

# Prediction of Cathay Pacific Airways Stock Price via Time Series Model for Year 2023 and 2024

Peijie Tan \*, Xinran Ren, Shaorui Xiong, Kaiqi Zhang

Hong Kong Metropolitan University, Hong Kong 999077, China;

\*Corresponding author E-mail: leorrrrtan@163.com

**Abstract.** This essay explains and demonstrates the prediction of the future trend of the stock of Cathay Pacific Airways Limited in the following two years using time series model. In order to correct the stationary of the primary data, the p-value in Augmented Dickey-Fuller test is used to check whether the model is stationary. After that, auto-correlation function (ACF) and partial auto-correlation function (PACF) which gives the lagged values from different series and find correlation of the residuals with the next lag value. Finally, use Autoregressive Integrated Moving Average model to function the corrected model to test which is the best fitted model. Afterwards, forecasting the confidence interval and the specific fluctuations by the best ARIMA model. The result doesn't show the actual stock price in the future time, the forecasting model, however, only shows the trend of its stock price. The results of the modelling and analysis are presented such that can be used as a study to the prediction of any non-stationary data.

**Keywords:** ACF; ARIMA; Forecast; PACF; Stationary; Time Series.

## 1. Introduction

Cathay Pacific Airways Limited (CPA) is an airline that provides civil aviation services to Hong Kong. Cathay Pacific Airways' stock was selected as the financial data to construct a time series model. Collecting the monthly stock price data of Cathay Pacific Airways from February 2000 to May 2022 from Yahoo! Finance. Then, there are few steps to lead to the future trend of the stock fluctuation. (Yahoo! Finance, 2022)

Due to the large-scale layoffs and work stoppages of Cathay Pacific due to the impact of the epidemic, the stock price of Cathay Pacific has experienced a serious downward trend in the past two years. Therefore, this trend can be estimated to obtain the approximate trend of Cathay Pacific in the next two years to determine whether Cathay Pacific is worth to invest in the future investment. Time series model could better help the investors to understand the influence of historical fluctuations to the future trend, based on that

The essay has a deeper understanding of the time series model and other auxiliary models and detection methods, so as to obtain more accurate future trends. Now there are many superimposed models based on different models to form a complex model to obtain more accurate predictions. The relative basis of time series models can be used as an investor wants to know whether the stocks of the companies, they invest in can reach the desired profit level, so such prediction models can well satisfy some investors who do not have a deep academic understanding.

## 2. Methodology

### 2.1 Data Collection

The work in this essay is focused on the data regarding the stock market. This study began with the stationary of the stock price of Cathay Pacific Airline Limited. The publicly available stock data sets contain historical data about its stock price has been collected from Yahoo finance. The dataset specifies the "adjusted closing price" against each month. The historical data of stock price collected through 20 years beginning from "February 2000 to May 2022" has been taken into consideration for this work. The data has been set up as follows the specific month and used for formulation of the model while the model needs to be corrected while tends to stationary before forecasting.

## 2.2 Data Analysis

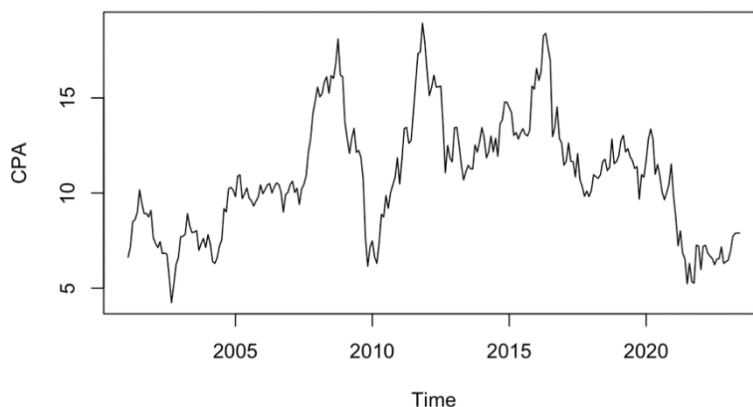
First, it is needed to determine whether the time series is stationary. The time series refers to a sequence of data of the same type sorted in chronological order. In financial data, the stock price trend sorted by occurrence time is a typical time series with non-stationary. Time series stationarity is related to data volatility and dispersion.( F.-M. Tseng, G.-H. Tzeng, H.-C. Yu, and B. J. C. Yuan, 2001) Usually, judging whether the data is stationary by whether the mean or variance or autocorrelation of the data change over time(E. G. Jain and B. Mallick, 2017). A stationary time series is plotted as the data fluctuates up and down around a fixed mean value. Conversely, non-stationary shows a continuous rise or a continuous decline or no fixed mean over a period of time. Then, stationary of time series model was needed some tests to check whether it is stationary. Including “augmented Dickey-Fuller (ADF) test (V. Tsioumas, S. Papadimitriou, Y. Smirlis, and S. Z. Zahran, 2017), L-jung-Box test (Hassani, H., & Yeganegi, M, 2019)”. In the ADF test, if the value of ‘p’ is less than 0.05 in 95% of confidence interval, then the series is stationary. However, there are some special cases when the series is non-stationary. After that, auto-correlation function (ACF) and partial auto-correlation function (PACF) are used to check whether which kind of model is the best to forest in the end.

## 2.3 Data Presentation

After analyzing and testing the given data from Cathay Pacific Airways Limited, the coding results by R will be shown as pictures or forms in the following paragraphs.

## 3. Result and Discussion

### 3.1 Time Series Plot Modelling of CPA’s Stock



**Figure 1.** Time series plot of Cathay Pacific Airline Limited’s stock price.

As indicated in Figure 1, the time series plot shows that there is no constant increase or decrease obviously. But the data fluctuates and the mean value changes greatly, so it initially judges that the time series is non-stationary.

In order to accurately determine whether the time series is stationary, it needs to detect through Augmented Dickey-Fuller Test (ADF Test).

**Table 1.** ADF test result.

Null & Alternative Hypothesis	<i>H0: the time series is non – stationary</i> <i>H1: the time series is stationary</i>
Title	Augmented Dickey-Fuller Test
Test Results:	PARAMETER: Lag Order: 1 STATISTIC: Dickey-Fuller: -0.6346 P VALUE: 0.4141

The essence of the ADF Test is a statistical significance test, which is to judge the hypothesis by the value of the p-value, and then accurately obtain the stationarity of the time series. The null hypothesis and alternative hypothesis are as follows (Table 1).

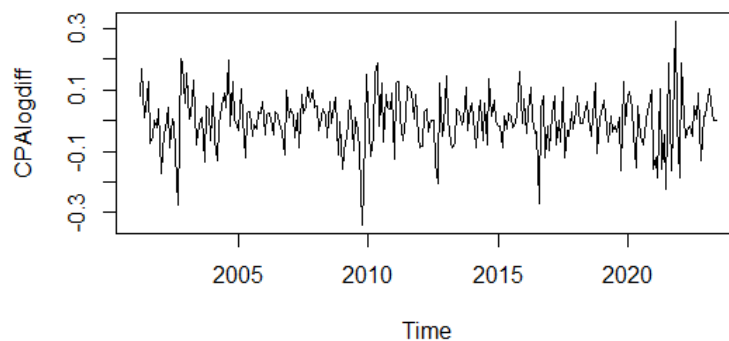
Since the p-value of the significance test is 0.4141 which is greater than 0.05, indicating that there is no evidence to reject the null hypothesis. Therefore, it determined that the time series is non-stationary. That is, the CPA stock changes over time and the statistical properties also change over time.

### 3.2 Differencing

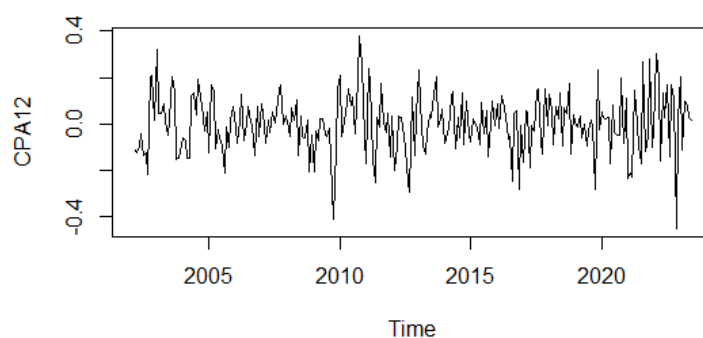
Non-stationary data are unpredictable and cannot be modeled or predicted, so it needs to convert non-stationary time series into stationary time series. To remove the time dependence of the series and stabilize the mean of the time series, transforming the series into a new time series using differencing method, where the trend and seasonality are reduced during the transformation. The data was corrected by taking one or many diffs, or taking logs, including fix variance, correct the mean and seasonality. After that, the mean should be completely constant in the data, the variance should be constant during the given time range, and the time series does not show seasonality.



**Figure 2.** Fixed variance of stock price.



**Figure 3.** Correct the mean of the fixed variance stock price.



**Figure 4.** Correct the seasonality of corrected stock price.

After using the log function in RStudio to derive the differencing results, the ADF Test was used again to test the stationary of the time series with the following null and alternative hypotheses.

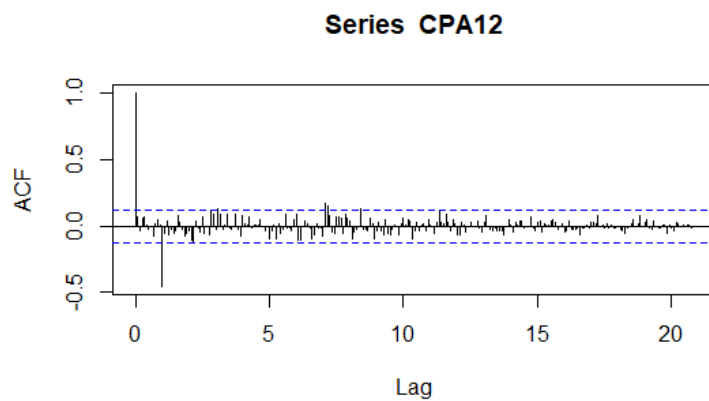
**Table 2.** Augmented Dickey-Fuller test result.

Null & Alternative Hypothesis	<i>H0: the time series is non – stationary</i> <i>H1: the time series is stationary</i>
Title	Augmented Dickey-Fuller Test
Test Results:	PARAMETER: Lag Order: 1 STATISTIC: Dickey-Fuller: -11.3521 P VALUE: 0.01

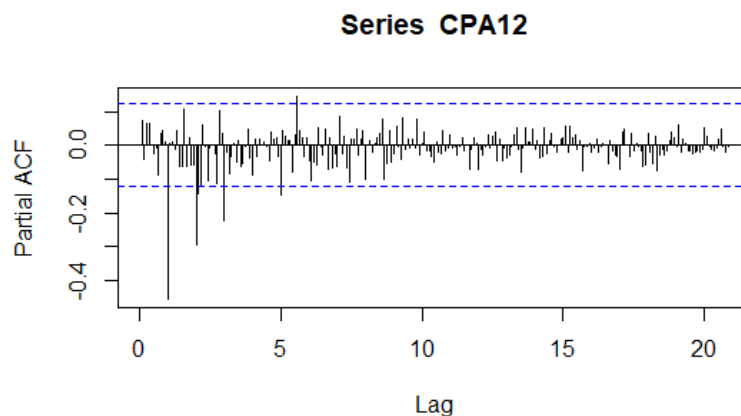
According to the ADF Test results from RStudio (Table 2), the p-value is equal to 0.01 in this test, which is less than 0.05. It means it has strong evidence to reject the null hypothesis. Thus, at the 5% significant level, it can concludes that time series is stationary after differencing.

### 3.3 ACF and PACF

After obtaining the stationary time series, the next step is to determine the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary time series. And they can form the basis for the determination of the ARIMA model, including the number of steps of the lagged terms in AR and MA.



**Figure 5.** The results of ACF plot.



**Figure 6.** The results of PACF plot.

ACF is a complete autocorrelation function that provides the autocorrelation value of any time series with lagged values. It describes the degree of correlation between the current value of the time series and its past values. In the Figure. 4, the ACF plot describes the autocorrelation between one observation and another, and includes both direct and indirect correlation information.

The PACF is a partial autocorrelation function. Instead of finding the correlation between a lagged value like ACF and the current one, it finds the correlation between the residuals and the next lagged

value. Thus, if there is any hidden information in the residuals that can be modelled by the next lag, it may obtain a good correlation. In the Figure. 5, the PACF plot describes only the direct relationship between the observations  $y_t$  and  $y_{t-k}$ , removing the effect of other lagged values ( $y_{t-1}$ ,  $y_{t-2}$  .....  $y_{t-k-1}$ ).

When deciding which process to use for modelling, the plots of ACF and PACF should be considered together to define the process. For the AR process, it expects the ACF plot to show a geometric or gradual decline, while the PACF plot declines sharply after a significant lag of  $p$ . The opposite is true for the MA process, which means that: the ACF plot should show a sharp decline after a certain number of  $q$  lags, while the PACF plot should show a geometric or gradual decline. However, if both the ACF and PACF plots show a gradually decreasing pattern, the ARMA process should be considered for modelling.

According to the ACF plot, it can be seen that most fall within a 2-fold standard deviation range, except for a few orders such as the first order, which are geometric decays of the ACF. This indicates that the ACF has a clear sinusoidal trajectory, which is a case of the autocorrelation coefficient decaying continuously to zero.

According to the PACF plot, the situation is roughly the same as the ACF plot, which also shows an autocorrelation coefficient decaying continuously to zero, which is a geometrically decaying PACF.

In summary, both ACF and PACF decay geometrically and it can be judged that was used the ARMA process for modelling.

### 3.4 Box-Ljung test

Before using ARIMA model to forecast the future trend, it better to use Box-Ljung test to check whether or not the autocorrelations for the errors or residuals are non-zero, which means is there any errors are sequence of independent, identically distributed (IID) random variables (i.e. white noise) or something more behind them. The null hypothesis and alternative hypothesis are as follow (Table 3).

**Table 3.** Box-Ljung test result.

Null & Alternative Hypothesis	$H_0$ : our model does not show lack of fit $H_1$ : our model does show a lack of fit
Box-Ljung test results:	<pre>&gt; Box.test(CPA12,type='Ljung',lag=log(length(CPA12)))</pre> Box-Ljung test  data: CPA12 X-squared = 3.8926, df = 5.5452, p-value = 0.6367

As the test results shows in Table 3, p-value is equal to 0.6367 which is greater than 0.05, indicating that there is no evidence to reject the null hypothesis. That is, our model does not show lack of fit and residuals are independent. Therefore, it can concludes that our model is Non- white noise stationary time series model. And ARIMA model can be used to forecast in the following analysis.

### 3.5 ARIMA Model:

According to the ACF and PACF test of the stationary series after differencing and set up an ARIMA model with the property that either  $p = 0$  or  $q = 0$  ( $p$ : the number of the autoregressive terms (AR),  $q$ : the number of lagged forecast errors in the prediction equation (MA)). Now it needs to use the ARIMA model to form the data and forecast the fluctuation in the future 24 months.

The ARIMA model is a type of regression analysis that measures the strength of one dependent variable in relation to other variables that change. Its goal is to forecast future securities or financial market movements by examining differences in time series values rather than actual values. By varying the values  $p$  and  $q$ , you can determine which model is best suited to forecast the future trend of the time series. Therefore, transforming the series and fit the following 3 models (Table 4) :

ARIMA (0,1,0), ARIMA (1,1,0), ARIMA (2,1,0) to test whether any model would fit in the appropriate model.

**Table 4.** ARIMA modelling result.

ARIMA Model	Test Result
ARIMA (0,1,0)	<pre>&gt; arima1 &lt;- Arima(CPA12, order=c(0,1,0)) &gt; summary(arima1) Series: CPA12 ARIMA(0,1,0)  sigma^2 = 0.02933: log likelihood = 88.15 AIC=-174.29 AICc=-174.28 BIC=-170.75  Training set error measures:               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Training set 0.000502197 0.1709188 0.1320834 84.86628 495.7529 0.7861009 -0.4426265 &gt;</pre>
ARIMA (1,1,0)	<pre>&gt; arima2 &lt;- Arima(CPA12, order=c(1,1,0)) &gt; summary(arima2) Series: CPA12 ARIMA(1,1,0)  Coefficients:       ar1     -0.4409 s.e.    0.0560  sigma^2 = 0.02367: log likelihood = 115.84 AIC=-227.68 AICc=-227.63 BIC=-220.6  Training set error measures:               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Training set 0.000746181 0.1532645 0.1165723 143.5371 273.6882 0.6937856 -0.1687618 &gt;</pre>
ARIMA (2,1,0)	<pre>&gt; arima3 &lt;- Arima(CPA12, order=c(2,1,0)) &gt; summary(arima3) Series: CPA12 ARIMA(2,1,0)  Coefficients:       ar1      ar2     -0.6080  -0.3761 s.e.    0.0579   0.0577  sigma^2 = 0.02036: log likelihood = 135.45 AIC=-264.89 AICc=-264.8 BIC=-254.27  Training set error measures:               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Training set 0.001122111 0.1418368 0.1096522 91.77597 249.09 0.6526003 -0.1071537 &gt;</pre>

After summarizing those models, the third model ARIMA (2,1,0) shows the best in terms of AICc (-264.8). And the other two models only slightly worse in terms of AICc (-227.63 and -174.28 respectively). However, the value of the first model slightly less bias than the second and the third model on ME scores. It's worth noting that the only difference between these models is whether p or q is set to zero, so the based on the summary of these three models, the first model ARIMA (0,1,0) is the best fit to form the ARIMA model for the stock price to forecast the future value.

**Table 5.** Auto-Arima modelling result.

ARIMA (0,0,0) & (2,0,0)	<pre>&gt; arima = auto.arima(CPA12) &gt; summary(arima) Series: CPA12 ARIMA(0,0,0)(2,0,0)[12] with zero mean  Coefficients:       sar1      sar2     -0.6753  -0.3556 s.e.    0.0614   0.0634  sigma^2 = 0.01058: log likelihood = 216.68 AIC=-427.36 AICc=-427.27 BIC=-416.73  Training set error measures:               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Training set -0.0007589053 0.1024508 0.07906853 90.15477 213.618 0.4705803 0.009213563 &gt;</pre>
-------------------------	---

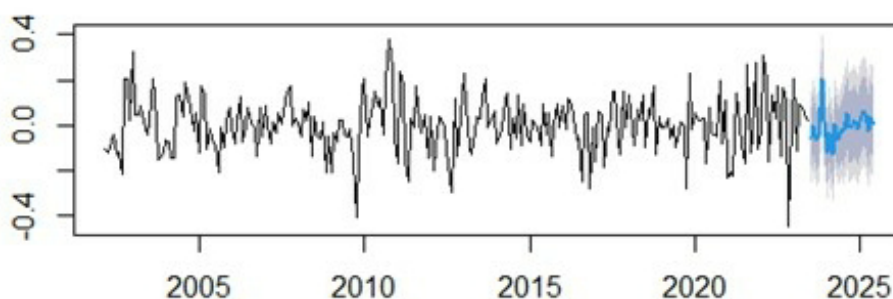
In order to check whether the above appropriate model is the best model that fit into the dataset to forecast the future trend. Running auto.arima function in RStudio and shows the best model(s) for the

data predicting. And the summary of the model which form by the auto.arima is ARIMA (0,0,0) and ARIMA (2,0,0) in the following table. (Table 5)

Both of these models show negative AICc (-427.27) and it is much smaller than the models which tests above. Also the absolute value of ME score also much smaller than the testing model. The lower ME score shows less errors in the forecast so that the forecast model would better close to the trend happens in the future and increase the accuracy of prediction. Thus, ARIMA (0,0,0) and ARIMA (2,0,0) are picked to form the forecast model.

### 3.6 Forecasting by ARIMA Model

After the parameters of the model are determined, it can be used as a prediction model to predict the future values of the time series while choosing the best-suited model is selected for time series data. And it needs to check that the forecast errors are not correlated, and normally distributed with mean zero and constant variance can better help us to forecast a correct model. Then using forecast function to test the time series data after guiding the stationarity.



**Figure. 7** The forecast results from ARIMA.

In the Figure. 6, it shows two different confidence intervals for the forecast line to fluctuate. Based on the analysis of different seasonality of the total time series, the graph shows the future trend of the stock price and moves in the interval that set up automatically. The light gray area shows the remarkable top and bottom of price limitation, and the dark gray area shows the normal area that the stock price will fluctuate normally in that area under the absence of any special circumstances which may influence to it. Therefore, even if the forecast graph does not show the accurate future stock price, it can help us to know about the general trend and fluctuate range of the stock price in the future.

## 4. Summary

According to the past 20 years of historical data and conducted ARIMA model to forecast the 2 months trend of Cathay Pacific Airways Limited which is only a consult for the investors to make decisions. However, the ARIMA model only conclude the time series and stock price to set up which may have influenced by any other unknown factors from the global environment. All in all, the future performance of Cathay Pacific Airways fluctuates in a stable interval which can be chosen as a great investment option.

## References

- [1] Alex Tech Bolg(2020). [Time Series] How to understand ACF and PACF. Retrieved from [https://blog.csdn.net/qq\\_41103204/article/details/105810742](https://blog.csdn.net/qq_41103204/article/details/105810742).
- [2] Amazonaws (2022). ARIMA Modeling. Retrieved from [https://rstudio-pubs-static.s3.amazonaws.com/126854\\_e21376aaeb324fbc802efc98ad17201c.html](https://rstudio-pubs-static.s3.amazonaws.com/126854_e21376aaeb324fbc802efc98ad17201c.html).
- [3] Enes Zvornicanin (2021). Choosing the best q and p from ACF and PACF plots in ARMA-type modeling. Retrieved from <https://www.baeldung.com/cs/acf-pacf-plots-arma-modeling>.
- [4] E. G. Jain and B. Mallick, "A study of time series models ARIMA and ETS," *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 4, pp. 5763, 2017. doi: 10.5815/ijmecs.2017.04.07.

- [5] F.-M. Tseng, G.-H. Tzeng, H.-C. Yu, and B. J. C. Yuan, "Fuzzy ARIMA model for forecasting the foreign exchange market," *Fuzzy Sets Syst.*, vol. 118, no. 1, pp. 919, 2001.
- [6] Hassani, H., & Yeganegi, M. (2019). Sum of squared ACF and the Ljung–Box statistics. *Physica A*, 520, 81-86.
- [7] Hayes (2021). What Is a Time Series? Retrieved from <https://www.investopedia.com/terms/t/timeseries.asp>.
- [8] LEONIE (2020). Time Series: Interpreting ACF and PACF. Retrieved from <https://www.kaggle.com/code/iamleonie/time-series-interpreting-acf-and-pacf/notebook>.
- [9] Mohammad Masum (2020). Time Series Analysis: Identifying AR and MA using ACF and PACF Plots. Retrieved from <https://towardsdatascience.com/identifying-ar-and-ma-terms-using-acf-and-pacf-plots-in-time-series-forecasting-ccb9fd073db8>.
- [10] Prabhakaran (2019). Augmented Dickey Fuller Test (ADF Test) – Must Read Guide. Retrieved from <https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/>.
- [11] Statisticshowto. (2022). Ljung Box Test: Definition. Retrieved from [https://www.statisticshowto.com/ljung-box-test/#google\\_vignette](https://www.statisticshowto.com/ljung-box-test/#google_vignette).
- [12] Subhasree (2018). Time Series Analysis Using ARIMA Model In R. Retrieved from <https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/>.
- [13] V. Tsioumas, S. Papadimitriou, Y. Smirlis, and S. Z. Zahran, "A novel approach to forecasting the bulk freight market," *Asian J. Shipping Logistics*, vol. 33, no. 1, pp. 3341, 2017.
- [14] Yahoo! Finance (2022). Cathay Pacific Airways Limited (0293.HK). Retrieved from <https://finance.yahoo.com/quote/0293.HK/history?period1=949363200&period2=1651363200&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true>.
- [15] Zach (2021). Augmented Dickey-Fuller Test in R. Retrieved from <https://www.statology.org/dickey-fuller-test-in-r/>.