Stationary Stochastic Time Series Models

Series de Tiempo

UTDT

January 2020

Introduction

- Uses of Univariate Times Series models:
  1. Forecasting: Need to find a model to characterize the series which then is used to produce a forecast.
  2. Finding which process an exogenous driving variable follows (e.g. A process for Dividends is usually assumed for calculating the fundamental value of Stock prices)

- Modeling a Time Series
- Consider a time series

$$(x_1, x_2....x_n)$$

as the realization of a stochastic process with distribution given by

$$p(x_1, x_2, ..., x_n).$$

- In general, fully specifying p is too ambitious: first and second moments

$$n \text{ means} : E(x_1), E(x_2), ..., E(x_n)$$
$$n \text{ variances} : V(x_1), V(x_2), ..., V(x_n)$$
$$\frac{n(n-1)}{2} \text{ covariances:} Cov(x_i, x_j), i < j.$$

- However, we only have n data points to estimate $2n + \frac{n(n-1)}{2}$ parameters!

Stationarity

Definition

A stochastic process is said to be **strictly stationary** *if its properties are unaffected by a change in the time origin, that is*

$$(\forall l)p(x_1, x_2, ..., x_n) = p(x_{1+l}, x_{2+l}, ..., x_{n+l}) \tag{1}$$

Definition

A stochastic process is said to be **weak stationary** *if the first and second moments exist and do not depend on time.*

$$E(x_1) = E(x_2) = ... = E(x_t) = \mu \tag{2}$$

$$V(x_1) = V(x_2) = ... = V(x_t) = \sigma^2 \tag{3}$$

$$Cov(x_t, x_{t-k}) = Cov(x_{t+l}, x_{t-k+l}) = \gamma_k \tag{4}$$

Comments

- Covariances are functions only of the lag k, and not of time. These are usually called **autocovariances**.
- From equations 3 and 4 we can obtain the **autocorrelations**

$$\rho_k = \frac{Cov(x_1, x_2)}{\sqrt[2]{V(x_1)V(x_2)}} = \frac{\gamma_k}{\sigma^2} = \frac{\gamma_k}{\gamma_o} \tag{5}$$

- The **autocorrelations** considered as a function of k are referred to as the autocorrelation function, **ACF**
- Notice:

$$\gamma_k = Cov(x_t, x_{t-k}) = Cov(x_{t-k}, x_t) = Cov(x_t, x_{t+k}) = \gamma_{-k} \tag{6}$$

Modelling a time Series using Moving Average and Autoregressive Processes: Using the ACF Function to identify a series.

- Lets define $y_t = (x_t - \mu)$ and consider the following first order M.A. Process, MA(1), for y_t:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Where ε_t is such that

$$\begin{aligned} E(\varepsilon_t) &= 0 \quad (7) \\ V(\varepsilon_t) &= \sigma^2 \quad (8) \\ Cov(\varepsilon_t, \varepsilon_{t-k}) &= 0 \text{ for all } k. \quad (9) \end{aligned}$$

- For this process we can:

1. Calculate the first and second moments of y_t and check the stationarity condition
2. Check its ACF in oder to use this function to Identify the process.

$$E(y_t) = 0$$

$$E(y_t)^2 = E(\varepsilon_t + \theta_1 \varepsilon_{t-1})^2 = \sigma^2(1 + \theta_1^2)$$

$$\begin{aligned} E(y_t y_{t-k}) &= E(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-k} + \theta_1 \varepsilon_{t-k-1}) \\ & \begin{cases} \sigma^2 \theta_1 & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases} \end{aligned}$$

$$\rho_k = \begin{cases} \theta_1/(1 + \theta_1^2) & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases}$$

- Notice:

1. The MA(1) satisfies the Stationarity Conditions
2. The autocorrelation function, ρ_k, for a MA(1) has only the first term different from zero.

A Moving Average of order q: MA(q)

- A moving average process of order q can be written as:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}, \quad t = 1, ..., T$$

  We can calculate the first and second moments to check stationarity and also use the ACF function to Identify the series:

-

$$E(y_t) = 0$$

$$E(y_t)^2 = \sigma^2(1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)$$

$$E(y_t y_{t-k}) = \begin{cases} \sum\limits_{j=0}^{q} \sigma^2 \theta_j \theta_{j+k} & \text{for } k = 1, 2, ..., q \\ 0 & \text{for } k > q \end{cases}$$

$$\rho_k = \begin{cases} \sum\limits_{j=0}^{q} \sigma^2 \theta_j \theta_{j+k} / \sum\limits_{j=0}^{q} \theta_j^2 & \text{for } k = 1, 2, ..., q \\ 0 & \text{for } k > q \end{cases}$$

- The ACF function of a MA(q) process has the first q terms different from zero and then all equal to zero.

Wold's Decomposition Theorem

Theorem

Every weakly stationary, purely non-deterministic, stochastic process $(x_t - \mu)$ can be written as a linear combination of uncorrelated random variables. The representation is given by:

$$\begin{aligned} (x_t - \mu) &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... \\ &= \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \quad \textit{where } \theta_0 = 1 \end{aligned} \tag{10}$$

Where ε_t is such that

$$E(\varepsilon_t) = 0 \tag{11}$$

$$V(\varepsilon_t) = \sigma^2 \tag{12}$$

$$Cov(\varepsilon_t, \varepsilon_{t-k}) = 0 \textit{ for all } k. \tag{13}$$

- NB, the MA(1) and the MA(q) are special cases of the theorem and therefore stationary.

Modelling a Time Series as an Autoregressive Model

An autoregressive process of order p, is written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t, \quad t = 1, ..., T$$

This will be denoted $y_t \tilde{} \mathrm{AR}(p)$

- We will check

1. Under which conditions this processes are stationary
2. Whether we can use the ACF function to identify AR(p) processes.

- Consider the following first order autorregressive process, AR(1),

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \qquad t = 1, \ldots\ldots, T.$$

Notice that if this relationship is valid for time t, it should also be valid for time $t-1$, that is

Example

$$y_{t-1} = \phi_1 y_{t-2} + \varepsilon_{t-1}$$

Substituting, we get:

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

and Iterating:

$$y_t = \phi_1^j y_{t-j} + \phi_1^{j-1} \varepsilon_{t-(j-1)} + \phi_1^{j-2} \varepsilon_{t-(j-2)} + \ldots + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

- If $|\phi| < 1$, the deterministic component of y_t is negligible if j is large enough

$$y_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} \tag{14}$$

- An $AR(1)$ may be written as a $MA(\infty)$ in which the coefficient of ε_{t-j} is ϕ_1^j.

AR(1) stationary process

Lemma

When $|\phi| < 1$, then the $AR\left(1\right)$ satisfies the conditions for stationarity

Proof.

The mean exists and does not depend on t

$$E(y_t) = E(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}) = 0$$

The Variance exists and does not depend on time

$$\begin{aligned} V(y_t) &= V(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}) \underset{E(y_t)=0}{=} E(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j})^2 \\ &= E(\sum_{j=0}^{\infty} \phi_1^{2j} \varepsilon_{t-j}^2) = \sigma^2 \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{\sigma^2}{(1-\phi_1^2)} \end{aligned}$$

□

Proof.

(continues)
The Autocovariances exist and only depend on lag k

$$\begin{aligned} E(y_t y_{t-k}) &= \phi_1 E(y_{t-1} y_{t-k}) + E(\varepsilon_t y_{t-k}) \\ \gamma_k &= \phi_1 \gamma_{k-1} + E(\varepsilon_t y_{t-k}) \end{aligned}$$

Notice that:

$$E(\varepsilon_t y_{t-k}) = E[\varepsilon_t(\phi_1^{j-1}\varepsilon_{t-k-(j-1)} + ... + \phi_1 \varepsilon_{t-(k-1)} + \varepsilon_{t-k})]$$

Given ε_t is white noise:

$$\gamma_k = \phi_1 \gamma_{k-1}$$

Thus

$$\rho_k = \phi_1 \rho_{k-1} \tag{15}$$

□

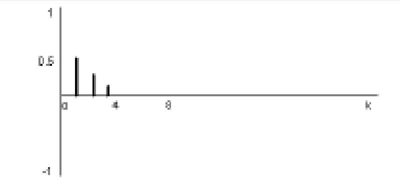
Comments

- Whenever the process is stationary the autocorrelation function declines exponentially:

$$\rho_k = \phi_1^k \rho_0.$$

Example

For $\phi = 0.5$:



General conditions for stationarity: The use of the lag operator

Definition

The lag operator, L, is defined by the transformation

$$Ly_t = y_{t-1}$$

- Notice that the lag operator may also be applied to y_{t-1} yielding

$$Ly_{t-1} = y_{t-2}$$

- Substitution, yields:

$$L^k y_t = y_{t-k} \qquad \text{for } k \geq 0$$

- The lag operator can be manipulated in a similar way to any algebraic quantity.

The lag operator

Example

Let us reproduce the $MA(\infty)$ *representation of the* $AR(1)$ *process:*

$$y_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}$$

where we assume $|\phi_1| < 1$ *. Using* L *this may be written as:*

$$y_t = \sum_{j=0}^{\infty} (\phi_1 L)^j \varepsilon_t = \varepsilon_t / (1 - \phi_1 L)$$

This can be rearranged in the following way

$$(1 - \phi_1 L) y_t = \varepsilon_t$$

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

Autoregressive processes using Lag Operators

An AR(p) process may be written as,

$$(1-\phi_1 L-\phi_2 L^2-...-\phi_p L^p)y_t=\varepsilon_t, \qquad\qquad t=1,....,T$$

or

$$\phi(L)y_t=\varepsilon_t, \qquad\qquad t=1,....,T$$

where $\phi(L)=(1-\phi_1 L-\phi_2 L^2-...-\phi_p L^p)$.

Stationarity Conditions for Autoregressive processes

- The Stationarity condition may be expressed in terms of the roots of the polynomial of order p in L. Consider an $AR(1)$

$$(1-\phi_1 L)y_t = \varepsilon_t, \qquad t = 1, ..., T.$$

When we consider the root of $(1-\phi_1 L) = 0$, that is $L = 1/\phi_1$, we see it is greater than 1 (in absolute value) whenever $|\phi_1| < 1$.Then the process seems to be stationay whenever $L > 1$.

- In general: an $AR(p)$ is said to be stationary when all the roots of the polynomial $(1-\phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$ lie outside the unit circle.

Moving Average processes using Lag Operators

Consider the following MA(q) process

$$y_t = (1 + \theta_1 L + \theta_2 L^2 + ... + \phi_q L^q)\varepsilon_t \qquad t = 1, ..., T$$

or

$$y_t = \theta(L)\varepsilon_t$$

where $\theta(L) = (1 + \theta_1 L + \theta_2 L^2 + ... + \phi_q L^q)$.

Remark A $MA(q)$ process is said to be **invertible** if all the roots of the polynomial $(1 + \theta_1 L + \theta_2 L^2 + ... + \phi_q L^q)$ lie outside the unit circle.

Autoregressive Moving Average Process

Definition

An autoregressive moving average process of order (p, q), denoted as ARMA(p, q) is written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

or

$$\Phi(L) y_t = \Theta(L) \varepsilon_t$$

Autoregressive Moving Average Process

- AR(p) and MA(q) are special cases of the ARMA(p, q) process.
- The stationarity of an ARMA process depends solely on its AR part
  - ARMA is stationary if $\Phi(L) = 0$ lies outside the unit circle
- Its invertibility depends only on its MA part
  - ARMA is invertible if $\Theta(L) = 0$ lies outside the unit circle
- If both conditions hold, ARMA can be written as $AR(\infty)$ or $MA(\infty)$

Example

ARMA (1,1)

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Autocovariance function

$$\begin{aligned}\gamma_k &= E(y_t y_{t-k}) = \phi_1 E(y_{t-1} y_{t-k}) + E(\varepsilon_t y_{t-k}) + \theta_1 E(\varepsilon_{t-1} y_{t-k}) \\ &= \phi_1 \gamma_{k-1} + E(\varepsilon_t y_{t-k}) + \theta_1 E(\varepsilon_{t-1} y_{t-k})\end{aligned}$$

Example

For $k = 0$

$$\gamma_0 = \phi_1\gamma_1 + E(\varepsilon_t(\phi_1 y_{t-1} + \varepsilon_t + \theta_1\varepsilon_{t-1})) + \theta_1 E(\varepsilon_{t-1}(\phi_1 y_{t-1} + \varepsilon_t + \phi_1\varepsilon_{t-1})$$

where

$$E(\varepsilon_t(\phi_1 y_{t-1} + \varepsilon_t + \theta_1\varepsilon_{t-1})) = \sigma^2$$

$$E(\varepsilon_{t-1}(\phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1\varepsilon_{t-2}) + \varepsilon_t + \theta_1\varepsilon_{t-1})) = (\phi_1 + \theta_1)\sigma^2$$

Then:

$$\gamma_0 = \phi_1\gamma_1 + \sigma^2 + \theta_1(\phi_1 + \theta_1)\sigma^2$$

Example

When $k = 1$

$$\gamma_1 = \phi_1 \gamma_0 + \theta_1 E(\varepsilon_{t-1}(\phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}))$$

Thus,

$$\gamma_1 = \phi_1 \gamma_0 + \theta_1 \sigma^2$$

For $k \geq 2$

$$\gamma_k = \phi_1 \gamma_{k-1}$$

Partial Autocorrelations

- The ACF can be used to check whether a process is a MA(q), and to determine q.
- The ACF for an AR declines exponentially
- However, we cannot guess the order of the AR from the plot of the ACF
- Thus, we need to use the Partial Autocorrelation Function, PACF, to determine the order of the autoregressive process.

Partial Autocorrelation

- Consider an AR(1) process, the correlation between y_t and y_{t-2} comes through the correlation each other has with y_{t-1}
- The k^{th} partial autocorrelation, $\phi_k = \phi_{kk}$, function measures the correlation which comes only from the direct effect of the k^{th} lag.
- To find how many partial autocorrelations are different from zero, we use the Yule Walker equations

Yule Walker equations

For an autoregressive process of order p the Yule-Walker equations are given by the following recursion formulae;

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_3 \rho_{k-3} + ... + \phi_p \rho_{k-p} \quad \textit{for } k = 1, ..., p.$$

Then we just need to set $k = p$ or $\phi_p = \phi_{kk}$ and solve the following system of equations:

$$\rho_k = \phi_{11} \rho_{k-1} + \phi_{22} \rho_{k-2} + \phi_{33} \rho_{k-3} + ... + \phi_{kk}.$$

Yule Walker equations

Giving values $k = 1....k$:

$$
\begin{aligned}
\rho_1 &= \phi_{11}\rho_0 + \phi_{22}\rho_1 + \phi_{33}\rho_2 + ... + \phi_{kk}\rho_{k-1} \quad \textit{for } k = 1 \\
\rho_2 &= \phi_{11}\rho_1 + \phi_{22}\rho_0 + \phi_{33}\rho_1 + ... + \phi_{kk}\rho_{k-2} \quad \textit{for } k = 2 \\
& ... \\
\rho_k &= \phi_{11}\rho_{k-1} + \phi_{22}\rho_{k-2} + \phi_{33}\rho_{k-3} + ... + \phi_{kk}\rho_0 \quad \textit{for } k = k
\end{aligned}
$$

- It is a k by k system which can be solved for $\phi_{ii}(\rho_1, ..., \rho_k)$ for $i = 1, 2, ...k$.

Yule Walker equations

- The PACF of an AR(p) has the first p terms different from zero and the rest equal to zero. The empirical methodology consists in finding which ϕ_{kk} are not significantly different from zero.

Example

AR (1)

$$\rho_k = \phi_1 \rho_{k-1}$$

Then,

$$\rho_k = \phi_{11} \rho_{k-1} \qquad \sin ce \ p = k = 1,$$

$$\begin{aligned} \rho_1 &= \phi_{11} \rho_0 \qquad \textit{for } k = 1. \\ & \textit{or} \\ \rho_1 &= \phi_{11} \end{aligned}$$

Example

AR (2)

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

Thus,

$$\rho_k = \phi_{11} \rho_{k-1} + \phi_{22} \rho_{k-2} \text{ since } p = k = 2,$$

Giving values to k

$$\begin{aligned} \rho_1 &= \phi_{11} + \phi_{22} \rho_1 && \text{for } k = 1, \\ \rho_2 &= \phi_{11} \rho_1 + \phi_{22} && \text{for } k = 2, \end{aligned}$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.$$

Identifying the series through ACF/PACF

- *For an AR(p) Process:* The ACF declines exponentially and the PACF is zero for lags greater than p
- *For a MA(p) Process:* The ACF is zero for lags greater than q and the PACF declines exponentially

Box Jenkins Methodology:

Using *sample* information, we might calculate sample ACF and PACF to try to identify the right model.

1. Transform the data, if necessary, so that the assumption of covariance stationarity is a reasonable one.
2. Make an initial guess of small values of p and q for an ARMA(p, q) model that might describe the transformed series.
3. Estimate the parameters in $\phi(L)$ and $\theta(L)$
4. Perform diagnostic analysis to confirm that the model is indeed consistent with the observed features of the data.

What's next:

- Empirical properties of sample analogs
- Estimation

Correlogram

- Basic tool
- An inspection of the correlogram may lead to the conclusion that the series is random, or that exhibits a pattern of serial correlation that which perhaps can be modeled by a particular stochastic process

Sample ACF and PACF

It can be proved that asymptotically:

$$
\begin{aligned}
\widehat{\rho_i} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \rho_i = 0 \\
\widehat{\phi_{ii}} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \phi_{ii} = 0
\end{aligned}
$$

where T is the sample size

Significance test

- In order to identify using ACF we should test whether the different parameters ρ_k are different from zero.
- Box Pierce Q- Statistic

$$Q = T \sum_{i=1}^{k} \widehat{\rho}_i^{\,2} \sim \chi^2(k) \text{ under } H_0 : \rho_1 = ... = \rho_k = 0$$

If I don't reject H_0, then, $\{\varepsilon_t\}$ is a white noise. This test has low power, even in large samples.

- Corrected version

$$Q^* = T(T+2)\sum_{i=1}^{k}\frac{\widehat{\rho}_i^{\,2}}{T-i}\,\tilde{}\,\chi^2(k) \text{ under } H_0 : \rho_1 = ... = \rho_k = 0$$

- When the k autocorrelations are calculated for an ARMA(p,q) model, we lose degrees of freedom. Thus

$$Q^*\,\tilde{}\,\chi^2(k-p-q)$$

Significance test

- When we use identifying tools such as :

$$\begin{aligned}\widehat{\rho_i} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \rho_i = 0 \\ \widehat{\phi_{ii}} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \phi_{ii} = 0\end{aligned}$$

  we find that these tools won't tell us neither whether the preferred model is mispecified, nor what to do when two different models seem to be equally valid.
- We will need to estimate these models.

Maximum Likelihood Estimation

- Usually when we estimate ARMA(p, q) models we evaluate the conditional maximum likelihood.
  - Assume that the first $max(p, q)$observations are known.
- Asymptotically, CMLE is equivalent to the MLE and is easier to calculate

Example

AR(1)

$$f(y_T, y_{T-1}, ..., y_2 | y_1, \phi_1, \sigma^2) = \prod_{i=2}^{T} f(y_i | y_{i-1}, \phi_1, \sigma^2)$$

where

$$f(y_i | y_{i-1}, \phi_1, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_t - \phi_1 y_{t-1})^2}{2\sigma^2}}$$

The objective is to maximize:

$$\mathcal{L} = -(T-1) log(2\pi)^{.5} - (T-1)\log\sigma - \frac{\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2}{2\sigma^2}$$

Model Selection Criteria

- Akaike Criteria (AIC):

$$AIC(p, q) = \log \hat{\sigma}^2 + 2(p + q)\,T^{-1}$$

- Shwartz Criteria (SC):

$$SC(p, q) = \log \hat{\sigma}^2 + (p + q)\,T^{-1}\log(T)$$

- The idea is to weigh likelihood and number of lags.
- It is best when we compare models with the same number of lags
- Overparametrized models have less forecast ability

- We choose the model with the lower AIC/SC
- SC picks more parsimonious models
- SC has better large sample properties, while AIC picks overparametrized models when $T \longrightarrow \infty$
- For small samples, AIC may outperform the others
- If they pick different models, as SC picks the most parsimonious you should check the residuals and with AIC you should check for the significance of the parameters

Forecast

- In practice, we only have an approximation to the "true GDP"
  - identification and estimation errors

Minimum Mean Square Error Forecast

We start with conditional mean

- Serves as benchmark
- It generates the forecast with the minimum mean square error
  ie, if the model is correct, there is no other statistic that produces smaller forecast errors

Given observations up to, and including y_T :

$$\widehat{y}_{T+l|T} = E(y_{T+l}|I_T)$$

where I_T is the information set up to T

Notice that for any other forecast $\widetilde{y}_{T+l|T}$, the forecast error maybe divided in two:

$$y_{T+l} - \widetilde{y}_{T+l|T} = [y_{T+l} - \widehat{y}_{T+l|T}] + [\widehat{y}_{T+l|T} - \widetilde{y}_{T+l|T}]$$

Squaring terms and conditional on T:

$$MSE(\widetilde{y}_{T+l|T}) = Var(y_{T+l}) + [\widetilde{y}_{T+l|T} - E(y_{t+l}|I_T)]^2$$

One step ahead forecast

Consider the following stationary and invertible ARMA (p, q):

$$\begin{aligned} y_{T+1} = \quad & \phi_1 y_T + \phi_2 y_{T-1} + \ldots.. + \phi_p y_{T-p+1} + \\ & \varepsilon_{T+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} + \ldots + \theta_q \varepsilon_{T-q+1} \end{aligned}$$

Then,

$$\begin{aligned} \widehat{y}_{t+1|T} = \quad & \phi_1 y_T + \phi_2 y_{T-1} + \ldots. + \phi_p y_{T-p+1} + \\ & \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} + \ldots + \theta_q \varepsilon_{T-q+1} \end{aligned}$$

We need values for $\{\varepsilon_t\}$ to obtain numerical values for the forecast

Example

AR(1)

$$y_{T+l} = \phi_1 y_{T+l-1} + \varepsilon_{T+l} \qquad \textit{at time } T+l$$

$$\widehat{y}_{T+l|T} = \phi_1 \widehat{y}_{T+l-1|T} \qquad l = 1, 2, ...$$

The initial value is $\widehat{y}_{T|T} = y_T$. *Then:*

$$\widehat{y}_{T+l|T} = \phi_1^l y_T$$

Example

(continues)
Let's calculate the forecasting error:

$$
\begin{aligned}
y_{T+l} - \widehat{y}_{T+l|T} &= \phi_1 y_{T+l-1} + \varepsilon_{T+l} - \phi_1^l y_T \\
&= \begin{array}{c} \phi_1^l y_T + \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + \\ \phi_1^2 \varepsilon_{T+l-2} + ... + \phi_1^{l-1} \varepsilon_{T+l} - \phi_1^l y_T \end{array}
\end{aligned}
$$

The forecasting error variance is given by:

$$
\begin{aligned}
V(y_{T+l} - \widehat{y}_{T+l|T}) &= V(\varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + \phi_1^2 \varepsilon_{T+l-2} + ... + \phi_1^{l-1} \varepsilon_{T+l}) \\
&= (1 + \phi_1^2 + \phi_1^4 + ... + \phi_1^{2(l-1)}) \sigma^2
\end{aligned}
$$

Example

MA(1)
On $T+1$, the equation for MA(1) is given by:

$$y_{T+1} = \varepsilon_{T+1} + \theta_1 \varepsilon_T$$

Thus, in general,

$$\widehat{y}_{T+l|T} = \hat{\varepsilon}_{T+l|T} + \theta_1 \hat{\varepsilon}_{T+l-1|T}$$

$$\begin{aligned} \widehat{y}_{T+l|T} &= \theta_1 \varepsilon_T \qquad \textit{for } l = 1 \\ &= 0 \qquad \textit{for } l > 1. \end{aligned}$$

Example

(*continues*)
The variance of the forecast error for a MA(1) is

$$\begin{aligned} V(y_{T+l} - \widehat{y}_{T+l|T}) &= \sigma^2 & \text{for } l = 1 \\ &= (1+\theta_1^2)\sigma^2 & \text{for } l > 1 \end{aligned}$$

Example

ARMA(1,1)

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\begin{aligned}
\widehat{y}_{T+l|T} &= \phi_1 y_T + \theta_1 \varepsilon_T && \text{for } l = 1 \\
&= \phi_1 \widehat{y}_{T+l-1|T} && \text{for } l > 1 \\
&= \phi_1^l y_T + \phi_1^{l-1} \theta_1 \varepsilon_T
\end{aligned}$$

Measuring the Accuracy of Forecasts

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{l}\sum_{i=T+1}^{T+l}(\widehat{Y}_{i|T} - Y_i)^2}$$

Mean Absolute Error

$$MAE = \frac{1}{l}\sum_{i=T+1}^{T+l}|\widehat{Y}_{i|T} - Y_i|$$

Remark Which indicator should be used depends of the purpose of the forecasting exercise. The RMSE will penalize big errors more than the MAE measure.