Time Series Analysis of China’s Air Passenger Traffic Amid the COVID-19 Pandemic

Muge Deng*

Department of Statistical Sciences, University of Toronto, Toronto, M4Y 0C3, Canada
*Corresponding author: muge.deng@mail.utoronto.ca

Abstract. China’s air transportation industry has a great development in recent decades along with economic growth and liberalization. However, the impact of COVID-19 pandemic on China’s civil aviation industry is severe and persistent. The paper discusses the development of China’s air transportation and examines the impact of the pandemic on airline industry. The Autoregressive Distributed Lag (ADL) Model and Granger Causality test will be used to investigate the relationships between China’s air passenger traffic and its potential factors including the new COVID-19 cases in China, the Consumer Price Index and unemployment rate in China. The investigation concludes that China’s air passenger traffic is closely related to its own past observations, and the past observations of new infected cases in China is significant in forecasting air passenger traffic. The ADL model forecasts China’s air passenger traffic will have an increasing trend in the following years, but it will still require longer time to recover from the COVID-19 impact.

Keywords: ADL model; Granger Causality; COVID-19; Air transportation.

1. Introduction

China’s air transport industry has developed rapidly and tremendously over these recent decades. In 2019, the annual volume of air passenger traffic is approximately 660 million, representing a seven-fold increase in the air passenger traffic over the last 16 years. Fu et. al noted that air transport liberalization has positive effects on the economy and traffic volume due to the increased competition and efficiency in the airline industry [1]. Young investigates the relationship between the demand of air transportation service and several factors including ticket price, journey time and income [2]. Zhang and Zhang examined the determinants of air passenger flows in China and noted that the continuing liberalization of the air transport industry and the role of low-cost carriers have great positive impacts on developing China’s domestic aviation markets [3].

However, the COVID-19 pandemic brought a huge impact on the global civil aviation industry and the total air passenger traffic in China decreased by 36.7% in 2020. The impact on international routes is deeper and more persistent and the air passenger traffic of international flights has decreased by more than 80% for two years in a row. The Civil Aviation Administration of China has imposed strict policies on international routes in which each airline company is only allowed to maintain one route to any country, and each route can only have one round-trip flight per week. The decrease in the number of international flights resulted in a supply-demand imbalance which led to sky-high ticket prices.

Investigating the relationship between the COVID-19 pandemic and China’s air passenger traffic will help to detect the changes in the pandemic situation, implement scientific, and precise policies and making achievements in developing the air transport industry safely. This paper will analyze the time series of China’s air passenger traffic and investigate the relationships between air passenger traffic and several potential factors including the new COVID-19 cases, Consumer Price Index, and unemployment rate using the Autoregressive Distributed Lag (ADL) Model and Granger Causality test. The ADL model will use the past observations of China’s air passenger traffic and other variables to forecast China’s air passenger traffic.
2. Methodology

2.1 ADL Model

This paper investigates the impact of the COVID-19 pandemic on China’s airline industry by examining the relationship between new COVID-19 cases in China and China’s air passenger traffic. The assumption of the pandemic has a hard hit on China’s air passenger traffic, especially on international flights. The Autoregressive Distributed Lag (ADL) Model will be used to analyze the relationship between China’s air passenger traffic and new COVID-19 cases in China, as well as other potential variables.

The ADL model is a multivariate time series model that makes a connection between the current observation of a variable with its past observations and other variables’ past observations. This approach has the advantage over other univariate models since there are no exogenous variables in the model, and it is able to capture and adapt the dynamic relationships among variables [4]. A similar approach has been used on studying the relationship between real exchange rate volatility and U.S. bilateral trade, and the relationship between stock market development and economic growth, which has the ability to investigate the dynamic interrelations [5, 6]. The variable of air passenger traffic has multiple determinants, such as macroeconomic environment and market liberalization, it is also an important indicator of China’s air transport industry. The number of variables in the model will affect the accuracy of the forecast, excluding a relevant variable may result in low forecast accuracy, but adding an irrelevant variable may make the model overcomplicated with excess noise [7]. Therefore, the ADL model that will be used in this paper can be expressed as:

\[
\ln AIRPASS_t' = \beta_0 + \beta_1 \ln AIRPASS'_{t-1} + \beta_2 \ln AIRPASS'_{t-2} + \gamma_1 COVID_{t-1} + \gamma_2 COVID_{t-2} \\
+ \delta_1 CPI_{t-1} + \delta_2 CPI_{t-2} + \theta_1 UNEMP'_{t-1} + \theta_2 UNEMP'_{t-2} + \mu_t
\]  

(1)

where \( \ln AIRPASS'_{t} \) is the first-order differenced of natural logs of China’s air passenger traffic at time \( t \); \( COVID_t \) is the monthly new COVID-19 cases in China at time \( t \); \( CPI_{t} \) is the seasonal-differenced monthly consumer price index in China; \( UNEMP'_{t} \) is the first-order differenced monthly unemployment rate in China at time \( t \); \( \mu_t \) are error terms.

In order to decide what variables should be included in the ADL model, the approach of Granger causality will be used to test the forecasting relevance of variables. As Granger suggested, the variable of new COVID-19 cases is causing China’s air passenger traffic if the prediction of air passenger traffic can be more accurate using the past observations of air passenger traffic and new COVID-19 cases than using the past observations of air passenger traffic alone [8, 9]. Therefore, Granger causality tests whether a variable is statistically significant for forecasting another variable. For example, an F-test will be performed with the null hypothesis of “new COVID-19 cases do not cause China’s air passenger traffic”, if the p-value is less than the significance level of 5%, the null hypothesis will be rejected.

2.2 Data Collection

The data of China’s air passenger traffic is obtained from the website of Civil Aviation Administration of China^2. Three time series of China’s air passenger traffic will be used in this paper, the first time series contains the monthly air passenger traffic of all flights ran by China’s civil aviation from July 1998 to August 2022, and the second time series contains the monthly data of air passenger traffic of China’s domestic flights from Jan 2005 to August 2022, the third time series contains the monthly data of China’s international flights from Jan 2005 to August 2022. The sum of the domestic and international passenger traffic is the observations in the first time series.

There are several indices of China’s air transportation industry, such as turnover volume, passenger load factor, and aircraft movements, the variable of passenger traffic is chosen since it can represent China’s air transport development and liberalization comprehensively. The supply of air
transportation service can be assessed by weekly flights and available seats, and the demand is assessed by number of passengers [10]. Zhang and Chen used several indices to discuss China’s air transport development, including labor productivity, load factor, and international inbound/outbound tourists [11]. Although these variables can also represent the development of China’s air transport industry, they can only demonstrate the growth from one aspect. Since this paper aims to investigate the impact of COVID-19 on China’s air transportation, a representative variable is required. Zhang and Zhang discuss the development of China’s air transport policy and choose air passenger traffic as the investigation object since it is an important indicator measuring the development of the air transport industry [3].

A few variables will be used as the potential determinants of China’s air passenger traffic, including monthly new cases of COVID-19 in China, the Consumer Price Index, and the unemployment rate in China. The time series of new cases of COVID-19 in China contains monthly observations from Feb 2020 to Aug 2022. The time series are collected from China’s National Health Commission. According to the 2020 annual statistical report of China’s Civil Aviation, the COVID-19 pandemic has brought tremendous shock to China’s civil aviation, the total volume of passenger traffic decreased by 36.7% compared with 2019, and the total volume of international passenger traffic decreased by 87.1%. In 2021, the statistical data show that the impact of COVID-19 on China’s civil aviation is more severe and persistent than expected. The passenger traffic of domestic flights has increased by 7.6% over last year, but the international routes have not recovered from the pandemic that the passenger traffic decreased by 84.6% compared with 2020.

The time series of the Consumer Price Index (CPI) contains monthly data from Jan 1995 to Aug 2022, obtained from the National Bureau of Statistics of China. CPI is defined as the measure reflecting the trend and degree of changes in prices of consumer goods and services purchased by urban and rural households during a given period. CPI is a useful measure of inflation and deflation, and therefore it is an important indicator of the economy. A paper by Fu et al. investigates the relationship between air transportation and the economy and noted that the demand of transportation service is driven by the overall economy, and it also affects the economy [1]. Fu et al. concluded that liberalization has a positive effect on the economy and traffic volume [1]. Therefore, CPI as an indicator of economy is included in the ADL model.

The time series of the unemployment rate contains monthly data from Jan 2017 to August 2022, obtained from the National Bureau of Statistics of China. The unemployment rate is an essential indicator of the performance of China’s labor market, it is also an indicator of the economy and should be included in the model.

3. Results

3.1 Descriptive Statistics

The graph below shows the three time series plots of China’s air passenger traffic. As the first plot of total air passenger traffic shows, the passenger traffic had an increasing trend before 2020, but it experienced a huge fall at the beginning of 2020. The observation of Feb 2020 is 8.34 million passengers, which is even lower than the level in 2004, which means the COVID-19 pandemic has caused a regression of 16 years. The observations after 2020 have some large fluctuations, but the crests are still lower than the level before COVID-19. Since the time series has an increasing trend and the variance is also increasing as the fluctuations are becoming larger in these recent years, the original time series is non-stationary and requires a logarithm transformation and first-order difference to become stationary.
The second plot shows the time series of air passenger traffic of domestic flights. The trend and pattern are similar to the first plot of total passenger traffic since most of the passengers carried are from domestic flights. The graph on the bottom shows the time series of air passenger traffic from international flights, it also has an increasing trend before the COVID-19 pandemic, but it experienced a dramatic drop and does not recover ever since. The volume of international air passenger traffic dropped to the lowest level of 0.082 million passengers in December 2021 which is even lower than the observations in 2005. The different patterns of air passenger traffic after 2020 among these three plots show that COVID-19 impact on international flights is much deeper and more persistent than on domestic flights due to several factors.

The three variables of China’s air passenger traffic have a different number of observations because the Civil Aviation Administration of China only published the total volume of air passenger traffic before 2005. The total air passenger traffic will be used as the response variable in the ADL model. Since the mean of domestic air passenger traffic is much larger than the mean of international air passenger traffic as shown in the table, a two-sample t-test will be performed and the test statistic is 28.975 with a p-value of 0.000 which is statistically significant at 5%. Therefore, the null hypothesis
is rejected and the conclusion of the true difference in means of the two groups is not zero could be made.

The variable of new COVID-19 cases has only 31 observations since it starts to collect observations in Feb 2020 after the COVID-19 pandemic. The difference between the maximum and minimum observation is 69,258 cases, and the value of standard deviation is also significantly large, which suggests that China experiences a few COVID-19 outbreaks. Figure 2 shows that there were two outbreaks of COVID-19 cases, and the number of new cases is stationary for the rest of the time.

![New COVID-19 Cases](image)

**Fig. 2** New COVID-19 Cases in China

The variable of the Consumer Price Index contains 332 observations. Figure 3 shows that the time series is approximately stationary, but there is a strong seasonality pattern as shown in the seasonal plot of quarterly CPI. Therefore, a seasonal difference is required to make the time series stationary.

![Consumer Price Index in China and seasonal plot of quarterly CPI](image)

**Fig. 3** Consumer Price Index in China and seasonal plot of quarterly CPI

The variable of unemployment rate contains only 68 monthly observations since it is the surveyed unemployment rate instead of the registered unemployment rate which is suspended since 2022. Figure 4 shows that the original time series is non-stationary and requires a first-order difference to stabilize the mean.
3.2 Regression Results from ADL Model

Table 2. Estimates of Regression Models

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln $AIRPASS'_{t-1}$</td>
<td>-0.101*</td>
<td>-0.101</td>
<td>-0.057</td>
<td>-0.294</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.245)</td>
<td>(0.256)</td>
<td>(0.292)</td>
</tr>
<tr>
<td>ln $AIRPASS'_{t-2}$</td>
<td>-0.211***</td>
<td>0.144</td>
<td>0.206</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.212)</td>
<td>(0.224)</td>
<td>(0.288)</td>
</tr>
<tr>
<td>$COVID_{t-1}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>$COVID_{t-2}$</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>$CPI_{t-1}$</td>
<td>0.097</td>
<td>-0.136</td>
<td>-0.167</td>
<td>-0.987</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td>(0.137)</td>
<td>(0.137)</td>
<td>(0.662)</td>
</tr>
<tr>
<td>$CPI_{t-2}$</td>
<td>(0.134)</td>
<td>0.087</td>
<td>(0.137)</td>
<td>(0.668)</td>
</tr>
<tr>
<td>$UNEMP'_{t-1}$</td>
<td>-0.987</td>
<td>(0.662)</td>
<td>-0.150</td>
<td></td>
</tr>
<tr>
<td>$UNEMP'_{t-2}$</td>
<td></td>
<td>(0.668)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.049</td>
<td>0.131</td>
<td>0.176</td>
<td>0.275</td>
</tr>
<tr>
<td>No. of Obs.</td>
<td>271</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>

Notes: Response variable is ln $AIRPASS'_{t}$; standard deviation in parenthesis; Significance level: * 0.1; ** 0.05; *** 0.01

Table 2 contains the estimates of coefficients for four ADL models, from the simplest one using only the air passenger traffic as an explanatory variable to the most complicated one with four explanatory variables.
Model 1 only uses the past observations of air passenger traffic to make predictions. Since \( \ln \text{AIRPASS}'_t \) is the first-order difference of natural logs of China's air passenger traffic, the value of \( \ln \text{AIRPASS}'_t \) will be negative if the air passenger traffic decreases, and it will be positive if the air passenger traffic increases compared with the previous month. The estimated coefficient for \( \ln \text{AIRPASS}'_{t-1} \) is -0.101, which means a 1% increase in the first-order difference of natural logs of China’s air passenger traffic at time \( t-1 \) will result in a -0.101% change in the first-order difference of natural logs of China’s air passenger traffic at time \( t \). The second estimate is -0.211, which means a 1% change in the first-order difference of natural logs of China’s air passenger traffic at time \( t-2 \) has an impact on the corresponding -0.211% change in the first-order difference of natural logs of China’s air passenger traffic at time \( t \). Since the two estimated coefficients are both negative, indicating the past observations \( \ln \text{AIRPASS}'_{t-1} \) and \( \ln \text{AIRPASS}'_{t-2} \) have an inverse effect on the current observation. For example, if both \( \ln \text{AIRPASS}'_{t-1} \) and \( \ln \text{AIRPASS}'_{t-2} \) are negative, which means China’s air passenger traffic has been decreasing for two months, the value of \( \ln \text{AIRPASS}'_t \) will be positive and the air passenger traffic will increase at time \( t \). According to the p-value, the two variables are both significant in forecasting the observation at time \( t \), especially the variable \( \ln \text{AIRPASS}'_{t-2} \) is closely related to the current observation.

Model 2 includes two variables to forecast China’s air passenger traffic. It only contains 28 observations since the time series of new COVID-19 cases has a limited amount of observations. By comparing the estimates of coefficients in model 1 and model 2, the first estimate remains unchanged, but the second estimate of parameter \( \ln \text{AIRPASS}'_{t-2} \) changes to a positive value of 0.144 and the absolute value of the estimate decreases, which means the observation of \( \ln \text{AIRPASS}'_{t-1} \) still has an inverse effect on the current observation, but \( \ln \text{AIRPASS}'_{t-2} \) now has a positive effect on \( \ln \text{AIRPASS}'_t \). The second estimate is 0.144, which suggests that a 1% change in \( \ln \text{AIRPASS}'_{t-2} \) will result in a 0.144% change in the first-order difference of natural logs of China’s air passenger traffic at time \( t \). The two estimates of coefficients of new COVID-19 cases are 0.000 since the actual value is very small and could not be shown in three decimals. However, it does not mean there is no relationship between China’s air passenger traffic and COVID-19 cases. The mean of new COVID-19 cases is much larger than the mean of air passenger traffic, the larger difference between the observations of the two time series results in the small estimate of coefficients. The variable \( \text{COVID}_{t-2} \) is significant in making predictions on \( \ln \text{AIRPASS}'_t \).

Model 3 and model 4 add the variables of CPI and unemployment rate to forecast China’s air passenger traffic. The values of the first two estimates do not change much among model 2, 3 and 4, but the standard error is increasing when more variable is included in the model. When adding extra variables into the model, the R-squared increases, but it does not mean the forecast on air passenger traffic is more accurate since adding additional variables may result in extra estimation error. In model 3 and 4, there are no significant parameters in forecasting the air passenger traffic. In order to investigate which variable should be included in the ADL model, the Granger Causality test will be conducted to test if each variable has the power in explaining China’s air passenger traffic.

<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New COVID-19 cases do not cause air passenger traffic</td>
<td>0.2397</td>
</tr>
<tr>
<td>2</td>
<td>CPI does not cause air passenger traffic</td>
<td>0.2078</td>
</tr>
<tr>
<td>3</td>
<td>Unemployment rate does not cause air passenger traffic</td>
<td>0.3852</td>
</tr>
</tbody>
</table>

Table 3 shows the results of three Grange Causality tests about the predicting power of three variables on China’s air passenger traffic. Since the p-value is greater than the significance level of 5%, the null hypothesis failed to be rejected. Therefore, the three variables of new COVID-19 cases, CPI, and unemployment rate do not have strong power in explaining and predicting China’s air passenger traffic.
3.3 Forecasting

The four models will be used to forecast the future observations of the first-order difference of natural logs of China’s air passenger traffic. Since the newest observation in some time series is the observation from July 2022, the four models are built using the observations up to July 2022, and they will forecast the future values starting from August 2022.

Figure 5 shows the forecasts of $\ln(\text{AIRPASS}_t')$ using model 1, indicating the air passenger traffic will decrease first then increase afterward. Since a first-order difference and log transformation are performed on the time series of China’s air passenger traffic to make it stationary, the value predicted will be transformed reversely to obtain the forecast of China’s air passenger traffic. The actual volume of air passenger traffic in August 2022 is 32.303 million, and the forecast volume using model 1 is 28.787 million which is lower than the actual value.

![Fig. 5 Forecasts of first-order difference of natural logs of China's air passenger traffic using Model 1](image1)

Figure 6 shows the forecasts of $\ln(\text{AIRPASS}_t')$ using model 2. Since the forecast values of the first-order difference of natural logs of air passenger traffic are positive, the forecast air passenger traffic will keep increasing from August 2022. The forecast volume of air passenger traffic in August 2022 is 34.781 million, which is closer to the actual observation of 32.303 million.

![Fig. 6 Forecasts of first-order difference of natural logs of China’s air passenger traffic using Model 2](image2)

Figure 7 shows the forecasts of $\ln(\text{AIRPASS}_t')$ using model 3. The forecast trend is similar to the trend in figure 7, and the forecast values are also positive, suggesting that the air passenger traffic will increase in the future. The forecast value of China’s air passenger traffic is 37.200 million, which is less accurate than the predictions of model 1 and model 2.
Fig. 7 Forecasts of first-order difference of natural logs of China’s air passenger traffic using Model 3

Figure 8 shows the forecasts of $\ln \text{AIRPASS}_t'$ using model 4. By reversing the difference and log transformation, the forecast China’s air passenger traffic will first increase, then decrease slightly for a few months, then increase again afterward. The forecast value of air passenger traffic in August 2022 is 36.153 million, which is closer to the actual value of 32.303 million than the predictions of model 3, but it is still less accurate than model 1 and model 2.

Fig. 8 Forecasts of first-order difference of natural logs of China's air passenger traffic using Model 4

3.4 Discussion

The Granger Causality test and ADL models are used to analyze China’s air passenger traffic and its potential determinants, but the results show that the three variables of new COVID-19 cases, CPI and unemployment rate do not have strong power in causing and explaining the air passenger traffic. Among the four ADL models, model 1 and model 2 contain more significant variables and make better predictions on air passenger traffic than model 3 and 4.

The COVID-19 impact on China’s air passenger traffic is obvious visually and statistically that the volume of air passenger traffic has dropped dramatically since 2020. However, the estimates of coefficients for new COVID-19 cases are all very small, and the variable of new COVID-19 cases does not have strong power in predicting the air passenger traffic. The results are different from the hypothesis may be due to the small sample size of new COVID-19 cases that model 2, 3 and 4 only contains 28 observations. The limited amount of sample size may result in an inefficient model and inaccurate forecasts.

CPI is included in the ADL model to investigate the relationship between the overall economy and air passenger traffic. Compared with CPI, GDP may be a better indicator of the overall economy, since CPI is focusing on inflation and deflation. However, GDP is not included in this model due to the mismatching of frequency between the time series of GDP and air passenger traffic. The time series of GDP of China is only available in quarterly frequency, and making the time series of air passenger traffic quarterly instead of monthly will reduce the number of observations. The original time series of China’s air passenger traffic contains 290 observations and it will have less than 100 observations if the frequency is changed to quarterly, which will be insufficient to build an ADL model.

China’s unemployment rate has the controversy of being inauthentic that the real unemployment rate would be much higher. The unemployment rate used in this paper is the surveyed unemployment rate but not the registered unemployment rate, which may be inaccurate.

China’s air passenger traffic during the COVID-19 pandemic have large fluctuations due to the strict epidemic prevention policy in China. After the pandemic started in 2020, the number of international flights decreased dramatically since one international air route can only have one round-trip flight every week. The air route will be suspended if there are confirmed cases of COVID-19 from the previous flights. The air passenger traffic of domestic flights will also be affected by the
epidemic situation in China. Zhang and Chen suggest that the development of China’s air transportation is depending on its overall economy and policy evolution, and the present air passenger traffic is also well affected by the pandemic prevention policy [11]. Therefore, the forecast error could be reduced by including a variable about China’s policy on air transportation.

4. Conclusion

China’s air transport industry has developed tremendously over these recent decades, four ADL models are built to investigate the relationship between China’s air passenger traffic and three potential determinants.

The results show that the current volume of China’s air passenger traffic is well related to its own past observations that the air passenger traffic at first and second lags are both significant in predicting the current observation. The negative estimates of coefficients suggest that the past observations of first-order difference of natural logs of China’s air passenger traffic have an inverse effect on the current observation. An increase in the first and second lags of first-order difference of natural logs of China’s air passenger traffic will result in a decrease in the current observation. The forecasts of China’s air passenger traffic using only its own past observations show that the air passenger traffic will first decrease then increase.

Although the result of the Granger Causality test does not support the hypothesis that the new COVID-19 cases are causing the air passenger traffic, the ADL model shows that the past observations of new COVID-19 cases have some power in explaining and predicting the current value of air passenger traffic. Model 3 and model 4 contain more variables which increase the R-squared, but it also brings in extra noise in the model. All the parameters become insignificant and the forecasts of air passenger traffic have larger errors.

This paper contributes to the study of China’s air transportation development and the COVID-19 impact on China’s civil aviation by using the ADL model and Granger Causality test. The ADL model built in this paper uses a limited amount of observations, and the forecast accuracy could be improved when the sample size of COVID-19 cases increases. The results may be useful to forecast the future trends in China’s air transportation during the COVID-19 pandemic. It also incurs thoughts and reflections on how to balance between strict pandemic prevention policy and a reasonable number of flights.

References


