ARIMA Procedures

Kyla Ayop | Jolia Keziah Balcita

December 7, 2023

Contents

I. INTRODUCTION

In this paper, we undertake a comprehensive exploration of ARIMA procedures which involves evaluating and identifying the most suitable ARIMA model for our dataset. The ARIMA model comprises AutoRegressive (AR), Integrated (I), and Moving Average (MA) components, denoted as (p, d, q). Subsequently, we will also be deriving its equation that encapsulates the mathematical relationship between past observations and future predictions.

By achieving this with meticulous analysis, we enhance our understanding of the underlying patterns and nuances of our data, providing a concise and predictive framework for time series forecasting.

II. THE DATA

The data we will be using is from the global economic indicators according to the World Bank throughout the year 1960-2017. Our main dataset is the exports of goods and services (% of GDP) of Australia.

```r
library(fpp3)
library(urca)
```

```r
global_economy
```

```
## # A tsibble: 15,150 x 9 [1Y]
## # Key:       Country [263]
##    Country     Code   Year         GDP Growth   CPI Imports Exports Population
##    <fct>       <fct> <dbl>       <dbl>  <dbl> <dbl>   <dbl>   <dbl>      <dbl>
##  1 Afghanistan AFG    1960  537777811.     NA    NA    7.02    4.13    8996351
##  2 Afghanistan AFG    1961  548888896.     NA    NA    8.10    4.45    9166764
##  3 Afghanistan AFG    1962  546666678.     NA    NA    9.35    4.88    9345868
##  4 Afghanistan AFG    1963  751111191.     NA    NA   16.9     9.17    9533954
##  5 Afghanistan AFG    1964  800000044.     NA    NA   18.1     8.89    9731361
##  6 Afghanistan AFG    1965 1006666638.     NA    NA   21.4    11.3     9938414
##  7 Afghanistan AFG    1966 1399999967.     NA    NA   18.6     8.57   10152331
##  8 Afghanistan AFG    1967 1673333418.     NA    NA   14.2     6.77   10372630
##  9 Afghanistan AFG    1968 1373333367.     NA    NA   15.2     8.90   10604346
## 10 Afghanistan AFG    1969 1408888922.     NA    NA   15.0    10.1    10854428
## # ... with 15,140 more rows
```

III. ARIMA MODELLING PROCEDURE

1 Historical Plot

The time plot shows some non-stationarity, with an overall increase of its export.

```
global_economy |>
  filter(Code == "AUS") |>
  autoplot(Exports) +
  labs(y = "% of GDP", x = "Year") + theme_bw()
```

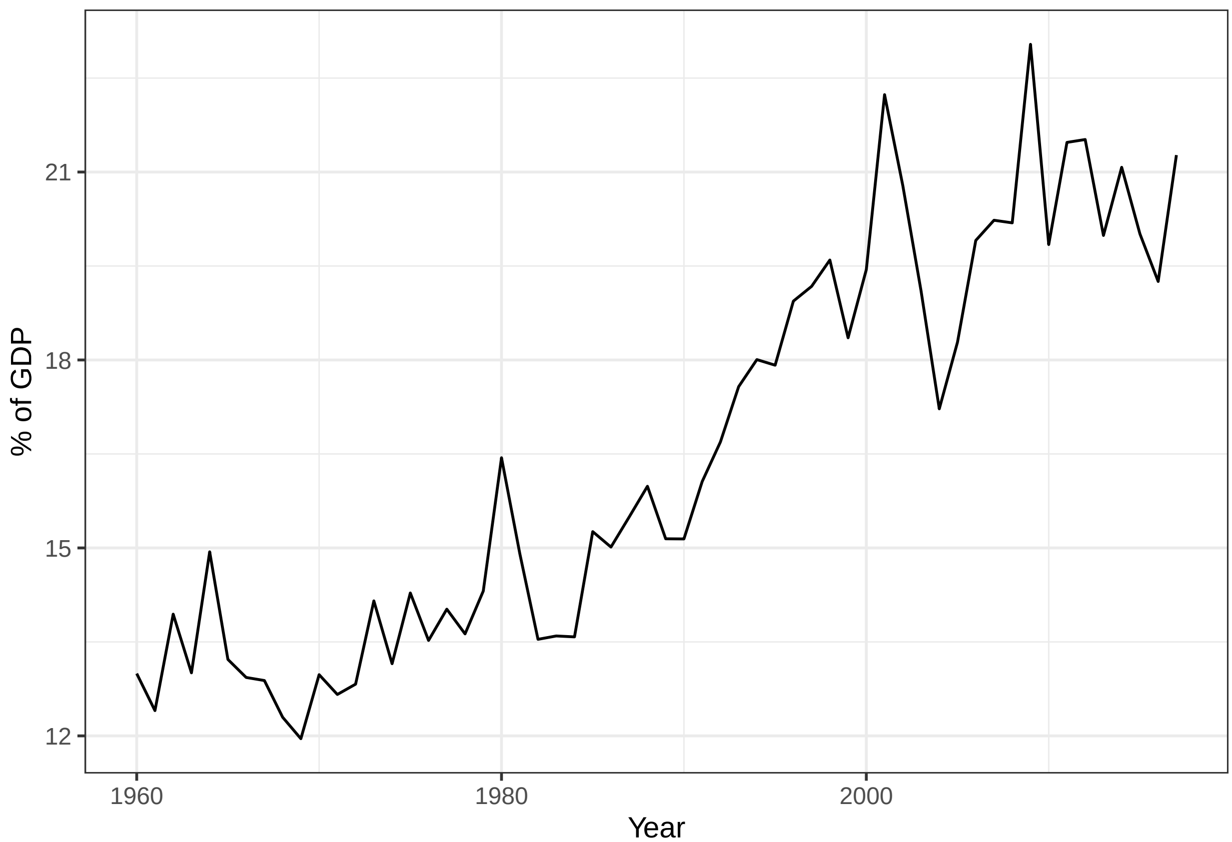


Figure 1: Australia's exports of goods and services (% of GDP) from 1960-2017.

2 Box-Cox Transformation

There is no evidence of changing variance, so we will not do a Box-Cox transformation.

3 Stationarity

Another way to determine more objectively whether differencing is required is to use a unit root test. In our analysis, we use the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. In this test, the null hypothesis is that the series is stationary. Consequently, small p-values (less than 0.05) suggest that differencing is required, otherwise is stationary.

```
global_economy |>
  filter(Code == "AUS") |>
  features(Exports, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Country   kpss_stat kpss_pvalue
##   <fct>         <dbl>       <dbl>
## 1 Australia      1.45        0.01
```

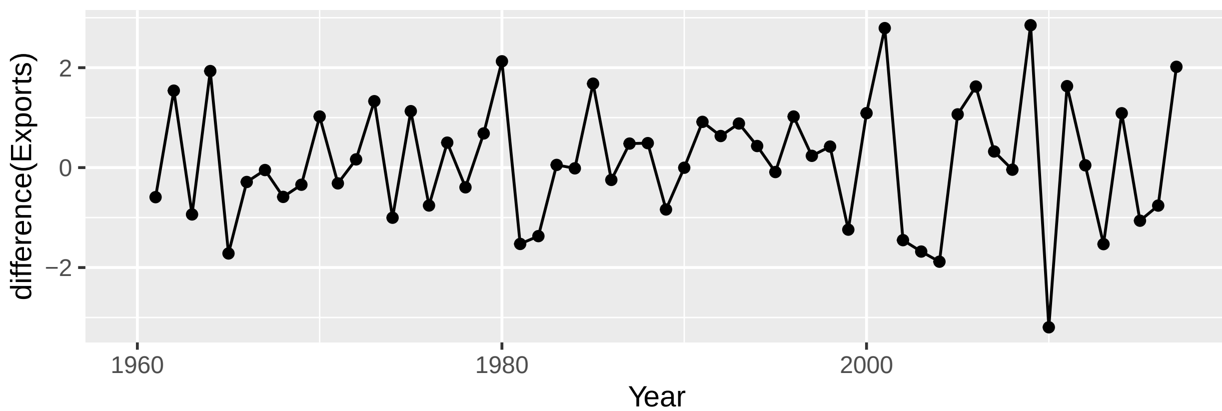
Here, with the original series of the Exports of goods and services of Australia, a p-value of 0.01 has been obtained and with the significance level of 0.05, the p-value is lower than the significance level with the KPSS statistic is 1.446615. Therefore, we reject the null hypothesis of stationarity. This result tells us that the series is non-stationary. Since the data is non-stationary, we take a first difference of the data.

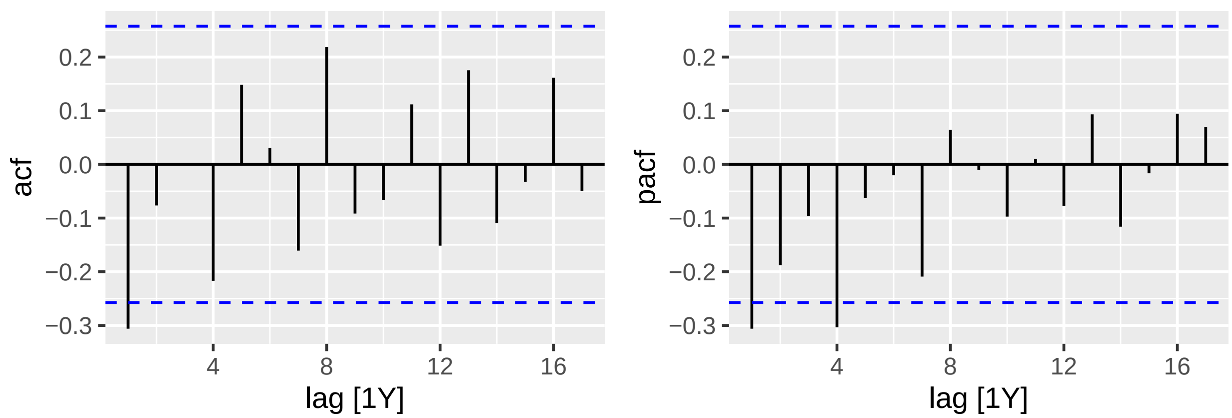
```
global_economy |>
  filter(Code == "AUS") |>
  features(difference(Exports), unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Country   kpss_stat kpss_pvalue
##   <fct>         <dbl>       <dbl>
## 1 Australia    0.0544         0.1
```

However, for the differenced series of the exports of goods and services of Australia, the KPSS statistic is 0.05440058. The associated p-value is 0.1, which is greater than the significance level of 0.05. With this, the p-value is not less than the significance level. As a result, there is not enough evidence to reject the null hypothesis of stationarity, hence the differenced series is stationary.

```
global_economy |>
  filter(Code == "AUS") |>
  gg_tsdisplay(difference(Exports), plot_type='partial')
```





Thus, taking the first difference resulted to a stationary data.

4 Examine ACF/PACF

From the PACF, suggests AR(4) model which results to ARIMA(4,1,0). From the ACF, it suggests MA(1) model which then be ARIMA(0,1,1). To explore more models we included the ARIMA(4,1,1) model.

```
aus_fit <- global_economy |>
  filter(Code == "AUS") |>
  model(arima410 = ARIMA(Exports ~ pdq(4,1,0)),
        arima011 = ARIMA(Exports ~ pdq(0,1,1)),
        arima411 = ARIMA(Exports ~ pdq(4,1,1)),
        stepwise = ARIMA(Exports),
        search = ARIMA(Exports, stepwise=FALSE))

aus_fit |> pivot_longer(!Country, names_to = "Model name",
                         values_to = "Orders")
```

```
## # A mable: 5 x 3
## # Key:     Country, Model name [5]
##   Country   `Model name`                    Orders
##   <fct>     <chr>                          <model>
## 1 Australia arima410     <ARIMA(4,1,0) w/ drift>
## 2 Australia arima011     <ARIMA(0,1,1) w/ drift>
## 3 Australia arima411     <ARIMA(4,1,1) w/ drift>
```

```
## 4 Australia stepwise     <ARIMA(1,1,1) w/ drift>
## 5 Australia search       <ARIMA(1,1,1) w/ drift>
```

5 Model Testing

We fit both an ARIMA(4,1,0), ARIMA(0,1,1) and an ARIMA(4,1,1) model along with two automated model selections, one using the default stepwise procedure, and one working harder to search a larger model space.

```
glance(aus_fit) |> arrange(AICc) |> select(.model:BIC)
```

```
## # A tibble: 5 x 6
##   .model   sigma2 log_lik   AIC  AICc   BIC
##   <chr>     <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 stepwise   1.27   -86.5  181.  182.  189.
## 2 search     1.27   -86.5  181.  182.  189.
## 3 arima011   1.32   -87.8  182.  182.  188.
## 4 arima410   1.28   -85.5  183.  185.  195.
## 5 arima411   1.30   -85.5  185.  187.  199.
```

It has been observed that both stepwise and search models of ARIMA(1,1,1) has the same values of AIC, AICc, and BIC.

Of the models fitted, the two automated model selection has found that an ARIMA(1,1,1) gives the lowest AICc value, closely followed by the ARIMA(0,1,1), ARIMA(4,1,0) and ARIMA(4,1,1) — the latter three being the models that we guessed from the ACF and PACF plots. The automated search and stepwise selection has identified an ARIMA(1,1,1) model, which has the same lowest AICc value and ARIMA(4,1,1), a guessed model, has the highest AICc value of the five models.

```
fit <- global_economy |>
  filter(Code == "AUS") |>
  model(ARIMA(Exports ~ pdq(1,1,1)))
report(fit)
```

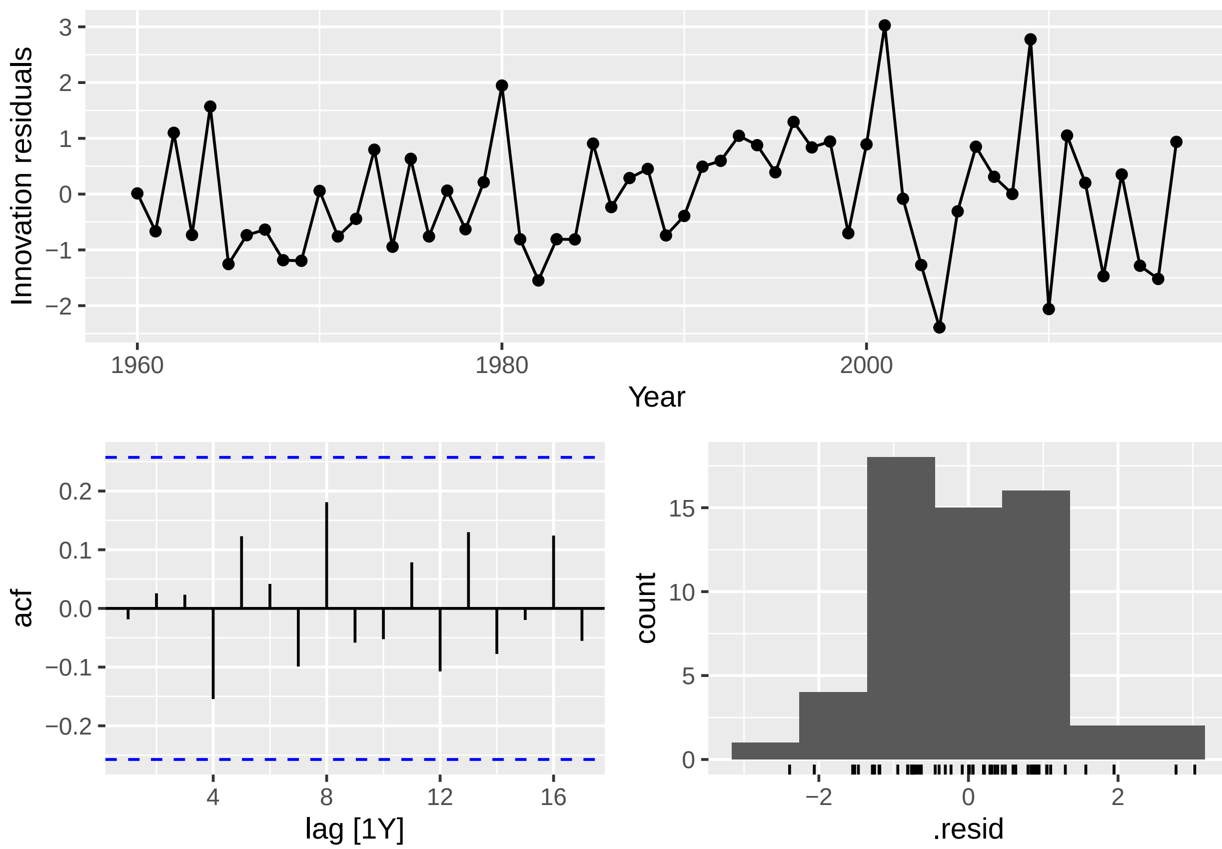
```
## Series: Exports
## Model: ARIMA(1,1,1) w/ drift
##
## Coefficients:
##          ar1      ma1  constant
##       0.3931  -0.8459    0.0924
## s.e.  0.2449   0.1790    0.0277
##
## sigma^2 estimated as 1.27:  log likelihood=-86.45
## AIC=180.91   AICc=181.68   BIC=189.08
```

The ARIMA(1,1,1) model has the equation:

y_t **= 0.09 + 0.39** y_{t-1} **- 0.85** ϵ_{t-1} **+** ϵ_t**, where** ϵ_t **is white noise with a standard deviation of** $1.1269 = \sqrt{1.27}$

6 Assessment of Residuals

```
aus_fit |>
  select(stepwise) |>
  gg_tsresiduals()
```



The ACF plot of the residuals from the ARIMA(1,1,1) model shows that all autocorrelations are within the threshold limits, indicating that the residuals are behaving like white noise.

Portmanteau test of the residuals

```
augment(aus_fit) |>
  filter(.model=='stepwise') |>
  features(.innov, ljung_box, lag = 10, dof = 2)
```

```
## # A tibble: 1 x 4
##   Country   .model   lb_stat lb_pvalue
##   <fct>     <chr>      <dbl>     <dbl>
## 1 Australia stepwise    6.13     0.632
```

A portmanteau test (setting K=2) returns a large p-value, also suggesting that the residuals are white noise. With the null hypothesis that states that the series is white noise or independent and identically distributed. With an alpha of 0.05, the Ljung-Box test results provide strong evidence that we fail to reject the null hypothesis that the series stepwise model of ARIMA(1,1,1) is white noise. The p-value of approximately 0.6323542 indicates that the series is random or white noise or independent and identically distributed.

7 Forecasting

Now, that the residuals look like white noise, we use the ARIMA(1,1,1) model to forecast future values. Here is a five-year forecast.

```
aus_fit |>
  forecast(h=5) |>
  filter(.model=='stepwise') |>
  autoplot(global_economy) +
   labs(y = "% of GDP", x = "Year") + theme_bw()
```

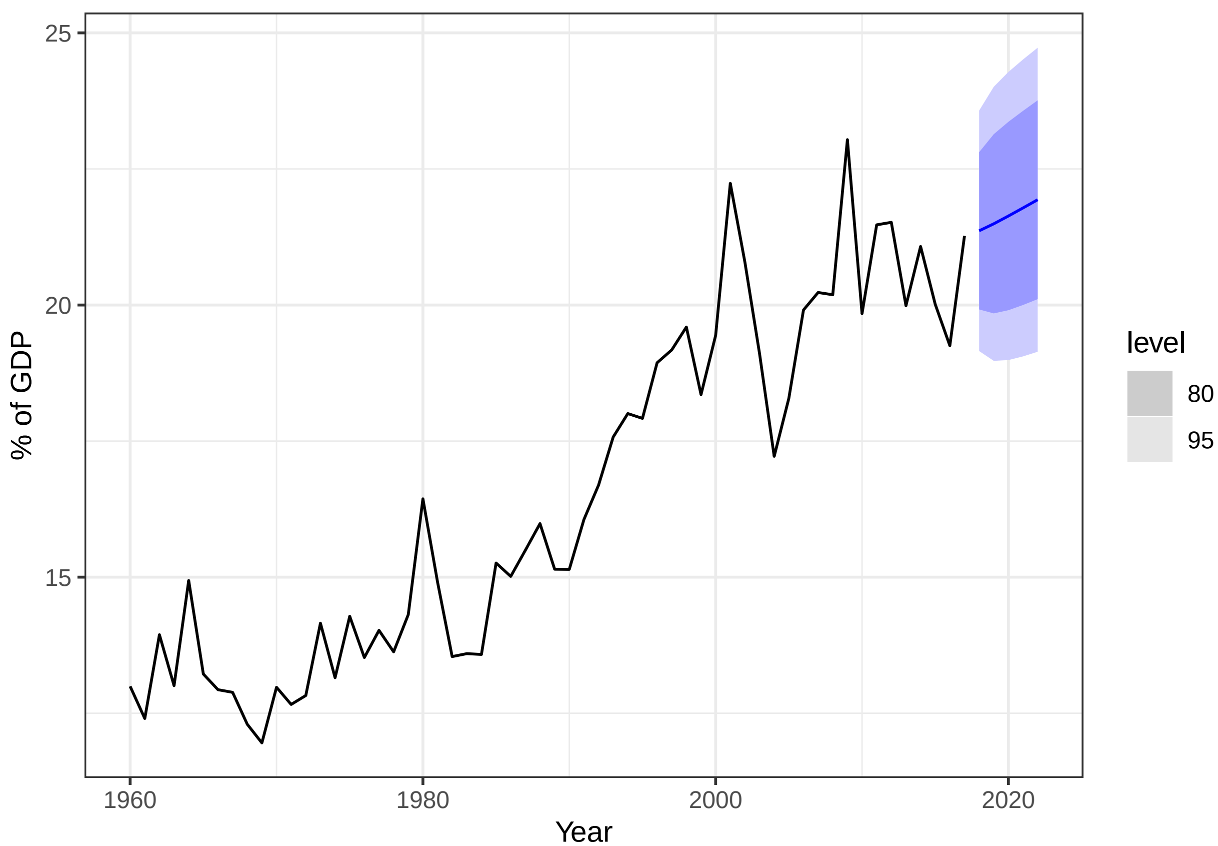


Figure 2: Forecast of goods and services (% of GDP) in Australia.