Financial Econometrics

Slides-04: ARMA models
Genera; Linear Process and characterization

Dr. Rachida Ouysse
School of Economics[1]

Plan.

Time Series Models (Mainly Theoretical Aspects)

- View time series as stochastic processes
- Notions of stationarity (Covariance Stationary)
- Models for stationary time series
  - General linear process (GLP): useful representation especially for computing expectations...
- Characteristics of models
  - Patterns in the AC and PAC of a model: White noise

Motivation

- Describe empirically relevant patterns in the data ??
- Obtain the distribution of future values, conditional on the past, in order to forecast the future values and evaluate the likelihood of certain events ??
- Provide insight in possible sources of non-stationarity

Characteristics of a Time Series

A **time series** $y_1, \ldots, y_T$ is a sequence of values a specific variable y has taken on at equal distances (e.g. daily, quarterly, yearly, ...) over some period of time.

These observations will be considered as being generated by some **stochastic Data Generating Process** (DGP)

- A time series $y_1, \ldots, y_T$ is generated by a stochastic process y_t, for $t = 1, \ldots, T$.
- A time series $y_1, \ldots, y_T$ is a collection of realizations of a random variable y_t ordered in time.

Univariate Time Series Models

A **time series model** tries to describe the stochastic process y_t b a relatively simple model. **Univariate time series models** are a class of models where one attempts to model and predict (economic) variables using only information contained in their ow past values and possibly current and past values of an error term.

These models are (mainly) atheoretical:

- not based upon any underlying theoretical model
- attempt to capture empirically relevant patterns in the data

$\Leftrightarrow$ **Structural models**

- generally based upon any underlying theoretical model
- attempt to model a variable from the current and/or past values of other explanatory variables (suggested by theory)

Defining stationarity and non-stationarity

A series y_t is **strictly stationary** if the distribution of its values is not affected by an arbitrary shift along the time axis:

$$f\left(y_t\right) = f\left(y_{t+k}\right) \qquad \forall k \tag{1}$$

$\rightarrow$ The entire distribution of y_t is not affected by an arbitrary shift along the time axis. See for example Figure 1.

A series y_t is **covariance** or **weakly stationary** if it satisfies:

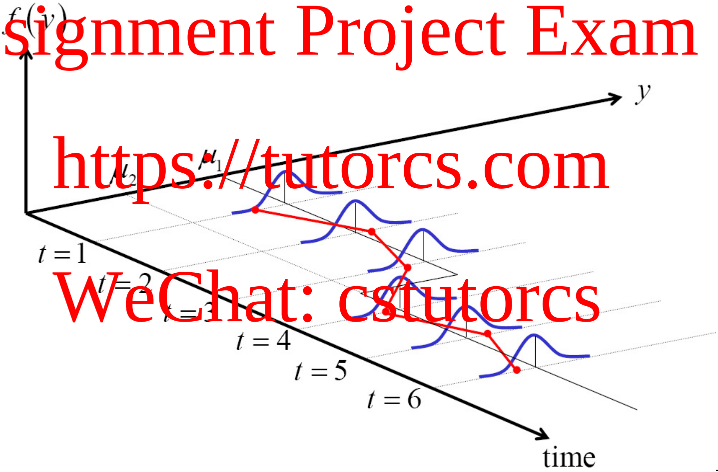
- $E\left(y_t\right) = \mu < \infty$
- $Var\left(y_t\right) = E\left(y_t - \mu\right)^2 = \sigma^2 < \infty$
- $Cov\left(y_t, y_{t-k}\right) = E\left(y_t - \mu\right)\left(y_{t-k} - \mu\right) = \gamma_k \quad \forall k$

$\rightarrow$ The first and the second moment of the distribution of y_t are finite and not affected by an arbitrary shift along the time axis.

- After being hit by a shock, a stationary series tends to return to its mean (called **mean reversion**) and fluctuations around this mean (measured by the variance) will have a broadly constant amplitude.

- If a time series is not stationary in the sense defined above, it is called **non-stationary**, i.e. non-stationary series will have a time-varying mean and/or a time-varying variance and/or time-varying covariances.
  - Non-stationarity can have different sources: linear trend, structural break, unit root, ...

Figure 5 : A non-stationary process (structural break)



- If the dependence structure is stable (stationary), it can be learned from historical data.
- Strict Stationarity
  - A time series is strictly stationary (SS) if its joint distribution at any set of points in time is invariant to any time shift.

$$\text{eg. dist}(y_{t_1}, y_{t_2}) = \text{ dist}(y_{t_1+s}, y_{t_2+s})$$

- Covariance Stationarity
  - A time series is covariance stationary (CS) if its mean, variance and autocovariance are all independent of the time index t, and its variance is finite.

$$E(y_t) = \mu, Var(y_t) = \gamma_0 < \infty, Cov(y_t, y_{t-j}) = \gamma_j \text{ for all } j$$

Assuming covariance stationarity, particular useful tools when building ARMA models are the so-called Autocorrelation and Partial Autocorrelation Function.

In general, the joint distribution of all values of y_t is characterised by the so-called **autocovariances**, i.e. the covariances between y_t and all of its lags y_{t-k}.

The **sample autocovariances** γ_k can be obtained as

$$\gamma_k = cov(y_t, y_{t-k}), \qquad k = 1, 2, \ldots \tag{2}$$

$$= \frac{1}{T-k} \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y}), \tag{3}$$

where $\bar{y} = T^{-1} \sum_{t=1}^{T} y_t$ is the sample mean.

As the autocovariances are not independent of the units in which the variables are measured, it is common to standardize by defining **autocorrelations** ρ_k as

$$\rho_k = \frac{cov\left(y_t, y_{t-k}\right)}{var\left(y_t\right)} = \frac{\gamma_k}{\gamma_0}. \tag{4}$$

Note that $\rho_0 = 1$ and $-1 \leq \rho_k \leq 1$.

The autocorrelations ρ_k considered as a function of k are referred to as the **autocorrelation function** (ACF) or correlogram of the series y_t. The ACF provides useful information on the properties of the DGP of a series as it describes the dependencies among observations.

If the data are generated from a **stationary process**, it can be shown that under the null hypothesis:

$$H_0 : \rho_k = 0 \quad \forall k > 0$$

the sample autocorrelation coefficients are **asymptotically normally distributed** with mean zero and variance $\frac{1}{T}$.

Therefore, in finite sample it holds:

$$\rho_k \sim N\left(0, T^{-1}\right)$$

The **individual significance** of an autocorrelation coefficient can be tested by constructing the 95% confidence interval:

$$\left[-1.96/\sqrt{T}; 1.96/\sqrt{T}\right]$$

Autocorrelation Function

Looking at a large number of autocorrelations, we will see that some exceed two standard deviations as a result of pure chance even though the true values in the DGP are zero (Type I error).

The **joint significance** of a group of m autocorrelation coefficients can be tested by the so-called Box-Pierce Q-statistic:

$$Q = T \sum_{k=1}^{m} \rho_k^2 \tag{5}$$

If the data are generated from a stationary process, Q is asymptotically χ^2 distributed with m degrees of freedom.

Superior small sample performance is obtained by modifying the q statistic (reported in EViews output):

$$Q^* = T(T+2) \sum_{k=1}^{m} \rho_k^2 / (T-k) \tag{6}$$

- An alternative piece of information is provided by the so-called partial autocorrelation function (PACF). The partial autocorrelation p_j s the correlation between y_t and $y_{t?k}$ conditional on $y_{t?1}, \cdots, y_{t?k+1}$. It measures the dependency between y_t and $y_{t?k}$ keeping constant in-between values.

- The sample partial autocorrelations can be calculated from OLS regressions:
  - $\hat{p}_1 = \hat{\phi}_{11}$ in $y_t = \phi_{10} + \phi_{11} y_{t-1} + e_{1t}$
  - $\hat{p}_2 = \hat{\phi}_{22}$ in $y_t = \phi_{20} + \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + e_{2t}$
  - $\hat{p}_3 = \hat{\phi}_{33}$ in $y_t = \phi_{30} + \phi_{31} y_{t-1} + \phi_{32} y_{t-2} + phi_{33} y_{t-2} + e_{3t}$

  $\ldots$

Defining a White Noise Process

A series y_t is called a **white noise process** if its DGP has a constant mean, a constant variance and is serially uncorrelated. Formally:

$$E(y_t) = E(y_{t-1}) = \ldots = \mu$$

$$Var(y_t) = Var(y_{t-1}) = \ldots = \sigma^2$$

$$Cov(y_t, y_{t-k}) = Cov(y_{t-j}, y_{t-j-k}) = \begin{cases} \sigma^2 & \text{if} \quad k = 0 \\ 0 & \text{otherwise} \end{cases}$$

y_t is a **zero-mean white noise process** if $\mu = 0$.

Defining a White Noise Process

- A time series ϵ_t is a white noise if its is covariance stationary with zero mean and no autocorrelation.
  - By definition:

$$E(\epsilon_t) = 0, Var(\epsilon_t) = \sigma^2, Cov(\epsilon_j, \epsilon_{t-j}) = 0, \text{ for all} j \neq 0.$$

  - A white noise is denoted as: $y_t \sim WN(0, \sigma^2)$
    A white noise is not necessarily $i.i.d$ (independent and identically distributed)
  - An $i.i.d$ white noise is denoted as: i.i.d $WN(0, \sigma^2)$
  - , White noises are building blocks of time series models.

- Key feature of a $WN(0, \sigma^2)$ is H_0: **no autocorrelation**
- The sampling distribution of the ACF and PACF for a WN is approximately $N(0, 1/T)$

▷ Reject H_0

if either ACF or PAC is outside the $\pm 1.96/\sqrt{T}$ bands; or

the Ljung-Box Q-stats have small p-values.

eg. NYSE Composite return squared r_t^2.

Correlogram of RC2

Sample: 1 1931
Included observations: 1930

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.179	0.179	62.044	0.000
		2	0.168	0.140	116.47	0.000
		3	0.175	0.131	176.00	0.000
		4	0.109	0.044	198.89	0.000
		5	0.184	0.130	264.80	0.000
		6	0.113	0.035	289.54	0.000
		7	0.136	0.067	325.24	0.000
		8	0.112	0.030	349.60	0.000
		9	0.108	0.039	372.11	0.000
		10	0.095	0.015	389.80	0.000
		11	0.107	0.042	411.90	0.000
		12	0.102	0.028	432.21	0.000
		13	0.066	-0.006	440.72	0.000
		14	0.053	-0.016	446.26	0.000
		15	0.041	-0.016	449.58	0.000
		16	0.097	0.052	467.86	0.000
		17	0.040	-0.018	470.96	0.000
		18	0.068	0.024	479.94	0.000
		19	0.063	0.014	487.77	0.000
		20	0.045	0.005	491.70	0.000

- **Why general linear process?**
  - ▷ **Wold decomposition.** Any covariance stationary process can be expressed as a general linear process.

$$y_t = \mu + \sum_{i=0}^{\infty} b_i \epsilon_{t-i}, \epsilon_t \sim WN(0, \sigma^2)$$

  - Because $b_i \to 0$ as $i \to \infty$,it is possible to use finite parameters to characterise CS time series. This leads to practical (parsimonious) models (ARMA).

- **Will mainly consider the cases with iid WN in this topic, for which "conditional"="unconditional"**
  - $E(\epsilon_t|\epsilon_{t-j}) = E(\epsilon_t)$ (ϵ_t is not predictable)
  - $Var(\epsilon_t|\epsilon_{t-j}) = Var(\epsilon_t)$ for all $j = 1, 2, , 3, \cdots$

General linear Process

- **Conditional Expectations**
  - The general linear process with i.i.d WN:

$$y_t = \mu + \sum_{i=0}^{\infty} b_i \epsilon_{t-i}, \epsilon_t \sim WN(0, \sigma^2)$$

  - Let Ω_t be the information set based on

$$\{y_t, y_{t-1}, \cdots, \epsilon_t, \epsilon t - 2, \cdots\}$$

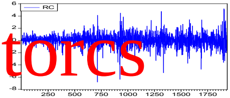
- Conditional mean and variance of y_{t+h} for $h = 1, 2, \cdots$:
  - $E(y_{t+h}|\Omega_t) = \mu + \sum_{i=h}^{\infty} b_i \epsilon_{t+h-i}$,
  - $Var(y_{t+h}|\Omega_t) = \sigma^2 \sum_{i=0}^{h-1} b_i^2$
  - What happens when $h \to \infty$?
  - ▷ Limited memory: info at t is not relevant to remote future.

- $y_{t+h} = \mu + \underbrace{b_0\varepsilon_{t+h} + \cdots + b_{h-1}\varepsilon_{t+1}}_{\text{not in } \Omega_t} + \underbrace{b_h\varepsilon_t + b_{h+1}\varepsilon_{t-1} + \cdots}_{\text{in } \Omega_t}.$
- $E(y_{t+h}|\Omega_t) = \mu + \sum_{i=h}^{\infty} b_i\varepsilon_{t+h-i}$,
- $\text{Var}(y_{t+h}|\Omega_t) = \sigma^2 \sum_{i=0}^{h-1} b_i^2$.

ε_{t+1} = 1-step forecast error

eg. When $h = 2$,

- $y_{t+2} = \mu + \underbrace{b_0\varepsilon_{t+2} + b_1\varepsilon_{t+1}}_{\text{not in } \Omega_t} + \underbrace{b_2\varepsilon_t + b_3\varepsilon_{t-1} + \cdots}_{\text{in } \Omega_t},$
- $E(y_{t+2}|\Omega_t) = \mu + b_2\varepsilon_t + b_3\varepsilon_{t-1} + \cdots$,
- $\text{Var}(y_{t+2}|\Omega_t) = \sigma^2(b_0^2 + b_1^2)$.



Conditional variance is smaller than unconditional variance. Variance being constant, not ideal to capture the **clustering** in return series. Need ARCH-type model.

- Use the information set Ω_t to forecast y_{t+h} for $h \geq 1$. Let $f_{t+h|t}$ be the forecast based on Ω_t.

- Choose $f_{t+h|t}$ to minimise the MSFE

$$\text{MSFE} = E[(y_{t+h} - f_{t+h|t})^2 | \Omega_t].$$

- The optimal point forecast is

$$f^*_{t+h|t} = E(y_{t+h}|\Omega_t).$$

Forecast error:

$$e_{t+h|t} = y_{t+h} - f^*_{t+h|t}\,,$$
$$\text{Var}(e_{t+h|t}|\Omega_t) = \text{Var}(y_{t+h}|\Omega_t).$$

- If μ, b_i, σ^2 are known, the 2-se interval forecast is

$$E(y_{t+h}|\Omega_t) \pm 2\sqrt{\text{Var}(y_{t+h}|\Omega_t)} \quad \text{or}$$

$$\left(\mu + \sum_{i=h}^{\infty} b_i \varepsilon_{t+h-i}\right) \pm 2\left(\sigma^2 \sum_{i=0}^{h-1} b_i^2\right)^{1/2}.$$

Summary: What to take from this lecture?

1. White noise is the building block of time series models
2. In order to model the dynamics of a time series, use the white noise process to piece together the dynamics: GLP

3. GLP useful representation: compute expectations, variance and ACF
4. Special models: AR, MA