Statistics on Unemployment in Hong Kong

1 Motivation

Generally speaking, macroeconomic policy objectives consist of four main objectives: economic growth, full employment, price stability and balance of payments. Obviously, the employment rate or unemployment rate, together with the economic growth (GDP), the price index (CPI, PPI) and the international balance of payments index, constitute the four major macroeconomic regulation indicators of the country.

The unemployment rate is the core index to reflect the use of labor resources in a country or region. All the governments of all countries have always regarded the unemployment rate as an important basis for judging the operating conditions of the macro-economy and the prosperity of the labor market, and then introduced or adjusted the relevant macroeconomic policies and employment policies.

Employment is the greatest livelihood. It is not only a common practice for the international community to publish the survey unemployment rate, but also an objective need for our government to serve the society. By announcing the survey unemployment rate, it can provide information support for the party and the government to strengthen and improve the macro control and make the employment policy scientifically, and can provide basic data for scientific research institutions and experts and scholars to analyze the situation of China's labor market, and provide basic data for the scientific research of employment and unemployment and employment policy consultation.

In order to provide more comprehensive, accurate, complete and timely information basis for the analysis of employment and unemployment in China, the registered unemployment rate and the investigation unemployment rate are used, and the data of two sources are published.

Above all, that's the reason that I choose this data set to research.

2 Data Set

This data set is from Census and Statistics Department of The Government of the Hong Kong Special Administrative Region from 1982 to 2018, however, I only choose data from January 1982 to January 2018 as the research object.

You can download the data set from [https://www.censtatd.gov.hk/home/index.jsp.](https://www.censtatd.gov.hk/home/index.jsp)

Firstly, I need to make some instructions on the data.

The unemployed population comprises all unemployed persons. For a person aged 15 or over to be classified as unemployed, that person should :

(a) not have had a job and should not have performed any work for pay or profit during the 7 days before enumeration;

(b) have been available for work during the 7 days before enumeration; and

(c) have sought work during the 30 days before enumeration.

However, if a person aged 15 or over fulfills the conditions (a) and (b) above but has not sought work during the 30 days before enumeration because he/she believes that work is not available, he/she is still classified as unemployed, being regarded as a so-called "discouraged worker".

Notwithstanding the above, the following types of persons are also classified as unemployed:

(a) persons without a job, have sought work but have not been available for work because of temporary sickness; and

(b) persons without a job, have been available for work but have not sought work because they:

(i) have made arrangements to take up a new job or to start business on a subsequent date; or

(ii) were expecting to return to their original jobs.

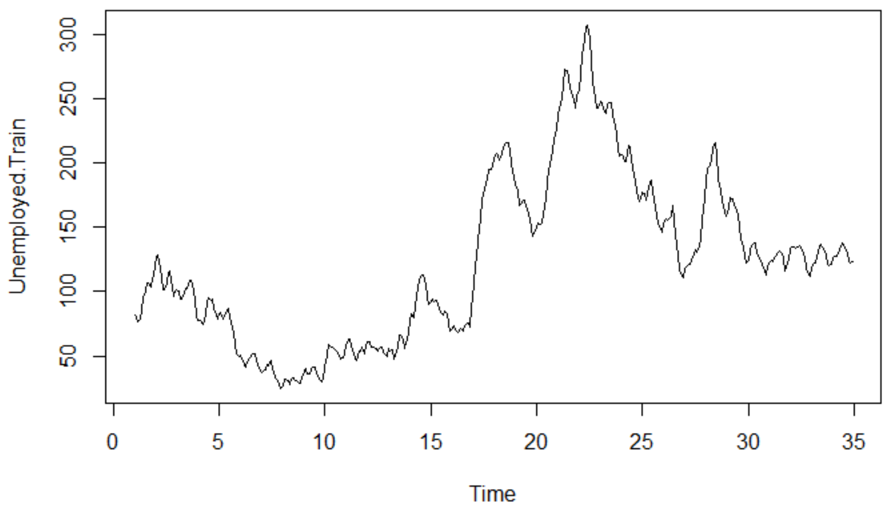
3 Time-series analyses

3.1 Split the data set

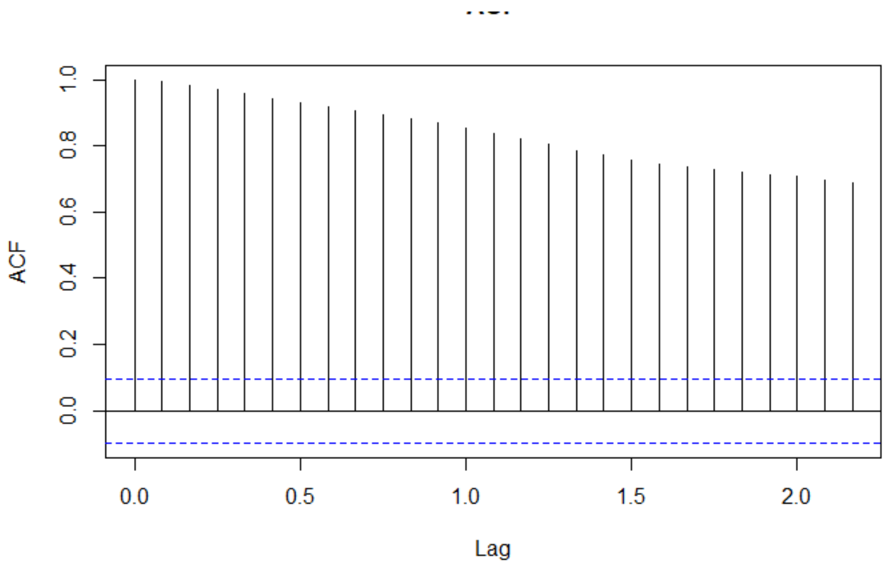
In order to judge the stability of the model and the quality of the prediction results, I choose to divide the data set into training set and test set. Fit the model on the training set, and test the effect of the model on the test set. I choose data from 1982 to 2015 as training set (408 observations), and 2016 and 2017 data as test sets (23 observations).

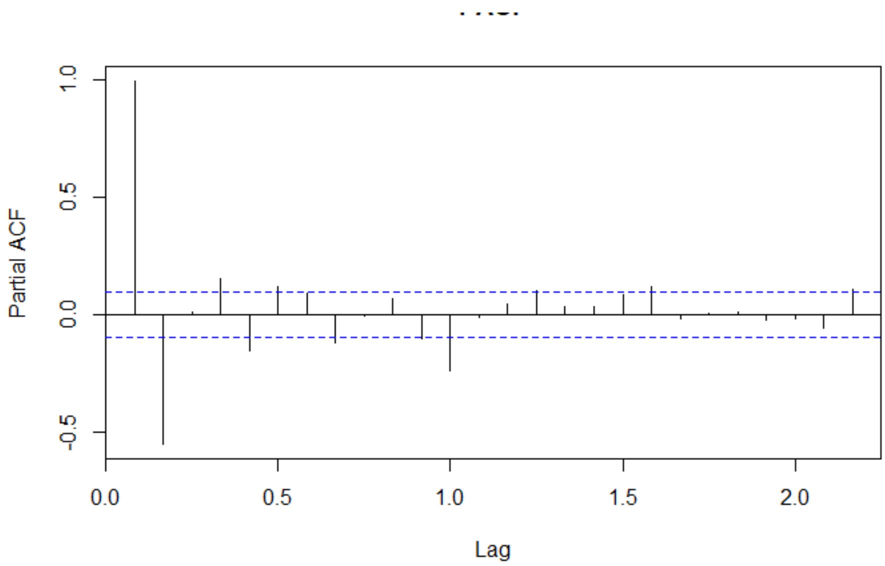
3.2 Drawing time sequence diagram and autocorrelation graph

Draw the sequence diagram of the sequence and observe whether the sequence is stable.



Draw the autocorrelation diagram of the sequence and determine whether the sequence is stable by autocorrelation diagram.If the autocorrelation coefficient of the autocorrelation diagram decays rapidly to 0, then the sequence is stationary, otherwise it is non-stationary.





From the time series graph, we can see that there is a upward trend in the sequence in mid-term, and the autocorrelation coefficient has not been reduced to 0 rapidly. (it is generally considered that the autocorrelation coefficient is lower than 2 times the standard difference, that is, the blue dotted line in the graph is 0), but presents the characteristics of the trailing. Therefore, the judgment sequence is a non-stationary sequence.

For non-stationary sequences, it is necessary to transform them into stationary sequences, which can be modeled by ARMA.

3.3 Stabilization of sequence

Because the sequence is non-stationary, the sequence must be transformed into a stationary sequence by differential. The difference can eliminate the linear trend of the sequence.

```
ndiffs(train)
```

```
## [1] 1
```

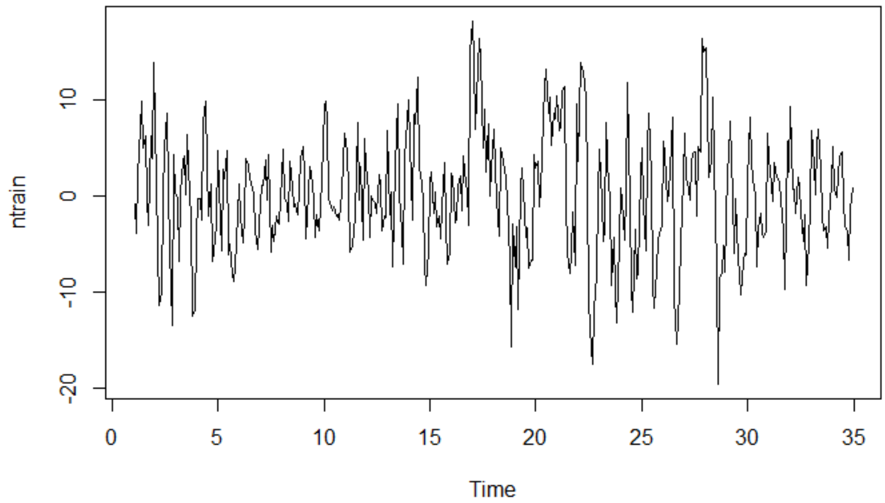
```
ntrain = diff(train,1)
```

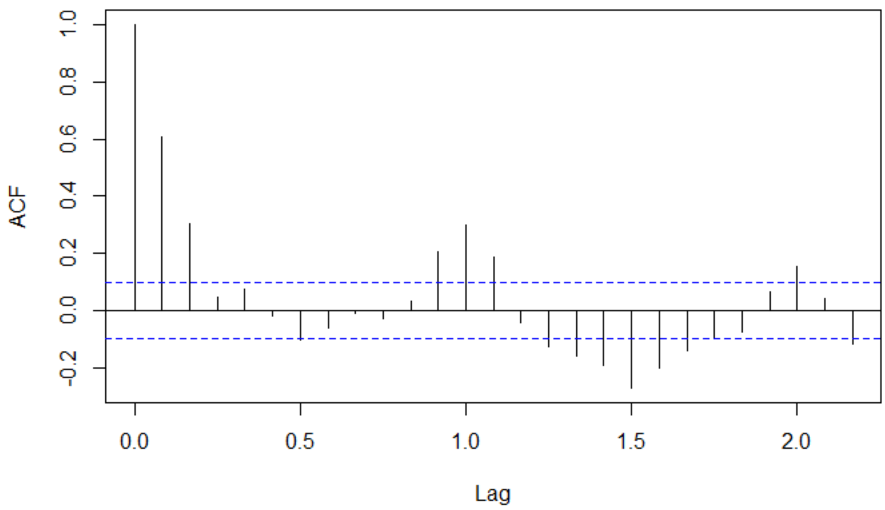
```
ndiffs(ntrain)
```

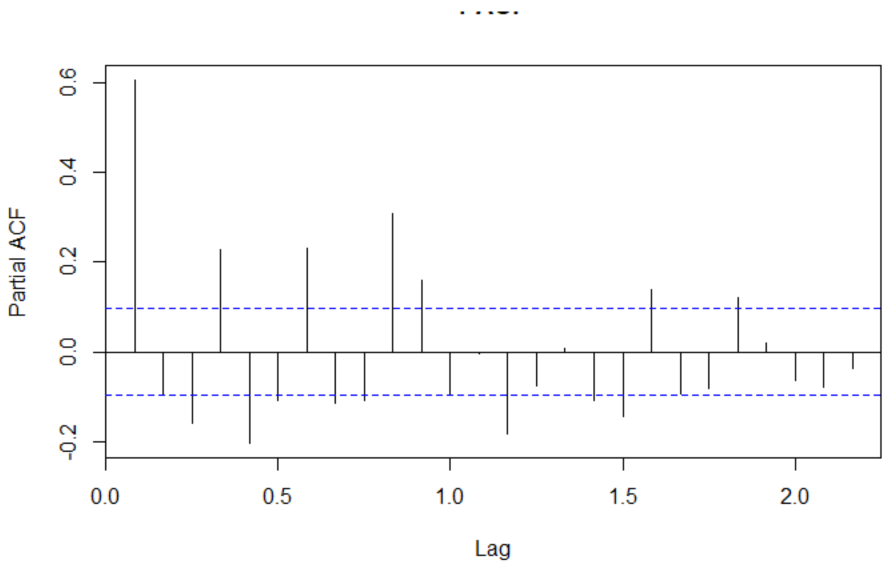
```
## [1] 0
```

The results show that the sequence needs to carry out the 1 order difference.

Now, draw the sequence diagram of the sequence and the autocorrelation diagram.







As you can see from the above picture, the 1 order difference sequence becomes a stationary sequence, and the autocorrelation graph shows that the autocorrelation coefficient is reduced to 0 in a special way after the delay of the 1 order. It is further indicated that the sequence is stable.

3.4 Order and Parameter Estimation of ARMA Model

ACF	PACF	Model
Trailing (gradually reduced to 0)	p order truncation (p order quickly reduced to 0)	ARIMA(p,d,0)
q order truncation	trailing	ARIMA(0,d,q)
trailing	trailing	ARIMA(p,d,q)

We can judge the order of the model through ACF and PACF diagrams. We can also select the model through the AIC and BIC values of the model. Here we select the model by selecting the minimum of AIC and BIC.

```
Fitting models using approximations to speed things up...

ARIMA(2,1,2)(1,0,1)[12] with drift         : 2343.382
ARIMA(0,1,0)            with drift         : 2635.693
ARIMA(1,1,0)(1,0,0)[12] with drift         : 2419.185
ARIMA(0,1,1)(0,0,1)[12] with drift         : 2488.506
ARIMA(0,1,0)                               : 2633.784
ARIMA(2,1,2)(0,0,1)[12] with drift         : 2322.378
ARIMA(2,1,2)            with drift         : 2332.407
ARIMA(2,1,2)(0,0,2)[12] with drift         : 2316.931
ARIMA(1,1,2)(0,0,2)[12] with drift         : 2319.963
ARIMA(3,1,2)(0,0,2)[12] with drift         : 2317.809
ARIMA(2,1,1)(0,0,2)[12] with drift         : 2421.931
ARIMA(2,1,3)(0,0,2)[12] with drift         : 2317.338
ARIMA(1,1,1)(0,0,2)[12] with drift         : 2430.177
ARIMA(3,1,3)(0,0,2)[12] with drift         : 2317.256
ARIMA(2,1,2)(0,0,2)[12]                    : 2314.907
ARIMA(2,1,2)(1,0,2)[12]                    : 2335.225
ARIMA(2,1,2)(0,0,1)[12]                    : 2320.361
ARIMA(1,1,2)(0,0,2)[12]                    : 2318.329
ARIMA(3,1,2)(0,0,2)[12]                    : 2317.57
ARIMA(2,1,1)(0,0,2)[12]                    : 2419.891
ARIMA(2,1,3)(0,0,2)[12]                    : 2315.29
ARIMA(1,1,1)(0,0,2)[12]                    : 2428.159
ARIMA(3,1,3)(0,0,2)[12]                    : 2315.406

Now re-fitting the best model(s) without approximations...

ARIMA(2,1,2)(0,0,2)[12]                    : 2313.996


Series: train
ARIMA(2,1,2)(0,0,2)[12]

Coefficients:
          ar1      ar2     ma1     ma2    sma1    sma2
      -0.1933  -0.0674  0.9959  0.9331  0.1331  0.1466
s.e.   0.0565   0.0560  0.0233  0.0177  0.0562  0.0503

sigma^2 estimated as 16.73:  log likelihood=-1149.86
AIC=2313.72   AICc=2314   BIC=2341.78
```

According to the result of R, we can choose the seasonal ARIMA model: ARIMA(2, 1, 2) (0, 0, 2)[12]. That is:

$$(1+0.1933B+0.0674B^2)(1-B)Z_t =$$

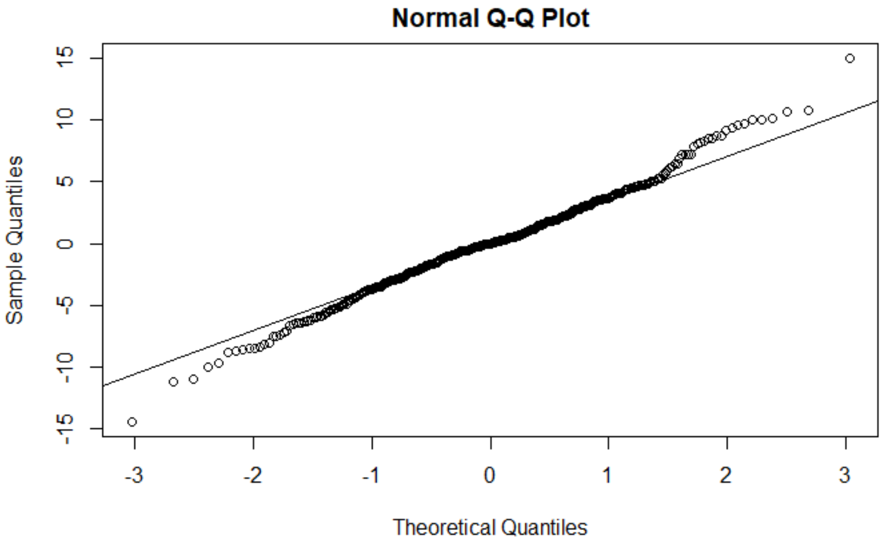
$$(1+0.9959B+0.9331B^2)(1+0.133B^{12}+0.1467B^{24})a_t$$

3.5 Model Testing

The test of model parameters includes two parts: the significance test of parameters and the normality and independence test of residuals.

Significance test of parameters: Depend on the 3.4, we know that all of the coefficient is not zero. Therefore, I skip this step.

Normality test of residuals: we can draw the Q-Q diagram of residuals to determine that the residuals of Q-Q graphs fall completely on the 45 degree line, which is consistent with normality assumption. Otherwise, the model may be wrong.



Independent test of residual: it is known from the definition of the front white noise that the residual (= estimated value - real value) should be an unrelated sequence. LB statistics are commonly used to test the residual.

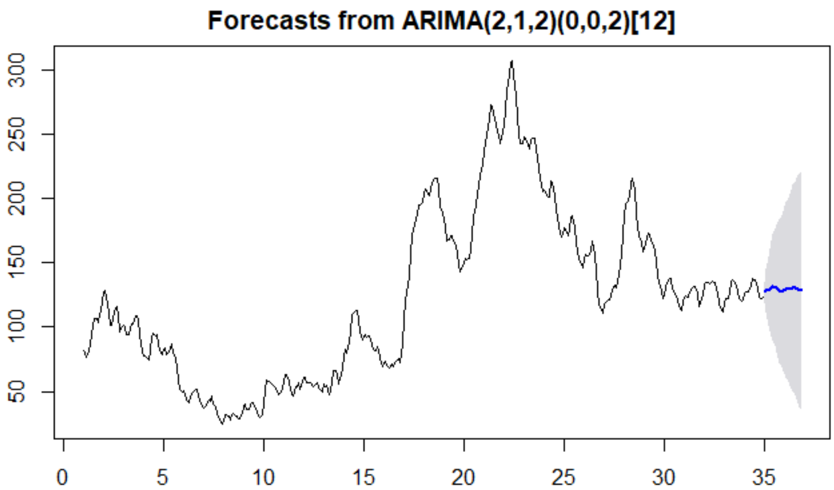
```
        Box-Ljung test

data:  trainarima$residuals
X-squared = 0.022461, df = 1, p-value = 0.8809
```

From the results of the Q-Q and LB tests (0.8809 > 0.05), the residual error conforms to the normal hypothesis and is not related, and the model fitting data is more sufficient and can be used for the next prediction.

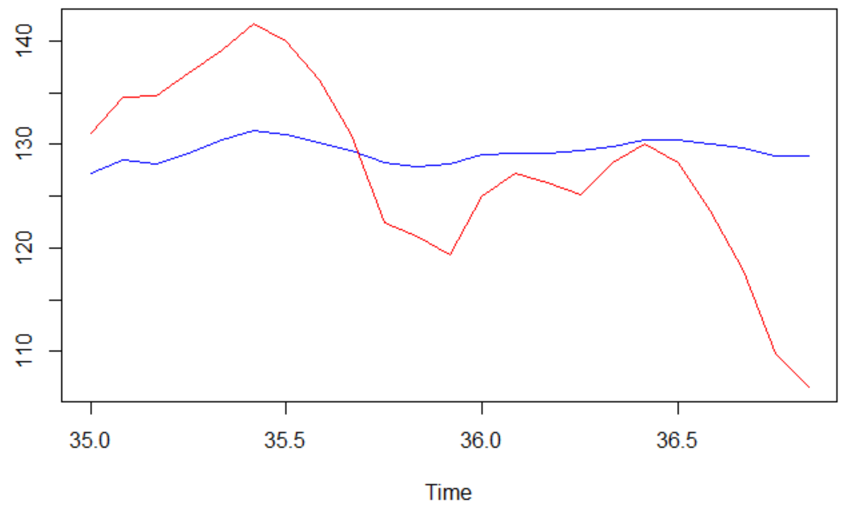
3.6 Forecast

We now have a seasonal ARIMA model that passes the required checks and is ready for forecasting. Forecasts from the model for the next two years and 95% prediction intervals are shown in Figure.



The point forecasts trend undulation, the prediction intervals allow for the data to trend upwards during the forecast period. It means that the number of people who are unemployed in Hond Kong tends to be stable.

Now, we can compare the test set with the prediction.



The red line is the real data and the blue line is predicted. From the prediction results, we can see that the predicted value in the last two years is more stable than the real value. To a certain extent, it reflects the change of unemployment rate. However, this model does not predict the final decline during this period. This problem often arises when time series are used to predict. If the data can be more, the model produced may predict better results. The time series model is only suitable for short-term forecasting, and is not suitable for long periods.

4 Conclusion

We calculated the mean value of the number of unemployed people from 2013 to 2017 and compared them with the mean of our forecast, and found that the difference was only around 1300 people(128013 and 129319), which means that we obtain a great model to forecast the unemployment rate in Hong Kong.

However, we must notice that in recent years, the trade and business activities between the mainland and Hong Kong are becoming more and more frequent, which will greatly increase the employment opportunities of Hong Kong, so it is necessary to take into account the factors in the introduction or adjustment of the relevant macroeconomic policies and employment policies.