Master in Financial Engineering
Financial Econometrics

Case study 1: Modeling a financial time series, cointegration, and pair trading (Statarb)

Pr. Florian Pelgrin - T.A. Maxime Couvert

Ecole Polytechnique Federale de Lausanne

Feb. 2019 - June 2019

Case study 1

- Exercise 1 : Modeling the 5-year interest rate
- Exercise 2 : The term structure of interest rates
- Exercise 3 : Pair trading or Statarb

Exercise 1 : Modeling the 5-year interest rate

Road map of exercise 1

1. Exercise 1
   - Quick reminder of the theory
   - Step 1 : Identification of orders
     - Question 1.
     - Question 2.a.
     - Question 2.b.
     - Question 3.
   - Step 2 : Estimation and selection of ARIMA Models
     - Question 4.
     - Question 5.
   - Step 3 : Diagnostic tests
     - Question 6.

Aim of the exercise

- Analyzing the main characteristics of financial time series
- Modeling and estimating these series
- Forecasting financial time series

Setting of the exercise

- X_t : 5-year interest rate series
- Period : 1960Q1-2013Q3
- Goal : propose an ARIMA specification
- Three steps :
  1. (Partial) identification of the autoregressive order (p), the moving average order (q) and the differentiation order (d) ;
  2. Estimation and selection of the ARIMA models ;
  3. Implementation of some diagnostic tests.

Quick reminder of the theory - stationarity

Definition

A stochastic process X_t is said to be **weakly stationary** if :

1) $\mathbb{E}(X_t) = \mu$ is independent of t;
2) $\mathbb{V}(X_t)$ is time invariant;
3) $\mathbb{C}\text{ov}(X_t, X_{t+h}) = \mathbb{E}\left[(X_t - \mu)\left(X_{t+h} - \mu\right)\right] = \gamma_X(h)$ is time invariant.

Quick reminder of the theory - white noise

Definition

A **white noise** is a stochastic process, (ϵ_t), such that :

$$1)\,\mathbb{E}(\epsilon_t) = 0 \text{ for all } t\,;$$
$$2)\,\mathbb{V}(\epsilon_t) = \sigma_\epsilon^2 < +\infty \text{ for all } t\,;$$
$$3)\,\mathbb{C}\text{ov}(\epsilon_t, \epsilon_{t-h}) = 0 \text{ for } h \neq 0$$

Quick reminder of the theory - ARIMA(p,d,q)

Definition

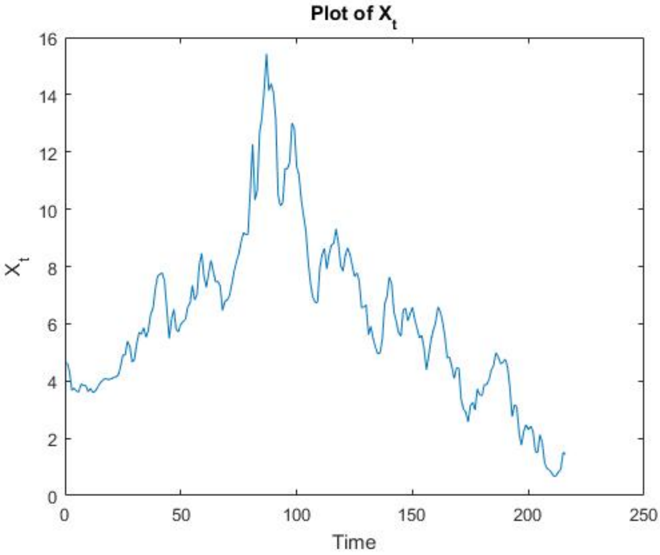
A stochastic process $(X_t)_{t\geq -p-d}$ is said to be an **ARIMA(p,d,q) - autoregressive integrated moving average - model** if it satisfies the following equation :

$$\Delta^d X_t = \mu + \phi_1 \Delta^d X_{t-1} + ... + \phi_p \Delta^d X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}$$

$$\Delta^d X_t = \mu + \sum_{k=1}^{p} \phi_k \Delta^d X_{t-k} + \sum_{k=0}^{q} \theta_k \epsilon_{t-k}$$

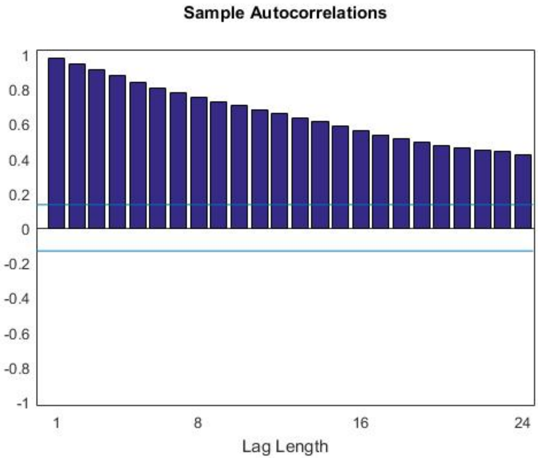
where d is the order of integration and $\Delta^d = (1-L)^d$ with L the lag operator.

Plot the time series X_t



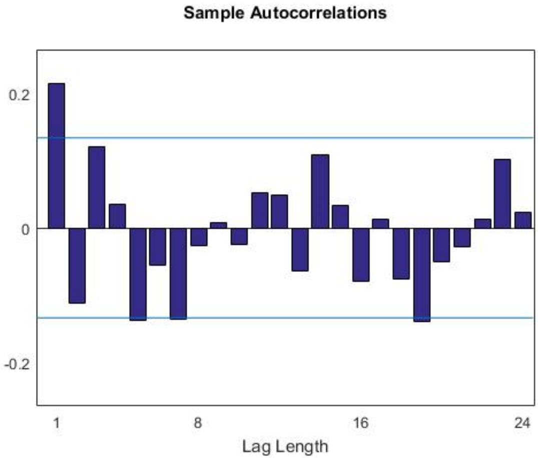
Display the autocorrelation function of (X_t) and ΔX_t

FIGURE – Autocorrelation of X_t



Display the autocorrelation function of (X_t) and ΔX_t

FIGURE – Autocorrelation of ΔX_t



Augmented Dickey-Fuller test (ADF)

- **Goal :** Assess whether there exists a root on the unit circle in the autoregressive lag polynomial of an AR(p) specification.
- **Design :** The ADF test is based on estimating the following test regression :

$$X_t = \sum_{j=1}^{p} \rho_j X_{t-j} + z_t' \delta + \epsilon_t$$

where (ϵ_t) is a weak white noise, z_t is a set of exogenous regressors ($z_t = \{1, t\}, z_t = \{t\},$ or $z_t = \{1\}$), p is the lag order (that can be determined by differen techniques) and $(\rho_1, \cdots, \rho_p)'$ and δ are parameters to estimate.

Augmented Dickey-Fuller test (ADF)

- **Design :** Using the so-called Beveridge-Nelson decomposition, the test regression can be rewritten as follows

$$X_t = (1 - \Phi(1))\, X_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta X_{t-j} + z_t' \delta + \epsilon_t$$

where $\Phi(1) = 1 - \sum_{j=1}^{p} \phi_j$.

- **Test specification :**

$$H_0 : \text{The series is non-stationary} \rightarrow \Phi(1) = 1$$
$$H_a : \text{The series is "weakly stationary"} \rightarrow \mid \Phi(1) \mid < 1.$$

Augmented Dickey-Fuller test (ADF)

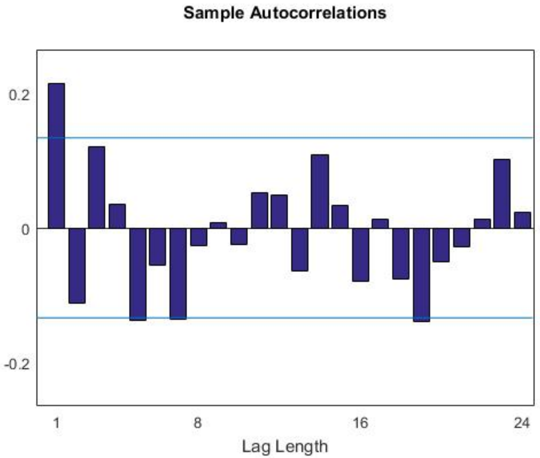
- **Assumptions :**
  - $z_t = 1$ for all t
  - p=5
- **Test :**

TABLE – ADF test

	X_t	ΔX_t
p-values	0.6277	0.001

Identification of (p ;q)

FIGURE – Autocorrelation Function of ΔX_t



Identification of (p ;q)

FIGURE – Partial Autocorrelation Function of ΔX_t

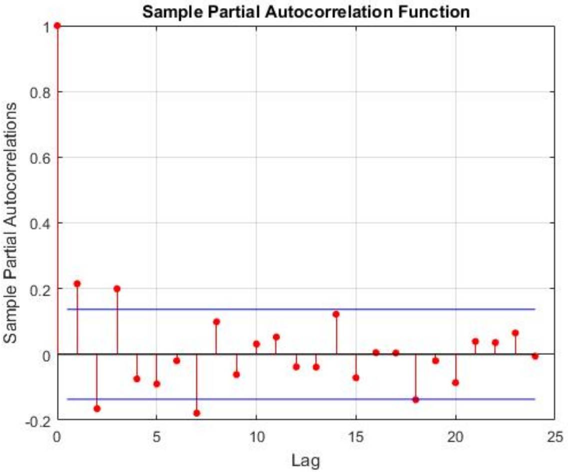


TABLE – AIC and BIC

Model	AIC	BIC
ARMA(1,1)	354.66	368.15
ARMA(1,2)	356.93	373.78
ARMA(1,3)	363.52	383.75
ARMA(1,4)	370.21	393.80
ARMA(2,1)	356.38	373.23
ARMA(2,2)	376.09	396.32
ARMA(2,3)	373.57	397.17
ARMA(2,4)	379.99	406.95
ARMA(3,1)	357.92	378.15
ARMA(3,2)	374.21	397.80
ARMA(3,3)	378.13	405.10
ARMA(3,4)	381.39	411.73
ARMA(4,1)	364.44	388.03
ARMA(4,2)	380.03	406.00
ARMA(4,3)	381.01	411.35
ARMA(4,4)	383.72	417.43

TABLE – Estimations : ARMA(1,1)

Parameter	Value	Std.Error	t-stat
Constant	-0.0193077	0.0602446	-0.320489
AR(1)	-0.275568	0.116785	-2.35961
MA(1)	0.577376	0.0985758	5.85718
Variance	0.293615	0.0190013	15.4524

TABLE – Estimations : ARMA(2,1)

Parameter	Value	Std.Error	t-stat
Constant	-0.017081	0.0492646	-0.34672
AR(2)	-0.147128	0.0469714	-3.13229
MA(1)	0.303605	0.0510443	5.94787
Variance	0.293227	0.0195068	15.032

TABLE – Estimations : ARMA(1,2)

Parameter	Value	Std.Error	t-stat
Constant	-0.0103245	0.0295008	-0.349974
AR(1)	0.299396	0.0505574	5.9219
MA(2)	-0.220013	0.0492289	-4.46918
Variance	0.293974	0.0192553	15.2672

Ljung-Box test for the first 10 lags