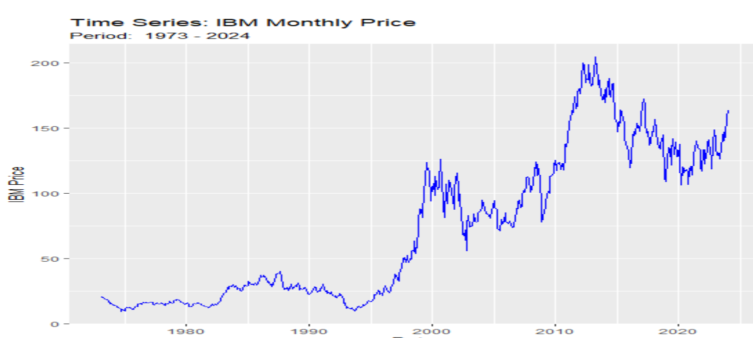
Lecture 13
Time Series:
Stationarity, AR(p) & MA(q)

Time Series: Introduction

- A time series y_t is a process observed in sequence over time, t = 1, ..., T $\Rightarrow Y_t = \{y_1, y_2, y_3, ..., y_T\}$.

Examples: IBM monthly stock prices from 1973:January till 2024:September (plot below); or USD/GBP daily exchange rates from February 15, 1923 to March 19, 1938.



Time Series: Introduction

Examples (continuation): Different ways to do the plot in R:

- Using plot.ts, creating a time series object in R:

```
# the function ts creates a timeseries object, start = 1973,1 (start of sample), frequency =
12(=monthly)
ts_ibm <- ts(x_ibm, start=c(1973,1), frequency=12)

plot.ts(ts_ibm,xlab="Time",ylab="IBM price", main="Time Series: IBM Stock Price")
```

- Using R package ggplot2

```
x_ibm <- SFX_da$IBM
x_date <- as.Date(SFX_da$Date, "%m/%d/%Y")
df <- data.frame(x_date, x_ibm)
ggplot(df, aes(x = x_date, y = x_ibm)) +
  geom_line(color="blue") +
  labs(x = "Date", y = "IBM Price", col = "blue", title = "Time Series: IBM Monthly Price",
       subtitle = "Period:  1973 - 2024")
```

Time Series: Introduction – Categories

- Usually, time series models are separated into two categories:
  - **Univariate** ($y_t \in R$, it is a scalar)

    Example: We are interested in the behavior of IBM stock prices as function of its past.

    ⇒ Primary model: Autoregressions (ARs).

  - **Multivariate** ($y_t \in R^m$, it is a vector-valued)

    Example: We are interested in the joint behavior of IBM returns, r_{IBM}, & bond yields, b_{IBM}, as function of their past

$$y_t = \begin{bmatrix} r_{IBM,t} \\ b_{IBM,t} \end{bmatrix}$$

    ⇒ Primary model: Vector autoregressions (VARs).

Time Series: Introduction – Dependence

• Given the sequential nature of y_t, we expect y_t & y_{t-1} to be dependent. This is the main feature of time series: **dependence**. It creates statistical problems.

• In classical statistics, we usually assume we observe several *i.i.d.* realizations of y_t. We use $\bar{y}$ to estimate the mean.

• With several independent realizations we are able to sample over the entire probability space and obtain a "good" –i.e., consistent or close to the population mean– estimator of the mean.

• But, if the samples are highly dependent, then it is likely that y_t is concentrated over a small part of the probability space. Then, the sample mean will not converge to the mean as the sample size grows.

Time Series: Introduction – Dependence

Technical note: With dependent observations, the classical results (based on LLN & CLT) are not to valid.

• We need new conditions in the DGP to make sure the sample moments (mean, variance, etc.) are good estimators population moments. The new assumptions and tools are needed: **stationarity**, **ergodicity**, CLT for martingale difference sequences (**MDS CLT**).

Roughly speaking, **stationarity** requires constant moments for y_t; **ergodicity** requires that the dependence is short-lived, eventually y_t has only a small influence on y_{t+k}, when k is relatively large.

Ergodicity describes a situation where the expectation of a random variable can be replaced by the time series expectation.

Time Series: Introduction – Dependence

An **MDS** is a discrete-time martingale with mean zero. In particular, its increments, ε_t's, are uncorrelated with any function of the available dataset at time t. To these ε_t's we will apply a CLT.

• The amount of dependence in y_t determines the 'quality' of the estimator. There are several ways to measure dependence. The most common measure: **Covariance**.

$$\text{Cov}(y_t, y_{t+k}) = E[(y_{t_t} - \mu)(y_{t+k} - \mu)]$$

Note: When μ = 0, then $\text{Cov}(y_t, y_{t+k}) = E[y_t\ y_{t+k}]$

Time Series: Introduction – Forecasting

• In a time series model, we describe how y_t depends on past y_t's. That is, the information set is $I_t = \{y_{t-1}, y_{t-2}, y_{t-3}, ...\}$

• The purpose of building a time series model: Forecasting.

• We estimate time series models to forecast out-of-sample. For example, the *l-step ahead* forecast: $\hat{y}_{T+l} = \text{E}_t[y_{t+l}\,|I_t]$.

Historical Note: In the 1970s it was found that very simple time series models out-forecasted very sophisticated (big) economic models.

This finding represented a big shock to the big multivariate models that were very popular then. It forced a re-evaluation of these big models.

Time Series: Introduction – White Noise

• In general, we assume the error term, ε_t, is uncorrelated with everything, with mean 0 and constant variance, σ^2. We call a process like this a **white noise (WN) process**.

• We denote a WN process as

$$\varepsilon_t \sim \text{WN}(0, \sigma^2)$$

• White noise is the basic building block of all time series. It can be written as simple function of a WN(0, 1) process:

$$z_t = \sigma\, u_t, \qquad u_t \sim i.i.d.\ \text{WN}(0, 1) \qquad \Rightarrow z_t \sim \text{WN}(0, \sigma^2)$$

• The z_t's are random shocks, with no dependence over time, representing unpredictable events. It represents a model of news.

Time Series: Introduction – Conditionality

• We make a key distinction: *Conditional* & *Unconditional* moments. In time series we model the conditional mean as a function of its past, for example in an AR(1) process, we have:

$$y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t.$$

Then, the **conditional mean** forecast at time t, conditioning on information at time I_{t-1}, is:

$$\text{E}_t[y_t \,|\, I_{t-1}] = \text{E}_t[y_t] = \alpha + \beta\, y_{t-1}$$

Notice that the **unconditional mean**, μ, is given by:

$$\text{E}[y_t] = \alpha + \beta\ \text{E}[y_{t-1}] = \frac{\alpha}{1-\beta} = \mu = \text{constant} \qquad (\beta \neq 1)$$

The conditional mean is time varying; the unconditional mean is not!

Key distinction: Conditional vs. Unconditional moments.

Time Series: Introduction – AR and MA models

• Two popular models for $E_t[y_t \mid I_t]$:

– An **autoregressive** (**AR**) **process** models $E_t[y_t \mid I_{t-1}]$ with lagged dependent variables:

$$E_t[y_t|I_t] = f(y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-p})$$

Example: AR(1) process, $y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t$.

– A **moving average** (**MA**) **process** models $E_t[y_t|I_t]$ with lagged errors, ε_t:

$$E_t[y_t \mid I_t] = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \dots, \varepsilon_{t-q})$$

Example: MA(1) process, $y_t = \mu + \theta_1\, \varepsilon_{t-1} + \varepsilon_t$

• There is a third model, **ARMA**, that combines lagged dependent variables and lagged errors.

Time Series: Introduction – Forecasting (again)

• We want to select an appropriate time series model to forecast y_t. In this class, we will use linear models, with choices: AR(p), MA(q) or ARMA(p, q).

• Steps for forecasting:

(1) Identify the appropriate model. That is, determine p, q.

(2) Estimate the model.

(3) Test the model.

(4) Forecast.

• In this lecture, we go over the statistical theory (stationarity, ergodicity), the main models (AR, MA & ARMA) and tools that will help us describe and identify a proper model.

CLM Revisited: Time Series Implications

• With autocorrelated data, we get dependent observations. For example, with autocorrelated errors:

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t,$$

the independence assumption is violated. The LLN and the CLT cannot be easily applied in this context. We need new tools.

• We introduce the concepts of **stationarity** and **ergodicity**. The ergodic theorem will give us a counterpart to the LLN.

To get asymptotic distributions, we also need a CLT for dependent variables, using new technical concepts: mixing and stationarity. Or we can rely on a new CLT: The **MDS CLT**.

• We will not cover these technical points in detail.

Time Series – Stationarity

• Consider the joint probability distribution of the collection of RVs:

$$F(y_{t_1}, y_{t_2}, \dots, y_{t_T}) = F(Y_{t_1} \le y_{t_1}, Y_{t_2} \le y_{t_2}, \dots, Y_{t_T} \le y_{t_T})$$

To do statistical analysis with dependent observations, we need extra assumptions. We need some form of invariance on the structure of the time series.

If the distribution F is changing with every observation, estimation and inference become very difficult.

• Stationarity is an invariant property: The statistical characteristics of the time series do not change over time.

• There different definitions of stationarity, they differ in how strong is the invariance of the distribution over time.

Time Series – Stationarity

- We say that a process is **stationary** of

1st order if $F(y_{t_1}) = F(y_{t_{1+k}})$ for any t_1, k

2nd order if $F(y_{t_1}, y_{t_2}) = F(y_{t_{1+k}}, y_{t_{2+k}})$ for any t_1, t_2, k

Nth-order if $F(y_{t_1}, \dots, y_{t_T}) = F(y_{t_{1+k}}, \dots, y_{t_{T+k}})$ for any t_1, ..., t_T, k

- *Nth-order* stationarity is a strong assumption (& difficult to verify in practice). *2nd order* (weak) stationarity is weaker. **Weak stationarity** only considers means & covariances (easier to verify in practice).

- Moments describe a distribution. We calculate moments as usual:

$$\text{E}[Y_t] = \mu$$

$$\text{Var}(Y_t) = \sigma^2 = E[(Y_t - \mu)^2]$$

$$\text{Cov}(Y_{t_1}, Y_{t_2}) = E[(Y_{t_1} - \mu)(Y_{t_2} - \mu)] = \gamma(t_1 - t_2)$$

Time Series – Stationarity & Autocovariances

- $\text{Cov}(Y_{t_1}, Y_{t_2}) = \gamma(t_1 - t_2)$ is called the **auto-covariance function**. It measures how y_t, measured at time t_1, and y_t, measured at time t_2, covary.

Notes: $\gamma(t_1 - t_2)$ is a function of $k = t_1 - t_2$
$\gamma(0)$ is the variance.

- The autocovariance function is symmetric. That is,

$$\gamma(t_1 - t_2) = \text{Cov}(Y_{t_1}, Y_{t_2}) = \text{Cov}(Y_{t_2}, Y_{t_1}) = \gamma(t_2 - t_1)$$

$$\Rightarrow \gamma(k) = \gamma(-k)$$

- Autocovariances are unit dependent. We have different values if we calculate the autocovariance for IBM returns in % or in decimal terms.

Remark: The autocovariance measures the (linear) dependence between two Y_t's separated by k periods.

Time Series – Stationarity & Autocorrelations

- From the autocovariances, we derive the **autocorrelations**:

$$\text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2}) = \frac{\gamma(t_1 - t_2)}{\sigma_{t_1}\sigma_{t_2}} = \frac{\gamma(t_1 - t_2)}{\gamma(0)}$$

the last step takes assumes: $\sigma_{t_1} = \sigma_{t_2} = \sqrt{\gamma(0)}$

- $\text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2})$ is called the **auto-correlation function** (**ACF**), –think of it as a function of $k = t_2 - t_1$. The ACF is also symmetric.

- Unlike autocovoriances, autocorrelations are not unit dependent. It is easier to compare dependencies across different time series.

- Stationarity requires all these moments to be independent of time. If the moments are time dependent, we say the series is **non-stationary**.

Time Series – Stationarity & Constant Moments

- For a strictly stationary process (constant moments), we need:

$$\mu_t = \mu$$
$$\sigma_t = \sigma$$

because $F(y_{t_1}) = F(y_{t_{1+k}}) \quad \Rightarrow \quad \mu_{t_1} = \mu_{t_{1+k}} = \mu$

$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$

Then,

$$F(y_{t_1}, y_{t_2}) = F(y_{t_{1+k}}, y_{t_{2+k}}) \Rightarrow \text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$

$$\Rightarrow \rho(t_1, t_2) = \rho(t_{1+k}, t_{2+k})$$

Let $t_1 = t - k$ & $t_2 = t$

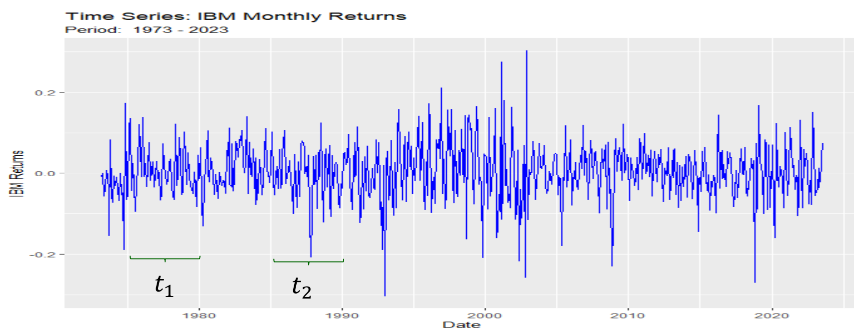
$$\Rightarrow \rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t - k) = \rho(k) = \rho_k$$

The correlation between any two RVs depends on the time difference. Given the symmetry, we have $\rho(k) = \rho(-k)$.

Time Series – Stationarity & Constant Moments

Example: Informally, we check if in any two periods separated by k observations, we have similar means, variances and covariances. That is,

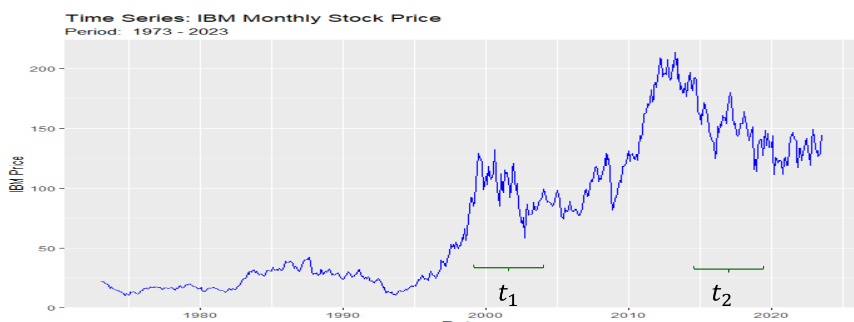
$$\mu_{t_1} = \mu_{t_{1+k}} = \mu$$
$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$
$$\text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$



Time Series – Stationarity & Constant Moments

Example: Informally, we check if in any two periods separated by k observations, we have similar means, variances and covariances. That is,

$$\mu_{t_1} = \mu_{t_{1+k}} = \mu$$
$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$
$$\text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$



Time Series – Weak Stationary

- A **Covariance stationary** process (or *2nd -order weakly stationary*) has:
  - constant mean, μ
  - constant variance, σ^2
  - covariance depends on time difference, k, between two RVs, $\gamma(k)$

That is, Z_t is covariance stationary if:

$$E(Z_t) = \text{constant} = \mu$$
$$\text{Var}(Z_t) = \text{constant} = \sigma^2$$
$$\text{Cov}(Z_{t_1}, Z_{t_2}) = \gamma(k = t_1 - t_2)$$

Remark: Covariance stationarity is only concerned with the covariance of a process, only the mean, variance and covariance are time-invariant.

Time Series – Stationarity: Example

Example: Assume y_t follows an AR(1) process:

$$y_t = \phi\, y_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

- **Mean**

Taking expectations on both side:

$$E[y_t] = \phi\, E[y_{t-1}] + E[\varepsilon_t]$$
$$\mu = \phi\, \mu + 0$$
$$E[y_t] = \mu = 0 \qquad \text{(assuming } \phi \neq 1)$$

- **Variance**

Applying the variance on both side:

$$\text{Var}[y_t] = \gamma(0) = \phi^2\, \text{Var}[y_{t-1}] + \text{Var}[\varepsilon_t]$$
$$\gamma(0) = \phi^2\, \gamma(0) + \sigma^2$$
$$\gamma(0) = \frac{\sigma^2}{1-\phi^2} \qquad \text{(assuming } |\phi| < 1)$$

Time Series – Stationarity: Example

Example (continuation): $y_t = \phi\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{WN}(0, \sigma^2)$

- **Covariance**

$$\begin{aligned}
\gamma(1) &= \text{Cov}[y_t, y_{t-1}] = \text{E}[y_t\, y_{t-1}] = \text{E}[(\phi\, y_{t-1} + \varepsilon_t)\, y_{t-1}] \\
&= \phi\, \text{E}[y_{t-1}\, y_{t-1}] + \text{E}[\varepsilon_t\, y_{t-1}] \\
&= \phi\, \text{E}[{y_{t-1}}^2] \\
&= \phi\, \text{Var}[{y_{t-1}}^2] \\
&= \phi\, \gamma(0)
\end{aligned}$$

$$\begin{aligned}
\gamma(2) &= \text{Cov}[y_t, y_{t-2}] = \text{E}[y_t\, y_{t-2}] = \text{E}[(\phi\, y_{t-1} + \varepsilon_t)\, y_{t-2}] \\
&= \phi\, \text{E}[y_{t-1}\, y_{t-2}] \\
&= \phi\, \text{Cov}[y_t, y_{t-1}] \\
&= \phi\, \gamma(1) \\
&= \phi^2\, \gamma(0)
\end{aligned}$$

$\vdots$

$$\gamma(k) = \text{Cov}[y_t, y_{t-k}] = \phi^k\, \gamma(0)$$

Time Series – Stationarity: Example

Example (continuation): $y_t = \phi\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{WN}(0, \sigma^2)$

- **Covariance**

$$\gamma(k) = \text{Cov}[y_t, y_{t-k}] = \phi^k\, \gamma(0)$$

$\Rightarrow$ If $|\phi| < 1$, y_t process is covariance stationary: mean, variance, and covariance are constant.

Remark: To establish stationarity, we need to impose conditions on the AR parameters. (Conditions are not needed for MA processes.)

Note: From the autocovariance function, we derive ACF:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi^k\, \gamma(0)}{\gamma(0)} = \phi^k$$

If $|\phi| < 1$, autocovariance function & ACF show exponential decay.

Time Series – Non-Stationarity: Example

Example: Assume y_t follows a Random Walk with drift process:

$$y_t = \mu + y_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

Doing backward substitution:

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

$$\Rightarrow y_t = \mu\ t + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$$

- **Mean** & **Variance**

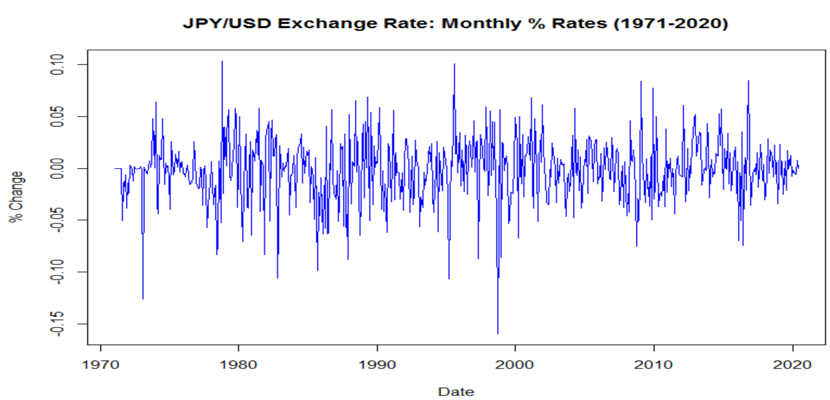
$$E[y_t] = \mu\ t + y_0$$
$$\text{Var}[y_t] = \gamma(0) = \sum_{j=0}^{t-1} \sigma^2 = \sigma^2\, t$$

$\Rightarrow$ the process y_t is non-stationary: moments are time dependent.

Stationary Series: Examples

Examples: Assume $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

$$y_t = 0.08 + \varepsilon_t + 0.4\,\varepsilon_{t-1} \qquad \text{- MA(1) process}$$
$$y_t = 0.13\, y_{t-1} + \varepsilon_t \qquad \text{- AR(1) process}$$

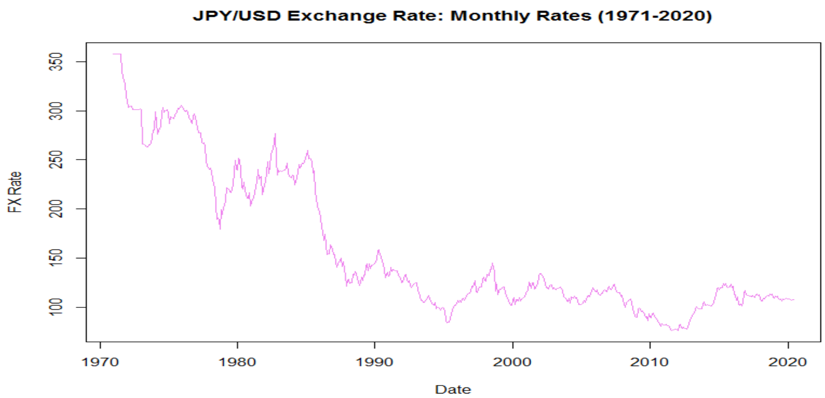


Non-Stationary Series: Examples

Examples: Assume $\varepsilon_t \sim WN(0, \sigma^2)$.

$y_t = \mu\, t + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \varepsilon_t$ - AR(2) with deterministic trend

$y_t = \mu + y_{t-1} + \varepsilon_t$ - Random Walk with drift



Time Series – Stationarity: Remarks

- Main characteristic of time series: Observations are **dependent**.

- If we have non-stationary series (say, mean or variance are changing with each observation), it is not possible to make inferences.

- Stationarity is an invariant property: the statistical characteristics of the time series do not vary over time.

- If IBM is weak stationary, then, the returns of IBM may change month to month or year to year, but the average return and the variance in two equal-length time intervals will be more or less the same.

Time Series – Stationarity (Again)

• In the long run, say 100-200 years, the stationarity assumption may not be realistic. After all, technological change has affected the return of IBM over the long run. But, in the short-run, stationarity seems likely to hold.

• In general, time series analysis is done under the stationarity assumption.

Time Series – Ergodicity of the Mean

• We want to estimate the mean of the process $\{Z_t\}$, $\mu(Z_t)$. But, we need to distinguishing between *ensemble average* (with m observations) and *time average* (with T observations):

- Ensemble Average: $\bar{\bar{z}} = \frac{\sum_{i=1}^{m} z_i}{m}$

- Time Series Average: $\bar{z} = \frac{\sum_{t=1}^{T} z_t}{T}$

Q: Which estimator is the most appropriate?
A: Ensemble Average. But, it is impossible to calculate for a time series. We only observe one Z_t, with dependent observations.

• Q: Under which circumstances we can use the time average (with only one realization of $\{Z_t\}$)? Is the time average an unbiased and consistent estimator of the mean? The **Ergodic Theorem** gives us the answer.

Time Series – Ergodicity

• Intuition behind Ergodicity:
We go to a casino to play a game with 20% return, but on average, one gambler out of 100 goes bankrupt. If 100 gamblers play the game, there is a 99% chance of winning and getting a 20% return. This is the *ensemble scenario*. Suppose that **gambler 35** is the one that goes bankrupt. Gambler 36 is not affected by the bankruptcy of gamble 35.

Suppose now that instead of 100 gamblers you play the game 100 times. This is the *time series* scenario. You win 20% every day until **day 35** when you go bankrupt. There is no day 36 for you (dependence at work!).

Result: The probability of success from the group (ensemble scenario) does not apply to one person (time series scenario).

Ergodicity describes a situation where the ensemble scenario outcome applies to the time series scenario.

Time Series – Ergodicity of the Mean

• Recall the sufficient conditions for consistency of an estimator: the estimator is asymptotically unbiased and its variance asymptotically collapses to zero.

1. Q: Is the time average is asymptotically unbiased? Yes.

$$\mathrm{E}[\bar{z}] = \frac{\sum_{t=1}^{T} E[Z_t]}{T} = \frac{\sum_{t=1}^{T} \mu}{T} = \mu$$

2. Q: Is the variance going to zero as T grows? It depends.

$$\mathrm{var}[\,\bar{z}\,] = \mathrm{var}[\,(z_1 + z_2 + \cdots + z_T)/T] =$$

$$= \frac{\sum_{t=1}^{T}\sum_{s=1}^{T} Cov[Z_t, Z_s]}{T^2} = \frac{\gamma_0}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\rho_{t-s}$$

$$= \frac{\gamma_0}{T^2}\sum_{t=1}^{T}\{\rho_{t-1} + \rho_{t-2} + \cdots + \rho_{t-T}\}$$

$$= \frac{\gamma_0}{T^2}\{(\rho_0 + \rho_1 + \cdots + \rho_{T-1}) + \cdots + (\rho_{T-1} + \rho_{T-2} + \cdots + \rho_0)\}$$

$$= \frac{\gamma_0}{T^2}\sum_{k=1}^{T-1)}(T - |k|)\rho_k = \frac{\gamma_0}{T}\sum_k(1 - \frac{|k|}{T})\rho_k$$

Time Series – Ergodicity of the Mean

$$\text{var}[\bar{z}] = \frac{\gamma_0}{T} \sum_k (1 - \frac{|k|}{T})\rho_k$$

$$\lim_{T\to\infty} \text{var}[\bar{z}] = \lim_{T\to\infty} \frac{\gamma_0}{T} \sum_k (1 - \frac{|k|}{T})\rho_k \overset{?}{\longrightarrow} 0$$

• If the Z_t were uncorrelated, the variance of the time average would be $O(\frac{1}{T})$. Since independent random variables are necessarily uncorrelated (but not vice versa), we have just recovered a form of the LLN for independent data.

Q: How can we make the remaining part, the sum over the upper triangle of the covariance matrix, go to zero as well?
A: We need to impose conditions on ρ_k. Conditions weaker than "they are all zero;" but, strong enough to exclude the sequence of identical copies.

Time Series – Ergodicity of the Mean

• We use two inequalities to put upper bounds on the variance of the time average:

$$\sum_{t=1}^{T-1} \sum_{k=1}^{T-t} \rho_K \le \sum_{t=1}^{T-1} \sum_{k=1}^{T-1} |\rho_K| \le \sum_{t=1}^{\infty} \sum_{k=1}^{\infty} |\rho_K|$$

Covariances can be negative, so we upper-bound the sum of the actual covariances by the sum of their magnitudes. Then, we extend the inner sum so it covers *all* lags. This might of course be infinite (sequence-of-identical-copies).

• **Definition**: A covariance-stationary process is *ergodic* for the mean if

$$\text{plim } \bar{z} = E[Z_t] = \mu$$

Ergodicity Theorem: Then, a sufficient condition for ergodicity for the mean is

$$\rho_k \to 0, \text{ as } k \to \infty$$

Time Series – Ergodicity of 2^{nd} Moments

• A sufficient condition to ensure ergodicity for second moments is:

$$\sum_{k=1}^{T-t} |\rho_K| < \infty$$

A process which is ergodic in the first and second moments is usually referred as **ergodic in the wide sense**.

• **Ergodicity under Gaussian Distribution**
If $\{Z_t\}$is a stationary Gaussian process, $\sum_{k=1}^{T-t} |\rho_K| < \infty$

is sufficient to ensure ergodicity for all moments.

Note: Recall that only the first two moments are needed to describe the normal distribution.

Time Series – Ergodicity – Theorems

• We state two essential theorems to the analysis of stationary time series. Difficult to prove in general.

Theorem I
If y_t is strictly stationary & ergodic and $x_t = f(y_t, y_{t-1}, y_{t-2}, y_{t-3},$ ....) is a RV, then x_t is strictly stationary and ergodic.

Theorem II (Ergodic Theorem)
If y_t is strictly stationary & ergodic and E[y_t] < ∞; then as $T \rightarrow \infty$;

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} y_t \xrightarrow{p} \text{E}[y_t]$$

• These results allow us to consistently estimate parameters using time-series moments.

Time Series - MDS

• **Definition**: ε_t is a martingale difference sequence (MDS) if

$$E[\varepsilon_t \,|\, I_{t-1}] = 0.$$

• Regression errors are naturally a MDS. Some time-series processes may be a MDS as a consequence of optimizing behaviour. For example, most asset pricing models imply that asset returns should be the sum of a constant plus a MDS.

• Useful property: ε_t is uncorrelated with any function of the lagged information I_{t-1}. Then, for $k > 0$ $\Rightarrow E[y_{t-k}\, \varepsilon_t] = 0$.

Time Series – MDS CLT

Theorem (MDS CLT)

If u_t is a strictly stationary and ergodic MDS and $E(u_t u_t') = \mathbf{\Omega} < \infty$; then as $T \to \infty$;

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} u_t \xrightarrow{d} N(0, \Omega)$$

• Application: Let $x_t = \{y_1, y_2, y_3, ..., y_T\}$, a vector of lagged y_t's. Then $(x_t \varepsilon_t)$ is a MDS. We can apply the MDS CLT Theorem. Then,

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \boldsymbol{x}_t' \, \varepsilon_t \xrightarrow{d} N(0, \Omega) \qquad \Omega = E[\boldsymbol{x}_t \boldsymbol{x}_t' \, \varepsilon_t^2]$$

• Like in the derivation of asymptotic distribution of OLS, the above result is the key to establish the asymptotic distribution in a time series context.

Time Series – Lag Operator

Define the operator L as

$$L^k\, z_t = z_{t-k}.$$

• It is usually called *Lag operator*. But it can produce lagged or forward variables (for negative values of k). For example:

$$L^{-3}\, z_t = z_{t+3}.$$

• Also note that if c is a constant $\Rightarrow L\, c = c$.

• Sometimes the notation for L when working as a lag operator is B *(backshift operator),* and when working as a forward operator is F.

• Important application: Differencing

$$\Delta\, z_t = (1 - L)\, z_t = z_t - z_{t-1}.$$
$$\Delta^2\, z_t = (1 - L)^2\, z_t = z_t - 2z_{t-1} + z_{t-2}.$$

Time Series – Useful Result: Geometric Series

• The function $f(x) = (1 - x)^{-1}$ can be written as an infinite geometric series (use a Maclaurin series around c =0):

$$f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots = \textstyle\sum_{n=0}^{\infty} x^n$$

• If we multiply $f(x)$ by a constant, a:

$$\textstyle\sum_{n=0}^{\infty} a x^n = \frac{a}{1-x} \rightarrow \quad \sum_{n=1}^{\infty} a x^n = a\left(\frac{1}{1-x} - 1\right)$$

Example: In Finance we have many applications of the above results.
- A stock price, P, equals the discounted some of all futures dividends. Assume dividends are constant, d, and the discount rate is r. Then:

$$P_t = \textstyle\sum_{t=1}^{\infty} \frac{d}{(1+r)^t} = d\left(\frac{1}{1-\frac{1}{1+r}} - 1\right) = d\left(\frac{1}{\frac{1+r-1}{1+r}} - 1\right) = \frac{d}{r}$$

where $x = \frac{1}{1+r}$

Time Series – Useful Result: Application

• We will use this result when, under certain conditions, we invert a lag polynomial (say, $\theta(L)$) to convert an AR (MA) process into an infinite MA (AR) process.

Example: Suppose we have an MA(1) process:

$$y_t = \mu + \theta_1\, \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t \qquad - \theta(L) = (1 + \theta_1 L)$$

Recall,

$$f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots = \sum_{n=0}^{\infty} x^n$$

Let $x = -\theta_1 L$. Then, assuming that $\theta(L)^{-1}$ is well defined,

$$\theta(L)^{-1} = \frac{1}{1-(-\theta_1 L)} = 1 + (-\theta_1 L) + (-\theta_1 L)^2 + (-\theta_1 L)^3 + (-\theta_1 L)^4 + \ldots$$

$$= \sum_{n=0}^{\infty} (-\theta_1 L)^n = 1 - \theta_1 L + \theta_1{}^2 L^2 - \theta_1{}^3 L^3 + \theta_1{}^4 L^4 + \cdots$$

Time Series – Useful Result: Application

Example (continuation):

$$\theta(L)^{-1} = \sum_{n=0}^{\infty} (-\theta_1 L)^n = 1 - \theta_1 L + \theta_1{}^2 L^2 - \theta_1{}^3 L^3 + \theta_1{}^4 L^4 + \cdots$$

Now, we multiply $\theta(L)^{-1}$ on both sides of the MA process

$$y_t = \mu + \theta(L)\, \varepsilon_t.$$

Then,

$$\theta(L)^{-1} y_t = \theta(L)^{-1}\mu + \theta(L)^{-1}\theta(L)\, \varepsilon_t = \mu^* + \varepsilon_t$$

$$\theta(L)^{-1} y_t = y_t - \theta_1 y_{t-1} + \theta_1^2\, y_{t-2} - \theta_1^3\, y_{t-3} + \theta_1^4\, y_{t-4} + \cdots$$
$$= \mu^* + \varepsilon_t$$

Then, solving for y_t:

$$y_t = \mu* + \theta_1 y_{t-1} - \theta_1^2\, y_{t-2} + \theta_1^3\, y_{t-3} - \theta_1^4\, y_{t-4} + \cdots + \varepsilon_t$$

That is, we get an AR(∞)!

Autoregressive (AR) Process

• We model the conditional expectation of y_t, $E_t[y_t | I_{t-1}]$, as a function of its past history. We assume $\varepsilon_t \sim WN(0, \sigma^2)$.

• The most common models are AR models. An AR(1) model involves a single lag, while an AR(p) model involves p lags. Then, the AR(p) process is given by:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

Using the lag operator we write the AR(p) process:

$$\phi(L)\, y_t = \varepsilon_t$$

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

Note: Inverting $\phi(L)$ delivers $y_t = \phi(L)^{-1} \varepsilon_t$ (an MA(∞)!).

AR Process – AR(1): Stability

• We can analyze the stability from the point of view of the roots of the lag polynomial. For the AR(1) process

$$\phi(z) = 1 - \phi_1 z = 0 \quad \Rightarrow |z| = \frac{1}{|\phi_1|} > 1$$

That is, the AR(1) process is stable if the root of $\phi(z)$ is greater than one (also said as "**the roots lie outside the unit circle**").

This result generalizes to AR(p) process. For the AR(3) process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t,$$

where $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3$

$\Rightarrow$ the roots, z_1, z_2 & z_3, should lie outside the *unit circle*.

For an AR(p), we need the roots of $\phi(z)$to be outside the unit circle

AR Process – AR(1): Stability

• For an AR(p), we need the roots of $\phi(z)$ to be outside the unit circle.

• For the AR(2), $y_t = \phi_1 \, y_{t-1} + \phi_2 \, y_{t-2}$.
we need the roots of $\phi(z)$ to be outside the unit circle.

The characteristic polynomial of the AR(2) can be written as:

$$\phi(z) = 1 - (\lambda_1 + \lambda_2)\, z - \lambda_1 \lambda_2 \, z^2 = (1 - \lambda_1 z)\,(1 - \lambda_2 z) = 0$$

where

$\phi_1 = \lambda_1 + \lambda_2$, & $\phi_2 = \lambda_1 \lambda_2$. ($\lambda_1$ & λ_2 = *eigenvalues/characteristic roots*.)

AR Process – AR(1): Stability

• Summary:

We say the process is globally (asymptotically) stable if the solution of the associated homogenous equation tends to 0, as $t \to \infty$.

Theorem

A necessary and sufficient condition for global asymptotical stability of a p^{th} order deterministic difference equation with constant coefficients is that *all roots* of the associated lag polynomial equation $\phi(z)=0$ have **moduli** strictly more than 1.

(For the case of real roots, moduli means "absolute values.")

AR(1) Process – Stationarity & ACF

- An AR(1) model:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN.$$

Recall that in a previous example, under the stationarity condition $|\phi_1| < 1$, we derived the mean, variance and auto-covariance function:

$E[y_t] = \mu = 0$ (assuming $\phi_1 \neq 1$)

$Var[y_t] = \gamma(0) = \sigma^2/(1 - \phi_1^2)$ (assuming $|\phi_1| < 1$)

$\gamma(k) = \phi_1^k \gamma(0)$

- We also derived the autocorrelations:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k$$

Remark: When $|\phi_1| < 1$, the autocorrelations do not explode as k increases. There is an exponential decay towards zero.

AR(1) Process – Stationarity & ACF

- ACF for an AR(1) process:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k$$

Then, the autocorrelogram –i.e., plot of $\rho(k)$ against k– shows

– when $0 < \phi_1 < 1$ ⇒ All autocorrelations are positive.

– when $-1 < \phi_1 < 0$ ⇒ The sign of $\rho(k)$ shows an alternating pattern beginning with a negative value.

– when $\phi_1 = 1$ ⇒ AR(1) is non-stationary, $\rho(k) = 1$, for all k. Present & past are always correlated!

AR Process – Stationarity

• Let's compute moments of y_t using the infinite sum (set $\mu = 0$):

$E[y_t] = \phi(L)^{-1}\, E[\varepsilon_t] = 0 \qquad (\Rightarrow \phi(L) \neq 0)$
$Var[y_t] = \phi(L)^{-2}\, Var[\varepsilon_t] = 0 \qquad (\Rightarrow \phi(L)^{-2} > 0)$
$E[y_t, y_{t-j}] = \gamma(t-j) = \phi_1\, \gamma(j-1) + \cdots + \phi_p\, \gamma(j-p)$

Using the *fundamental theorem of algebra*, $\phi(z)$ can be factored as

$$\phi(z) = (1 - r_1{}^{-1}z)\,(1 - r_2{}^{-1}z) \ldots\, (1 - r_p{}^{-1}z)$$

where the r_1, ..., $r_p \in C$ are the roots of $\phi(z)$. If $\phi_1's$ coefficients are all real, the roots are either real or come in complex conjugate pairs.

Theorem: The linear AR(p) process is strictly stationary and ergodic if and only if $|r_j|>1$ for all j, where $|r_j|$ is the modulus of the complex number r_j.

• We usually say "*all roots lie outside the unit circle*."

AR Process – Stationarity

• We usually say "*all roots lie outside the unit circle*."

Note: If one of the r_j's equals 1, $\phi(L)$ (& y_t) has a unit root –i.e., $\phi(L = 1) = 0$. This is a special case of *non-stationarity*.

• Recall $\phi(L)^{-1}$produces an infinite sum on the ε_{t-j}'s. If this sum does not explode, we say the process is *stable*.

• If the process is stable, we can calculate $\frac{\delta y_t}{\delta \varepsilon_{t-j}}$.

$\frac{\delta y_t}{\delta \varepsilon_{t-j}}$ = How much y_t is affected today by an innovation (a shock) $t - j$ periods ago. When expressed as a function of j, we call this the **impulse response function** (**IRF**).

AR Process – Example: AR(1)

Example: AR(1) process

$y_t \ = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN.$

E[y_t] = μ = 0 $\qquad \Rightarrow \phi_1 \neq 1 \qquad (r_1 \neq 0)$

Var[y_t] = $\gamma(0) = \sigma^2/(1-\phi_1^2)$ $\qquad \Rightarrow |\phi_1| < 1 \qquad (r_1 > 1)$

$$\frac{1}{1-\phi_1^i} = \sum_{i=0}^{\infty} \phi_1{}^{ij} \qquad i = 1, 2$$

Note: These infinite sums will not explode (*stable* process) if

$|\phi_1| < 1 \qquad \Rightarrow$ stationarity condition.

Under this condition, we can calculate the impulse response function:

$$\frac{\delta y_{t+j}}{\delta \varepsilon_t} = \frac{\delta y_t}{\delta \varepsilon_0} = \phi^j.$$

AR Process – Example: AR(1)

• The autocovariance function for an AR(p) process is:

$$\gamma(t-j) = Cov\left[y_t, y_{t-j}\right] = \phi_1\, \gamma(j-1) + \cdots + \phi_p\, \gamma(j-p)$$

For the AR(1) process:

$$\gamma(k) = \phi_1\, \gamma(k-1)$$

• There is a recursive formula for $\gamma(k)$:

$$\gamma(k) = \phi_1^k\, \gamma(0)$$

• Again, when $|\phi_1| < 1$, the autocovariance do not explode as k increases. There is an exponential decay towards zero.

AR Process – Example: AR(1)

• Note: $\gamma(k) = \phi_1^k\, \gamma(0)$
– when $0 < \phi_1 < 1$ $\Rightarrow$ All autocovariances are positive.
– when $-1 < \phi_1 < 0 \Rightarrow$ The sign of $\gamma(k)$ shows an alternating pattern beginning a negative value.

• The AR(1) process has the Markov property:
The distribution of y_t given $\{y_{t-1}, y_{t-2}, \ldots\}$ is the same as the distribution of y_t given $\{y_{t-1}\}$.

AR Process – Example: AR(2)

Example: AR(2) process
$y_t = \mu + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \varepsilon_t \;\; \Rightarrow (1 - \phi_1\, L - \phi_2 L^2) y_t = \mu + \varepsilon_t$

We can invert $(1 - \phi_1 L - \phi_2\, L^2)$ to get the MA(∞) process.

• Stationarity Check
- E$[y_t] = \mu/(1 - \phi_1 - \phi_2) = \mu^*$ $\Rightarrow \phi_1 + \phi_2 \neq 1.$
- Var$[y_t] = \sigma^2/(1 - \phi_1^2 - \phi_2^2)$ $\Rightarrow \phi_1^2 + \phi_2^2 < 1$

Stationarity condition: $|\phi_1 + \phi_2| < 1$

• Things can be simpler by rewriting the AR(2) in matrix AR(1) form:

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \;\; \Rightarrow \;\; \tilde{y}_t = \tilde{\mu} + A\tilde{y}_{t-1} + \tilde{\varepsilon}_t$$

Note: Now, we check $[\mathbf{I} - \boldsymbol{A}^i]$ $(i = 1, 2)$ for stationarity conditions

AR(2) Process – Stationarity & VAR

• We can derived a matrix lag polynomial A(L):

$$\tilde{y}_t = \tilde{\mu} + \boldsymbol{A}\,\tilde{y}_{t-1} + \tilde{\varepsilon}_t \qquad \Rightarrow A(L)\tilde{y}_t = [I - AL]\,\tilde{y}_t = \tilde{\varepsilon}_t$$

Note: Recall $(I - F)^{-1} = \sum_{j=0}^{\infty} F^j = I + F + F^2 + \ldots$

Checking that [I – $\boldsymbol{A}L$] is not singular, same as checking that $\boldsymbol{A}^i$ does not explode. The stability of the system can be determined by the eigenvalues of $\boldsymbol{A}$. That is, get the λ_i's and check if $|\lambda_i|<1$ for all $\boldsymbol{i}$.

$$\boldsymbol{A} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \Rightarrow |\boldsymbol{A} - \lambda I| = \det\begin{bmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{bmatrix} =$$
$$= -(\phi_1 - \lambda)\lambda - \phi_2 = \phi_2 - \phi_1\lambda + \lambda^2$$

• Solution to quadratic equation: $\lambda_i = \frac{\phi_1 \pm \sqrt{\phi_1{}^2 - 4\phi_2}}{2}$

• If $|\lambda_i| < 1$ for all i =1, 2, y_t is stable (not explode) & stationary.

AR(2) Process – Stationarity & VAR

• If $|\lambda_i| < 1$ for all i =1, 2, y_t is stable (not explode) & stationary.

For the AR(2) process, we derive relations between λ_i's & ϕ_i's:

$$\lambda_1\lambda_2 = \phi_2 \qquad \Rightarrow |\lambda_1\lambda_2| = |\phi_2| < 1$$
$$\lambda_1 + \lambda_2 = \phi_1 \qquad \Rightarrow |\lambda_1 + \lambda_2| = |\phi_1| < 2$$

• We derived autocovariance function, $\gamma(k)$, before, getting a recursive formula. Let's write the first autocovariances:

($k = 0$) $\qquad \gamma(0) = \phi_1\,\gamma(1) + \phi_2\gamma(2) + \sigma^2$

($k = 1$) $\qquad \gamma(1) = \frac{\phi_1}{1-\phi_2}\gamma(0)$

($k = 2$) $\qquad \gamma(2) = [\frac{\phi_1{}^2}{1-\phi_2} + \phi_2]\,\gamma(0)$

With $|\phi_2| < 1$, we get well defined $\gamma(1)$, $\gamma(2)$ & $\gamma(0)$.

AR(2) Process – Stationarity & VAR

• The AR(2) in matrix AR(1) form is called **Vector AR(1)** or **VAR(1).**

Nice property: The VAR(1) is Markov -i.e., forecasts depend only on today's data.

• It is straightforward to apply the VAR formulation to any AR(p) processes. We can also use the same eigenvalue conditions to check the stationarity of AR(p) processes.

AR(2) Process – Stationarity & ACF

• An AR(2) model:

$$y_t = \mu + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim WN.$$

• **Moments:** (μ =0)

$$\mathrm{E}[y_t] = \frac{\mu}{(1-\phi_1-\phi_2)} = 0 \qquad \text{(assuming } \phi_1 + \phi_2 \neq 1\text{)}$$

$$\mathrm{Var}[y_t] = \frac{\sigma^2}{(1-\phi_1^2-\phi_2^2)} \qquad \text{(assuming } \phi_1^2 + \phi_2^2 < 1\text{)}$$

• **Autocovariance function**

$$\begin{aligned}\gamma(k) &= \mathrm{Cov}[y_t, y_{t-k}] = \mathrm{E}[(\phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \varepsilon_t)\, y_{t-k}] \\ &= \phi_1\, \mathrm{E}[y_{t-1}\, y_{t-k}] + \phi_2\, \mathrm{E}[y_{t-2}\, y_{t-k}] + \mathrm{E}[\varepsilon_t\, y_{t-k}] \\ &= \phi_1\, \gamma(k-1) + \phi_2\, \gamma(k-2) + \mathrm{E}[\varepsilon_t\, y_{t-k}]\end{aligned}$$

We have a **recursive formula**.

AR(2) Process – Stationarity & ACF

- Recursive formula: $\gamma(k) = \phi_1\, \gamma(k-1) + \phi_2\, \gamma(k-2) + \mathrm{E}[\varepsilon_t\, y_{t-k}]$

(k=0) $\gamma(0) = \phi_1\, \gamma(-1) + \phi_2\, \gamma(-2) + \mathrm{E}[\varepsilon_t\, y_t]$
$\quad = \phi_1\, \gamma(1) + \phi_2\, \gamma(2) + \sigma^2$

(k=1) $\gamma(1) = \phi_1\, \gamma(0) + \phi_2\, \gamma(1) + \mathrm{E}[\varepsilon_t\, y_{t-1}]$
$\quad = \phi_1\, \gamma(0) + \phi_2\, \gamma(1) + 0$
$\Rightarrow \gamma(1) = [\phi_1/(1-\phi_2)]\, \gamma(0)$

(k=2) $\gamma(2) = \phi_1\, \gamma(1) + \phi_2\, \gamma(0) + \mathrm{E}[\varepsilon_t\, y_{t-2}]$
$\quad = \phi_1\, \gamma(1) + \phi_2\, \gamma(0) + 0$
$\Rightarrow \gamma(2) = [\frac{\phi_1^2}{1-\phi_2} + \phi_2]\, \gamma(0)$

Replacing $\gamma(1)$ and $\gamma(2)$ back to $\gamma(0)$:

$$\gamma(0) = [\phi_1^2/(1-\phi_2)]\, \gamma(0) + [\phi_2\, \phi_1^2/(1-\phi_2) + \phi_2^2]\, \gamma(0) + \sigma^2$$
$$= \frac{\sigma^2(1-\phi_2)}{(1-\phi_2) - \phi_1^2\,(1+\phi_2) + \phi_2^2\,(1-\phi_2)} \qquad \Rightarrow |\phi_2| < 1$$

AR(2) Process – Stationarity & ACF

- Dividing the recursive formula for $\gamma(k)$ by $\gamma(0)$, we get the ACF:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1\, \rho(k-1) + \phi_2\, \rho(k-2) + \frac{\mathrm{E}[\varepsilon_t\, y_{t-k}]}{\gamma(0)}$$

(k=0) $\rho(0) = 1$

(k=1) $\rho(1) = \phi_1/(1-\phi_2)$

(k=2) $\rho(2) = \phi_1\, \rho(1) + \phi_2\, \rho(0) = \phi_1^2/(1-\phi_2) + \phi_2$

(k=3) $\rho(3) = \phi_1\, \rho(2) + \phi_2\, \rho(1) =$
$\quad = \phi_1^3/(1-\phi_2) + \phi_1\, \phi_2 + \phi_2\, \phi_1/(1-\phi_2)$

Remark: Again, we see exponential decay in the ACF.

From the work above, for stationarity, we need:
$\phi_1 + \phi_2 \neq 1$.
$\phi_1^2 + \phi_2^2 < 1$.
$|\phi_2| < 1$.

AR Process – Causality

• The AR(p) model: $\phi(L)y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim WN.$

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

Then, $y_t = \phi(L)^{-1}(\mu + \varepsilon_t)$, $\Rightarrow$ an MA(∞) process!

• But, we need to make sure that we can invert the polynomial $\phi(L)$. When $\phi(L) \neq 0$, we say the process y_t is *causal* (strictly speaking, a *causal function of* $\{\varepsilon_t\}$).

Definition: A linear process $\{y\}$ is *causal* if there is a

$$\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \cdots$$
$$\sum_{j=0}^{\infty} |\psi_j(L)| < \infty$$
$$y_t = \psi(L)\varepsilon_t.$$

AR Process – Causality

Example: AR(1) process:

$$\phi(L)y_t = \mu + \varepsilon_t, \qquad \text{where } \phi(L) = 1 - \phi_1 L$$

Then, y_t is causal if and only if:

$|\phi_1| < 1$ (same condition as stationarity)

or

the root r_1 of the polynomial $\phi(z) = 1 - \phi_1 z$ satisfies $|r_1| > 1$.

Question: How do we calculate the ψ_i's coefficients for an AR(p)?
A: Matching coefficients (μ =0):

$$Y_t = \frac{1}{(1 - \phi_1 L)}\varepsilon_t \overset{|\phi_1|<1}{\cong} \sum_{i=0}^{\infty} \phi_1{}^i L^i \varepsilon_t$$
$$= (1 + \phi_1 L + \phi_1{}^2 L^2 + \cdots)\,\varepsilon_t \qquad \Rightarrow \psi_i = \phi_1{}^i, \quad i \geq 0$$

AR Process – Calculating the ψ_i's

Example: AR(2) - Calculating the ψ_i's by matching coefficients.

$$\underbrace{\left(1 - \phi_1 L - \phi_2 L^2\right)}_{\Phi(L)}\left(y_t - \mu\right) = \varepsilon_t \quad \Rightarrow \quad \Phi(L)\Psi(L) = 1$$

$\psi_0 = 1$
$\psi_1 = \phi_1$
$\psi_2 = \phi_1{}^2 + \phi_2$
$\psi_3 = \phi_1{}^3 + 2\phi_1\phi_2$
$\psi_j = \phi_1\psi_{j-1} + \phi_2\,\psi_{j-2} \qquad j \geq 2.$

We can solve these *linear difference equations* in several ways:
- Numerically
- Guess the form of a solution and using an inductive proof
- Using the theory of linear difference equations.

AR Process – Estimation and Properties

• Define

$$\boldsymbol{x}_t = \left(1\ y_{t-1}\ y_{t-2} \ldots .\, y_{t-p}\right)$$
$$\boldsymbol{\beta} = (\mu\ \phi_1\ \phi_2\ \ldots .\ \phi_p)$$

• Then the model can be written as $\quad y_t = \boldsymbol{x}_t{}'\boldsymbol{\beta} + \varepsilon_t$

• The OLS estimator is $\quad \mathbf{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

• Recall $u_t = x_t\varepsilon_t$ is a MDS. It is also strictly stationary & ergodic.

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{T} u_t = \frac{1}{\sqrt{T}}\sum_{i=1}^{T} \boldsymbol{x}_t{}'\varepsilon_t \xrightarrow{d} N(0, \boldsymbol{\Omega}) \qquad \boldsymbol{\Omega} = E[\boldsymbol{x}_t\boldsymbol{x}_t{}'\varepsilon_t{}^2]$$

• The vector $x_t$is strictly stationary and ergodic, and by Theorem I so is $\boldsymbol{x}_t\boldsymbol{x}_t{}'$. Then, by the Ergodic Theorem

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t{}' \xrightarrow{p} \mathrm{E}[\boldsymbol{x}_t\boldsymbol{x}_t{}'] = \boldsymbol{Q}$$

AR Process – Estimation and Properties

• **Consistency**

Putting together the previous results, the OLS estimator can be rewritten as:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \boldsymbol{\beta} + \left(\sum_{i=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \left(\sum_{i=1}^{T} \boldsymbol{x}_t' \varepsilon_t\right)$$

Then,

$$\mathbf{b} = \boldsymbol{\beta} + \left(\sum_{i=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \left(\sum_{i=1}^{T} \boldsymbol{x}_t' \varepsilon_t\right) \xrightarrow{p} \boldsymbol{\beta} + \boldsymbol{Q}^{-1}\mathbf{0} = \boldsymbol{\beta}$$

⇒ the OLS estimator is consistent.

AR Process – Asymptotic Distribution

• **Asymptotic Normality**

We apply the MDS CLT to $\boldsymbol{x}_t \varepsilon_t$. Then, it is straightforward to derive the asymptotic distribution of the estimator (similar to the OLS case):

Theorem If the AR(p) process y_t is strictly stationary and ergodic and E[y_t^4], then as $T \rightarrow \infty$;

$$\sqrt{T}(\mathbf{b} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{Q}^{-1} \boldsymbol{\Omega}\, \boldsymbol{Q}^{-1}) \qquad \boldsymbol{\Omega} = E[\boldsymbol{x}_t \boldsymbol{x}_t' \varepsilon_t^2]$$

• Identical in form to the asymptotic distribution of OLS in cross-section regression ⇒ asymptotic inference is the same.

• The asymptotic covariance matrix is estimated just as in the cross-section case: The sandwich estimator.

AR Process – Bootstrap

• So far, we constructed the bootstrap sample by randomly resampling from the data values $(y_t, \boldsymbol{x}_t)$. This created an *i.i.d* bootstrap sample.

• This is inappropriate for time-series, since we have dependence.

• There are two popular methods to bootstrap time series.

(1) *Model-Based (Parametric) Bootstrap*

(2) *Block Resampling Bootstrap*

AR Process – Bootstrap

(1) Model-Based (Parametric) Bootstrap

1. Estimate **b** and residuals $\boldsymbol{e}$:

2. Fix an initial condition $\{y_{t-k+1}, y_{t-k+2}, y_{t-k+3}, ..., y_0\}$

3. Simulate *i.i.d.* draws $\boldsymbol{e}$* from the empirical distribution of the residuals $\{e_1, e_2, e_3, ..., e_T\}$.

4. Create the bootstrap series y_t by the recursive formula

$$y_t{}^* = \hat{\mu} + \hat{\phi}_1 y_{t-1}{}^* + \hat{\phi}_2 y_{t-2}{}^* + ... + \hat{\phi}_p y_{t-p}{}^* + \varepsilon_t{}^*$$

<u>Pros</u>: Simple. Similar to the usual bootstrap.

<u>Cons</u>: This construction imposes homoskedasticity on the errors $\boldsymbol{e}$* ; which may be different than the properties of the actual $\boldsymbol{e}$. It also imposes the AR*(p)* as the DGP.

AR Process – Bootstrap

(2) Block Resampling

1. Divide the sample into T/m blocks of length m.
2. Resample complete blocks. For each simulated sample, draw T/m blocks.
3. Paste the blocks together to create the bootstrap time-series y_t*.

Pros: It allows for arbitrary stationary serial correlation, heteroskedasticity, and for model misspecification.

Cons: It may be sensitive to the block length, and the way that the data are partitioned into blocks. May not work well in small samples.

Moving Average Process

- An MA process models $\text{E}_t[y_t \,|\, I_{t-1}]$ with lagged error terms. An MA(q) model involves q lags.

- We keep the white noise assumption for ε_t: ε_t ~ WN(0, σ^2)

Example: A linear MA(q) model:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t,$$

where

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_2 L^3 + \ldots + \theta_q L^q$$

- In time series, the constant does not affect the properties of AR and MA process. It is usually removed (think of the data analyzed as demeaned). Thus, in this situation we say "without loss of generalization", we assume μ = 0.

MA Process – MA(1): Stationarity

Example: MA(1) process:

$$y_t = \theta_1\, \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t, \qquad \text{with } \theta(L) = (1 + \theta_1 L)$$

- **Mean**

  $E[y_t] = 0$

- **Variance**

  $\text{Var}[y_t] = \gamma(0) = \sigma^2 + \theta_1^2\, \sigma^2 = \sigma^2\ (1 + \theta_1^2)$

- **Covariance**

  $\text{Cov}[y_t, y_{t-1}] = \gamma(1) = E[y_t\, y_{t-1}]$
  $\qquad = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t)*(\theta_1\varepsilon_{t-2} + \varepsilon_{t-1})] = \theta_1\sigma^2$

  $\text{Cov}[y_t, y_{t-2}] = \gamma(2) = E[y_t\, y_{t-2}]$
  $\qquad = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t) * (\theta_1\, \varepsilon_{t-3} + \varepsilon_{t-2})] = 0$

MA Process – MA(1): Stationarity

Example (continuation): MA(1) process:

- **Covariance**

  $\text{Cov}[y_t, y_{t-1}] = \gamma(1) = E[y_t\, y_{t-1}]$
  $\qquad = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t)*(\theta_1\varepsilon_{t-2} + \varepsilon_{t-1})] = \theta_1\sigma^2$

  $\text{Cov}[y_t, y_{t-2}] = \gamma(2) = E[y_t\, y_{t-2}]$
  $\qquad = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t) * (\theta_1\, \varepsilon_{t-3} + \varepsilon_{t-2})] = 0$

  $\vdots$

  $\gamma(k) = E[y_t\, y_{t-k}] = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t) * (\theta_1\varepsilon_{t-(k+1)} + \varepsilon_{t-k})] = 0 \quad (\text{for } k>1)$

That is, for $|k| > 1$, $\quad \gamma(k) = 0$.

$\Rightarrow$ MA(1) is always stationary –i.e., independent of values of θ_1.

<u>Remark</u>: The MA(q=1) process has $\gamma(q) = 0$, for $q > 1$. This result generalizes to MA(q) process: after lag q, the autocovariances are 0.

MA(1) Process – ACF

Example (continuation): To get the ACF, we divide the autocovariances by $\gamma(0)$. Then, the autocorrelation function (ACF):

$$\rho(0) = \gamma(0)/\gamma(0) = 1$$

$$\rho(1) = \gamma(1)/\gamma(0) = \frac{\theta_1\,\sigma^2}{\sigma^2\,(1+\theta_1^2)} = \frac{\theta_1}{(1+\theta_1^2)}$$

$$\vdots$$

$$\rho(k) = \gamma(k)/\gamma(0) = 0 \qquad\qquad (\text{for } k > 1)$$

Remark: The autocovariance function is **zero** after lag 1. Similarly, the ACF is also **zero** after lag 1, that is, y_t is correlated with itself (y_t) and y_{t-1}, but not y_{t-2}, y_{t-3}, ... Contrast this with the AR(1) model, where the correlation between y_t and y_{t-k} is never zero.

The ACF is usually shown in a plot, the **autocorrelogram**. When we plot $\rho(k)$ against k, we plot also $\rho(0)$ which is 1.

MA(1) Process – ACF

Example (continuation):

$$\rho(1) = \frac{\theta_1}{(1+\theta_1^2)}$$

Note that $|\rho(1)| \leq 0.5$.

When $\theta_1 = \mathbf{0.5}$ $\Rightarrow \rho(1) = \mathbf{0.4}$.

$\theta_1 = -0.9$ $\Rightarrow \rho(1) = -0.497238$.

$\theta_1 = -2$ $\Rightarrow \rho(1) = -0.4$.

$\theta_1 = \mathbf{2}$ $\Rightarrow \rho(1) = \mathbf{0.4}$. (same $\rho(1)$ for θ_1 & $\frac{1}{\theta_1}$.)

Note: Both MA(1) processes, with $\theta_1 = \mathbf{0.5}$ and $\theta_1 = \mathbf{2}$, have the same ACF. That is, ACFs are not unique. This is a problem: we deduce the order and the coefficients through the ACF, which is what we observe.

MA Process – MA(q): Stationarity

- Q: Is MA(q) stationary? Check the moments (assume μ = 0).

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

- **Mean**

E[y_t] = E[ε_t] + θ_1 E[ε_{t-1}] + θ_2 E[ε_{t-2}] + ... + θ_q E[ε_{t-q}] = 0

- **Variance**

$$\text{Var}[y_t] = \text{Var}[\varepsilon_t] + \theta_1^{\,2}\, \text{Var}[\varepsilon_{t-1}] + \theta_2^{\,2}\, \text{Var}[\varepsilon_{t-2}] + \ldots + \theta_q^{\,2}\, \text{Var}[\varepsilon_{t-q}]$$
$$= (1 + \theta_1^{\,2} + \theta_2^{\,2} + \ldots + \theta_q^{\,2})\, \sigma^2.$$

To get a positive variance, we require

$$(1 + \theta_1^{\,2} + \theta_2^{\,2} + \ldots + \theta_q^{\,2}) > 0. \qquad \text{(always positive)}$$

- **Covariance**

It can shown (check book) for the k *autocovariance:*

$$\gamma(k) = \sigma^2 \textstyle\sum_{j=k}^{q} \theta_j\, \theta_{j-k} \qquad \text{for } |\, k\, | \leq q \quad (\text{where } \theta_0 = 1)$$
$$\gamma(k) = 0 \qquad \text{for } |\, k\, | > q$$

MA Process – MA(q): Stationarity

- **Covariance**

$$\gamma(k) = \sigma^2 \textstyle\sum_{j=k}^{q} \theta_j\, \theta_{j-k} \qquad \text{for } |\, k\, | \leq q \quad (\text{where } \theta_0 = 1)$$
$$\gamma(k) = 0 \qquad \text{for } |\, k\, | > q$$

Remark: After lag q, the autocovariances are 0.

Applying formula:

$$\gamma(1) = \sigma^2 \textstyle\sum_{j=1}^{q} \theta_j\, \theta_{j-1}$$
$$= \sigma^2\, \theta_1 + \sigma^2\, \theta_2\, \theta_1 + \sigma^2\, \theta_3\, \theta_2 + \ldots + \sigma^2 \theta_q\, \theta_{q-1}$$

$$\gamma(2) = \sigma^2 \textstyle\sum_{j=2}^{q} \theta_j\, \theta_{j-2}$$
$$= \sigma^2\, \theta_2 + \sigma^2\, \theta_3\, \theta_1 + \sigma^2\, \theta_4\, \theta_2 + \ldots + \sigma^2 \theta_q\, \theta_{q-2}$$

$$\vdots$$

$$\gamma(q) = \sigma^2 \textstyle\sum_{j=q}^{q} \theta_{j-q} = \sigma^2 \theta_q$$

MA Process – MA(q): Stationarity

• In general, for the k *autocovariance:*

$$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \, \theta_{j-k} \qquad \text{for } |\,k\,| \leq q \quad (\text{where } \theta_0 = 1)$$

$$\gamma(k) = 0 \qquad \text{for } |\,k\,| > q$$

• It is easy to verify that the sums $\sum_{j=k}^{q} \theta_j \, \theta_{j-k}$ are finite. Then, mean, variance and covariance are constant.

$\Rightarrow$ MA(q) is always stationary –i.e., independent of values of θ_j's.

• Check for MA(1):

$k = 0 \qquad \gamma(0) = \sigma^2 \sum_{j=0}^{1} \theta_j \, \theta_{j-0} = \sigma^2(1 + \theta_1^2)$

$k = 1 \qquad \gamma(1) = \sigma^2 \sum_{j=1}^{1} \theta_j \, \theta_{j-1} = \sigma^2 \theta_1$

$k > 1 \qquad \gamma(k) = 0$

Remark: After lag $q = 1$, the autocovariances of an MA(1) are 0.

MA Process – Invertibility

• As mentioned above, the autocovariances are non-unique.

Example: Two MA(1) processes that produce the same $\gamma(k)$:

$y_t = \varepsilon_t + \mathbf{0.2}\, \varepsilon_{t-1}, \qquad \varepsilon_t \sim$ *i.i.d.* N(0, 25)

$z_t = \upsilon_t + \mathbf{5}\, \upsilon_{t-1}, \qquad \upsilon_t \sim$ *i.i.d.* N(0; 1)

We only observe the time series, y_t or z_t, and not the noise, ε_t or υ_t. We cannot distinguish between the models using the autocovariances.

We want to select one process to forecast: We select the model with an AR(∞) representation. That is, we select the process that is invertible.

• Assuming $\theta(L) \neq 1$, we invert $\theta(L)$:

$$y_t = \mu + \theta(L)\, \varepsilon_t \qquad \Rightarrow \theta(L)^{-1} y_t = \Pi(L)\, y_t = \mu^* + \varepsilon_t.$$

$$\Rightarrow y_t = \mu^* + \sum_{j=1}^{\infty} \pi_j \, y_{t-j} + \varepsilon_t$$

MA Process – Invertibility

- We convert an MA(q) into an AR(∞):

$$y_t = \mu* + \sum_{j=1}^{\infty} \pi_j \, y_{t-j} \; + \varepsilon_t$$

We need to make sure that $\Pi(L) = \theta(L)^{-1}$ is defined: We require $\theta(L) \neq 0$. When this condition is met, we can write ε_t as a causal function of y_t. We say the MA is *invertible.* For this to hold, we require:

$$\sum_{j=0}^{\infty} |\pi_j(L)| < \infty$$

Technical note: An invertible MA(q) is typically required to have roots of the lag polynomial equation $\theta(z) = 0$ greater than one in absolute value (**outside the unit circle**). In the MA(1) case,

$$\theta(z) = (1 + \theta_1 \, z) = 0 \quad \Rightarrow \text{root: } z = -\frac{1}{\theta_1} \; (\Rightarrow |\theta_1| < 1)$$

In the previous example, we select the model with $\theta_1 = \mathbf{0.2}$.

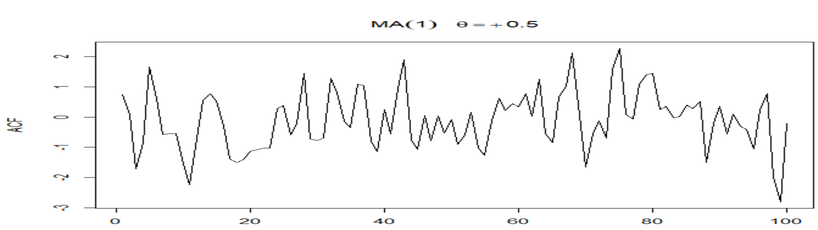
MA(1) Process – ACF: Simulations

Simulated Example: We simulate with R function *arima.sim* (& plot) three MA(1) processes, with standard normal ε_t -i.e., $\mu = 0$ & $\sigma^2 = 1$:

$$y_t = \varepsilon_t + 0.5\, \varepsilon_{t-1}$$
$$y_t = \varepsilon_t - 0.9\, \varepsilon_{t-1}$$
$$y_t = \varepsilon_t - 2\, \varepsilon_{t-1}$$

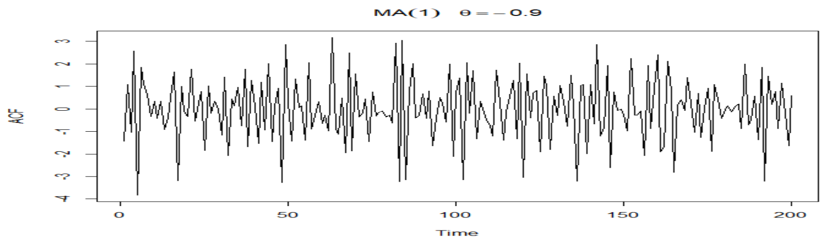
R script to plot $y_t = \varepsilon_t + \mathbf{0.5}\, \varepsilon_{t-1}$ with **200** simulations

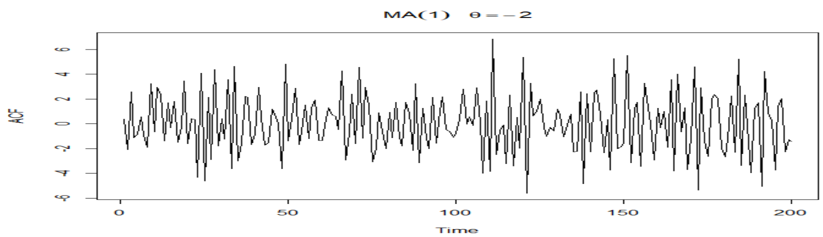
> plot(arima.sim(list(order=c(0,0,1), ma = **0.5**), n = **200**), ylab="ACF", main=(expression(MA(1)~~~theta==+.5)))



MA(1) Process – ACF: Simulations

Simulated Example (continuation):





Note: The process $\theta_1 > 0$ is smoother than the ones with $\theta_1 < 0$.

MA(1) Process – ACF: Simulations

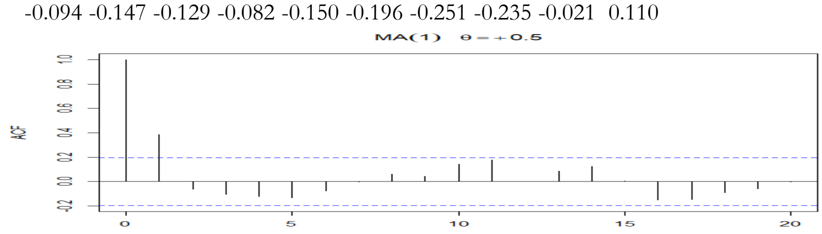
Simulated Example (continuation): Below, we compute and plot the ACF for the 3 simulated process.

1) $y_t = \varepsilon_t + 0.5\, \varepsilon_{t-1}$

sim_ma1_5 <- arima.sim(list(order=c(0,0,1), ma = **0.5**), n = 200)
acf_ma1_5 <- acf(sim_ma1_5, main=(expression(MA(1)~~~theta==+.5)))
> acf_ma1_5

Autocorrelations of series 'sim_ma1_5', by lag

0	1	2	3	4	5	6	7	8	9	10	11	12	13
1.000	**0.438**	0.069	0.014	0.103	0.173	0.107	0.015	-0.080	-0.054	0.011	-0.006	0.041	0.000
14	15	16	17	18	19	20	21	22	23				
-0.094	-0.147	-0.129	-0.082	-0.150	-0.196	-0.251	-0.235	-0.021	0.110				



MA(1) Process – ACF: Simulations

Simulated Example (continuation):

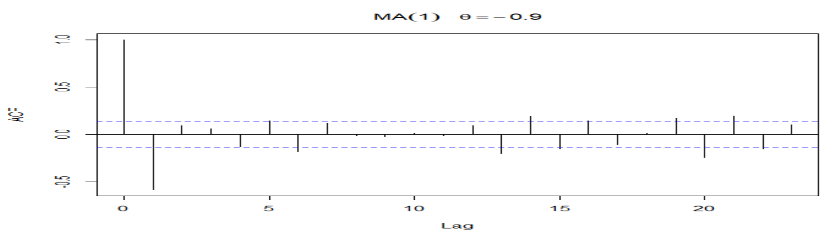
2) $y_t = \varepsilon_t - 0.9\ \varepsilon_{t-1}$

```
sim_ma1_9 <- arima.sim(list(order=c(0,0,1), ma = -0.9), n = 200)
acf_ma1_9 <- acf(sim_ma1_5, main=(expression(MA(1)~~~theta==+.5)))
> acf_ma1_9
```

Autocorrelations of series 'sim_ma1_9', by lag

0	1	2	3	4	5	6	7	8	9	10	11	12	13
1.000	**-0.584**	0.093	0.061	-0.132	0.147	-0.181	0.122	-0.013	-0.023	0.014	-0.012	0.092	-0.199

14	15	16	17	18	19	20	21	22	23
0.193	-0.155	0.143	-0.107	0.014	0.174	-0.244	0.196	-0.154	0.105



MA(1) Process – ACF: Simulations

Simulated Example (continuation):

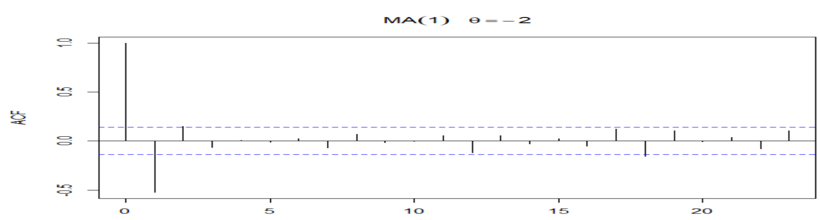
3) $y_t = \varepsilon_t - 2\ \varepsilon_{t-1}$

```
sim_ma1_2 <- arima.sim(list(order=c(0,0,1), ma = -2), n = 200)
acf_ma1_2 <- acf(sim_ma1_2, main=(expression(MA(1)~~~theta==-2)))
> acf_ma1_2
```

Autocorrelations of series 'sim_ma1_2', by lag

0	1	2	3	4	5	6	7	8	9	10	11	12	13
1.000	**-0.524**	0.150	-0.064	0.006	-0.014	0.022	-0.070	0.068	-0.015	-0.002	0.054	-0.121	0.055

14	15	16	17	18	19	20	21	22	23
-0.029	0.026	-0.054	0.121	-0.156	0.106	-0.009	0.037	-0.080	0.104



MA Process – Example: MA(1)

Simulated Example (continuation):

– Invertibility: If $|\theta_1| < 1$, we can write $(1 + \theta_1 L)^{-1} \;\; y_t + \mu^* = \varepsilon_t$

$$\Rightarrow \left(1 - \theta_1\, L + \theta_2\, L^2 - \theta_2\, L^3 + \ldots + \theta_q\, L^q + \cdots\right) y_t + \mu^* =$$
$$= \mu* + \textstyle\sum_{i=1}^{\infty} \pi_{\,i}(L)\;y_t = \varepsilon_t$$

That is, $\pi_i = {\theta_1}^i$.

The simulated process with $\theta_1 = -2$ is non-invertible, the infinite sum of π_i would explode. We would select the MA(1) with $\theta_1 = -.5$.

MA Process – Estimation

• MA processes are more complicated to estimate. Consider an MA(1):

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

We cannot do OLS, since we do not observe ε_{t-1}. But, based on the ACF, we estimate θ_1.

• The auto-correlation of order one is:

$$\rho(1) = \theta_1/(1 + {\theta_1}^2)$$

Then, we can use the **Method of Moments** (MM), which sets the theoretical moment equal to the estimated sample moment $\rho(1)$, r_1. Then, we solve for the parameter of interest, θ_1:

$$r_1 = \frac{\hat{\theta}_1}{(1 + {\hat{\theta}_1}^{\,2})} \qquad \Rightarrow \qquad \theta_1 = \frac{1 \pm \sqrt{1 - 4r_1^2}}{2r_1}$$

• A nonlinear solution and difficult to solve.

MA Process – Estimation

• Alternatively, if $|\theta_1| < 1$, we can invert the MA(1) process. Then, based on the AR representation, we can try finding a ∈(-1; 1):

$$\varepsilon_t(a) = y_t + a\, y_{t-1} + a^2\, y_{t-2} + a^3\, y_{t-3} + \ldots.$$

and look (numerically) for the least-square estimator

$$\hat{\boldsymbol{\theta}} = \arg\min_\theta \{S(\boldsymbol{y}; \boldsymbol{\theta}) = \textstyle\sum_{t=1}^{T} \varepsilon_t(a)^2\}$$

where $a^t = \theta_1{}^t$.

The Wold Decomposition

Theorem - Wold (1938).
Any covariance stationary $\{y_t\}$ has infinite order, moving-average representation:

$$y_t = S_t + \kappa_t,$$

where
κ_t is a deterministic term –i.e., completely predictable. For example, $\kappa_t = \mu$ or a linear combination of past (known) values of κ_t.
$S_t = \sum_{j=0}^{\infty} \psi_j\, \varepsilon_{t-j}$ (= $\psi(L)\varepsilon_t$, with $\psi(L)$ = infinite lag polynomial)
$\sum_{j=0}^{\infty} \psi_j^2 < \infty$ (for stability of polynomial, **square summability**)
ψ_j only depend on j (weights of innovations are not time dependent)
$\psi_0 = 1$ (a convenient assumption)
$\varepsilon_t \sim$ WN(0, σ^2) (ε_t independent and uncorrelated with S_t)

• y_t is a linear combination of innovations over time plus a deterministic part.

The Wold Decomposition

• A stationary process can be represented as an MA(∞) plus a deterministic "trend."

$$y_t = \sum_{j=0}^{\infty} \psi L^j \varepsilon_{t-j} + \kappa_t, \qquad \psi_0 = 1$$

Example:

Let $x_t = y_t - \kappa_t$. (x_t = MA(∞) part) Then, check moments:

$$E[x_t] = E[y_t - \kappa_t] = \sum_{j=0}^{\infty} \psi_j E[\varepsilon_{t-j}] = 0.$$

$$E[x_t^{\,2}] = \sum_{j=0}^{\infty} \psi_j^2 E[\varepsilon_{t-j}^2] = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

$$E[x_t, x_{t-j}] = E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots)(\varepsilon_{t-j} + \psi_1 \varepsilon_{t-j-1} + \psi_2 \varepsilon_{t-j-2} + \ldots)$$

$$= \sigma^2 (\psi_j + \psi_1 \psi_{j+1} + \psi_2 \psi_{j+2} + \ldots) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+2}$$

x_t is a covariance stationary process.

ARMA Process

• The Wold theorem is the backbone of time series analysis. We will approximate the Wold infinite lag polynomial $\psi(L)$ with a ratio of two finite lag polynomials. This approximation is the basis of ARMA modeling.

• A combination of AR(p) and MA(q) processes produces an ARMA(p, q) process:

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

$$= \mu + \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i L^i \varepsilon_t + \varepsilon_t$$

$$\Rightarrow \phi(L) y_t = \mu + \theta(L) \varepsilon_t$$

• Usually, we insist that $\phi(L) \neq 0$, $\theta(L) \neq 0$ & that the polynomials $\phi(L)$, $\theta(L)$ have no **common factors**. This implies it is not a lower order ARMA model.

ARMA Process – Common Factors

An ARMA(p, q) model with common factors has a lower order ARMA model. That is, a lower p and q.

Example: Common factors.
Suppose we have the following ARMA(2, 3) model

$$y_t = 0.6\, y_{t-1} - 0.3\, y_{t-2} + \varepsilon_t - 1.4\, \varepsilon_{t-1} + 0.9\, \varepsilon_{t-2} + 0.3\, \varepsilon_{t-3}$$

with

$$\phi(L) = 1 - .6L + .3L^2$$
$$\theta(L) = 1 - 1.4L + .9L^2 - .3L^3 = (1 - .6L + .3L^2)(1 - L)$$

This model simplifies to: $y_t = (1 - L)\varepsilon_t$
$= \varepsilon_t - \varepsilon_{t-1}$ $\Rightarrow$ an MA(1) process.

- Simplify the common factors and keep the simpler representation.

ARMA Process – Representation

- ARMA(p, q) model:

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

- Cases:

Pure AR Representation: $\Pi(L)(y_t - \mu) = \varepsilon_t \Rightarrow \Pi(L) = \frac{\phi_p(L)}{\theta_q(L)}$

Pure MA Representation: $(y_t - \mu) = \Psi(L)\varepsilon_t \Rightarrow \Psi(L) = \frac{\theta_q(L)}{\phi_p(L)}$

Special cases: – $p = 0$: MA(q)
– $q = 0$: AR(p).

ARMA(1, 1) – Stationarity & ACF

- For an ARMA(1, 1) we have:.

$$y_t = \mu + \phi_1\, y_{t-1} + \theta_1\, \varepsilon_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

- **Moments:** ($\mu = 0$)

$E[y_t] = \mu\, /\, (1 - \phi_1) = 0$ (assuming $\phi_1 \neq 1$)

$Var[y_t] = \sigma^2\, (1 + \theta_1^2\,)\, /\, (1 - \phi_1^2)$ (assuming $|\phi_1| < 1$)

- **Autocovariance function** ($\mu = 0$)

$$\begin{aligned}\gamma(k) &= Cov[y_t,\ y_{t-k}] \\ &= E[\{\phi_1\, y_{t-1} + \theta_1\ \varepsilon_{t-1} + \varepsilon_t\}\, y_{t-k}] \\ &= \phi_1\, E[y_{t-1}\, y_{t-k}] + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}] \\ &= \phi_1\, \gamma(k-1) + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}]\end{aligned}$$

- Again, we have a **recursive formula**.

$$\gamma(k) = \phi_1\, \gamma(k-1) + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}] + E[\varepsilon_t\, y_{t-k}]$$

ARMA(1, 1) – Stationarity & ACF

- We have a recursive formula:

$$\gamma(k) = \phi_1\, \gamma(k-1) + E[\varepsilon_t\, y_{t-k}] + \theta_1\, E[\varepsilon_{t-1}\, y_{t-k}]$$

It can be shown, after a lot of algebra:

For $k = 0$,

$$\gamma(0) = \phi_1\, \gamma(1) + \sigma^2 + \theta_1\, (\phi_1\, \sigma^2 + \theta_1 \sigma^2)$$

For $k = 1$,

$$\gamma(1) = \phi_1\, \gamma(0) + \theta_1\, \gamma(1)$$

For $k = 2$,

$$\gamma(2) = \phi_1\, \gamma(1)$$

For k,

$$\gamma(k) = \phi_1^{k-1}\, \gamma(1), \quad k > 1$$

$\Rightarrow$ If $|\phi_1| < 1$, exponential decay.

ARMA(1, 1) – Stationarity & ACF

- Two equations for $\gamma(0)$ and $\gamma(1)$:

$$\gamma(0) = \phi_1\,\gamma(1) + \sigma^2 + \theta_1\,(\phi_1\sigma^2 + \theta_1\sigma^2)$$
$$\gamma(1) = \phi_1\,\gamma(0) + \theta_1\,\gamma(1)$$

Solving for $\gamma(0)$ & $\gamma(1)$:

$$\gamma(0) = \sigma^2\,\frac{1+\theta_1{}^2+2\,\phi_1\theta_1}{1-\phi_1{}^2}$$
$$\gamma(1) = \sigma^2\,\frac{(1+\phi_1\theta_1)*(\phi_1+\theta_1)}{1-\phi_1{}^2}$$
$$\vdots$$
$$\gamma(k) = \phi_1{}^{k-1}\,\gamma(1), \quad k>1 \quad \Rightarrow \text{If } |\phi_1|<1, \text{ exponential decay.}$$

Note: If stationary, ARMA(1,1) & AR(1) show exponential decay. Difficult to distinguish one from the other through autocovariances.

ARMA: Stationarity, Causality and Invertibility

Theorem: If ϕ(L) and $\theta(L)$ have no common factors, a (unique) *stationary* solution to $\phi(L)y_t = \theta(L)\varepsilon_t$ exists if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1\,z - \phi_2\,z^2 - \ldots - \phi_p\,z^p \neq 0.$$

(i.e., roots of $\phi(z) = 0$ need to be **outside the unit circle**, $|z| > 1$.)

This ARMA(p, q) model is causal if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$$

This ARMA(p, q) model is invertible if and only if

$$|z| \leq 1 \Rightarrow \theta(z) = 1 + \theta_1 z - \theta_2 z^2 + \ldots + \theta_p z^p \neq 0.$$

Note: Real data cannot be *exactly* modeled using a finite number of parameters. We choose p, q to create a good approximated model.

ARMA Process – SDE Representation

• Consider the ARMA(p, q) model:

$$\phi(L)\ x_t = \theta(L)\ \varepsilon_t$$

Let $x_t = y_t - \mu$ & $w_t = \theta(L)$

Then, $x_t = \mu + \phi_1\, x_{t-1} + \phi_2\, x_{t-2} + \cdots + \phi_p\, x_{t-p} + w_t$,

$\Rightarrow$ x_t is a p -th-order linear **stochastic difference equation** (SDE).

Example: 1st-order SDE (AR(1)): $x_t = \phi\, x_{t-1} + \varepsilon_t$,

Recursive solution (Wold form):

$$x_t = \phi^{t+1}\, x_{-1} + \textstyle\sum_{j=0}^{\infty} \phi^j\, \varepsilon_{t-j} = \phi^{t+1}\, x_{-1} + \sum_{j=0}^{\infty} \psi_i\, \varepsilon_{t-j}$$

where x_{-1} is an initial condition.

ARMA Process – Dynamic Multiplier

• The dynamic multiplier measurers the effect of ε_t on subsequent values of x_t: That is, the first derivative on the Wold representation:

$$\delta x_{t+j}/\delta\varepsilon_t = \delta x_j/\delta\varepsilon_0 = \psi_j.$$

For an AR(1) process:

$$\delta x_{t+j}/\delta\varepsilon_t = \delta x_j/\delta\varepsilon_0 = \phi^j$$

• That is, the dynamic multiplier for any linear SDE depends only on the length of time j, not on time t.

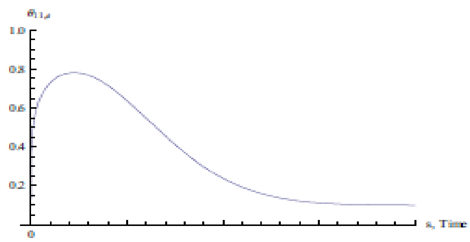
ARMA Process – Impulse Response Function

• The *impulse-response function* (IRF) gives a sequence of dynamic multipliers as a function of time from the one time change in the innovation, ε_t.

• Usually, IRF are represented with a graph, that measures the effect of the innovation, ε_t, on y_t over time:

$\delta y_{t+j}/\delta\varepsilon_t + \delta y_{t+j+1}/\delta\varepsilon_t + \delta y_{t+j+2}/\delta\varepsilon_t + ... = \psi_j + \psi_{j+1} + \psi_{j+2} +$

• Once we estimate the ARMA coefficients, it is easy to draw an IRF.



ARMA Process – Addition

• Q: We add two ARMA process, what order do we get?

• Adding MA processes

$$x_t = A(L)\,\varepsilon_t$$
$$z_t = C(L)\,u_t$$
$$y_t = x_t + z_t = A(L)\,\varepsilon_t + C(L)\,u_t$$

- Under independence:

$$\gamma_y(j) = E[y_t y_{t-j}] = E[(x_t + z_t)(x_{t-j} + z_{t-j})]$$
$$= E[(x_t x_{t-j} + z_t z_{t-j})] = \gamma_x(j) + \gamma_z(j)$$

- Then, $\gamma(j) = 0$ for $j >$ Max(q_x, q_z) $\Rightarrow y_t$ is ARMA(0, *max*(q_x, q_z))

- Implication: MA(2) + MA(1) = MA(2)

$$\phi(L)\;x_t = \theta(L)\;\varepsilon_t$$

ARMA Process – Addition

• Q: We add two ARMA process, what order do we get?

• Adding AR processes

$$(1 - A(L))\ x_t = \varepsilon_t$$
$$(1 - C(L))\ z_t = u_t$$
$$y_t = x_t + z_t = ?$$

- Rewrite system as:

$$\big(1 - C(L)\big)(1 - A(L))\ x_t = (1 - C(L))\varepsilon_t$$
$$\big(1 - A(L)\big)(1 - C(L))\ z_t = (1 - A(L))u_t$$
$$\big(1 - A(L)\big)(1 - C(L))\ y_t = x_t + z_t = (1 - C(L))\varepsilon_t + (1 - A(L))u_t$$
$$= \varepsilon_t + u_t - [C(L)\ \varepsilon_t + A(L))u_t]$$

- Then, y_t is ARMA($p_x + p_z$), max(p_x, p_z).

ARMA Process: Identification and Estimation

• We defined the ARMA(p, q) model:

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

The mean does not affect the order of the ARMA. Then, if $\mu \neq 0$, we demean the data: $x_t = y_t - \mu$.

Then, $\phi(L)\ x_t = \theta(L)\ \varepsilon_t \quad \Rightarrow\ x_t$ is a *demeaned* ARMA process.

• Next lecture, we will study:
  - Identification of p, q.
  - Estimation of ARMA(p, q)