Box-Jenkins Approach

PD Dr. Ralf Stecking and **Abigail Opokua Asare**

Department of Business Adminstration,
Economics and Law

Institute of Economics
Carl von Ossietzky University Oldenburg

20 January, 2026

Overview

1. Determination of model order
2. Estimation of parameters
3. Model fit
4. Model diagnosis

Determination of the model order (p, q)

- The model order (p, q) of an ARMA model is determined by graphical inspection of the autocorrelation function (ACF) and the partial autocorrelation function (PACF).
- The **autocorrelation coefficient** is computed by:

$$ACF_k = \frac{\sum_{t=1}^{T-k}(X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

Carl von Ossietzky Universität Oldenburg

Determination of the model order (p, q)

- The **partial autocorrelation coefficient** measures the correlation between X_t and X_{t-k}, after eliminating the influence of the values between X_t and X_{t-k}.
- The partial autocorrelation function PACF shows correlations between X_t and X_{t-k}, *controlled* for $X_{t-1}, X_{t-2}, ..., X_{t-k+1}$.
- **Computation:** If time series X_t is predicted by the regression function of k previous time series, the partial autocorrelation is given by the regression coefficient b_k of the k-th (the last) predicting time series, thus

$$X_t = b_0 + b_1 X_{t-1} + b_2 X_{t-2} + b_k X_{t-k}$$

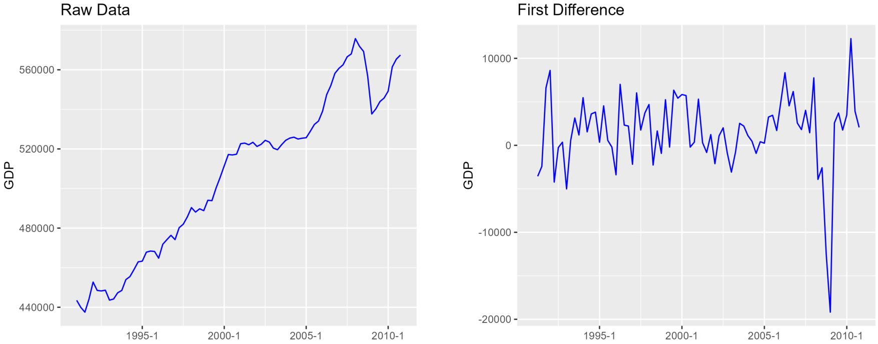
- Finally $PACF_k = b_k$ is the **partial autocorrelation coefficient** of lag k.

Determination of the model order (p, q)

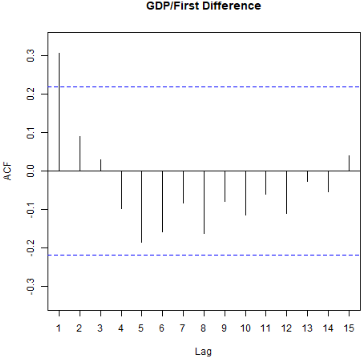
- The model order (p, q) of an ARMA process is identified by inspection of ACF and PACF.

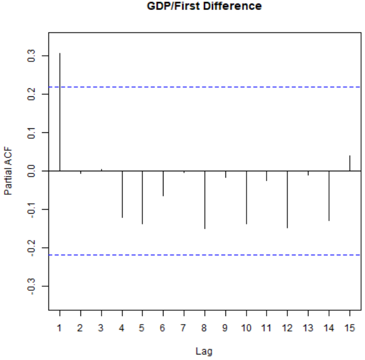
	ACF	PACF
$AR(p)$	• infinite • Exponential decay or damped or sine wave	• infinite • $\alpha_\tau^{(\tau)} = 0$ for $\tau > p$ Significant spikes at p lag but none beyond.
$MA(q)$	• infinite • $\rho_\tau = 0$ for $\tau > q$ Significant spikes at q lag but none beyond.	• infinite • Exponential decay or damped or Sine waves
$ARMA(p,q)$	• When $AR(p)$ $\tau > q$ Exponential decay	• When $MA(q)$ $\tau > p$ Exponential decay

GDP and First Difference



ACF and PACF Plot of GDP first difference





Determination of the model order (p, q)

- In practical application the determination of p and q by ACF and PACF often can be difficult.
- In many cases subjective decisions must be made.
- Alternative approach:
  - Choose upper limits p_{MAX} and q_{MAX} for p and q
  - Estimate every possible model with $0 \leq p \leq p_{MAX}$ and $0 \leq q \leq q_{MAX}$.
  - Choose the best model by using a proper measure for the goodness of fit!

Carl von Ossietzky Universität Oldenburg

Parameter estimation

- ARMA(p, q) model:

$$X_t = c_0 + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

- The estimation of the parameters of an ARMA(p, q) model is carried out by **iterative estimation procedures**. The following procedures are commonly used:
  - (Non linear) least squares method
  - Maximum likelihood estimation

- Iterative estimation procedures require the specification of (random) initial values for parameters c_0, ϕ_k and θ_l.

  Subsequently: Stepwise modification of the initial values until a predefined criterion (e.g. minimizing the error sum of squares of the model) is met.

The model fit

- In order to assess the model fit the residual variance $\hat{\sigma}^2(k)$ can be used:

$$\hat{\sigma}^2_{p,q} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^2$$

  with T number of observations, p and q are parameters of the ARMA(p, q) model and $\hat{\varepsilon}_t$ is estimating the residual at time point t.

- Problem: the residual variance (the *error of the model*) is inversely proportional to the model size, i.e. models with a large number of parameters are favored (similar to the determination coefficient R^2 of the linear regression).

- In many cases the forecasting performance of such models is poor: they are **overfitted!**

Information criteria

An information criterion measures the model fit, considering a penalty term for the number of parameters p and q. Two well-known information criteria are:

1. AIC criterion (Akaike's information criterion)

$$AIC(p,q) = ln\ \hat{\sigma}^2_{p,q} + 2 \cdot \frac{(p+q)}{T}$$

2. BIC criterion (Bayesian information criterion)

$$BIC(p,q) = ln\ \hat{\sigma}^2_{p,q} + \frac{(p+q)ln\ T}{T}$$

Always choose the ARMA(p, q) model for which the criterion is **minimal**.

Model diagnosis

- Question: Is the selected model adequate?
- In an ideal case the ARMA(p, q) model reproduces the relevant characteristics of time series X_t. This means: the unexplained remainder (i.e. the residuals $\hat{\varepsilon}_t$) does not contain any information, structures or systematic $\rightarrow$ white noise!
- There are two different procedures for model diagnosis:
  1. Graphical:
     - plot of the residuals over time t: are there any systematic figures detectable?
     - ACF and PACF plot of the residuals: are there any **significant** correlations?
  2. Statistical: Box-Pierce test, Ljung-Box test

Carl von Ossietzky Universität Oldenburg

Box-Pierce test

- Box-Pierce test is testing the hypothesis, that the first K auto-correlations of the residuals are equal to zero:

$$H_0 : \rho_{\hat{\varepsilon}}(1) = \rho_{\hat{\varepsilon}}(2) = ... = \rho_{\hat{\varepsilon}}(K) = 0$$

- The test statistic reads:

$$Q_{BP} = T \sum_{k=1}^{K} \rho_{\hat{\varepsilon}}^2(k)$$

Q_{BP} is approximately χ^2 distributed with $K - p - q$ degrees of freedom.

Ljung-Box test

- Ljung-Box test is also testing the hypothesis, that the first K autocorrelations of the residuals are equal to zero:

$$H_0 : \rho_{\hat{\varepsilon}}(1) = \rho_{\hat{\varepsilon}}(2) = ... = \rho_{\hat{\varepsilon}}(K) = 0$$

- The test statistic reads:

$$Q_{LB} = T(T+2)\sum_{k=1}^{K} \frac{\rho_{\hat{\varepsilon}}^2(k)}{T-k}$$

  Q_{LB} is approximately χ^2 distributed with $K-p-q$ degrees of freedom.

- For small sample sizes or more robust results, the Ljung-Box test is generally preferred.

GDP: ARIMA (0,1,0)

```
Series: df1$gdp
ARIMA(0,1,0) with drift

Coefficients:
          drift
      1568.9019
s.e.   487.6669

sigma^2 = 19031079:  log likelihood = -773.67
AIC=1551.34   AICc=1551.5   BIC=1556.08

Training set error measures:
                   ME     RMSE      MAE          MPE      MAPE      MASE      ACF1
Training set 5.524162 4307.587 2964.663 -0.002124409 0.5852234 0.8769136 0.3049348
```

GDP: ARIMA (1,1,0)

```
Series: df1$gdp
ARIMA(1,1,0) with drift

Coefficients:
         ar1      drift
      0.3082  1543.0566
s.e.  0.1071   666.7094

sigma^2 = 17434466:  log likelihood = -769.75
AIC=1545.5   AICc=1545.82   BIC=1552.61

Training set error measures:
                   ME     RMSE      MAE         MPE      MAPE      MASE         ACF1
Training set 28.28396 4096.422 2933.614 0.004053159 0.5811208 0.8677297 -0.002197978
```

Carl von Ossietzky Universität Oldenburg

GDP: ARIMA (0,1,1)

```
Series: df1$gdp
ARIMA(0,1,1) with drift

Coefficients:
         ma1      drift
      0.2916  1557.5321
s.e.  0.1025   599.4028

sigma^2 = 17557870:  log likelihood = -770.02
AIC=1546.04   AICc=1546.36   BIC=1553.15

Training set error measures:
                   ME     RMSE     MAE         MPE      MAPE      MASE      ACF1
Training set 17.60785 4110.894 2924.23 0.001432736 0.5795691 0.8649541 0.0185144
```

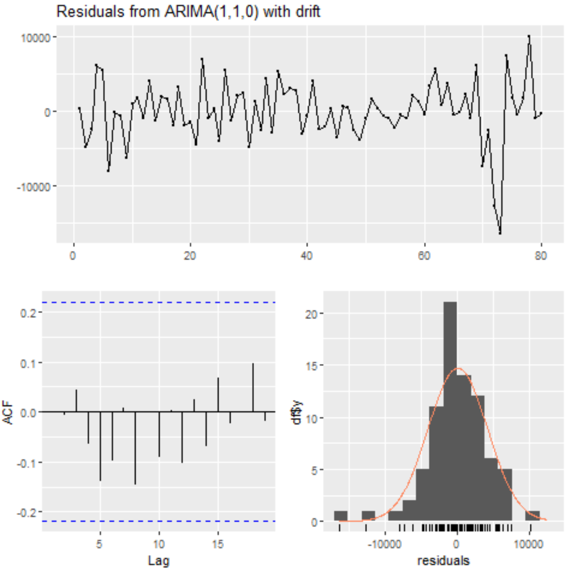
Carl von Ossietzky Universität Oldenburg

GDP: ARIMA (1,1,1)

```
Series: df1$gdp
ARIMA(1,1,1) with drift

Coefficients:
         ar1     ma1      drift
      0.3021  0.0061  1543.9456
s.e.  0.3498  0.3663   665.0868

sigma^2 = 17663908:  log likelihood = -769.75
AIC=1547.5   AICc=1548.04   BIC=1556.98

Training set error measures:
                   ME     RMSE      MAE         MPE      MAPE      MASE        ACF1
Training set 27.60567 4096.427 2932.842 0.003907974 0.5809815 0.8675012 -0.0021814
```

Carl von Ossietzky Universität Oldenburg

Model diagnosis: ARIMA (1,1,0)



Model diagnosis: ARIMA (1,1,0)

```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,0) with drift
Q* = 5.8105, df = 9, p-value = 0.7587

Model df: 1.   Total lags used: 10
```