Forecasting & Predictive Analytics

Guillaume Chevillon
chevillon@essec.edu

ESSEC Business School

October-December 2017
3rd set of slides
Univariate Time Series

Overview

- Univariate Time Series
- ARMA models

STOCHASTIC UNIVARIATE TIME SERIES

1. Stationarity
2. Autocovariance and Partial autocovariance

Stochastic Processes

- Stochastic processes

$$\{y_t\}$$

- Examples
  - IID

$$y_t \stackrel{iid}{\sim} \mathsf{N}(0,1)$$

  - Random walk

$$y_t = y_{t-1} + \epsilon_t$$

  - ARMA$(1,1)$

$$y_t = \phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$$

  - GARCH$(1,1)$

$$\begin{aligned} y_t &\sim \mathsf{N}(0,h_t) \\ h_t &= \omega + \alpha y_{t-1}^2 + \beta h_{t-1} \end{aligned}$$

  - Many more....
- Today's focus: ARMA

1) Stationarity

- Key concept
- Stationarity is statistically meaningful form of regularity
- Two types:
  Definition: Covariance Stationarity
  A stochastic process $\{y_t\}$ is covariance (or weakly) stationary if

$$\begin{aligned} \mathsf{E}\left[y_t\right] &= \mu \quad \forall t \\ \mathsf{V}\left[y_t\right] &= \sigma^2 \quad \forall t \\ \mathsf{E}\left[\left(y_t - m\right)\left(y_{t-s} - \mu\right)\right] &= \gamma_s \quad \forall t \end{aligned}$$

  where σ is finite
  Definition: Strict Stationarity
  A stochastic process $\{y_t\}$ is strictly stationary if the joint distribution of $\{y_t, y_{t-1}, ..., y_{t-h}\}$ only depends on h and not on t.

2) Stationarity examples

- Stationary time series
  - IID: always strict, covariance if $\sigma^2 < \infty$
  - $\mathrm{AR}(1): y_t = \phi y_{t-1} + \epsilon_t$, strict if $|\phi| < 1$, covariance if $\mathsf{V}[\epsilon_t] < \infty$
  - $\mathrm{ARCH}(1): y_t \sim \mathsf{N}(0, h_t), h_t = \omega + \alpha y_{t-1}^2$, both if $\alpha < 1$
- Non-stationary time series
  - linearly trending: $y_t = \alpha + \beta t + \epsilon_t$
    - $\mathsf{E}[y_t] = \alpha + \beta t$
  - Random walks: $y_t = y_{t-1} + \epsilon_t$
    - $\mathsf{V}[y_t] = \sigma^2 t$
  - Models with structural breaks: $y_t = \mu_1 + \epsilon_t$ if $t < 1000$, $y_t = \mu_2 + \epsilon_t, t \geq 1000$.
    - $\mathsf{E}[y_t] = \mu_1 + I_{\{t \geq 1000\}}(\mu_2 - \mu_1)$

Ergodicity

measure of asymptotic independence

Theorem: Ergodic theorem

- If $\{y_t\}$ is ergodic and the r^{th} moment is finite, then

$$T^{-1}\sum_{t=1}^{T} y_t^r \xrightarrow{p} \mu_r$$

non-ergodic

$$y_t = u_0 + \epsilon_t$$

Definitions

- White noise: the simplest process with no inertia (memoryless)
- Linear 1-st order difference equation:

$$y_t = \phi y_{t-1} + w_t$$

- Dynamic multiplier: the effect of a change in w_t on y_{t+j} (i.e., j periods ahead), holding everything else fixed (including y_{t-1}) is

$$\frac{\partial y_{t+j}}{\partial w_t} = \phi^j.$$

  The quantity $\partial y_{t+j} / \partial w_t$ is called the (jth) *dynamic multiplier* of w_t on y_t. It represents the response of the process $\{y_t\}$ to a temporary change or 'impulse' in w_t. Thus, $\partial y_{t+j} / \partial w_t$ as a function of j is also referred to as the *impulse response function.*
- Lag operator:

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

ARMA MODELS

1. Wold Decomposition Theorem
2. AR Models
3. MA Models
4. Asymptotic Theory & Estimation
5. ARMA
6. Box-Jenkins Methodology

Stationarity

- $\{y_t\}$ is **weakly** or **covariance** stationary if the first and second moments of the process exist and are time-invariant.

$$\mathsf{E}\left[y_t\right] = \mu < \infty \quad \forall t$$
$$\mathsf{E}\left[\left(y_t - \mu\right)\left(y_{t-h} - \mu\right)\right] = \gamma_t\left(h\right) = \gamma_t\left(-h\right) = \gamma\left(h\right) < \infty \quad \forall t$$

- $\{y_t\}$ is **strictly** stationary if for any values of $h_1, h_2, ..., h_n$ the joint distribution of $y_t, y_{t+h_1}, ..., y_{t+h_n}$ depends only on the intervals $h_1, ..., h_n$ and not on t:

$$f\left(y_t, y_{t+h_1}, ..., y_{t+h_n}\right) = f\left(y_\tau, y_{\tau+h_1}, ..., y_{\tau+h_n}\right) \quad \forall t, \tau$$

- Linear Filter transforms an input series $\{x_t\}$ into an output series $\{y_t\}$ using a lag polynomial $A(L)$:

$$y_t = A(L)\, x_t = \left(\sum_{j=-n}^{m} a_j L^j \right) x_t = \sum_{j=-n}^{m} a_j x_{t-j}$$
$$= a_{-n} x_{t+n} + ... + a_0 x_t + ... + a_m x_{t-m}$$

- Linear Process

$$y_t = A(L)\, \epsilon_t = \left(\sum_{j=-n}^{m} a_j L^j \right) \epsilon_t = \sum_{j=-n}^{m} a_j \epsilon_{t-j}$$

where $\epsilon_t \sim \mathsf{WN}\left(0, \sigma^2\right)$.
Note: for $|x| < 1$

$$1 + x + x^2 + ... + x^n = \sum_{j=0}^{n} x^j = \frac{1 - x^{n+1}}{1 - x} \to \frac{1}{1-x} = \sum_{j=0}^{\infty} x^j$$

Wold Decomposition Theorem

Wold Decomposition– any zero-mean covariance stationary process $\{y_t\}$ can be represented in the form:

$$y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + \kappa_t$$

where $\psi_0 = 1$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.
ϵ_t is white noise and represents the error made in forecasting y_t based on a linear function of its past $Y_{t-1} = \left\{y_{t-j}\right\}_{j=1}^{\infty}$:

$$\epsilon_t = y_t - \mathsf{L}\left(y_t | Y_{t-1}\right)$$

κ_t is a deterministic term $\kappa_t = \mathsf{L}\left(\kappa_t | Y_{t-1}\right)$.

Box-Jenkins approach

- Approximate the infinite lag polynomial with the ratio of two finite-order polynomials $\phi(L)$ and $\theta(L)$:

$$\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j \approx \frac{\theta(L)}{\phi(L)} = \frac{1 + \theta_1 L + ... + \theta_q L^q}{1 - \phi_1 L - ... - \phi_p L^p}$$

- Type of time series models

Type	Model	p	q
$AR(p)$	$\phi(L)\, y_t = \epsilon_t$	$p > 0$	$q = 0$
$MA(q)$	$y_t = \theta(L)\, \epsilon_t$	$p = 0$	$q > 0$
$ARMA(p, q)$	$\phi(L)\, y_t = \theta(L)\, \epsilon_t$	$p > 0$	$q > 0$

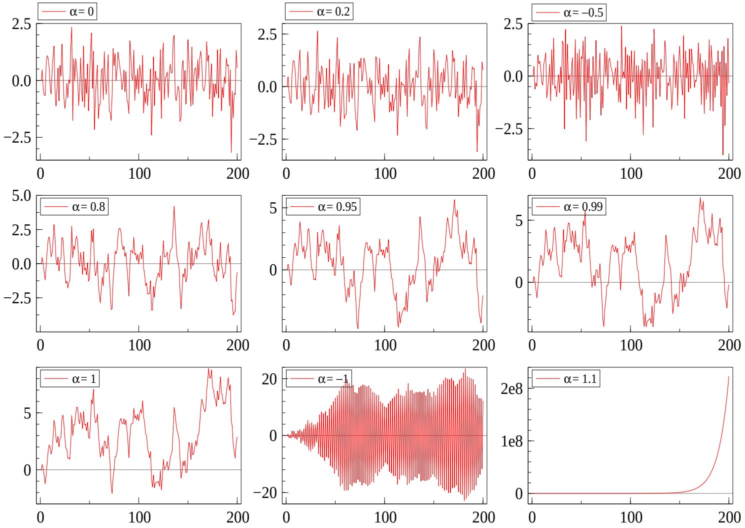
1) Autoregressive processes

- AR processes: univariate and companion form, stability
- Autocovariance:

$$\gamma_t(k) = \mathsf{E}\left[(y_t - \mathsf{E}[y_t])(y_{t-k} - \mathsf{E}[y_{t-k}])\right]$$

  where $\gamma_t(k) \equiv \gamma_k$ if $\mathsf{E}[y_t] \equiv \mu$
- The AR(1) : expectation, variance, backward substitution, lag-polynomial
- AR(p) : stationarity, invertibility, long-run multiplier
- ACF and PACF
- Estimation

Stability of AR(1) processes

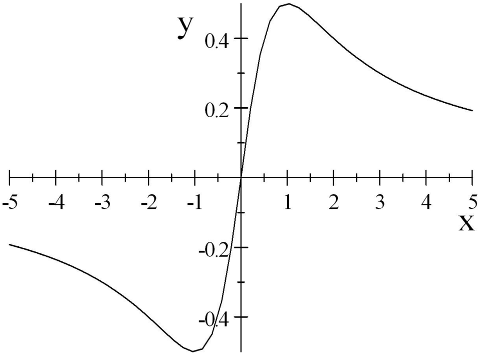


Simulated time series from an AR(1) $y_t = \alpha y_{t-1} + \varepsilon_t$ for various values of α.

2) Moving average processes

- MA processes: $\mathrm{MA}(1)$, $\mathrm{MA}(q)$, $\mathrm{MA}(\infty)$
- Expectation and variance
- Stationarity, invertibility and long term multiplier
- Estimation: maximum likelihood

MA(1) cannot yield first-order autocorrelation greater than 0.5:



MA(1) model $y_t = \epsilon_t + \theta\epsilon_{t-1}$ θ is on the x-axis, and $\rho_1 = \theta / \left(1 + \theta^2\right)$ is on the y-axis.

3) Asymptotic theory and estimation

Theorem

Let Y_t be a covariance stationary process with absolutely summable autocovariances. Then

1. *LLN:* $\overline{Y}_T \overset{m.s.}{\rightarrow} \mu$
2. $\lim_{T\to\infty} \left\{ T\, \mathsf{E}\left[\left(\overline{Y}_T - \mu \right)^2 \right] \right\} = \sum_{j=-\infty}^{+\infty} \gamma_j$ *(long run variance).*

Any ARMA process has absolutely summable autocovariances, and hence follows a LLN. Besides, it follows a CLT.

Theorem

Let $Y_t - \mu \sim MA(\infty)$, then

$$\sqrt{T}\left(\overline{Y}_T - \mu \right) \overset{L}{\rightarrow} \mathsf{N}\left(0, \sum_{j=-\infty}^{+\infty} \gamma_j \right).$$

OLS estimation

Hence OLS estimation of a covariance stationary AR(p) process is asymptotically consistent. Indeed the OLS estimator $\widehat{\Phi}$ of Φ in

$$\begin{aligned} Y_t &= c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{WN}\left(0, \sigma^2\right) \\ &= \left(1, Y_{t-1}, ..., Y_{t-p}\right) \Phi + \varepsilon_t \end{aligned}$$

is biased, yet not asymptotically so, since

$$\sqrt{T}\left(\widehat{\Phi} - \Phi\right) \xrightarrow{L} \mathsf{N}\left(0, \sigma^2 \mathbf{Q}^{-1}\right)$$
$$\mathbf{Q} = \mathsf{E}\left[\left(1, Y_{t-1}, ..., Y_{t-p}\right)' \left(1, Y_{t-1}, ..., Y_{t-p}\right)\right]$$

Maximum Likelihood Estimation of Gaussian MA processes

- Likelihood function coincides with the Joint density:

$$\mathcal{L}\left(\Phi; y_1, \ldots, y_T\right) = f_{Y_1,\ldots,Y_T}\left(y_1, \ldots, y_T; \Phi\right). \tag{1}$$

  but as a function of the parameter Φ
- Independence of y_t implies it can be factored

$$\mathcal{L}\left(\Phi; y_1, \ldots, y_T\right) = \prod_{t=1}^{T} f_{Y_t}\left(y_t; \Phi\right)$$

- At least conditionally

$$\mathcal{L}\left(\Phi; y_1, \ldots, y_T\right) = \prod_{t=p+1}^{T} f_{Y_t|\mathbf{Y}_{t-1}}\left(y_t|\mathbf{y}_{t-1}; \Phi\right) f_{\mathbf{Y}_p}\left(\mathbf{y}_p; \Phi\right)$$

- application to $\mathrm{AR}(1)$, $\mathrm{AR}(p)$, $\mathrm{MA}(1)$, $\mathrm{MA}(q)$

Wold's decomposition theorem: Any covariance-stationary process Y_t can be represented in the form

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t, \tag{2}$$

where $\psi_0 = 1$ and $\{\psi_j\}$ is absolutely summable. The term ε_t is white noise, and equals the forecast error associated with the optimal linear one-step-ahead forecast of Y_t namely,

$$\varepsilon_t = Y_t - \mathsf{L}\left[Y_t | Y_{t-1}, Y_{t-2}, \ldots\right].$$

The term κ_t can be predicted arbitrarily well by a linear function of *past* values of Y, and is uncorrelated with ε_{t-j} for any j.

ARMA models

Emergence

- Approximate $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ via the parsimonious rational fraction

$$\psi(L) = \frac{1 + \theta_1 L + \ldots + \theta_q L^q}{1 - \phi_1 L - \ldots - \phi_p L^p}$$

- Hence

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

  is an ARMA(p, q).
- Common factors
- Stationarity and invertibility
- Estimation via Maximum Likelihood
- ACF and PACF

5) Box-Jenkins Methodology

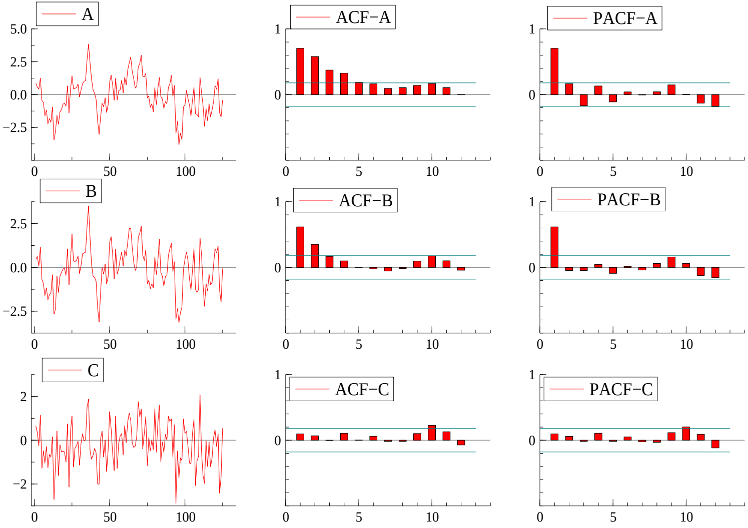
1. Transform the data if necessary, e.g., take differences, so that the assumption of covariance stationarity is a reasonable one.
2. Plot the (possibly transformed) series together with its correlogram and partial correlogram, ACF and PACF. Use this to make an initial guess about p and q.
3. Estimate the parameters in $\phi(L)$ and $\theta(L)$ of an ARMA(p, q).
4. Perform diagnostic tests to confirm the model is consistent with the observed data, e.g., test that the fitted errors $\widehat{\varepsilon}_t$ are white noise.
5. If necessary, select the model using Information criteria.

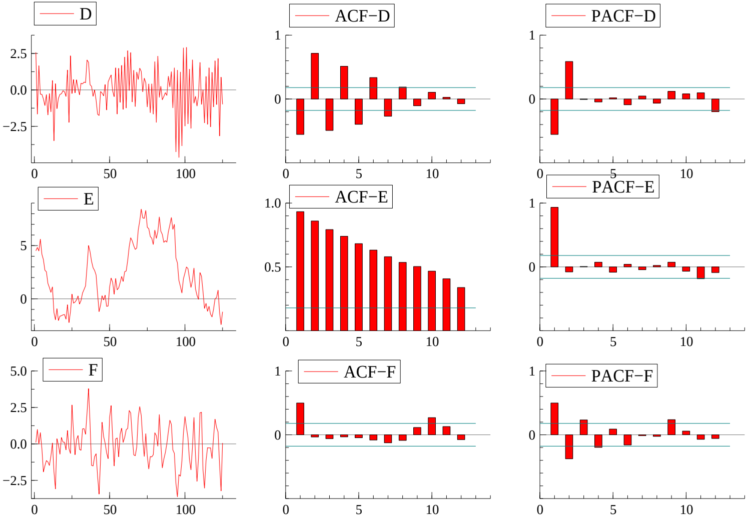
Example

Let $u_t \sim \mathsf{GWN}\left(0, \sigma^2\right)$. The following $\{y_t\}$ stochastic processes are driven a common innovation process $\{u_t\}$:

$$
\begin{aligned}
(i) \quad & y_t = 0.64 y_{t-1} + u_t \\
(ii) \quad & y_t = -0.2 y_{t-1} + 0.64 y_{t-2} + u_t \\
(iii) \quad & y_t = u_t + u_{t-1} \\
(iv) \quad & y_t = 0.64 y_{t-2} + u_t + 0.64 u_{t-1} \\
(v) \quad & y_t = y_{t-1} + u_t \\
(vi) \quad & y_t = y_{t-1} + u_t - 0.9 u_{t-1}
\end{aligned}
$$

A realization of each of these is presented in figure **??**, together with estimated ACF and PACF. Can you tell which ARMA process generated which graph?

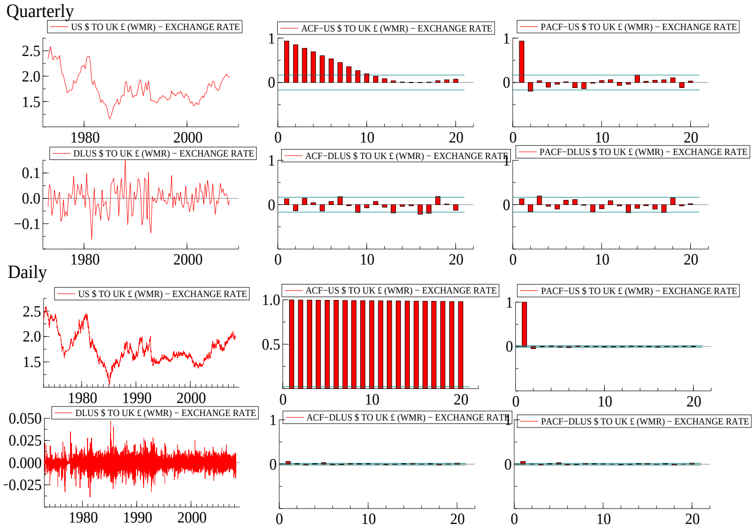

A
ACF−A
PACF−A
B
ACF−B
PACF−B
C
ACF−C
PACF−C
5.0
2.5
0.0
−2.5
2
0
−2
1
0
5
10
50
100


D
2.5
0.0
−2.5
0
50
100
ACF−D
1
0
0
5
10
PACF−D
1
0
0
5
10
E
5
0
0
50
100
ACF−E
1.0
0.5
0
5
10
PACF−E
1
0
0
5
10
F
5.0
2.5
0.0
−2.5
0
50
100
ACF−F
1
0
0
5
10
PACF−F
1
0
0
5
10

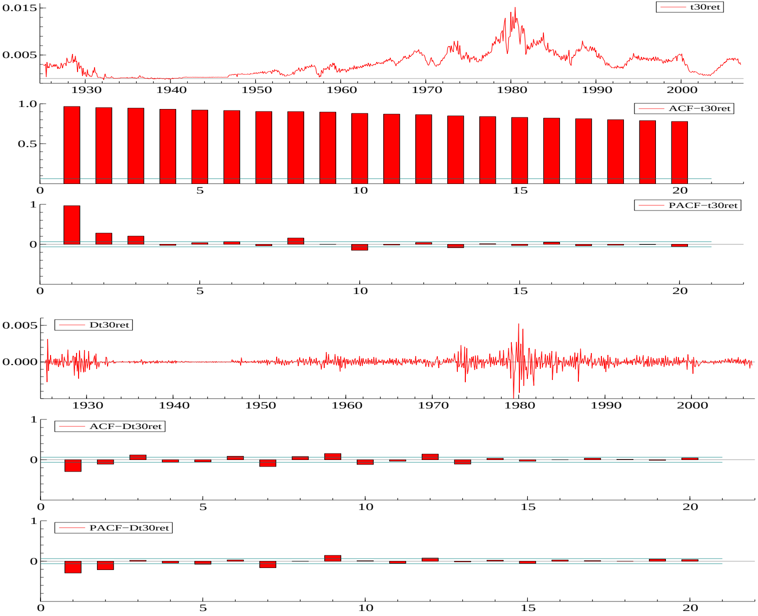
Data series

- monthly VWM
- Moody's Baa-Aaa monthly corporate bond spread (*the default spread*)
- Sterling-Dollar Exchange rate (daily/quarterly)
- US treasury's 30 day bill interest rate (monthly)

Exchange Rates



Interest Rates



Stock Absolute Returns