Business Data Science
Course “Prediction and Forecasting”
Part II Box-Jenkins Forecasting Method

Siem Jan Koopman

http://sjkoopman.net

Vrije Universiteit Amsterdam / Tinbergen Institute

2024

Plan

Program continues:

- Statistical Modeling
- Time Series Processes and Their Properties
- Stationary and Non-Stationary Time Series
- ARMA Models – A Review
    - Autoregressive Model
    - Moving Average Model
    - ARMA model
- Forecasting with ARMA models
- Box-Jenkins method
- Assignments

Time Series Processes,

Stationary and Non-Stationary Time Series

Time Series Processes and Properties

Time Series

$$Y_1\,,\ Y_2\,,\ Y_3\,,\ \ldots\,,\ Y_t\,,\ \ldots$$

can be regarded as the result of a process.

In a statistical modeling framework, we regard the observed time series as realizations of a stochastic process:
a process that is subject to random noise.

The properties of the stochastic process can be described in terms of

$$\text{mean}:\ \mu_t = \mathbb{E}(Y_t)$$

$$\text{variance}:\ \sigma_t^2 = \mathbb{V}\text{ar}(Y_t)$$

$$\text{autocovariance}:\ \gamma_{t,s} = \mathbb{C}\text{ov}(Y_t, Y_s)$$

Stationary Time Series Processes

The time series process

$$Y_1\,,\;Y_2\,,\;Y_3\,,\;\ldots\,,\;Y_t\,,\;\ldots$$

is said to be weak stationary when

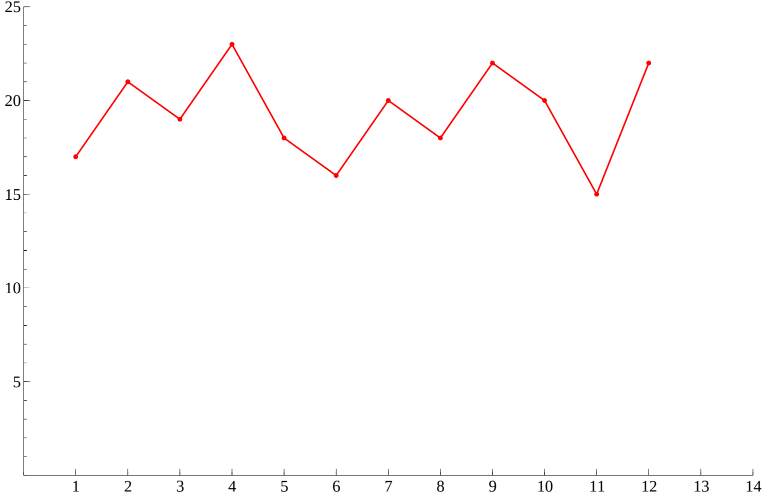
$$\text{mean}:\;\mu_t=\mu\,,\qquad \text{variance}:\;\sigma_t^2=\sigma^2\,,$$

$$\text{autocovariance}:\;\gamma_{t,s}=\gamma_\tau,\quad \text{for }\tau=t-s>0$$

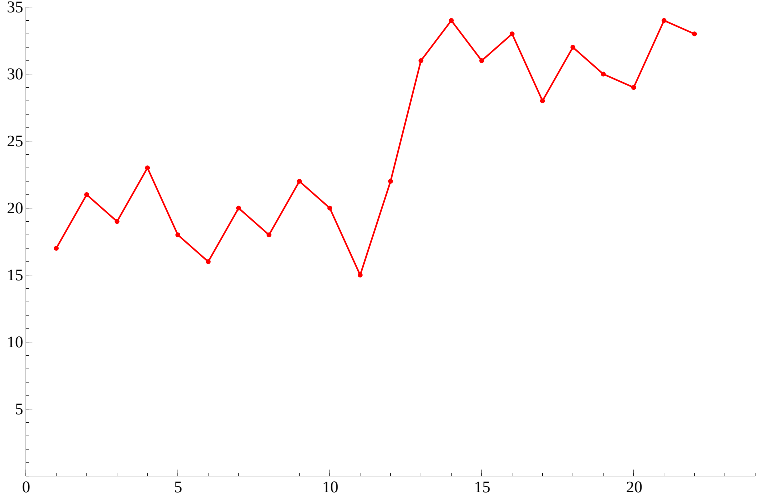
that is, *the first two moments are constant over time.*

The process is said to be strict stationary when the complete distribution is constant over time.

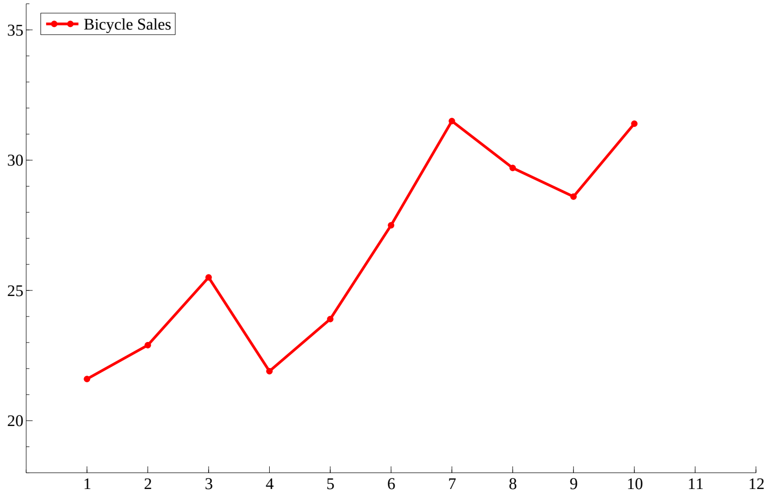
Gasoline Sales Data: (Non-)Stationary ?



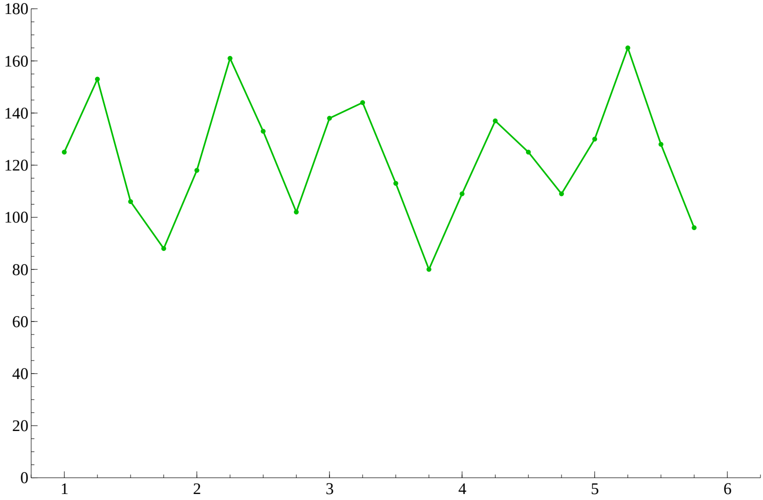
Gasoline Sales Data: (Non-)Stationary ?



Bicycle Sales Data: (Non-)Stationary ?



Umbrella Sales: (Non-)Stationary ?

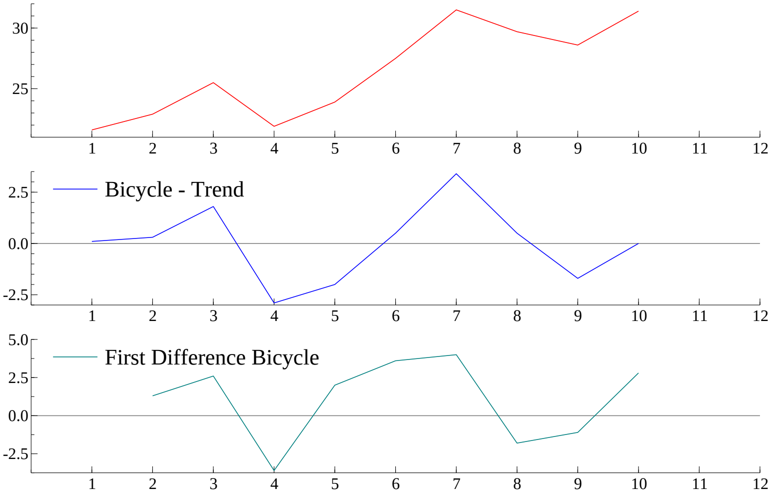


Non-Stationary: What to do about it ?

- Adjustment via regression
  - Fit a trend : Y_t process is referred to as Trend-Stationary
  - Fit a seasonal pattern, etc.
- Adjustment via data transformation
  - First differencing :
    $\Delta Y_t = Y_t - Y_{t-1}$, we focus on the growth of Y_t
  - Second differencing :
    $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$
  - Seasonal differencing

  When differencing leads to a stationary process, we have a Difference-Stationary process.
- When analysis is done for adjusted time series, reverse the transformation for forecasting. Focus remains on Y_t.

Adjust to Stationary: Bicycle Sales



Fitting a trend or Differencing ?

What is best to do ? The observed time series should tell you !

We can formally test for difference-stationary:

- Dicky-Fuller test
- Augmented Dickey-Fuller (ADF) test
- KPSS test, etc.

We will leave this for other courses.

In practice, we can rely on graphical tools:

- Time Series Graph itself
    - Scan for a global constant mean
    - Scan for a global constant variance
    - How to scan for constant autocovariances ?
- Sample autocorrelation function : correlogram
- Sample partial autocorrelation function

Sample statistics

$$\text{mean : } \mu \Leftrightarrow \bar{Y} = T^{-1} \sum_{t=1}^{T} Y_t$$

$$\text{variance : } \sigma^2 \Leftrightarrow s^2 = (T-1)^{-1} \sum_{t=1}^{T} (Y_t - \bar{Y})^2$$

$$\text{autocovariance : } \gamma_\tau \Leftrightarrow c_\tau = (T-1-\tau)^{-1} \sum_{t=\tau+1}^{T} (Y_t - \bar{Y})(Y_{t-\tau} - \bar{Y})$$

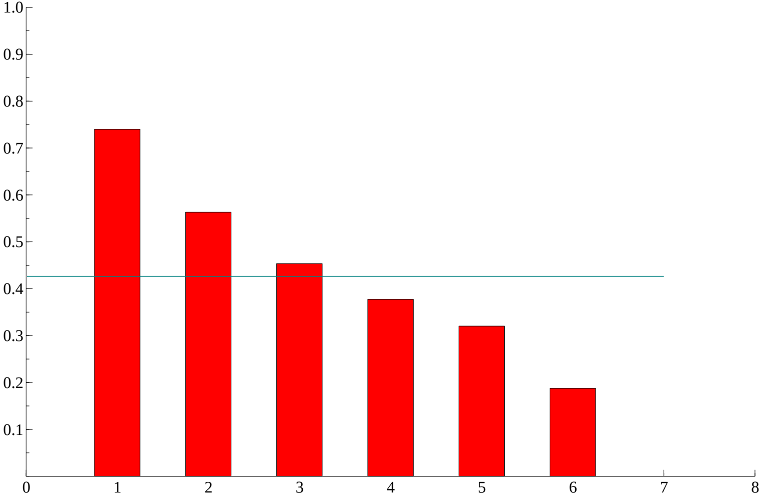
$$\text{autocorrelation : } \rho_\tau = \gamma_\tau / \sigma^2 \Leftrightarrow r_\tau = c_\tau \, / \, s^2, \qquad \text{for } \tau = 1, 2, 3, \ldots$$

The sample autocorrelation function r_τ is also referred to as the

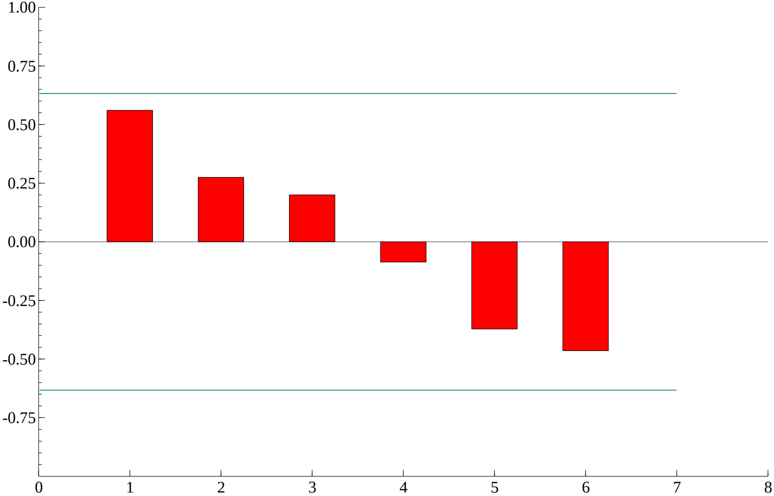
correlogram

Correlogram Gasoline Sales ($T = 12$)

Correlogram Gasoline Sales ($T = 24$)



Correlogram Bicycle Sales



Autoregressive (AR) Processes

What do we learn from Correlogram?

The autocorrelation function indicates the serial dependence within the time series. When autocorrelations are sufficiently non-zero, we say that there is evidence of serial dependence in the time series (memory, persistence).

We have more abilities to predict and forecast when there is “memory” in the time series !

Different kinds of memory are reflected by the correlogram.

Different time series processes imply different autocorrelation functions.

Different observed time series show different correlograms.

Example : Gaussian white noise

Consider the model assumption

$$Y_t \sim \mathcal{NID}(0, \sigma^2)$$

where $\mathcal{NID}$ refers to normally, independent and identically distributed, from which it follows that

$$\mathbb{E}(Y_t) = \mu = 0, \qquad \mathbb{V}\text{ar}(Y_t) = \sigma^2, \qquad \gamma_\tau = 0,$$

for $\tau = 1, 2, \ldots$, that is

$$\rho_0 = 1, \qquad \rho_\tau = \gamma_\tau \,/\, \sigma^2 = 0.$$

There is no memory in this process of Y_t.

The appropriate forecast for Y_{T+1} is $Y_{T+1}^F = \mu = 0$.

Example : Random Walk process

Consider the random walk model

$$Y_{t+1} = Y_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2)$$

and

$$Y_{t+1} = Y_t + \varepsilon_t = Y_{t-1} + \varepsilon_{t-1} + \varepsilon_t = Y_{t-s} + \varepsilon_{t-s} + \ldots + \varepsilon_t$$

for any $1 < s < t$. For $s = t - 1$, it follows that

$$Y_{t+1} = Y_1 + \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_t$$

$$\mathbb{E}(Y_{t+1}) = \mathbb{E}(Y_1), \qquad \mathbb{V}\mathrm{ar}(Y_{t+1}) = \mathbb{V}\mathrm{ar}(Y_1) + t\,\sigma^2.$$

It follows that Random Walk process is Non-Stationary.

All autocorrelations are close to 1, for large T. Please verify !

Example: Autoregressive Process

The Autoregressive process of order 1, the AR(1) process, is given by

$$Y_t = \mu + \phi\, Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2),$$

with constant μ and autoregressive coefficient ϕ for lag Y_{t-1}.

When we set $\mu = 0$, recursive substitution yields

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \varepsilon_t \ = \ \phi^2 Y_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t \\ &= \phi^3 Y_{t-3} + \phi^2 \varepsilon_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t \ = \ \ldots \\ &= \phi^s Y_{t-s} + \phi^{s-1} \varepsilon_{t-(s-1)} + \ldots + \phi\varepsilon_{t-1} + \varepsilon_t. \end{aligned}$$

for $s > 2$. Notice that (with $\mu = 0$) we have $Y_{t-1} = \phi Y_{t-2} + \varepsilon_{t-1}$ and $Y_{t-j} = \phi Y_{t-j-1} + \varepsilon_{t-j}$ for $j = 2, 3, \ldots$

Stationary Autoregressive Process

The AR(1) process $Y_t = \mu + \phi Y_{t-1} + \varepsilon_t$ with $\mu = 0$ and $\phi = 1$ reduces to a Random Walk process.

The assumption for Stationarity is $|\phi| < 1$.

The stationary AR(1) process $Y_t = \mu + \phi Y_{t-1} + \varepsilon_t$, for $\mu = 0$, can be represented by

$$Y_t = \phi^t Y_0 + \phi^{t-1}\varepsilon_1 + \phi^{t-2}\varepsilon_2 + \ldots + \phi\varepsilon_{t-1} + \varepsilon_t$$

Since $|\phi| < 1$, and for large enough t, we have

$$Y_t = \phi^{t-1}\varepsilon_1 + \phi^{t-2}\varepsilon_2 + \ldots + \phi^2\varepsilon_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t$$

From this representation of the AR(1) process, we can derive its properties, see the Appendix.

Stationary Autoregressive Process

Properties of the stationary AR(1) process $Y_t = \mu + \phi\, Y_{t-1} + \varepsilon_t$ are

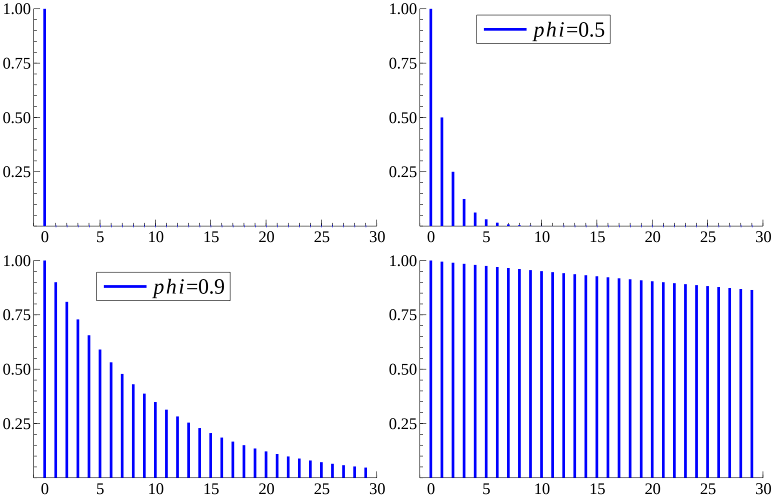
- Mean is $\mathbb{E}(Y_t) = \mu \,/\, (1 - \phi)$;
- Variance is $\mathbb{V}\mathrm{ar}(Y_t) = \sigma^2 \,/\, (1 - \phi^2)$;
- Autocovariance for lag τ is $\mathbb{C}\mathrm{ov}(Y_t, Y_{t-\tau}) = \phi^\tau\, \sigma^2 \,/\, (1 - \phi^2)$;
- Autocorrelation for lag τ is $\mathbb{C}\mathrm{orr}(Y_t, Y_{t-\tau}) = \phi^\tau$.

The autocorrelation function is given by

$$\rho_\tau = \phi^\tau.$$

Can you verify these properties of the AR(1) process ?
See the Appendix.

ACFs for WN, AR(1), RW



AR(1) parameter estimation

The AR(1) model is given by

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2),$$

We assume stationarity $|\phi| < 1$.

- The coefficients μ, ϕ and σ^2 are unknown;
- We estimate unknown coefficients by maximum likelihood;
- Estimation can be based on full maximum likelihood estimation or on conditional likelihood estimation;
- In the latter case, estimation reduces to regression analysis;
- In case of the Gasoline time series ($T = 12$), we obtain : $\hat{\mu} = 27.12$, $\hat{\phi} = -0.40$, $\hat{\sigma} = 2.46$.

AR(1) forecasting

The AR(1) model is given by

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2).$$

The AR(1) forecast is

$$\begin{aligned} Y^F_{T+1} &= \mathbb{E}(Y_{T+1} | Y_1, \ldots, Y_T) \\ &= \mathbb{E}(\mu + \phi Y_T + \varepsilon_{T+1} | Y_1, \ldots, Y_T) \ = \ \mu + \phi Y_T. \end{aligned}$$

The forecast error $U_{T+1} = Y_{T+1} - Y^F_{T+1}$ is

$$U_{T+1} = Y_{T+1} - \mu - \phi Y_T \equiv \varepsilon_{T+1},$$

with properties

$$\begin{aligned} \mathbb{E}(U_{T+1}) &= \mathbb{E}(\varepsilon_{T+1}) = 0, \\ \mathbb{V}\mathrm{ar}(U_{T+1}) &= \mathbb{V}\mathrm{ar}(\varepsilon_{T+1}) = \sigma^2. \end{aligned}$$

AR(1) forecasting: 2-steps

The AR(1) model is given by

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2).$$

The AR(1) two-step ahead forecast is

$$\begin{aligned} Y^F_{T+2} &= \mathbb{E}(Y_{T+2} | Y_1, \ldots, Y_T) \\ &= \mathbb{E}(\mu + \phi Y_{T+1} + \varepsilon_{T+2} | Y_1, \ldots, Y_T) \\ &= \mu + \phi\, \mathbb{E}(Y_{T+1} | Y_1, \ldots, Y_T) \\ &= \mu + \phi Y^F_{T+1} \;=\; (1+\phi)\mu + \phi^2 Y_T. \end{aligned}$$

The two-step ahead forecast error $U_{T+2} = Y_{T+2} - Y^F_{T+2}$ is

$$\begin{aligned} U_{T+2} &= \mu + \phi Y_{T+1} + \varepsilon_{T+2} - (1+\phi)\mu - \phi^2 Y_T \\ &= (1+\phi)\mu + \phi^2 Y_T + \phi \varepsilon_{T+1} + \varepsilon_{T+2} - (1+\phi)\mu - \phi^2 Y_T \\ &= \phi \varepsilon_{T+1} + \varepsilon_{T+2}, \end{aligned}$$

with properties $\mathbb{E}(U_{T+2}) = 0, \qquad \mathbb{V}\mathrm{ar}(U_{T+2}) = (1+\phi^2)\sigma^2.$

AR(1) forecasting: h-steps

The AR(1) h-step ahead forecast, for $h = 1, 2, 3, \ldots$, is

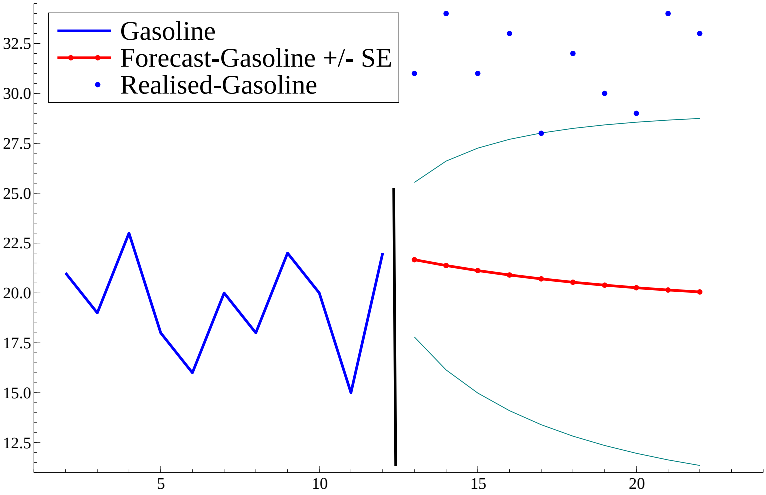
$$\begin{aligned} Y_{T+h}^{F} &= \mathbb{E}(Y_{T+h} | Y_1, \ldots, Y_T) \\ &= \mathbb{E}(\mu + \phi Y_{T+h-1} + \varepsilon_{T+h} | Y_1, \ldots, Y_T) \\ &= \ldots \\ &= \mu \cdot \sum_{j=0}^{h-1} \phi^j + \phi^h Y_T. \end{aligned}$$

The h-step ahead forecast error $U_{T+h} = Y_{T+h} - Y_{T+h}^{F}$ is

$$U_{T+h} = \sum_{j=0}^{h-1} \phi^j \, \varepsilon_{T+h-j},$$

with properties $\mathbb{E}(U_{T+h}) = 0$, $\qquad \mathbb{V}\mathrm{ar}(U_{T+h}) = \sigma^2 \cdot \sum_{j=0}^{h-1} \phi^{2j}$.

Gasoline AR(1) forecasts



Model-based forecasting

Our treatment of the AR(1) model shows some aspects of statistical forecasting, or forecasting based on a statistical dynamic model.

Model-based forecasting is more elaborate, it requires a solid basis of knowledge in statistical theory.

Also, it is more work: more involved computations are needed.

The analysis is also subject (or restricted) to model assumptions which can be unrealistic and difficult to verify.

On the other hand, the analysis provides more information such as the confidence bounds of forecasts. It also enables the time series analyst to gain more insights and a better understanding of the underlying dynamic process of the time series.

A better understanding of the dynamics also allows the formulation of better forecast functions !!

AR(1) forecasting: Gasoline Sales

The AR(1) model :

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2)$$

In case of the Gasoline Sales time series ($T = 12$), we have

- Estimates $\hat{\mu} = 27.12$, $\hat{\phi} = -0.40$, $\hat{\sigma} = 2.46$.
- The one-step ahead forecasting function of the AR(1) model is $Y^F_{T+1} = \mathbb{E}(Y_{T+1} | Y_1, \ldots, Y_T) = \mu + \phi Y_T$;
- The forecast based on AR(1) model and parameter estimates is $\hat{Y}^F_{T+1} = 27.12 - 0.40 \times Y_T$;
- The last observation for the Gasoline Sales data is $Y_T = 22$;
- We obtain the forecast $27.12 - 0.40 \times 22 = 18.32$.

AR(1) forecasting precision

To assess the forecast precision of the AR(1) model, we need to generate 11 forecasts :

$$\hat{Y}_t^F = \hat{\mu} + \hat{\phi} \times Y_{t-1},$$

for $t = 2, \ldots, T$ with $T = 12$.
We can take the estimated values $\hat{\mu} = 27.12$ and $\hat{\phi} = -0.40$.

But is it the correct choice for $\hat{\mu}$ and $\hat{\phi}$?
These estimates are based on the full sample.
The forecast $\hat{Y}_t^F$ is not solely based on past data, it is also based on present and future data !!

For each forecast $\hat{Y}_t^F$, we need to re-estimate the unknown coefficients using the observation set $\{Y_1, \ldots, Y_{t-1}\}$.

AR(1) forecasting precision

For each forecast $\hat{Y}_t^F$, we need to re-estimate the unknown coefficients using the observation set $\{Y_1, \ldots, Y_{t-1}\}$.

However, to estimate the coefficients for computing the forecast $\hat{Y}_2^F$ based on the AR(1) model is challenging. How to estimate three unknown coefficients using one observation ?

We need to have at least THREE observations, but it is better to start with a few more.

Forecast Comparisons, 6 forecasts

Method	ME	MAE	MAPE	MSE
Running Avg	0.35	2.20	11.88	6.69
Random Walk	1.00	4.00	20.74	19.00
Expert	0.50	2.17	11.58	6.17
AR(1)	-0.15	1.89	10.53	5.29

ME = Mean Error
MAE = Mean Absolute Error
MAPE = Mean Absolute Percentage Error
MSE = Mean Squared Error

Autoregressive Processes : AR(p)

Autoregressive Process AR(p)

The AR(1) process can be extended with more lags :

$$Y_t = \mu + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2).$$

We refer to it as the AR(p) process, for possible values $p = 1, 2, 3, \ldots$.

The parameters $\phi_1, \ldots, \phi_p$ can be made subject to restrictions such that the AR(p) process is Stationary.
For this course, we just assume that the ϕ parameters are subject to the stationary restriction.

For modeling and forecasting of an observed time series Y_t, we will treat the AR(p) equation as a regression model and apply the *least squares* method for the estimation of ϕ and σ^2 parameters.

Partial Autocorrelation function

The partial autocorrelation measure the relationship between Y_t and Y_{t-k} after removing the effects of lags $1, 2, \ldots, k-1$, that is Y_{t-1}, $Y_{t-2}, \ldots, Y_{t-k+1}$.

The first partial autocorrelation is identical to the first autocorrelation, because there is nothing between them to remove.

Each partial autocorrelation can be estimated as the last coefficient in an autoregressive AR(k) model. Specifically, the kth partial autocorrelation coefficient, is equal to the estimate of ϕ_k in an AR(k) model, for $k = 2, 3, 4, \ldots$.

It requires the estimation of the ϕ coefficients in $k-1$ models : AR(2), AR(3), ... , AR(k).

In practice, there are more efficient algorithms for computing the partial autocorrelations than fitting all of these autoregressions, but they give the same results.

PACF in practice: choice of p

Given the definition and construction of the partial autocorrelation function (PACF), the PACF is designed to determine the appropriate order of the AR(p) model.

By plotting the PACF (in a similar way as the correlogram), one can determine what a reasonable value is for p.

There are more formal statistical test procedures available for determining the optimal p value for the AR(p) model, but the PACF provides clearly useful information in this respect.

Forecasting with AR(p) model

The forecasts of an AR(p) model for different forecast windows h can be obtained recursively as follows

$$\hat{Y}^F_{T+h} = \mu + \phi_1 Y^F_{T+h-1} + \cdots + \phi_p Y^F_{T+h-p}$$

More specifically

$$\begin{aligned}
Y^F_{T+1} &= \mu + \phi_1 Y_T + \phi_2 Y_{T-1} + \cdots + \phi_p Y_{T+1-p} \\
Y^F_{T+2} &= \mu + \phi_1 Y^F_{T+1} + \phi_2 Y_T + \cdots + \phi_p Y_{T+2-p} \\
Y^F_{T+3} &= \mu + \phi_1 Y^F_{T+2} + \phi_2 Y^F_{T+1} + \cdots + \phi_p Y_{T+3-p} \\
&\cdots
\end{aligned}$$

The AR(p) forecasts have the ability to capture very rich stationary dynamics in the time series including cyclical and seasonal processes.

Moving Average (MA) Processes

Moving Average Processes

The Moving Average process of order q, the MA(q) process, is given by

$$Y_t = \mu + \varepsilon_t + \theta_1 \, \varepsilon_{t-1} + \ldots + \theta_q \, \varepsilon_{t-q}, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2),$$

with constant μ and moving average coefficients θ_j for $j = 1, \ldots, q$.

The moving average process MA(q) is more straightforward.

The moving average process is stationary by its construction.

Moving Average Process MA(1)

Properties of the MA(1) process $Y_t = \mu + \varepsilon_t + \theta\,\varepsilon_{t-1}$ are

- Mean is $\mathbb{E}(Y_t) = \mu$;
- Variance is $\mathbb{V}\mathrm{ar}(Y_t) = \sigma^2\,(1 + \theta^2)$;
- Autocovariance is $\mathbb{C}\mathrm{ov}(Y_t, Y_{t-1}) = \theta\,\sigma^2$ and for lag $\tau > 1$ is $\mathbb{C}\mathrm{ov}(Y_t, Y_{t-\tau}) = 0$;
- Autocorrelation is $\mathbb{C}\mathrm{orr}(Y_t, Y_{t-1}) = \theta\,/(1 + \theta^2)$ and for lag $\tau > 1$ is $\mathbb{C}\mathrm{orr}(Y_t, Y_{t-\tau}) = 0$.

The autocorrelation function is given by

$$\rho_1 = \theta\,/(1 + \theta^2), \qquad \rho_\tau = 0, \quad \tau = 2, 3, 4, \ldots .$$

Can you verify these properties of the MA(1) process ?

Invertibility restriction : since ρ_1 is the same for θ and $1/\theta$, we have $|\theta| < 1$.

Correlogram in practice: choice of q

The Moving Average process of order q, the MA(q) process, is given by

$$Y_t = \mu + \varepsilon_t + \theta_1 \, \varepsilon_{t-1} + \ldots + \theta_q \, \varepsilon_{t-q}, \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma^2),$$

with autocorrelation function (ACF)

$$\rho_j = \gamma_j^* \, / (1 + \sum_{k=1}^{q} \theta_k^2), \quad j = 1, \ldots, q, \qquad \rho_j = 0, \quad j > q$$

where $\gamma_j^* = \theta_j + \sum_{k=1}^{q-j} \theta_k \theta_{k+j}$. By plotting the correlogram (or the sample ACF) we can determine directly what a reasonable value is for q.

There are more formal statistical test procedures available for determining the optimal q value for the MA(q) model, but the correlogram provides clearly useful information in this respect.

Forecasting with MA(1) model

Consider the MA(1) model

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad t = 1, \ldots, T,$$

where θ is known. How can we forecast disturbances ε_t that we never observe ! But we do observe $Y_1, \ldots, Y_T$ and they should give information about ε_t !

When we assume that $\varepsilon_0 = 0$, we can compute $\varepsilon_1, \ldots, \varepsilon_T$ without error for a realised series $Y_1, \ldots, Y_T$. We do this as follows.

Given $\varepsilon_0 = 0$, observing $Y_1, \ldots, Y_T$ is observing $\varepsilon_1, \ldots, \varepsilon_T$:

$$\begin{aligned}
Y_1 &= \varepsilon_1 \\
&\Rightarrow \varepsilon_1 = Y_1, \\
Y_2 &= \varepsilon_2 + \theta\varepsilon_1 = \varepsilon_2 + \theta Y_1 \\
&\Rightarrow \varepsilon_2 = Y_2 - \theta Y_1,
\end{aligned}$$

etc.

Forecasting with MA(1) model

By having $\varepsilon_0 = 0$, the MA(1) forecast is

$$\begin{aligned} Y_{T+1}^F &= \mathbb{E}(Y_{T+1}|Y_1, \ldots, Y_T) \\ &= \mathbb{E}(\mu + \varepsilon_{T+1} + \theta\varepsilon_T|Y_1, \ldots, Y_T) \ = \ \mu + \theta\varepsilon_T. \end{aligned}$$

The forecast error is

$$U_{T+1} = Y_{T+1} - Y_{T+1}^F = Y_{T+1} - \mu - \theta\varepsilon_T = \varepsilon_{T+1},$$

with properties

$$\begin{aligned} \mathbb{E}(U_{T+1}) &= \mathbb{E}(\varepsilon_{T+1}) = 0, \\ \mathbb{V}\mathrm{ar}(U_{T+1}) &= \mathbb{V}\mathrm{ar}(\varepsilon_{T+1}) = \sigma^2. \end{aligned}$$

Forecasting with ARMA(1, 1) model

Consider the ARMA(1, 1) model as given by

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \qquad .$$

By having $\varepsilon_0 = 0$, the forecast is given by

$$\begin{aligned} Y_{T+1}^F &= \mathrm{E}(Y_{T+1} | Y_1, \ldots, Y_T) \\ &= \mathrm{E}(\mu + \phi Y_T + \varepsilon_{T+1} + \theta \varepsilon_T | Y_1, \ldots, Y_T) \\ &= \mu + \phi Y_T + \theta \varepsilon_T. \end{aligned}$$

The forecast error is $U_{T+1} = Y_{T+1} - Y_{T+1}^F = \varepsilon_{T+1}$.

Forecasting with ARMA Model

We have considered the forecast function of the AR(1), AR(p), MA(1) and ARMA($1, 1$) models.

These models belong to the class of Autoregressive Moving Average (ARMA) models.

For stationary time series, the ARMA class has an excellent record of forecasting with high-precision.

Forecasting with AR models is relatively straightforward.

Forecasting with MA and ARMA models is a bit more challenging.

Box-Jenkins Forecasting Method

Box-Jenkins Forecasting: 4 steps

1. *Identification*: determine appropriate values for p, d and q for ARMA(p, q) model, after possible differencing, d.

   The main tools are time series plot, correlogram, PACF.
2. *Estimation*: estimate parameters of selected ARMA model.
3. *Diagnostic Checking*: in-sample analysis, verify whether prediction residuals are white noise.
   Also, look at R-square, and information criterion such as Akaike Information Criteria (AIC).

   When satisfied, continue, otherwise go back to 1.
4. *Forecasting*: out-of-sample forecasting and validate forecast precision.

Assignment II

Assignment Tasks (by group)
deadline Friday 2 March

Assignment II (1/2)

Reconsider the complete Assignment I in full (including re-estimation of models when sample changes), but without the references to seasonal issues,

- (a) Implement the AR(1) forecasts for Assignment I and compare it with performances of other methods;
- (b) Implement the MA(1) forecasts for Assignment I and compare it with performances of other methods;
- (c) Implement the ARMA(1,1) forecasts for Assignment I and compare it with performances of other methods;
- (d) Comment on these additional results and give a recommendation whether more lags need to be added to the AR/MA/ARMA models.

The estimation of parameters in AR models can be done via regression as discussed above. The estimation of parameters in MA and ARMA models can be done by minimizing an in-sample forecast error statistic such as MSE and/or MAE.

Assignment II (2/2)

Reconsider the complete Assignment I in full (including re-estimation of models when sample changes), but without the references to seasonal issues,

- (e) Consider multi-step forecasting in Assignment I.
  It requires re-visiting all methods (including ARMA forecasting, see above) and formulate forecast functions for $\hat{Y}^F_{T+h}$, that is the h-step ahead forecast.
  The implementation is only needed for $h = 2$, but you are free to also carry out Assignment I again for other values of $h = 3, 4, 5, \ldots$.

Appendix : Properties AR(1)

Mean of AR(1) Process

Let $\{Y_t\}$ be generated by an AR(1) process with $\mu = 0$ and $|\phi| < 1$. Then, it follows that $\mathbb{E}(Y_t) = 0$.

Given the assumption $|\phi| < 1$, we have that Y_t admits the representation $Y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$ and,

$$\mathbb{E}(Y_t) = \sum_{j=0}^{\infty} \phi^j \, \mathbb{E}(\varepsilon_{t-j}) = 0$$

So the mean is zero and does not change over time!

When $\mu \neq 0$, we have

$$\mathbb{E}(Y_t) = \mu \cdot \sum_{j=0}^{\infty} \phi^j + \sum_{j=0}^{\infty} \phi^j \, \mathbb{E}(\varepsilon_{t-j}) = \mu \, / \, (1 - \phi),$$

since $\sum_{j=0}^{\infty} \phi^j = 1 \, / \, (1 - \phi)$.

Variance of AR(1) Process

Let $\{Y_t\}$ be generated by an AR(1) process with $|\phi| < 1$ and $\mu = 0$. Then, it follows that $\mathbb{V}\mathrm{ar}(Y_t) = \sigma^2 \,/\, (1 - \phi^2)$.

Given the assumption $|\phi| < 1$ and for $\mu = 0$, we have that Y_t admits the representation $X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$ and,

$$\mathbb{V}\mathrm{ar}(Y_t) = \mathbb{V}\mathrm{ar}\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \phi^{2j}\, \mathbb{V}\mathrm{ar}(\varepsilon_{t-j}) = \frac{\sigma^2}{1 - \phi^2}.$$

Notice that ϕ is a fixed coefficient and, for $|a| < 1$, $\sum_{j=0}^{\infty} a^j = (1 - a)^{-1}$. So variance does also not change over time!

Autocovariance function of AR(1)

Let $\{Y_t\}$ be generated by an AR(1) process with $|\phi| < 1$ and $\mu = 0$. Then, it follows that $\gamma_h := \mathbb{C}\mathrm{ov}(Y_t, Y_{t+h}) = \phi^h \gamma_0$ with $\gamma_0 = \mathbb{V}\mathrm{ar}(Y_t)$.

Given the assumption $|\phi| < 1$, we have that Y_t admits the representation $X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$ and,

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}(Y_t, Y_{t-h}) &= \mathbb{C}\mathrm{ov}\left(\sum_{k=0}^{\infty} \phi^k \varepsilon_{t-k} \,,\, \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-h-j} \right) \\
&= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \phi^k \phi^j \,\mathbb{C}\mathrm{ov}\left(\varepsilon_{t-k} \,,\, \varepsilon_{t-h-j} \right) = \sum_{j=0}^{\infty} \phi^{h+j} \phi^j \sigma^2 \\
&= \sum_{j=0}^{\infty} \phi^h \phi^{2j} \sigma^2 = \phi^h \sum_{j=0}^{\infty} \phi^{2j} \sigma^2 = \phi^h \gamma_0.
\end{aligned}
$$

Note: $\mathbb{C}\mathrm{ov}(\varepsilon_{t-k}, \varepsilon_{t-h-j}) = \sigma^2$ when $k = h + j$, and 0 otherwise. Also, $a^{b+c} \cdot a^d = a^b \cdot a^{c+d}$. Autocovs do not change over time!

Autocorrelation function of AR(1)

Let $\{Y_t\}$ *be generated by an AR(1) process with* $|\phi| < 1$ *and* $\mu = 0$*. Then, it follows that* $\rho_h = \rho(Y_t, Y_{t-h}) = \phi^h$*.*

By definition (correlation = covariance / variance)

$$\rho_h := \frac{\gamma_h}{\gamma_0}.$$

Hence, it holds trivially that,

$$\rho_h := \frac{\gamma_h}{\gamma_0} = \frac{\phi^h\,\gamma_0}{\gamma_0} = \phi^h \qquad \text{for} \quad h \geq 1.$$

It follows that mean, variance, autocovariance and autocorrelations do not change over time.
Therefore, the AR(1) process with $|\phi| < 1$ is weakly stationary.