ARIMA

Charlene P. Garridos & Ken Andrea Bahian

2023-12-07

Contents

1 R Libraries 2

2 Time Series Data 2

3 R code selects a non-seasonal ARIMA model automatically. 3

4 Forecast 4

5 ACF 5

6 PACF 6

7 Residuals 8

8 Forecasts from the chosen model 10

1 R Libraries

```
library(urca)   # Unit Root and Cointegration Tests for Time Series Data
library(fpp3)   #"Forecasting: Principles and Practice" package
```

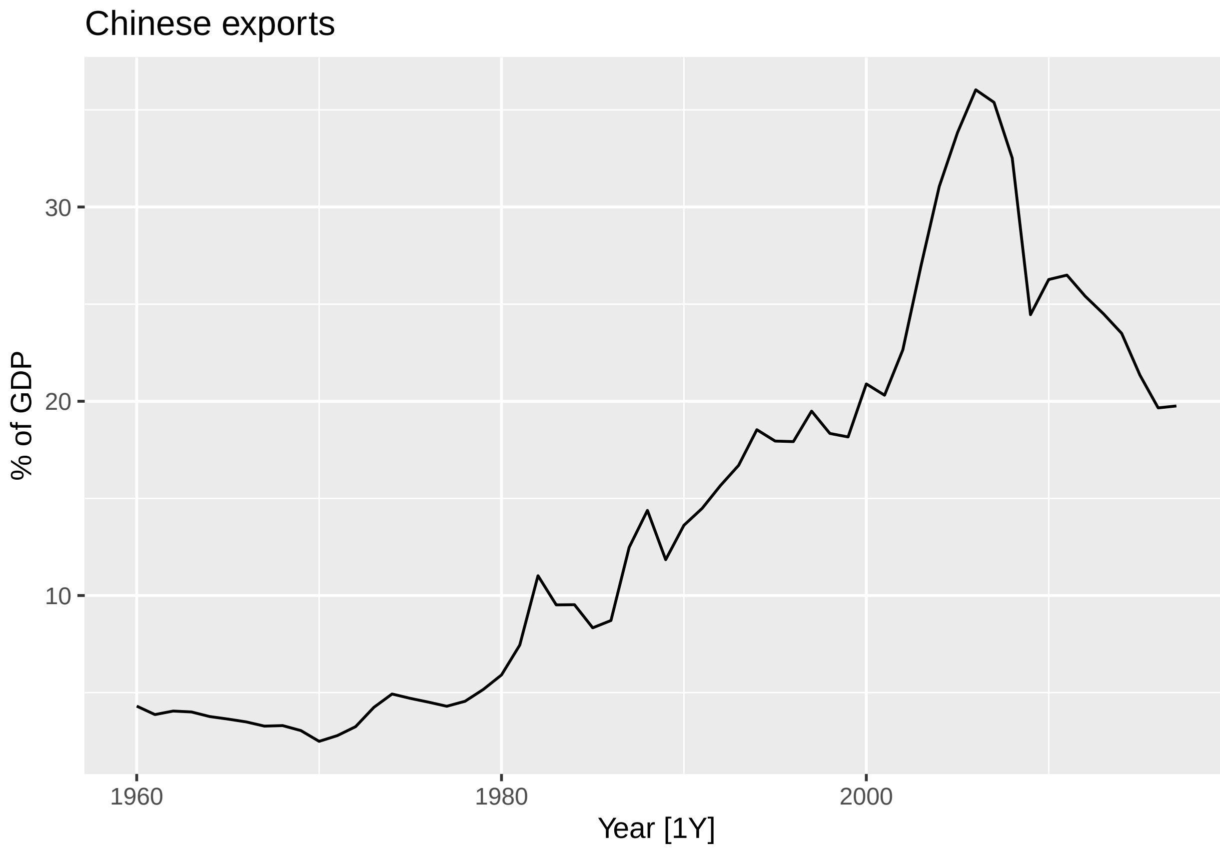
```
## -- Attaching packages ---------------------------------------------- fpp3 0.5 --
```

```
## v tibble      3.2.1     v tsibble     1.1.3
## v dplyr       1.1.2     v tsibbledata 0.4.1
## v tidyr       1.3.0     v feasts      0.3.1
## v lubridate   1.9.2     v fable       0.3.3
## v ggplot2     3.4.2     v fabletools  0.3.3
```

```
## -- Conflicts ------------------------------------------------- fpp3_conflicts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()
```

2 Time Series Data

```
global_economy |>
  filter(Code == "CHN") |>
  autoplot(Exports) +
  labs(y = "% of GDP", title = "Chinese exports")
```



Interpretation

The time plot shows some non-stationarity, initially showing growth followed by a noticeable decline. Higher exports usually boost a country's GDP, particularly for export-oriented economies like China. The trend of increasing exports typically aligns with GDP expansion. However, a decline in GDP was observed around 2007, despite this relationship.

3 R code selects a non-seasonal ARIMA model automatically.

```
fit <- global_economy |>
  filter(Code == "CHN") |>
  model(ARIMA(Exports))
report(fit)
```

```
## Series: Exports
## Model: ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##        0.297
```

```
## s.e.  0.125
##
## sigma^2 estimated as 3.398:  log likelihood=-115.28
## AIC=234.57   AICc=234.79   BIC=238.66
```

4 Forecast

```
fit |> forecast(h=10) |>
  autoplot(global_economy) +
  labs(y = "% of GDP", title = "Chinese exports")
```

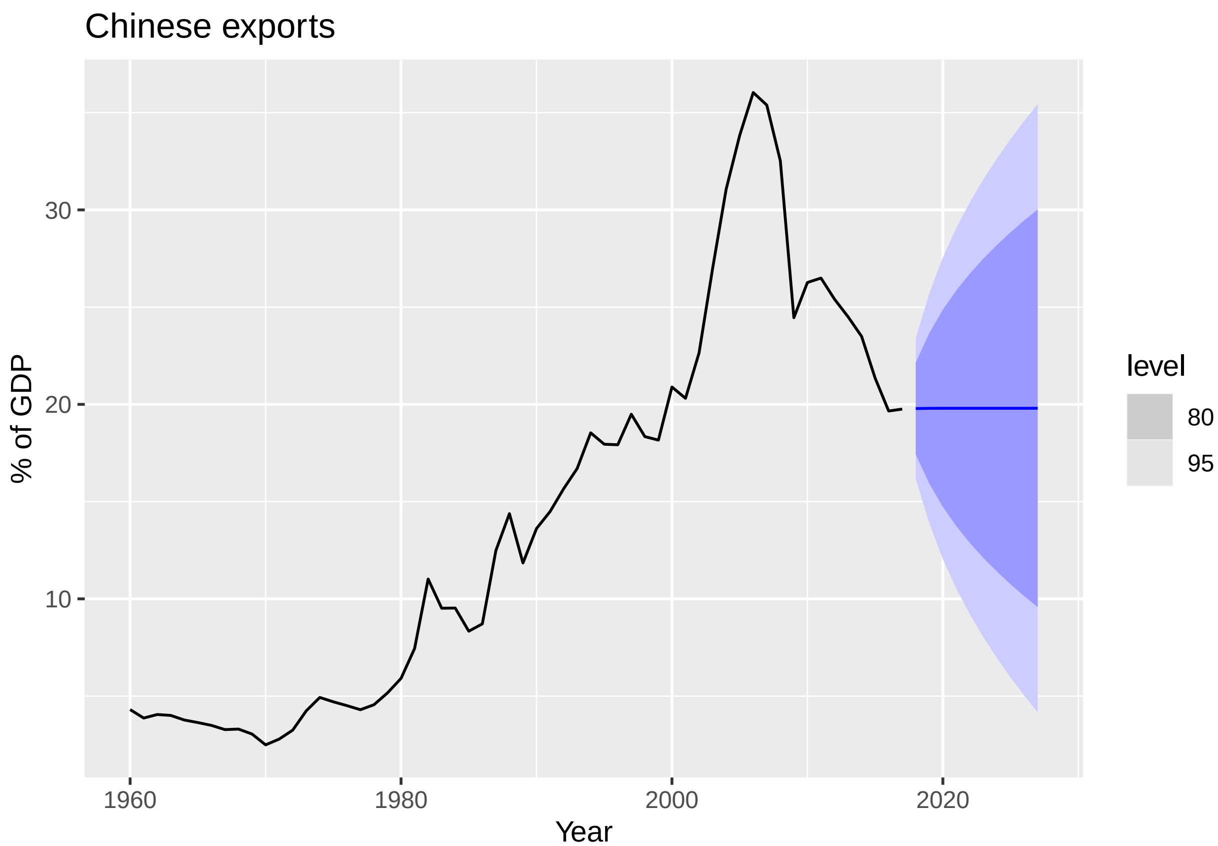


Figure 1: Forcasted Exports of the China as a percentage of GDP.

5 ACF

```
global_economy |>
  filter(Code == "CHN") |>
  ACF(Exports) |>
  autoplot()
```

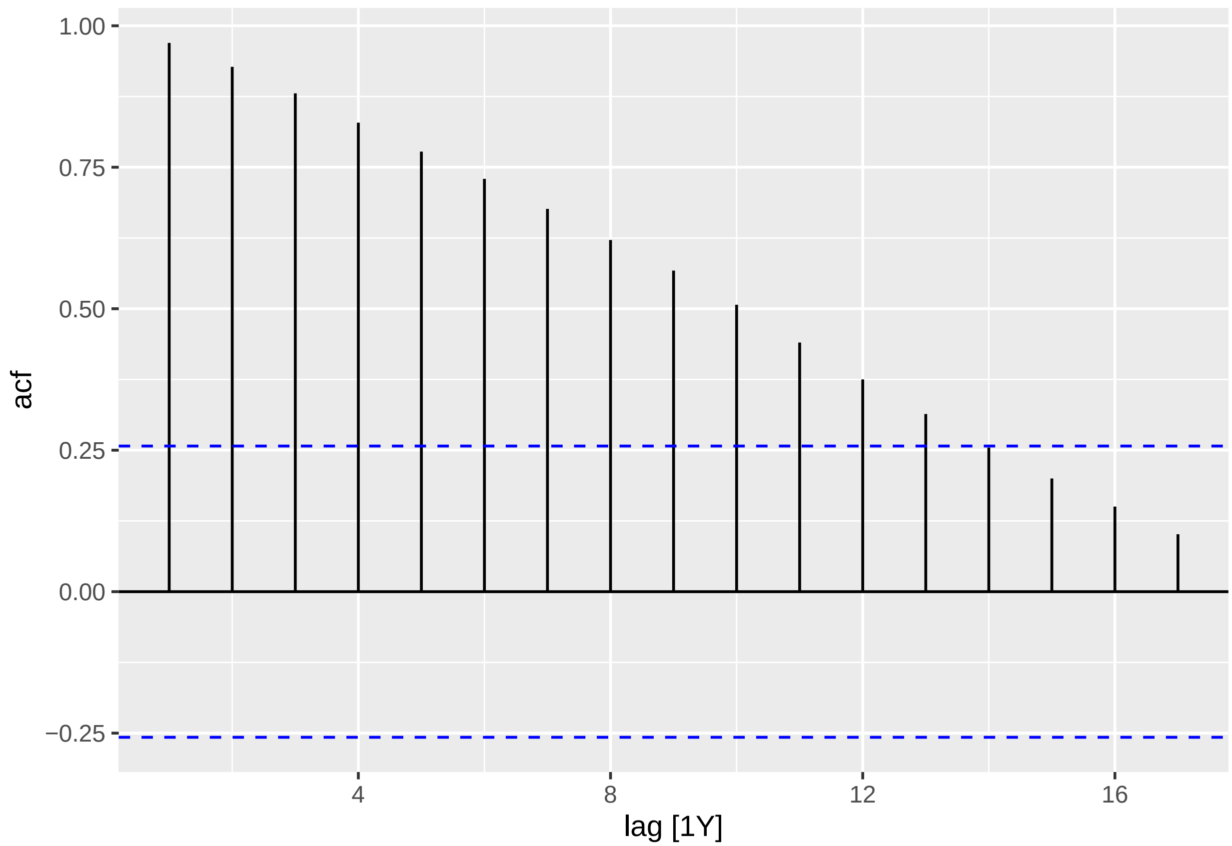


Figure 2: ACF plot for the China Exports.

6 PACF

```
global_economy |>
  filter(Code == "CHN") |>
  PACF(Exports) |>
  autoplot()
```

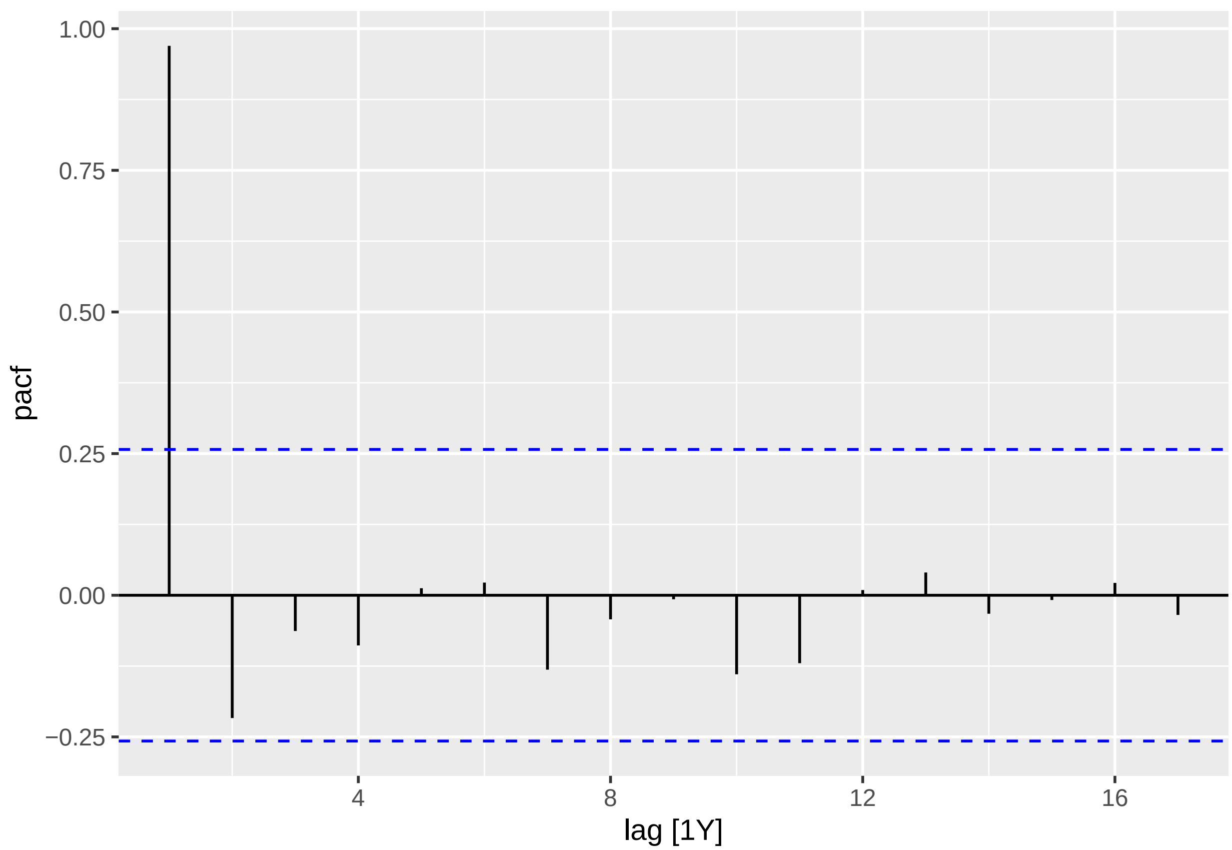


Figure 3: PACF plot for the China Exports.

Interpretation

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

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

```
## Warning: Removed 1 row containing missing values (`geom_line()`).

## Warning: Removed 1 rows containing missing values (`geom_point()`).
```

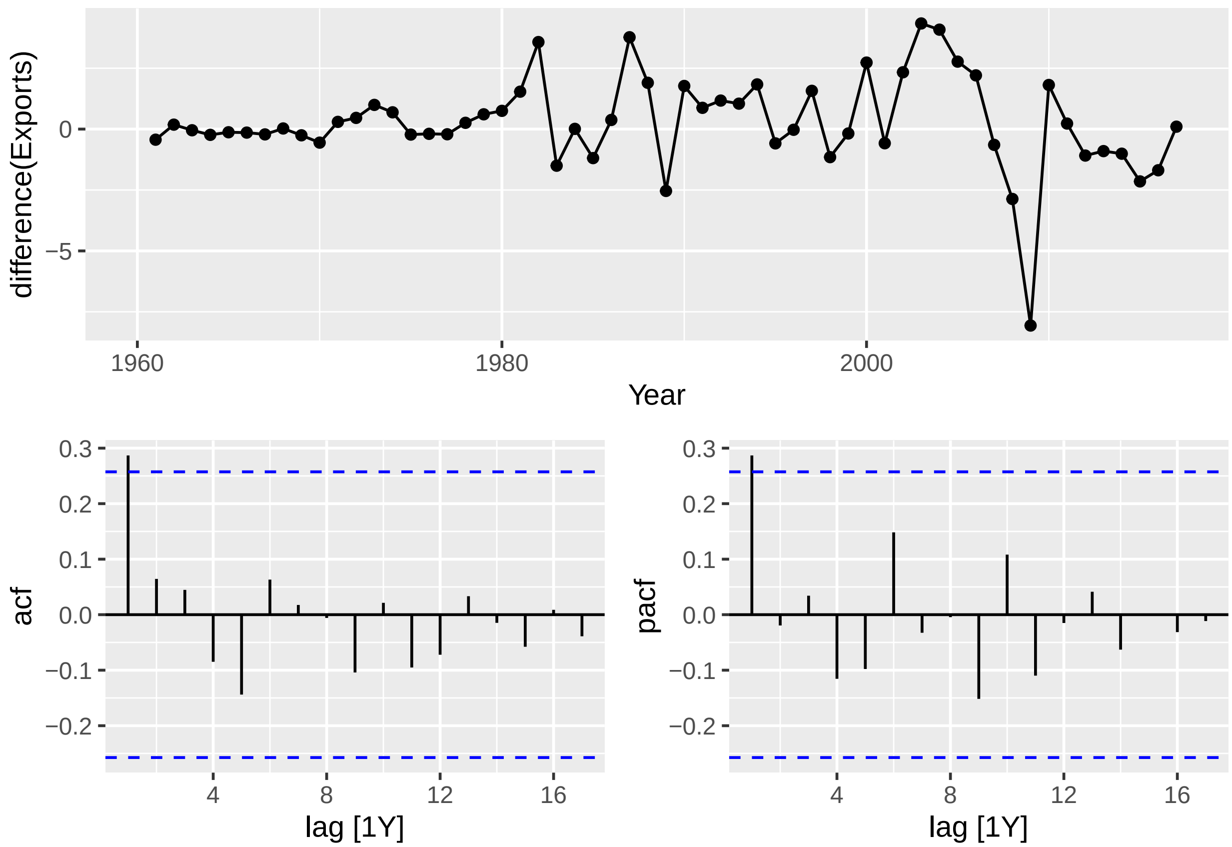


Figure 4: Time plot and ACF and PACF plots for the differenced China Exports.

Interpretation

The ACF (AutoCorrelation Function) of the original data indicates that the data is not stationary. To address the non-stationarity, we will take a first difference of the data. The differenced data are shown in the figure above.

```
fit2 <- global_economy |>
  filter(Code == "CHN") |>
  model(ARIMA(Exports ~ pdq(0,1,1)))
report(fit2)
```

```
## Series: Exports
## Model: ARIMA(0,1,1)
```

```
##
## Coefficients:
##          ma1
##       0.2915
## s.e.  0.1217
##
## sigma^2 estimated as 3.411:  log likelihood=-115.39
## AIC=234.79   AICc=235.01   BIC=238.87
```

```
fit4 <- global_economy |>
  filter(Code == "CHN") |>
  model(ARIMA(Exports ~ pdq(1,1,0)))
report(fit4)
```

```
## Series: Exports
## Model: ARIMA(1,1,0)
##
## Coefficients:
##         ar1
##       0.297
## s.e.  0.125
##
## sigma^2 estimated as 3.398:  log likelihood=-115.28
## AIC=234.57   AICc=234.79   BIC=238.66
```

Interpretation

> The Model 1 or fit2 model indicates that a model with a first-order moving average (MA) component. The estimated coefficient for the MA(1) term is 0.2915, and the standard error is 0.1217. The estimated variance σ^2 is 3.411, and the information criteria (AIC, AICc, and BIC) are provided for model evaluation.
>
> Model 2 or fit4 model has a first-order autoregressive (AR) component. The estimated coefficient for the AR(1) term is 0.297, with a standard error of 0.125. The estimated variance σ^2 is 3.398, and similar information criteria are provided for model evaluation.
>
> Based on the information provided, where the Partial Autocorrelation Function (PACF) suggests an ARIMA(0,1,1) model or IMA(1,1), and the fitted model fit4 and the Autocorrelation Function (ACF) suggest an ARIMA(1,1,0) model or ARI(1,1) which is an alternative candidate model.

7 Residuals

```
chn_fit <- global_economy |>
  filter(Code == "CHN") |>
  model(arima011 = ARIMA(Exports ~ pdq(0,1,1)),
        arima110 = ARIMA(Exports ~ pdq(1,1,0)),
        stepwise = ARIMA(Exports),
        search = ARIMA(Exports, stepwise=FALSE))

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

```
## # A mable: 4 x 3
## # Key:     Country, Model name [4]
##   Country `Model name`         Orders
##   <fct>   <chr>               <model>
## 1 China   arima011     <ARIMA(0,1,1)>
## 2 China   arima110     <ARIMA(1,1,0)>
## 3 China   stepwise     <ARIMA(1,1,0)>
## 4 China   search       <ARIMA(1,1,0)>
```

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

```
## # A tibble: 4 x 6
##   .model   sigma2 log_lik   AIC  AICc   BIC
##   <chr>     <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 arima110   3.40   -115.  235.  235.  239.
## 2 stepwise   3.40   -115.  235.  235.  239.
## 3 search     3.40   -115.  235.  235.  239.
## 4 arima011   3.41   -115.  235.  235.  239.
```

```
chn_fit |>
  select(search) |>
  gg_tsresiduals()
```

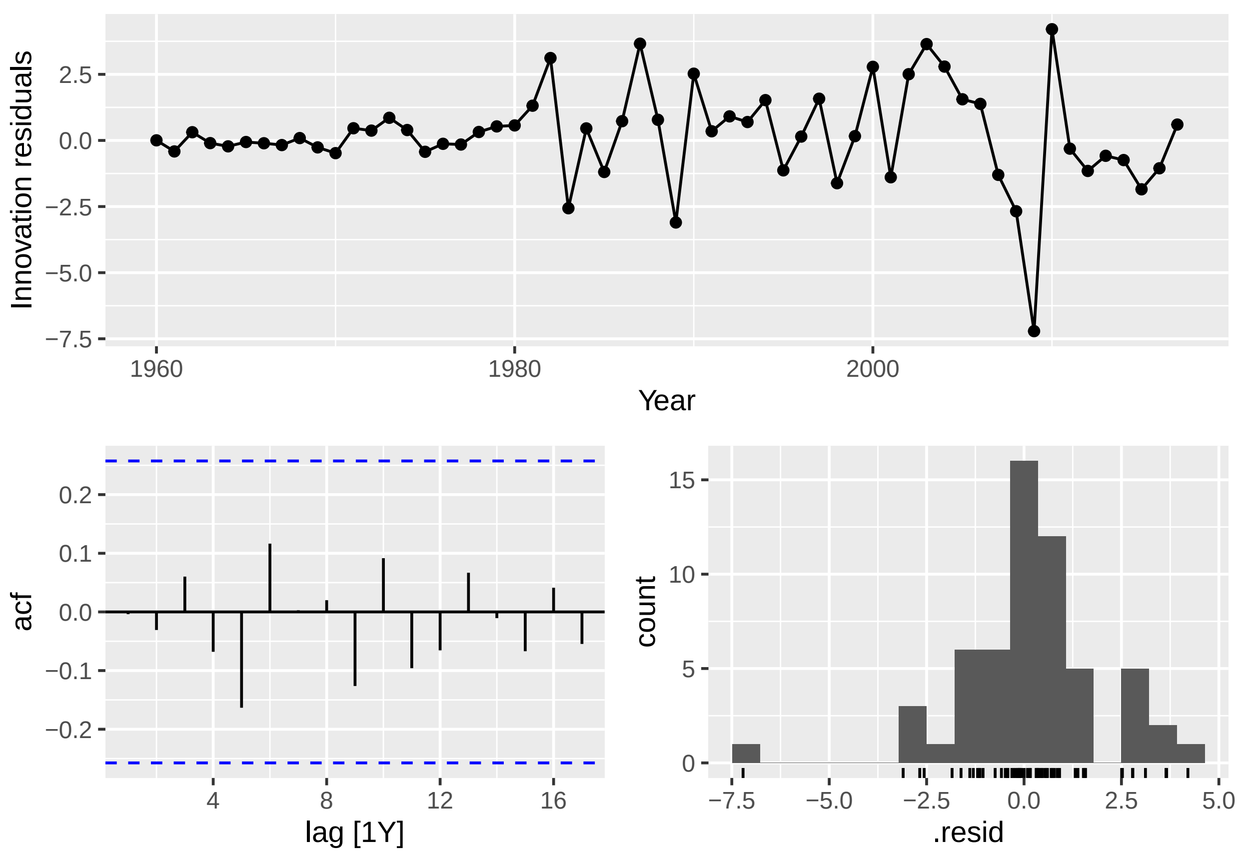


Figure 5: Residual plots.

```
augment(chn_fit) |>
  filter(.model=='search') |>
  features(.innov, ljung_box, lag = 10, dof = 1)
```

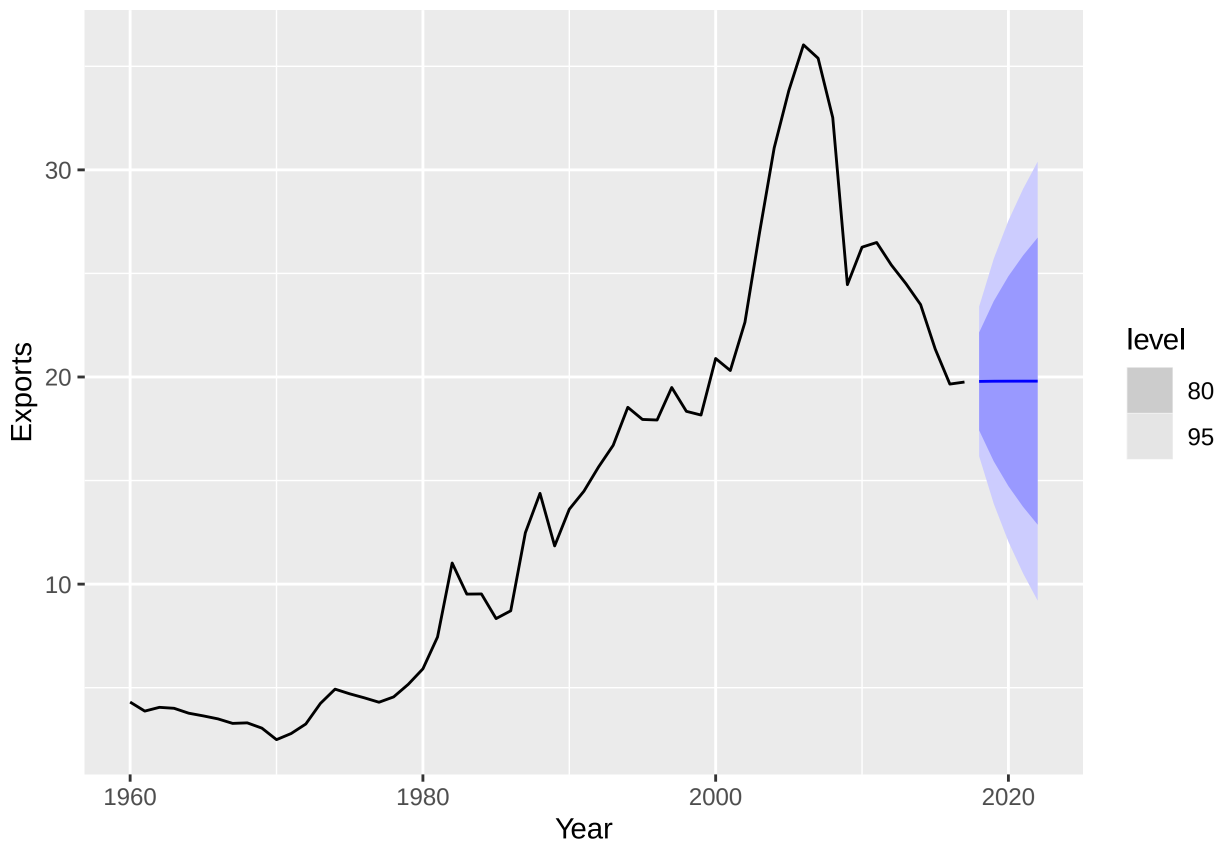
```
## # A tibble: 1 x 4
##   Country .model lb_stat lb_pvalue
##   <fct>   <chr>    <dbl>     <dbl>
## 1 China   search    5.01     0.833
```

Interpretation

> The model ARIMA(1,1,0) appears to have the lowest AIC, AICc, and BIC values among the models considered. The Ljung-Box test for autocorrelation in the residuals of the search model shows a p-value of 0.833, suggesting that there is no significant autocorrelation in the residuals up to lag 10.
>
> The model, ARIMA(1,1,0), seems to be the preferred model based on the lowest AIC, AICc, and BIC values. The Ljung-Box test indicates that the residuals of the selected model do not show significant autocorrelation up to lag 10.
>
> The ACF plot of the residuals from the ARIMA(1,1,0) model shows that all autocorrelations are within the threshold limits, indicating that the residuals are behaving like white noise. A portmanteau test (setting K=10) returns a p-value of 0.833, suggesting that the residuals are white noise, and there is no significant autocorrelation up to lag 10.
>
> In summary, the exhaustive search identified the ARIMA(1,1,0) model as providing a good fit to the data. This conclusion is based on almost identical AICc values among the models considered, with diagnostic checks, including the ACF plot of residuals and a portmanteau test (K=10), supporting the selection of ARIMA(1,1,0) as a suitable choice for forecasting the time series data for China.

8 Forecasts from the chosen model

```
chn_fit |>
  forecast(h=5) |>
  filter(.model=='search') |>
  autoplot(global_economy)
```



Interpretation

The figure shows the chosen forecast model which is ARIMA(1,1,0) and can be represented by the following equation:

$$Z_t = \phi_1 Z_{t-1} + \varepsilon_t$$

In this case, with ARIMA(1,1,0):

$$Z_t = 0.297 Z_{t-1} + \varepsilon_t$$

Here:

- Z_t is the differenced time series at time t.
- ϕ_1 is the autoregressive coefficient for lag 1, and in your case, it's 0.297.
- Z_{t-1} is the differenced time series at lag 1.
- ε_t is the white noise error term at time t.

This equation expresses the relationship between the differenced values of the time series and the lagged differenced values, with an autoregressive term of order 1. It's a concise representation of the dynamics of the ARIMA(1,1,0) model fitted to the data.