

Rob J Hyndman
George Athanasopoulos

FORECASTING

PRINCIPLES AND PRACTICE

A comprehensive introduction to the latest forecasting methods using R. Learn to improve your forecast accuracy using dozens of real data examples.



3RD EDITION

 **OTexts**
Oxford Texts in Finance and Probability

9. ARIMA models

9.5 Non-seasonal ARIMA models

OTexts.org/fpp3/

ARIMA models

AR: autoregressive (lagged observations as inputs)

I: integrated (differencing to make series stationary)

MA: moving average (lagged errors as inputs)

ARIMA models

AR: autoregressive (lagged observations as inputs)

I: integrated (differencing to make series stationary)

MA: moving average (lagged errors as inputs)

An ARIMA model is rarely interpretable in terms of visible data structures like trend and seasonality. But it can capture a huge range of time series patterns.

ARIMA models

Autoregressive Moving Average models:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

ARIMA models

Autoregressive Moving Average models:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

- Predictors include both **lagged values of y_t and lagged errors.**
- Conditions on AR coefficients ensure stationarity.
- Conditions on MA coefficients ensure invertibility.

ARIMA models

Autoregressive Moving Average models:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

- Predictors include both **lagged values of y_t** and **lagged errors**.
- Conditions on AR coefficients ensure stationarity.
- Conditions on MA coefficients ensure invertibility.

Autoregressive Integrated Moving Average models

- Combine ARMA model with **differencing**.
- $(1 - B)^d y_t$ follows an ARMA model.

ARIMA models

Autoregressive Integrated Moving Average models

ARIMA(p, d, q) model

AR: p = order of the autoregressive part

I: d = degree of first differencing involved

MA: q = order of the moving average part.

- White noise model: ARIMA(0,0,0)
- Random walk: ARIMA(0,1,0) with no constant
- Random walk with drift: ARIMA(0,1,0) with const.
- AR(p): ARIMA($p,0,0$)
- MA(q): ARIMA(0,0, q)

Backshift notation for ARIMA

■ ARMA model:

$$y_t = c + \phi_1 B y_t + \dots + \phi_p B^p y_t + \varepsilon_t + \theta_1 B \varepsilon_t + \dots + \theta_q B^q \varepsilon_t$$

or $(1 - \phi_1 B - \dots - \phi_p B^p) y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t$

■ ARIMA(1,1,1) model:

$$\begin{array}{ccccc} (1 - \phi_1 B) & (1 - B) y_t & = & c + (1 + \theta_1 B) \varepsilon_t \\ \uparrow & \uparrow & & \uparrow \\ \text{AR}(1) & \text{First} & & \text{MA}(1) \\ & \text{difference} & & \end{array}$$

Backshift notation for ARIMA

■ ARMA model:

$$y_t = c + \phi_1 B y_t + \dots + \phi_p B^p y_t + \varepsilon_t + \theta_1 B \varepsilon_t + \dots + \theta_q B^q \varepsilon_t$$

$$\text{or } (1 - \phi_1 B - \dots - \phi_p B^p) y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t$$

■ ARIMA(1,1,1) model:

$$(1 - \phi_1 B) (1 - B) y_t = c + (1 + \theta_1 B) \varepsilon_t$$

↑
AR(1)

↑
First
difference

↑
MA(1)

Expand: $y_t = c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$

R model

Intercept form

$$(1 - \phi_1 B - \dots - \phi_p B^p) y'_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t$$

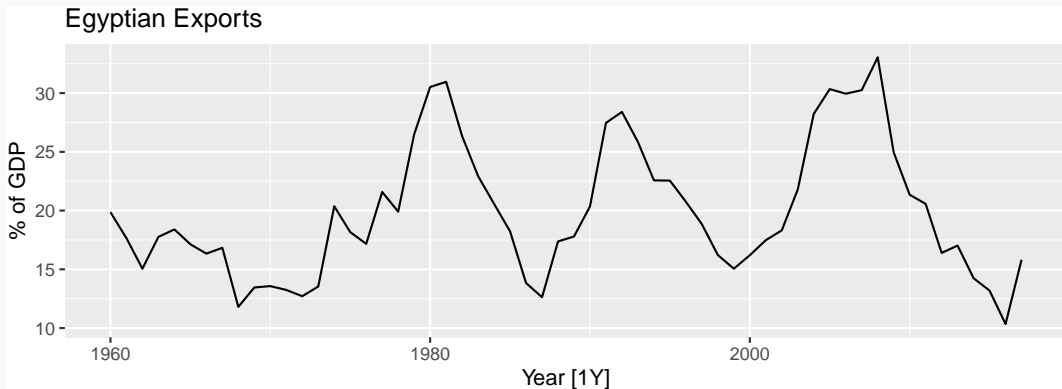
Mean form

$$(1 - \phi_1 B - \dots - \phi_p B^p)(y'_t - \mu) = (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t$$

- $y'_t = (1 - B)^d y_t$
- μ is the mean of y'_t .
- $c = \mu(1 - \phi_1 - \dots - \phi_p)$.
- fable uses intercept form

Egyptian exports

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



Egyptian exports

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

```
## Series: Exports
## Model: ARIMA(2,0,1) w/ mean
##
## Coefficients:
##          ar1      ar2      ma1  constant
##          1.676 -0.8034 -0.690      2.562
## s.e.    0.111   0.0928   0.149      0.116
##
## sigma^2 estimated as 8.046:  log likelihood=-142
## AIC=293   AICc=294   BIC=303
```

Egyptian exports

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

```
## Series: Exports
## Model: ARIMA(2,0,1) w/ mean
##
## Coefficients:
##          ar1      ar2      ma1  constant
##          1.676 -0.8034 -0.690      2.562
## s.e.    0.111   0.0928   0.149      0.116
##
## sigma^2 estimated as 8.046:  log likelihood=-142
```

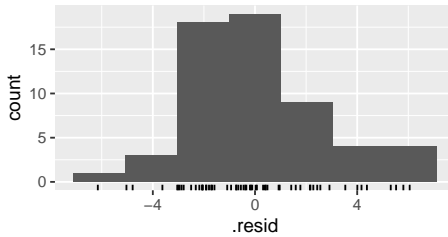
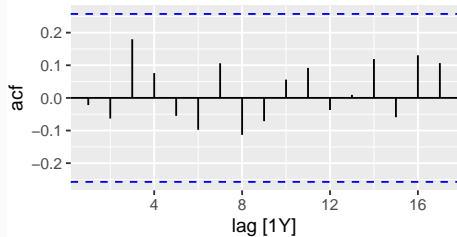
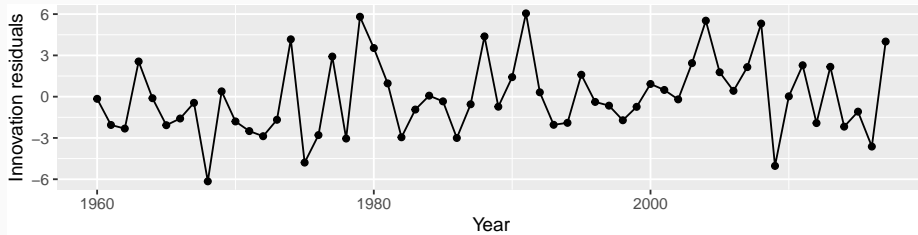
ARIMA(2,0,1) model:

$$y_t = 2.56 + 1.68y_{t-1} - 0.80y_{t-2} - 0.69\varepsilon_{t-1} + \varepsilon_t,$$

where ε_t is white noise with a standard deviation of $2.837 = \sqrt{8.046}$.

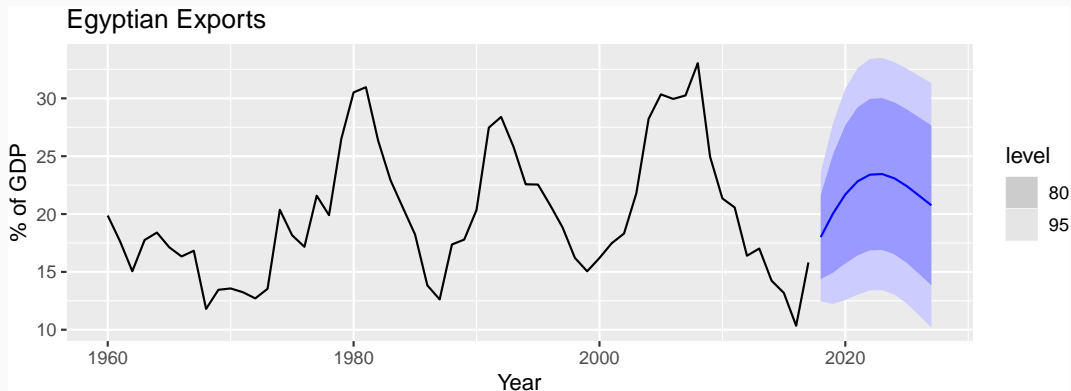
Egyptian exports

```
gg_tsresiduals(fit)
```



Egyptian exports

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



Understanding ARIMA models

- If $c = 0$ and $d = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d = 1$, the long-term forecasts will go to a non-zero constant.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 0$, the long-term forecasts will go to the mean of the data.
- If $c \neq 0$ and $d = 1$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 2$, the long-term forecasts will follow a quadratic trend.

Understanding ARIMA models

Forecast variance and d

- The higher the value of d , the more rapidly the prediction intervals increase in size.
- For $d = 0$, the long-term forecast standard deviation will go to the standard deviation of the historical data.

Cyclic behaviour

- For cyclic forecasts, $p \geq 2$ and some restrictions on coefficients are required.
- If $p = 2$, we need $\phi_1^2 + 4\phi_2 < 0$. Then average cycle of length
$$(2\pi) / \left[\arccos(-\phi_1(1 - \phi_2)/(4\phi_2)) \right] .$$

Partial autocorrelations

Partial autocorrelations measure relationship between y_t and y_{t-k} , when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

Partial autocorrelations

Partial autocorrelations measure relationship between y_t and y_{t-k} , when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

α_k = k th partial autocorrelation coefficient

= equal to the estimate of ϕ_k in regression:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k} + \varepsilon_t.$$

Partial autocorrelations

Partial autocorrelations measure relationship between y_t and y_{t-k} , when the effects of other time lags $1, 2, 3, \dots, k-1$ — are removed.

α_k = k th partial autocorrelation coefficient

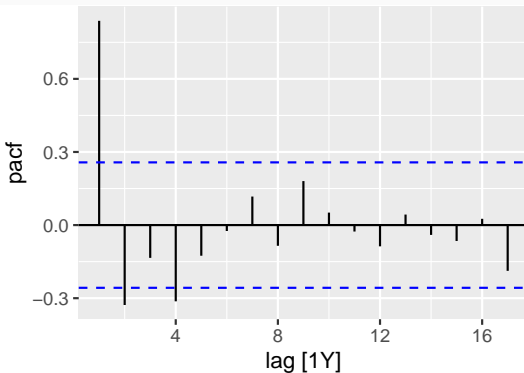
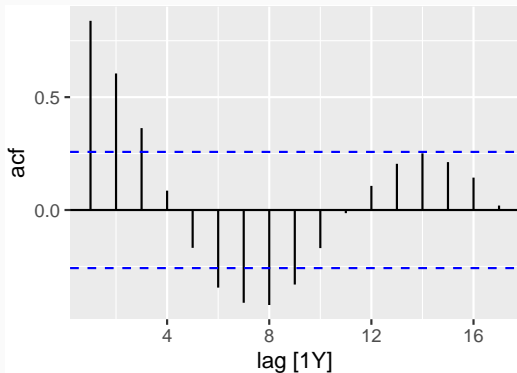
= equal to the estimate of ϕ_k in regression:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k} + \varepsilon_t.$$

- Varying number of terms on RHS gives α_k for different values of k .
- $\alpha_1 = \rho_1$
- same critical values of $\pm 1.96/\sqrt{T}$ as for ACF.
- Last significant α_k indicates the order of an AR model.

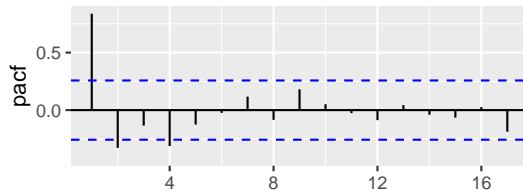
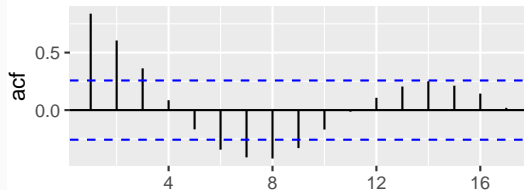
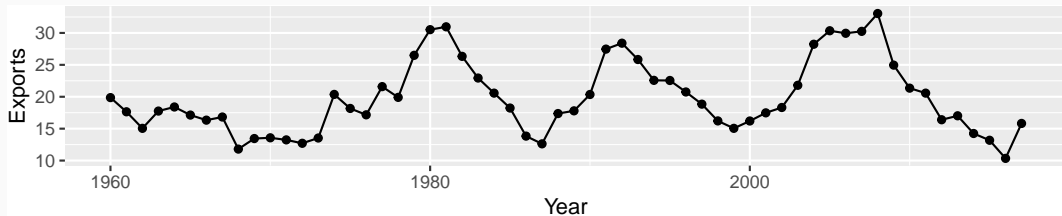
Egyptian exports

```
egypt <- global_economy |> filter(Code == "EGY")  
egypt |> ACF(Exports) |> autoplot()  
egypt |> PACF(Exports) |> autoplot()
```



Egyptian exports

```
global_economy |>  
  filter(Code == "EGY") |>  
  gg_tsdisplay(Exports, plot_type = "partial")
```



ACF and PACF interpretation

AR(1)

$$\begin{aligned}\rho_k &= \phi_1^k && \text{for } k = 1, 2, \dots; \\ \alpha_1 &= \phi_1 && \alpha_k = 0 \quad \text{for } k = 2, 3, \dots\end{aligned}$$

So we have an AR(1) model when

- autocorrelations exponentially decay
- there is a single significant partial autocorrelation.

ACF and PACF interpretation

AR(p)

- ACF dies out in an exponential or damped sine-wave manner
- PACF has all zero spikes beyond the p th spike

So we have an AR(p) model when

- the ACF is exponentially decaying or sinusoidal
- there is a significant spike at lag p in PACF, but none beyond p

ACF and PACF interpretation

MA(1)

$$\begin{aligned}\rho_1 &= \theta_1 / (1 + \theta_1^2) & \rho_k &= 0 & \text{for } k = 2, 3, \dots; \\ \alpha_k &= -(-\theta_1)^k / (1 + \theta_1^2 + \dots + \theta_1^{2k})\end{aligned}$$

So we have an MA(1) model when

- the PACF is exponentially decaying and
- there is a single significant spike in ACF

ACF and PACF interpretation

MA(q)

- PACF dies out in an exponential or damped sine-wave manner
- ACF has all zero spikes beyond the q th spike

So we have an MA(q) model when

- the PACF is exponentially decaying or sinusoidal
- there is a significant spike at lag q in ACF, but none beyond q