery plans and prevent (or minimise) any catastrophe in handling passengers during and post pandemic. The purpose of this study is to investigate COVID-19's impact on the US domestic air passengers demand, identify the most influential features on air passenger demand, and design more accurate forecast models. In addition, we address a computational challenge in developing forecasting models due to the volatility of the recent data as a result of the COVID-19 crisis. We use both traditional and artificial intelligence methods and discuss their capabilities to handle the challenge.

Keywords: air passenger demand; the US airlines market; seasonal time series forecasting; deep learning; gated recurrent units.

Reference to this paper should be made as follows: Jafari, N. (2022) 'The chaos on US domestic airline passenger demand forecasting caused by COVID-19', *Int. J. Business Forecasting and Marketing Intelligence*, Vol. 7, No. 3, pp.241–258.

Biographical notes: Nahid Jafari is an Operations Research Scientist who serves as an Assistant Professor in the School of Business at the State University of New York-Farmingdale, NY. She received her PhD in Statistics and Operations Research from the RMIT University, Melbourne, Australia in 2013. She has completed her BSc and MSc in Applied Mathematics. Her expertise is in the area of discrete optimisation (mainly integer programming and network design problems) and time series forecasting.

1 Introduction

The COVID-19 pandemic has devastated the aviation industry worldwide and has had significant economic and social impacts, including hospitality industry (Maneenop and Kotcharin, 2020; Sun et al., 2020; Warnock-Smith et al., 2021). Demand for air travel is essentially driven by business and leisure activities. While the passenger aviation industry expands during periods of economic growth, it contracts during economic downturns.

According to the US Department of Transportation's Bureau of Transportation Statistics (BTS), from 2003 to 2017, US air revenue passenger-miles for domestic travel increased by approximately 38% (https://rosap.ntl.bts.gov/view/dot/42912). Due to the impact of COVID-19 on US domestic air travel, the passengers demand was reduced because of working remotely from home instead of business travel and in person meetings, economic crisis, and travel restrictions. Hotle and Mumbower (2021) have stated that performed departures decreased by 71.5% in May 2020 compared to May 2019.

Air passenger demand forecasting is receiving increased attention because of intrinsic difficulties and practical applications. A forecasting system that gives accurate estimates of passenger demand is an essential tool for airline decision making. The forecasting of air passenger demand over time can be analysed as time series forecast in which demand is complex due to its irregularity, high volatility, and seasonality. Air passenger demand, however, exhibits consistently complex nonlinearity and non-stationarity.

Time series forecasting is an important decision making and planning tool in multiple fields such as finance, economics, physics, medicine, and engineering. It uses observed time series in the past to predict an unseen time series in a look-ahead horizon. The sequence of the input data, its systematic pattern (e.g., seasonality and non-stationary), the length of prediction time horizon, and random noise add complexity to the predictive modelling. Among quantitative forecasting approaches, econometric, statistical, and artificial intelligence approaches are well known. There are studies that discuss advantages and disadvantages of the mentioned approaches (Januschowski et al., 2020; Makridakis et al., 2018).

The primary purpose of this paper was to develop highly accurate forecasting models to forecast the US domestic air passenger demand. We were, specifically, working on developing neural deep network models and compare their performance with statistical models before the pandemic hits. The COVID-19 outbreak in March 2020 resulted, in unprecedented disruption to air transportation activity [see in Figure 1(a)]. Due to the dramatic change in patterns and trends, the direction and aim of our work has expanded from focusing on developing well-performing forecasting models to collecting appropriate data on additional related factors, then analysing their influence on air passenger demand and developing solutions for accurate forecasts.

As mentioned above, we study the US domestic air passenger demand in 21st century, particularly from 2001 to 2021. Our contribution to this literature is to investigate COVID-19's impact on the US domestic air passengers demand, identify the most influential features on air passenger demand, and design more accurate forecast models. There has been no previously reported study associated with this problem.

In the following section, we first describe the data which we have collected by introducing their sources, then we discuss preparing the data, and finally we add the background of the study. In Section 3, we present regression analysis on the data to recognise relationship of air passenger demand and other presented variables. The last part of the section represents the result of predictions based on the traditional models which have been introduced. In Section 4, we develop several deep learning models on our dataset, then conduct air passenger demand prediction based on the best fitted model. We discuss the obtained result in each section. In the last section, we summarise the performance of our developed models, discuss their strengths and drawbacks.

2 The empirical study

2.1 Data description and preparation

In this research, the monthly data of the US domestic airlines passenger traffic and related factors between January 2001 and April 2021 were obtained from publicly available sources as following. The dataset consists of air passenger demand from US domestic airlines and other factors that may have a potential impact on air passenger demand. The main variables used in this study are: number of passengers, available seats, airlines fuel cost, average air fares, unemployment rate, stock market return, COVID-19 positivity ratio, and travel restriction. We obtained the data related to aviation such as number of passengers and available seats from the US Department of Transportation Bureau of Transportation Statistics, which provides comprehensive data on airlines in the US domestic and international markets (https://www.transtats.bts.gov/). Also we obtained inflation-adjusted air fares from https://www.bts.gov/air-fares and airlines fuel cost (as cost per gallon) from https://www.transtats.bts.gov/fuel.asp. The economic factors used in this research included unemployment rate from US Bureau of Labor Statistics (https://www.bls.gov/eag/eag.us.htm) and stock market return (S&P 500 index) from http://www.officialdata.org/ us/stocks/s-p-500/1900. We extracted COVID-19 positivity ratio from https://covidtracking.com. We, also, generated a binary variable referring to travel restriction during the pandemic.

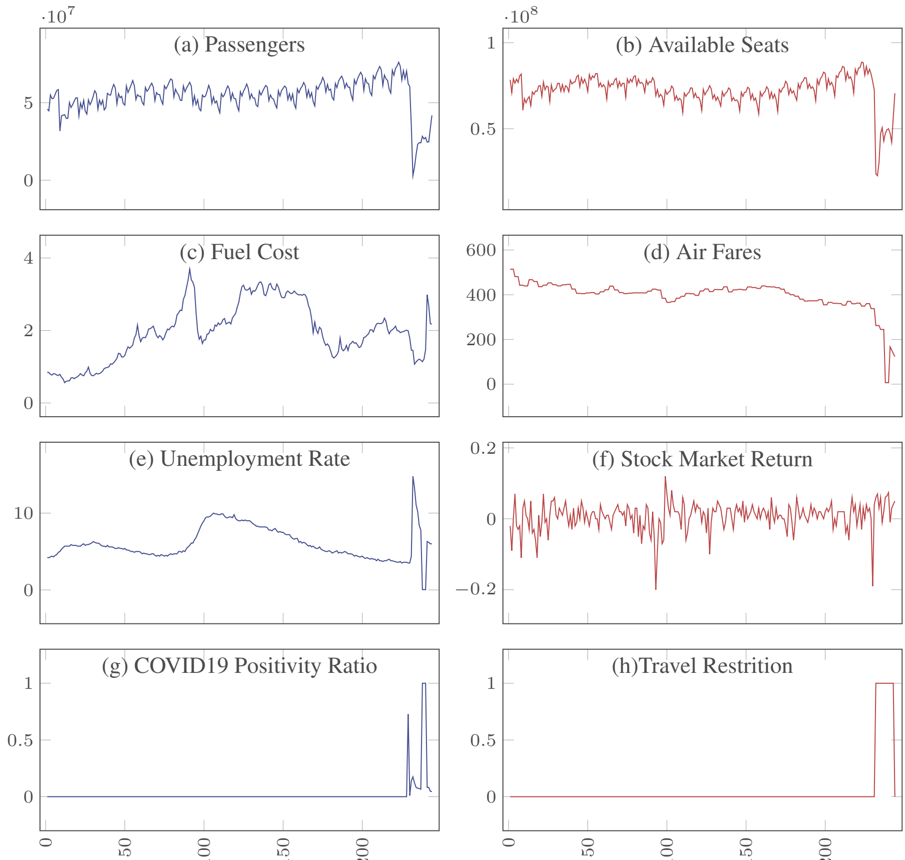
To begin our analysis, we plot all variables of the dataset to present their details and trends. The dataset contains 244 observations (20 years × 12 months plus 4 months). Figure 1(a) shows that the demand for air travel spikes, as expected, in the summer months June, July, and August. It shows that the number of passengers has a gradual upward trend except for two minor drops and one major drop. The first minor drop is related to the terror attacks of 11 September (2001), and the other during the Great Recession (2007–2009). Afterward, the demand for air travel grew slowly through 2013, and then the growth trend has accelerated significantly till a dramatic decline occurred around when COVID-19 was declared by the World Health Organization (WHO) a pandemic in March 2020. Figures 1(b) to 1(h) show the evolution of other variables in the dataset over the period examined in the present study. To conduct accurate analysis and forecast with multiple variables, we re-scale all variables by $[0,1]$ normalisation using the equation

$$x_i^{scaled} = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

To do normalisation in Python, we use the MinMaxScaler transform object from the scikit--learn library. To measure the accuracy of the fitted models in this study, we compute root-mean-square error (RMSE) and mean absolute percentage error (MAPE) by the following formula where y_i denotes the actual value and $\hat{y}_i$ is the predicted value of passengers, and N is the size of dataset.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \qquad MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%.$$

Figure 1 Line plot of air passenger demand, available seats, airlines fuel cost, average air fares, unemployment rate, stock market return, COVID-19 positivity rate, and travel restriction during 2001–2021 (see online version for colours)



Note that RMSE is useful when we compare the performance of different forecast models, while MAPE expresses the percentage of forecast variance from the actual.

To begin, we did a correlation analysis on our data taking (number of) passengers as a dependent variable and the rest as independent variables. Passengers has a strong positive association ($r = 0.8919$) with (number of available) Seats and strong negative association ($r = -0.7073$) with travel restriction. Passengers and unemployment rate is negatively and moderately correlated ($r = -0.4810$). Between passengers and air fares and fuel cost, there is a weak positive correlation and stock market return and COVID-19 positivity ratio a weak negative correlation.

2.2 Background on air passenger demand forecasting

There are many different methodological approaches to forecasting air passenger demand. For example, time series forecasting problems in commercial aviation can be examined as standard regression problems with time-varying parameters. Duval and Schiff (2011) and Abed et al. (2001) developed regression analysis models for forecasting the number of air passengers using various financial measures. Kim and Shin (2016) used basic regression analysis to consider causal relationships between air passenger demand and other variables. Carmona-Benitez et al. (2017) proposed a forecasting approach using an econometric dynamic model (EDM) to estimate passenger demand in the Mexican air transport industry. They applied the panel data Arellano-Bover method to calibrate the EDM, which was validated by the Sargan test and the Arellano-Bond autocorrelation test. Hsiao and Hansen (2011) modelled city-pair air passenger demand at the route level using a type of discrete choice method which are widely used for the analysis of individual choice behaviour. In addition, linear multivariate time series can be modelled by vector autoregression (VAR) (Cao and Tay, 2003) and nonlinear time-series models nonlinear VAR (Samadi et al., 2019) and more (Karimuzzaman and Hossain, 2020; Comi et al., 2020).

Autoregressive integrated moving average (ARIMA) models is a practical method among statistical time series forecasting models. The ARIMA model (Box and Pierce, 1970) is one of the most prominent among univariate time series forecasting models which include other models like autoregression (AR), moving average (MA), and autoregressive moving average (ARMA), another model which adapts various exponential smoothing techniques (McKenzie, 1984). ARIMA models are rarely used in high dimensional multivariate time series forecasting due to their high computational cost, but Tsui et al. (2014) employed the Box-Jenkins ARIMA methodology for forecasting Hong Kong' s passenger demand.

Jin et al. (2020) proposes a hybrid approach for air passenger demand forecasting, which consists of variational mode decomposition (VMD), autoregressive moving average model (ARMA) and kernel extreme learning machine (KELM). Tascon and Olariaga (2021) seeks to evaluate the impact of demand forecasts on the management of runway capacity, taking Bogotá-El Dorado International Airport in Bogota, Colombia as a case study. The situation described was faced with the use of simulation under the system dynamics approach, to research the management of air transport.

Artificial intelligence techniques have been investigated in air transportation forecasting. Neural networks are capable of learning patterns and the tendencies of the series. Brazilian air transport demand forecasting using neural networks has been studied by Alekseev et al. (2002) and Alekseev and Seixas (2009). The studies found that neural processing outperforms the traditional econometric approaches. Srisaeng et al. (2015) developed and empirically examined genetic algorithm optimisation models (an alternative artificial intelligence-based approach) for forecasting Australia's quarterly domestic airline passenger demand. To predict short-term air passenger traffic, Xiao et al. (2014) combined singular spectrum analysis for identifying and extracting the trend and seasonality of air transport demand with an adaptive-network-based fuzzy inference system (another artificial intelligence technology) to deal with the irregularity and volatility of demand.

Papageorgiou and Poczeta (2017) proposed a two-stage model for multivariate time series prediction based on the efficient capabilities of evolutionary fuzzy cognitive maps (FCMs) enhanced by structure optimisation algorithms and artificial neural networks (ANNs). Chai and Lim (2016) presented a forecasting model of cyclical fluctuations of the economy based on the time delay coordinate embedding method. The model uses a neuro-fuzzy network with weighted fuzzy membership functions. Martinez et al. (2018) proposed a new strategy that forecasts every different season using a different specialised k-nearest neighbours (kNN) learner. Wang et al. (2015) presented an improved extreme learning machine, which has simple structure and good performance, for online sequential prediction of multivariate time series.

Deep learning, which has received an increasing amount of attention in time series analysis, is a branch of machine learning. It encompasses models and architectures that learn optimal features from data by capturing increasingly complex representations of the data with combinations of layers of nonlinear data transformations (LeCun et al., 2015; Goodfellow et al., 2016).

3 The methodology

3.1 Regression analysis of air passenger demand

We begin our analysis by deploying a multiple linear regression model to examine the relationship of the air passenger demand to the other listed variables above. Multiple linear regression attempts to model a relationship between endogenous (or dependent) variables and exogenous (or independent) variables by fitting a linear equation to observed data. Given k exogenous variables $x_1, x_2, \ldots, x_k$) and an endogenous y, the line equation can be written as

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{it} + \epsilon_t,$$

where β_0 is a constant term (intercept), $\beta_i,\ i = 1, \ldots, k$ are regression (or slope) coefficients corresponding each exogenous, ϵ_t is the error term. Note that in this approach, the ordinary least squares, β_i is estimated subject to that the sum of squares of residuals to be minimised,

$$\underset{\beta \in R^p}{\arg\min} \frac{1}{N} \|Y - X\beta\|^2 = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2$$

Table 1 Result of a multiple linear regression, a Lasso, and a ridge regression models on our dataset

Variables (x_i)	*Multiple (linear) regression*		*Lasso regression*		*Ridge regression*	
	Coefficients (β_i)	*P-value*	*Coefficients* (β_i)	*P-value*	*Coefficients* (β_i)	*P-value*
Seats	0.8831317*	0.000	0.8399527		0.8712428*	0.000
Fuel cost	0.1341389*	0.000	0.094619		0.1338457*	0.000
Air fares	–0.1690892*	0.000	0		–0.1689806*	0.000
Unemployment rate	0.0106171	0.688	0		0.0063535	0.810
Stock market return	0.0316597	0.243	0		0.0314513	0.247
COVID-19 positivity ratio	–0.1266668*	0.010	0		–0.1266731*	0.010
Travel restriction	–0.1142123*	0.000	0		–0.1183052*	0.000
Intercept	0.0656067	0.223	0.0325006		0.0759226	0.158
R-squared	0.8863		0.8468		0.8862	
RMSE	0.04903		0.0034		0.0490	

Notes: We add star sign in front of coefficients which are significant at 5%, i.e., their p-value < 0.05. The number of observations is 244 and passengers is the dependent variable.

The first part of Table 1 shows the result of running multiple linear regression model on our dataset. The obtained equation is

$$P = 0.0656067 + 0.8831317 \times S + 0.1341389 \times FC - 0.1690892 \times AF + 0.0106171 \times UR + 0.0316597 \times SMR - 0.1266668 \times CPR - 0.1142123 \times TR$$

where the written abbreviation are as passengers (P), seats (S), fuel cost (FC), air fares (AF), unemployment rate (UR), stock market return (SMR), COVID-19 positivity ratio (CPR), and travel restriction (TR). Each coefficient, β_i, $i = 1, \ldots, k$, indicates that an endogenous variable has strong (close to 1) versus weak (close to 0) relationship or positive vs negative relationship with an exogenous variable. Therefore, passengers has strong positive, weak positive, and weak negative relationship with seats, fuel cost and air fares, respectively at 5% significant level. Also it has negative relationship with COVID-19 positivity ratio and travel restriction at 5% significant level. However, the relationship of passengers with unemployment rate and stock market return is non-significant at 5%. We better to add that lower unemployment rates and higher stock market return correspond to growth in customers' purchasing power, and a better economy.

3.2 Lasso and ridge regression

Lasso regression is a type of linear regression that shrinks data values towards a central point such as the mean. The LASSO stands for least absolute shrinkage and selection operator. In this model, the loss function is augmented such that we not only minimise the sum of squared residuals but also penalise the size of parameter estimates.

$$\arg\min_{\beta \in R^p} \frac{1}{N} \|Y - X\beta\|^2 + \lambda \|\beta\|^2 = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

The term $\lambda \sum_{j=1}^{p} |\beta_j|$, is a penalty that increases in value the more complex the model. It causes Lasso to omit variables, that is, it tends to make coefficients to absolute zero as shown in middle columns of Table 1. The lasso regression can cause a small bias in the model in which the prediction is too dependent upon a particular variables. Note that if $\lambda = 0$ corresponds to simple linear regression model.

Ridge regression shrinks the coefficients and it helps to reduce the model complexity and multi-collinearity. The difference between ridge and lasso regression is that it tends to make coefficients to absolute zero as compared to Ridge which never sets the value of coefficient to absolute zero (see the last part of Table 1). We need to add that we have done multi-collinearity diagnostic tests on our dataset and resulted that there is multi-collinearity in our data.

3.3 Nonlinear regression

To examine the behaviour of the exogenous variables on the endogenous variable in our dataset, we undertake two nonlinear regression analysis. First, because of the obvious impact of the COVID-19 pandemic on the air passenger demand, we compute the regression equation using exponentialised of the two variables, COVID-19 positivity ratio and travel restriction. See the equation bellow and the results in the middle part of Table 1.

$$P = \beta_0 + \beta_1 \times S + \beta_2 \times FC + \beta_3 \times AF + \beta_4 \times UR + \beta_5 \times SMR$$
$$+ \beta_6 \times \exp(CPR) \times \beta_7 \times \exp(TR)$$

Table 2 Result of a nonlinear exponential and a nonlinear logarithmic regression models on our dataset

	Exponential (nonlinear) regression	
	Dependent variable: passengers	
Variables (x_i)	*Coefficients* (β_i)	*P-value*
Seats	0.8863851*	0.000
Fuel cost	0.1345925*	0.000
Air fares	–0.1654807*	0.000
Unemployment rate	0.0111441	0.42
Stock market return	0.0317513	0.244
exp(COVID-19 positivity ratio)	–0.0617754*	0.033
exp(travel restriction)	–0.0684434*	0.000
Intercept	0.1905597*	0.014
R-squared	0.8852	
RMSE	0.04926	
	Logarithmic (nonlinear) regression	
	Dependent variable: log(passengers)	
Variables (xi)	*Coefficients (βi)*	*P-value*
Seats	1.513692*	0.000
Fuel cost	0.2499268*	0.000
Air fares	–0.1246112*	0.019
Unemployment rate	–0.0137104	0.779
Stock market return	0.0376589	0.449
COVID-19 positivity ratio	–0.2604267*	0.004
Travel restriction	–0.3351179*	0.000
Intercept	–1.555337*	0.000
R-squared	*0.8912*	
RMSE	*0.08988*	

Notes: We add star sign in front of coefficients which are significant at 5%, i.e., their p-value < 0.05.

The last columns of Table 1 shows the result which we obtained from running a logarithmic regression on our dataset. This time, we set endogenous variable is equal to exponentiate sum of the exogenous variables as $y = \exp\left(\sum x_i\right)$ which is equivalent to $\log(y) = \sum x_i$. Comparing the performance of the computed regression models based on the R-squared and RMSE (see both Tables 1 and 2), despite, RMSE is higher in Logarithmic regression model, the sign of coefficients are more relative. For instance, unemployment rate has negative relationship with passengers. Moreover, the coefficients of COVID-19 positivity ratio, and travel restriction are higher than the other two models.

3.4 Vector autoregression

We continue our experiments by implementing a multiple variable time series model which is so-called the vector autoregression (VAR). In this model, we set a vector of endogenous (dependent) variables with two elements as Passengers and Seats and the remaining are the exogenous variables. The general equation of the two-variable VAR is as

$$y_{1t} = c_1 + \sum_{i=1}^{k} \beta_{1i} x_{it} + \epsilon_{1t},$$

$$y_{2t} = c_2 \sum_{i=1}^{k} \beta_{2i} x_{it} + \epsilon_{2t}.$$

We run the model with various lag value, and we could get a best fit with lag 1 to 4 as shown in Table 3.

Table 3 Result of the vector autoregression (VAR) model on our dataset with lag = 4

Variables (x_i)	*Dependent variable: passengers*		*Dependent variable: seats*	
	Coefficients (β_{1_i})	*P-value*	*Coefficients* (β_{2_i})	*P-value*
Fuel cost	0.0388153	0.069	0.030934	0.110
Air fares	–0.1791918*	0.000	–0.2290704*	0.000
Unemployment rate	–0.1774682*	0.000	–0.2305138*	0.000
Stock market return	0.0016228	0.961	0.0055454	0.853
COVID-19 positivity ratio	–0.1580346*	0.006	–0.085554	0.103
Travel restriction	–0.3204319*	0.000	–0.3985939*	0.000
Intercept	0.6272289*	0.000	0.7514974*	0.000
R-squared	0.8400		0.8562	
RMSE	0.059366		0.053814	

Notes: In this experiment, passengers and seats are dependent variables. Coefficients with star sign means that they are significant at 5%, i.e., their p-value < 0.05.

3.5 Seasonal ARIMA regression

We analyse air passenger demand using univariate time series forecasting, autoregressive integrated moving average (ARIMA). This model is based on the concept that more recently observed values have a greater effect on the forecast than older observed values. When data exhibits a periodic (seasonal) feature, it is common to use seasonal ARIMA model, often abbreviated by SARIMA.

Table 4 Result of ARIMA regression for ARIMA$(p, d, q) \times (P, D, Q)_S$ while we set fixed values of $(P, D, Q)_S$ as $(1, 1, 1)_{12}$ for various values for (p, d, q)

(p, d, q)	(1, 1, 5)	(1, 1, 6)	(3, 1, 1)	(5, 1, 1)	(8, 1, 1)	(7, 1, 4)	(3, 0, 3)
AR coefficients							
ϕ_1	–0.5337059	0.3446433**	–0.4729978*	–1.245342*	0.624294*	0.7546515*	0.9436532*
ϕ_2			–0.5654004*	–0.7241379*	–0.198319**	–1.274407*	–0.7536254*
ϕ_3			–0.4759691*	–0.8089902*	0.0337308**	0.7567043*	0.785918*
ϕ_4				–0.2178835**	0.4596826*	–0.6205241*	
ϕ_5				0.1323879**	–0.0599637**	–0.2534946**	
ϕ_6					–0.1715543**	–0.0186314**	
ϕ_7					0.1071643**	–0.297548*	
ϕ_8					-0.1604164**		
MA coefficients							
θ_1	0.3620903**	–0.6392048*	0.1841963**	1	–0.8776107*	–1.040512	–0.1233878**
θ_2	–0.3588104*	–0.2502467**				1.158526	0.4044564*
θ_3	–0.2892561*	–0.0434982**				–1.040511	-0.3368399*
θ_4	0.2845543*	0.4204105*				0.9999985	
θ_5	0.3396193*	0.053188**					
θ_6		–0.3116162*					
Log likelihood	530.3946	537.5158	536.7762	537.7231	540.8658	545.2199	538.8701
AIC	–1,026.789	–1,039.032	–1,043.552	–1,043.446	–1,041.732	–1,050.44	–1,043.74
BIC	–968.2681	–977.0682	–991.9162	-988.3675	-972.883	-981.5915	–985.1456

Notes: Coefficients with star sign means that they are significant at 5% and two star sign goes with coefficients which are close to 0.
The best fitted model goes with more AR and MA coefficients having star and two stars besides high value of the log-likelihood and low value of AIC and BIC.

Table 5 Calculated errors for developed model evaluation

Error	*Linear reg.*	*Lasso reg.*	*Ridge reg.*	*Exp. reg.*	*Log. reg.*	*Vector autoreg.*	$ARIMA(8,1,1) \times (1,1,1)_{12}$
RMSE	4.822e-2	5.597e-2	4.823e-2	4.844e-2	3.751e-1	5.748e-2	2.323e-2
MAPE (%)	4.982	7.069	6.124	4.98	20.698	6.097	2.357

A non-seasonal ARIMA model is combination of autoregression and a moving average model that can be stated as $ARIMA(p, d, q)$, where p is the autoregressive (AR) order, d is the degree of differencing , and q is the moving average (MA) order. $ARIMA(p, 0, 0)$ is an autoregression model and $ARIMA(0, 0, q)$ is a moving average model. The general equation of $ARIMA$ can be written as

$$\hat{Y}_t = c + \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Here φ_1 denotes the estimated autoregressive for lag one (the AR(1) coefficient), c denotes the constant in the forecasting equation, and β is the vector of regression coefficients for the independent variables.

In this analysis, we follow steps including developing a model, diagnostic checking, forecasting, and validating the forecast. To explore a best fitting SARIMA model, we use grid search method by tuning various values of ARIMA models' parameters. Table 4 shows several of the fitted models which seem reasonable. In the shown sample models, a fixed values $P = 1$, $D = 1$, $Q = 1$, and $S = 12$ are chosen. To run diagnostic checking, we use Akaike information criteria (AIC) and Bayesian information criterion (BIC). In the process of selecting best SARIMA model based on the shown result in Table 4, we add star sign to the coefficient which are at 5% level of significance (i.e., p-value < 0.05). Also, we add two star sign to the coefficients which are close to 0. We compare different SARIMA models presented in Table's columns based on the coefficients of AR and MA, the log-likelihood, and the value of AIC and BIC. The best fitted model goes with more AR and MA coefficients having star and two stars besides high value of the log-likelihood, low value of AIC and BIC. Hence, $ARIMA(8, 1, 1) \times (1, 1, 1)_{12}$ is the best SARIMA for our data. The equation of the model is presented as

$$\begin{aligned}\hat{P}t = &-0.0003135 + 1.052695 S_t - 0.0027889 AF_t + 0.0149053 FC_t \\ &+ 0.0896706 UR_t - 0.0240424 SMR_t - 0.0446792 CPR_t + 0.1500731 TR_t \\ &+ 0.624294 P_{t-1} - 0.198319 P_{t-2} + 0.0337308 P_{t-3} + 0.4596826 P_{t-4} \\ &- 0.0599637 P_{t-5} - 0.1715543 P_{t-6} + 0.1071643 P_{t-7} - 0.1604164 P_{t-8} \\ &- 0.8776107 \epsilon_{t-1}\end{aligned}$$

3.6 Predictive analysis of air passenger demand

To conduct air passenger demand forecasting, we use all the five developed models above to run prediction on the dependent variable (passengers). Table 5 show the obtained loss values for each model. Comparing the loss function RMSE shows that ARIMA $(8, 1, 1) \times (1, 1, 1)_{12}$ outperforms others and has lowest error in predicting Passengers using other six independent variables.

4 Neural deep network models

The two major architectures of deep neural networks are convolution neural networks (CNNs), which are appropriate for spatial data, object recognition and image modelling, and recurrent neural networks (RNNs), which are suitable for sequence modelling. RNNs significantly enhance the capabilities of the feed forward network with recurrent memory loops, which take the input from the previous and/or same layers or states. This gives them a unique ability to model along the dimension of time and learn from arbitrary sequences of events and inputs. Long short-term memory (LSTM) and gated recurrent units (GRU) are modifications of the RNN model and gradient-based learning algorithms (Hochreiter and Schmidhuber, 1997; Cho et al., 2014).

Time series forecasting (as supervised learning) using deep neural networks has been studied since early work was done using naive RNNs (Connor et al., 1991) and hybrid models (Zhang et al., 1998; Zhang, 2003). Models have progressed from Zhang (2003) combining ARIMA and multilayer perceptions (MLPs), to the recent combination of vanilla RNN and Dynamic Boltzmann Machines in time series forecasting (Dasgupta and Osogami, 2017). Yu et al. (2017) also proposed a deep learning approach to forecast short-term and long-term traffic patterns. They applied a deep neural network based on LSTM to forecast peak-hour traffic and managed to identify unique characteristics of traffic data. They further improved the model for post-accident forecasting with a Mixture Deep LSTM model. Lai et al. (2017) also proposed a deep learning framework designed for multivariate time series forecasting, namely long- and short-term time-series network (LSTNet). LSTNet uses CNNs and RNNs to extract short-term local dependency patterns among variables and to discover long-term patterns for time series trends.

4.1 Model development

To design a best fitting deep learning model using Keras on our dataset, we begin with investigating various CNNs and RNNs models and combination of both. In one model, we had a convolutional layer (Conv1D) plus max pooling and flatten layers. In another model we only used RNNs such as a LSTM layer or a GRU layer. In another case we use a combination of Conv1D and LSTM layes. And we developed a model containing a ConvLSTM layer that is a variant of LSTM.

Besides exploring different type of models, we encountered a problem. As the COVID-19 pandemic caused a significant disruption in air transport patterns, prediction becomes more challenging and less accurate. The noise (or volatility) happens at the end of our data, including from the COVID-19 era, while training data never learn that. To address this problem, we create two sets of data, one so-called pre-COVID-19 data from 2001-2019 and the other is so-claled include-COVID-19 data from 2001-2021 and COVID-19 related features. To conduct prediction, we, first, develop a best fit deep learning model on the pre-COVID-19 data, then run the selected model on include-COVID-19 data.

Table 6 shows a comparison of the three developed RNN models on the pre-COVID-19 data. The first model is a RNN model which consists of a GRU layer, and in the second model, we define with another RNN model that we define with an LSTM layer. Model 3 contains a GRU and a LSTM layers. Furthermore, all models have Dense(1) layer as a single output layer for making one-step predictions. The table shows the parameters of each model, including the input layers, activation function, name of optimiser, number of neurons, number of epochs, batch size, and loss measures. Comparing the errors (after 10 runs), model 3 has the lowest loss among all of the tested models.

Table 6 The outcome of tuning hyper parameters to find a best first model examined on three RNN models on our pre-COVID-19 data

	Layers	*Activation*	*Optimiser*	*Epoches*	*Neurons*	*Batch size = 30*		*Batch size = 50*		*Batch size = 100*		*Batch size = 200*	
						RMSE	*MAPE %*	*RMSE*	*MAPE %*	*RMSE*	*MAPE %*	*RMSE*	*MAPE %*
Model 1	GRU	Softplus	Adam	200	32	1,470,042	3.275	1,433,743	3.09	1,450,625	3.082	1,439,149	3.109
					50	1,455,832	3.182	1,433,087	3.103	1,447,062	3.081	1,438,049	3.119
					64	1,461,939	3.21	1,433,920	3.105	1,437,248	3.081	1,482,093	3.196
					128	1,474,217	3.284	1,454,665	3.164	1,442,138	3.088	1,416,034	3.048
				100	32	1,448,220	3.082	1,462,106	3.107	1,462,636	3.115	1,494,496	3.186
					50	1,456,226	3.119	1,461,727	3.104	1,458,594	3.099	1,464,072	3.109
					64	1,453,334	3.125	1,458,954	3.105	1,459,010	3.092	1,461,115	3.1
					128	1,471,051	3.214	1,467,977	3.153	1,462,273	3.112	1,458,778	3.12
Model 2	LSTM	Softplus	Adam	200	32	1,456,587	3.221	1,490,647	3.317	1,471,044	3.151	1,436,827	3.117
					50	1,516,419	3.448	1,482,434	3.32	1,450,455	3.092	1,425,491	3.079
					64	1,483,841	3.337	1,453,754	3.233	1,442,528	3.099	1,420,425	3.078
					128	1,472,289	3.282	1,454,873	3.23	1,441,335	3.136	1,444,530	3.087
				100	32	1,447,235	3.096	1,452,120	3.075	1,457,626	3.087	1,477,110	3.145
					50	1,456,546	3.117	1,459,929	3.082	1,464,967	3.11	1,462,939	3.097
					64	1,472,948	3.129	1,462,188	3.103	1,461,406	3.102	1,462,342	3.118
					128	1,472,023	3.225	1,467,549	3.134	1,463,123	3.105	1,464,916	3.106
Model 3	GRU and LSTM	Softplus	Adam	200	32	1,421,228	3.202	1,431,891	3.16	1,412,002	3.042	1,427,438	3.052
					50	1,447,396	3.272	1,442,600	3.191	1,470,431	3.17	1,466,645	3.103
					64	1,464,789	3.258	1,462,792	3.281	1,449,709	3.095	1,425,265	3.066
					128	1,431,974	3.202	1,480,920	3.313	1,446,294	3.148	1,437,262	3.098
				100	32	1,452,575	3.167	1,455,569	3.102	1,479,406	3.133	1,491,811	3.157
					50	1,483,563	3.283	1,460,842	3.156	1,448,617	3.073	1,442,352	3.06
					64	1,517,528	3.409	1,487,499	3.254	1,451,130	3.083	1,461,512	3.125
					128	1,574,476	3.591	1,547,057	3.486	1,471,006	3.136	1,452,329	3.104

Notes: We tried CNNs and RNNs models, but here we only present well-performed ones. The errors obtained after 10 runs.

Based upon the obtained results from our experiments in fitting a model, we select model 3 as the fitted model to predict air passenger demand. Recall that in model 3 we define a GRU layer and a LSTM layer with 32 neurons and a dense layer with 1 neuron in the output layer for predicting passengers. The activation function for both GRU and LSTM layers is Softplus (i.e., $softplus(x) = \log(1 + \exp(x)))$ and for the dense layer is ReLU (i.e., $ReLU(x) = \max(0, x)$). We use the efficient Adam (derives from adaptive moments) which is an adaptive learning rate optimisation algorithm. The model fits for 200 training epochs with a batch size of 200.

Applying the selected model, model 3, on our include-COVID19 dataset (2001–2021) produces unacceptable (or inadequate) result. It is clear that the noise at the end of the data caused by the pandemic would affect the forecast. The outcome of prediction on the training set is close to the actual data and acceptable, but on the test set is too far from the actual, with substantial error of about 50%. To address this, we manipulate our data with duplicating same data several times to let the pattern to be seen in the training as well. We duplicated our data for 5 times and worked on fitting a new model on this data. Our best fitted model on include-COVID19 data consists of a GRU layer and Dense layer each with 64 neurons which provides better prediction results. The MAPE of this model on training set is 5.191% and on test set is 6.324% on 100 runs.

5 Concluding remarks

This research investigates COVID-19's impact on the US domestic air passenger demand by considering influence of various factors. After identifying the most influential features, we developed and tested several traditional and artificial intelligence methods to forecast air passenger demand. Our collected data consists of the US domestic air passenger demand from January 2001 to April 2021 that includes the volatility resulting from COVID-19 pandemic, which caused interruptions in forecasting. Our regression analysis demonstrated that (number of) passengers has a strong negative relationship with travel restriction. Passengers has moderate negative and weak positive correlation with the states of economy, as lower unemployment rate and higher stock market returns correspond to growth in customers' purchasing power, and a better economy.

Conducting predictive analysis, we developed traditional forecast models such as multiple linear regression, LASSO and Ridge regression, nonlinear regression in the form of exponential and logarithmic, vector autoregression, and seasonal ARIMA regression. Among neural deep networks forecasting models, we developed RNNs models including a LSTM layer or a GRU layer or both. Among our developed models, the seasonal ARIMA model, ARIMA $(8, 1, 1) \times (1, 1, 1)_{12}$, outperforms others. Clearly, the deep learning model is not able to handle the noise at the end of the data caused by the COVID-19 pandemic. To resolve this, we feed the training set with the noise as well.

We implemented regression models mentioned in Section 3 using 64-bit Stata 14 and deep learning models mentioned in Section 4 in Python 3.8 (64-bit), where the deep learning methods are implemented using Tensorflow 2.5 on an ordinary desktop computer.

References

Abed, S.Y., Ba-Fail, A.O. and Jasimuddin, S.M. (2001) 'An econometric analysis of international air travel demand in Saudi Arabia', *Journal of Air Transport Management*, Vol. 7, No. 3, pp.143–148.

Alekseev, K., Caloba, L. and Seixas, J. (2002) 'A neural forecasting model for the air transport passenger demand', *22nd International Symposium on Forecasting*, International Institute of Forecasters, Dublin.

Alekseev, K. and Seixas, J. (2009) 'A multivariate neural forecasting modeling for air transport–preprocessed by decomposition: a Brazilian application', *Journal of Air Transport Management*, Vol. 15, No. 5, pp.212–216.

Box, G.E.P. and Pierce, D.A. (1970) 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of the American Statistical Association*, Vol. 65, No. 332, pp.1509–1526.

Cao, L.J. and Tay, F.E.H. (2003) 'Support vector machine with adaptive parameters in financial time series forecasting', *IEEE Transactions on Neural Networks*, Vol. 14, No. 6, pp.1506–1518.

Carmona-Benitez, R.B., Nieto, M.R. and Miranda, D. (2017) 'An econometric dynamic model to estimate passenger demand for air transport industry', *Transportation Research Procedia*, Vol. 25, pp.17–29.

Chai, S.H. and Lim, J.S. (2016) 'Forecasting business cycle with chaotic time series based on neural network with weighted fuzzy membership functions', *Chaos, Solitons & Fractals*, Vol. 90, pp.118–126.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014) 'Learning phrase representations using rnn encoder–decoder for statistical machine translation', *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, pp.1724–1734.

Comi, A., Zheng, X., Liu, C.M. and Wei, P. (2020) 'Air transportation direct share analysis and forecast', *Journal of Advanced Transportation*.

Connor, J., Atlas, L.E. and Martin, D.R. (1991) 'Recurrent networks and narma modeling', *Proceedings of the 4th International Conference on Neural Information Processing Systems*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp.301–308.

Dasgupta, S. and Osogami, T. (2017) 'Nonlinear dynamic Boltzmann machines for time-series prediction', *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*.

Duval, D.T. and Schiff, A. (2011) 'Effect of air services availability on international visitors to New Zealand', *Journal of Air Transport Management*, Vol. 17, No. 3, pp.175–180.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*, MIT Press.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Comput.*, Vol. 9, No. 8, pp.1735–1780.

Hotle, S. and Mumbower, S. (2021) 'The impact of COVID-19 on domestic US air travel operations and commercial airport service', *Transportation Research Interdisciplinary Perspectives*, Vol. 9, p.100277.

Hsiao, C.Y. and Hansen, M. (2011) 'A passenger demand model for air transportation in a hub-and-spoke network', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 47, No. 6, pp.1112–1125.

Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M. and Callot, L. (2020) 'Criteria for classifying forecasting methods', *International Journal of Forecasting*, Vol. 36, No. 1, pp.167–177.

Jin, F., Li, Y., Sun, S. and Li, H. (2020) 'Forecasting air passenger demand with a new hybrid ensemble approach', *Journal of Air Transport Management*, Vol. 83, p.101744.

Karimuzzaman, M. and Hossain, M.M. (2020) 'Forecasting performance of nonlinear time-series models: an application to weather variable', *Modeling Earth Systems and Environment*, Vol. 6, No. 4, pp.2451–2463.

Kim, S. and Shin, D.H. (2016) 'Forecasting short-term air passenger demand using big data from search engine queries', *Automation in Construction*, Vol. 70.

Lai, G., Chang, W., Yang, Y. and Liu, H. (2017) *Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks*, CoRR abs/1703.07015.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, Vol. 521, pp.436–444.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018) 'Statistical and machine learning forecasting methods: concerns and ways forward', *PLOS ONE*, Vol. 13, No. 3, pp.1–26.

Maneenop, S. and Kotcharin, S. (2020) 'The impacts of covid-19 on the global airline industry: an event study approach', *Journal of Air Transport Management*, Vol. 89, p.101920.

Martinez, F., Frias, M.P., Perez-Godoy, M.D. and Rivera, A.J. (2018) 'Dealing with seasonality by narrowing the training set in time series forecasting with KNN', *Expert Systems with Applications*, Vol. 103, pp.38–48.

McKenzie, E. (1984) 'General exponential smoothing and the equivalent arma process', *Journal of Forecasting*, Vol. 3, No. 3, pp.333–344.

Papageorgiou, E.I. and Poczeta, K. (2017) 'A two-stage model for time series prediction based on fuzzy cognitive maps and neural networks', *Neurocomputing*, Vol. 232, pp.113–121.

Samadi, S.Y., Hajebi, M. and Farnoosh, R. (2019) 'A semiparametric approach for modelling multivariate nonlinear time series', *Canadian Journal of Statistics*, Vol. 47, No. 4, pp.668–687.

Srisaeng, P., Baxter, G., Richardson, S. and Wild, G. (2015) 'A forecasting tool for predicting Australia's domestic airline passenger demand using a genetic algorithm', *Journal of Aerospace Technology and Management*, Vol. 7, pp.476–489.

Sun, X., Wandelt, S. and Zhang, A. (2020) 'How did COVID-19 impact air transportation? A first peek through the lens of complex networks', *Journal of Air Transport Management*, Vol. 89, p.101928.

Tascon, D.C. and Olariaga, O.D. (2021) 'Air traffic forecast and its impact on runway capacity. a system dynamics approach', *Journal of Air Transport Management*, Vol. 90, p.101946.

Tsui, K., Balli, H., Gilbey, A. and Gow, H. (2014) 'Forecasting of Hong Kong airport's passenger throughput', *Tourism Management*, Vol. 42, pp.62–76.

Wang, W., Pedrycz, W. and Liu, X. (2015) 'Time series long-term forecasting model based on information granules and fuzzy clustering', *Engineering Applications of Artificial Intelligence*, Vol. 41, pp.17–24.

Warnock-Smith, D., Graham, A., O'Connell, J.F. and Efthymiou, M. (2021) 'Impact of COVID-19 on air transport passenger markets: examining evidence from the Chinese market', *Journal of Air Transport Management*, Vol. 94, p.102085.

Xiao, Y., Liu, J.J., Hu, Y., Wang, Y., Lai, K.K. and Wang, S. (2014) 'A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting', *Journal of Air Transport Management*, Vol. 39, pp.1–11.

Yu, R., Li, Y., Shahabi, C., Demiryurek, U. and Liu, Y. (2017) *Deep Learning: A Generic Approach for Extreme Condition Traffic Forecasting*, pp.777–785.

Zhang, G. (2003) 'Time series forecasting using a hybrid arima and neural network model', *Neurocomputing*, Vol. 50, pp.159–175.

Zhang, G., Patuwo, B.E. and Hu, M.Y. (1998) 'Forecasting with artificial neural networks: the state of the art', *International Journal of Forecasting*, Vol. 14, No. 1, pp.35–62.