Forecasting Methods and Applications
Applied Data Analysis School

Lecture 4

Univariate Time Series Models

November 2021

Cristina Amado
University of Minho

Outline of the lecture:

- Basic concepts
- Autoregressive models
- Moving average models
- ARIMA models
- Model building
- Applications

References:

- Diebold, F. X. (2007), Elements of Forecasting, 4th edition, South-Western College Publishing.
- Montgomery, D. C., Jennings, C. L. and Kulahci, M. (2015), Introduction to Time Series Analysis and Forecasting, 2nd edition, Wiley.
- Brockwell, P. J. and Davis, R. A. (2002), Introduction to Time Series and Forecasting, 2nd edition, Springer-Verlag, New York.

Basic concepts in time series

Definition: A time series y_t is a sequence of observations measuring certain quantity over time $t = 1, 2, ..., T$. The subscript t denotes the individual observation, corresponding to (equally spaced) periods of time.

Remark: In applications, a time series is a sequence of data points, typically measured at successive time instants spaced at uniform time intervals. Thus, the observations have a natural temporal ordering. Usually, the measurements are observed at equally spaced time intervals which results in discrete time series.

Time series analysis: Time series analysis is a multidisciplinary field. Time series occur in economics, finance, life sciences, earth sciences, etc.

Objectives of time series analysis:

- **Modelling:** Understanding the stochastic mechanism underlying a series.
- **Forecasting:** Predicting future values of a series based on its own history.
- **Statistical inference:** Is the linear model the most suitable for modelling the series?

Notation:

- Sequence of observations over time generally observed at equally spaced time intervals (annual, monthly, daily, etc.).
- y_t = observed value of the series y at time t.
- Sample of T observations: $\{y_1, y_2, ..., y_T\}$

Basic concepts

- **Stationarity:** A stationary time series is one whose properties do not depend on the time at which the series is observed.

- **Characteristics** of a stationary series:
  - Roughly horizontal
  - Constant variance
  - No patterns predictable in the long-term

- This means that time series with trends or seasonality are not stationary. In general, a stationary time series will have no predictable patterns in the long-term.

- Note that time series with aperiodic cycles are stationary given that, in the long-term, the timing of these cycles is not predictable.

Basic concepts in time series

If $\{y_t\}$ is **strictly stationary**, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

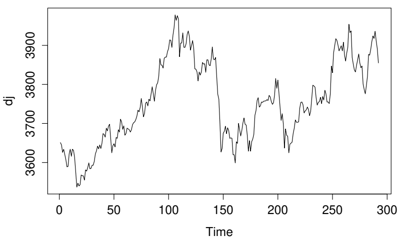
A time series $\{y_t\}$ is **weakly stationary** (or stationary in covariance) if, $\forall t$

1. $\mathrm{E}(y_t) = \mu$,
2. $\mathrm{Var}(y_t) = \sigma^2$,
3. $\mathrm{Cov}(y_t, y_{t-s}) = \gamma_s < \infty, \;\; s = 1, \ldots$

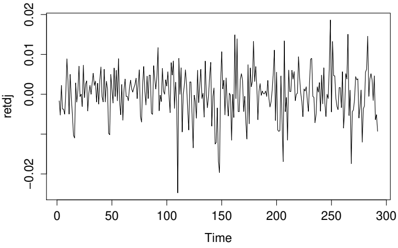
Remark: Strict stationarity implies weak stationarity provided that the first two moments of the series exist. On the other hand, a weakly stationary series may not be strictly stationary.

Remark: $\{y_t\}$ i.i.d. $N(0,1) \Longrightarrow \{y_t\}$ is strictly stationary.

Smoothing methods



(a) DJIA



(b) DJIA returns

Figure: Time series data: (a) Dow-Jones price index; (b) Daily returns of the Dow-Jones index.

Basic concepts in time series

- Transformations such as logarithms can help to stabilize the variance of a time series.
- Alternatively, we can make a time series stationary is by differencing. Differencing can help stabilize the mean of a time series by removing changes in the level, and therefore eliminating trend and seasonality.
- **Differencing:** The process of computing the differences between consecutive observations

$$\Delta y_t = y_t - y_{t-1}$$

Basic concepts in time series

- **Identification of non-stationary series:**
  - Time series plot
  - The autocorrelation function (ACF) of non-stationary data decreases slowly
  - For non-stationary data, the value of the first-order autocorrelation $\widehat{\rho}_1$ is often large and positive.
  - Statistical tests

Definition: ε_t is a **white noise** process or $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ when

(i) $\text{E}\varepsilon_t = 0$, $\forall t$

(ii) $\text{E}\varepsilon_t^2 = \sigma^2$, $\forall t$

(iii) $\text{E}\varepsilon_t\varepsilon_s = 0$, $t \neq s$

Thus, a white noise series is stationary.

Definition: $\{\varepsilon_t\} \sim \text{iid}N(0, \sigma^2)$ is a Gaussian white noise process.

- When the differenced series is white noise, the model for the original series can be written as

$$y_t - y_{t-1} = \varepsilon_t \quad \text{or} \quad y_t = y_{t-1} + \varepsilon_t$$

  This model is known as **random walk** without drift.

- A random walk model is very widely used for non-stationary data, particularly finance and economic data. Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction

- The forecasts from a random walk model are equal to the last observation, given that future movements are unpredictable, and are equally likely to go up or down.

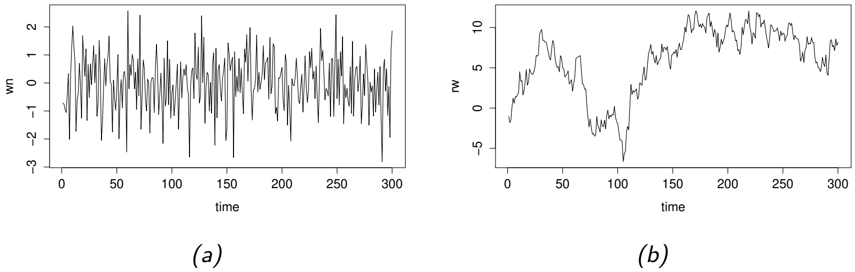


Figure: (a) White noise; (b) Random walk without a drift.

Basic concepts in time series

Lag operator L:

$$\begin{aligned} Ly_t &= y_{t-1} \\ L(Ly_t) = L^2 y_t &= y_{t-2} \\ &\dots \end{aligned}$$

More generally: $L^j y_t = y_{t-j}$.

First-order differences:

$$\Delta y_t = y_t - y_{t-1} = y_t - Ly_t = (1 - L)y_t$$

Second-order differences:

$$\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2} = (1 - 2L + L^2)y_t = (1 - L)^2 y_t$$

In general, the dth-order differencing:

$$\Delta^d y_t = (1 - L)^d y_t$$

Remark: The ACF is a measure of linear dependence between the process and its past, i.e., it measures the memory of the process. If $\rho_j = 0$ there can be another type of relationship between y_t and y_{t-j} (e.g. nonlinear).

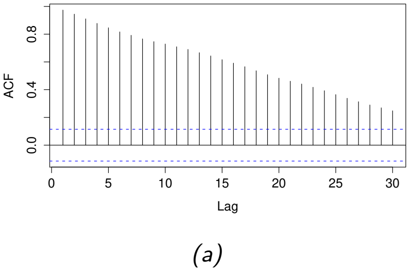
Definition: The j-th autocorrelation of a weakly stationary process $\{y_t\}$ is

$$\rho_j = \frac{\gamma_j}{\gamma_0}$$

onde γ_j is the lag-j autocovariance function of y_t.

The **autocorrelation function (ACF) or correlogram** is the graphically representation of $\rho_j,\ j = 0, 1, \ldots$.

Basic concepts in time series



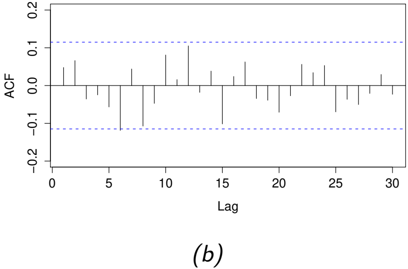


Figure: (a) ACF of the Dow-Jones index; (b) ACF of the daily returns of the Dow-Jones index.

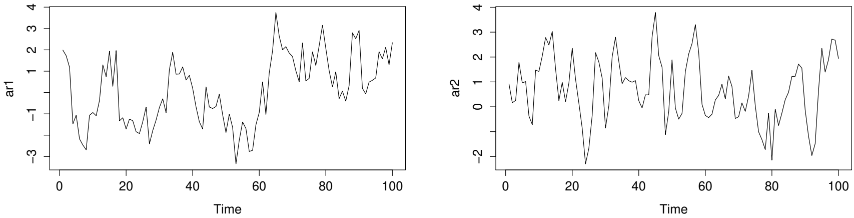
Autoregressive models

- In a multiple regression model, we forecast the variable of interest using a linear combination of predictors.
- In an autoregression model, we forecast the variable of interest using a linear combination of *past values of the variable*.
- Thus, an **autoregressive model of order** p **or an AR**(p) **model** can be written as

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t,$$

where c is a constant and ε_t is white noise. This is a multiple regression with lagged values of y_t as predictors.

Autoregressive models



(a) $y_t = 0.8y_{t-1} + \varepsilon_t$

(b) $y_t = 0.9y_{t-1} - 0.2y_{t-1} + \varepsilon_t$

Figure: Simulated AR processes.

Autoregressive models

- For an AR(1) model:
  - When $\phi_1 = 0$, then y_t is a white noise.
  - When $\phi_1 = 1$ and $c = 0$, then y_t is a random walk.
  - When $\phi_1 = 1$ and $c \neq 0$, then y_t is a random walk with drift.
  - When $\phi_1 < 0$, then y_t tends to oscillate between positive and negative values.

- We normally restrict AR models to stationary data, and then some **stationarity conditions** are required on the parameters:
  - AR(1) model: $-1 < \phi_1 < 1$.
  - AR(2) model: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.
  - AR(p) when $p \geq 3$: restrictions on the parameters are much more complicated.

Moving average models

- A moving average model uses past forecast errors in a regression model.

- Thus, a **moving average model of order q or an MA(q) model** can be written as

$$y_t = c + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q} + \varepsilon_t,$$

  where c is a constant and ε_t is white noise. This is a multiple regression with lagged values of y_t as predictors.

- Each value of y_t can be seen as a weighted moving average of the past few forecast errors.

Moving average models

- As with AR models, the variance of the error term ε_t will only change the scale, not the pattern of the series.

- **Moving average models vs. moving average smoothing:**

  - A moving average model is used for forecasting future values while moving average smoothing is used for estimating the trend-cycle of past values.

Moving average models

- Note that any stationary AR(p) model can be written as an MA(∞) model.
  - For example, consider an AR(1) model

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

  and so on.

- When $-1 < \phi_1 < 1$, the value of $\phi_1^k \to 0$ as $k \to \infty$, so we will obtain

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \ldots,$$

  i.e., an MA(∞) process.

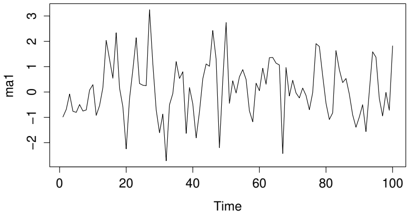
Moving average models

- **Invertibility:** In the same way, we can write any invertible MA(q) process as an AR(∞) process.

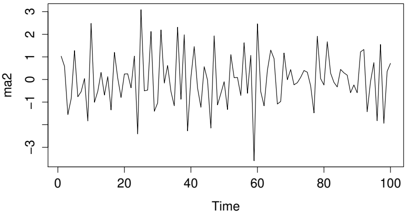
- **Invertibility constrains:**

  - MA(1) model: $-1 < \theta_1 < 1$.
  - MA(2) model: $-1 < \theta_2 < 1$, $\theta_1 + \theta_2 > -1$, $\theta_1 - \theta_2 < 1$.
  - MA(q) when $q \geq 3$: more complicated conditions on the parameters.

Moving average models



(a) $y_t = 0.1 + 0.4\varepsilon_{t-1} + \varepsilon_t$



(b) $y_t = -0.7\varepsilon_{t-1} + 0.1\varepsilon_{t-1} + \varepsilon_t$

Figure: Simulated MA processes.

ARIMA models

- When we include on the right hand of the model both lagged values of y_t and lagged errors.as "predictors" we have a non-seasonal **AutoRegressive Integrated Moving Average model or ARIMA model.**

- The full form of an ARIMA model is

$$\Delta^d y_t = c + \phi_1 \Delta^d y_{t-1} + \ldots + \phi_p \Delta^d y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

  where ε_t is a white noise.

- We call this an ARIMA(p, d, q) model, where:
  - $p =$ order of the autoregressive component
  - $d =$ degree of the first differencing
  - $q =$ order of the moving average component

ARIMA models

- Using the lag operator notation:

$$\underbrace{(1-\phi_1 L - \ldots - \phi_p L^p)}_{\text{AR}(p)} \underbrace{(1-L)^d}_{d \text{ differences}} y_t = c + \underbrace{(1+\theta_1 L + \ldots + \theta_q L^q)}_{\text{MA}(q)} \varepsilon_t$$

- **Some special models:**
  - White noise: ARIMA(0,0,0)
  - Random walk: ARIMA(0,1,0) without constant
  - Random walk with drift: ARIMA(0,1,0) with a constant
  - Autoregression: ARIMA(p, 0, 0)
  - Moving average: ARIMA(0, 0, q)

Box and Jenkins (1970) methodology

Box and Jenkins (1970) were the first to consider a procedure for building empirical time series models. It consists of three stages (i) Model specification; (ii) Estimation; and (iii) Diagnostic checking.

1. Model specification: Based on the inspection of the empirical ACF's and PACF's and information criteria.
   - MA(q) process: $\rho_j = 0$ (ACF), for all $k > q$
   - AR(p) process: $\phi_{jj} = 0$ (PACF), for all $k > p$
   - ARMA(p,q): if none of the functions have lags equal to zero, then choose an ARMA. Due to the identification problem the choice of p and q requires some care.
2. Estimation: The main approaches to estimate ARMA models are the nonlinear least squares and the maximum likelihood methods.
3. Diagnostic checking: Do the residuals satisfy the underlying assumptions? Do the residuals look like a white noise process?

Model building

- Using the ACF and the PACF the model specification can be very subjective.
- An alternative is to use a (data driven) criterion as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) or Hannan-Quinn Information Criterion (HQIC) for determining the order of an ARIMA model or for selecting predictors in the regression model.
- Good models are obtained by minimizing either the AIC, BIC or HQIC:

$$
\begin{aligned}
\mathrm{AIC}(k) &= \ln \hat{\sigma}^2(k) + \frac{2}{T}k && \text{(Akaike)} \\
\mathrm{BIC}(k) &= \ln \hat{\sigma}^2(k) + \frac{\ln(T)}{T}k && \text{(Schwarz)} \\
\mathrm{HQIC}(k) &= \ln \hat{\sigma}^2(k) + \frac{2\ln\ln(T)}{T}k && \text{(Hannan-Quinn)}
\end{aligned}
$$

Model building

Diagnostic checking: Informal methods

- Visual inspection of the series of residuals: standardisation of the residuals is recommended in order to detect outliers, structural breaks, heteroskedasticity, etc.
- Calculate the sample ACF and PACF for the residuals (to detect remaining serial correlation). The ACF and PACF for the residuals should be in the interval $\pm 2/\sqrt{T}$ (rule of thumb), otherwise we might suspect that the model is misspecified.

Diagnostic checking: Statistical tests

- Tests against remaining residual autocorrelation: Tests against remaining residual autocorrelation: Box and Pierce portmanteau test, Ljung and Box test, and the Breusch-Godfrey LM test.
- Other tests: Lomnicki-Jarque-Bera test for normality, testing against conditional heteroskedasticity, RESET test, stability analysis (CUSUM test, Chow test, etc.), testing against nonlinearity, etc.

Forecasting

The main purpose of ARMA models is the projection or extrapolation of the observed time series y_t, $t = 1, \dots, T$, into the future, having observed y_t until moment t.

Predictions are made conditionally on the available information until period t

$$\Omega_t = \{y_t, y_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots\}$$

In order to evaluate the precision of the predictions one may use the mean squared error (MSE):

$$MSE(\widehat{y}_{t+h}) = \mathsf{E}(y_{t+h} - \widehat{y}_{t+h})^2$$

which optimal prediction is

$$\widehat{y}_{t+h} = \mathsf{E}(y_{t+h}|\Omega_t) \equiv \mathsf{E}_t y_{t+h} = y_{t+h|t}.$$

Forecasting

Remark: In the end of period t, the random variables have already been observed: $\mathrm{E}_t y_t = y_t$ and $\mathrm{E}_t \varepsilon_t = \varepsilon_t$. For future values, we have $\mathrm{E}_t \varepsilon_t = \mathrm{E}\varepsilon_t = 0$ (given the independence of $\{\varepsilon_t\}$).

AR(p) process: $y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t$

1-step ahead forecast at forecast origin t :
$y_{t+1|t} = \phi_1 y_t + ... + \phi_p y_{t+1-p}$

h-step ahead forecast at forecast origin t :
$y_{t+h|t} = \phi_1 y_{t+h-1|t} + ... + \phi_p y_{t+h-p|t}$
(compute forecasts recursively for $h = 1, 2, ...$)

Forecasting

Forecast error:

$$y_{t-h} - y_{t+h|t} = \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \ldots + \theta_{h-1} \varepsilon_{t+1} \sim (0, \sigma_y^2(h))$$

where θ_i are the coefficients of a MA representation,

$$\sigma_y^2(h) = \mathrm{E}(y_{t+h} - y_{t+h|t})^2 = \sigma^2 \sum_{j=0}^{h-1} \theta_j^2$$

and $\varepsilon_t \sim N(0, \sigma^2)$.

95% forecast interval:

$$\left[y_{t+h|t} - 1.96\sigma_y(h), y_{t+h|t} + 1.96\sigma_y(h) \right]$$

Note: In practice, replace the unknown parameters by estimates. All results hold asymptotically or approximately under suitable assumptions.