Time series Econometrics (Theory)

Romain Lafarguette, Ph.D. Amine Raboun, Ph.D.

Quants & IMF External Experts

romainlafarguette.github.io/ amineraboun.github.io/

Singapore Training Institute, 17 April 2023

Table of Contents

Why Time series Modeling is different from other statistical models

Why AR model is fundamentally different from an OLS?
Based on the model equation, AR is a special case of OLS where the explanatory variables are the lagged time series

$$AR(1): \quad y_t = \alpha + \beta \times y_{t-1} + \epsilon_t$$

$$OLS: \quad y_t = \alpha + \beta \times X_t + \epsilon_t$$

Why Time series Modeling is different from other statistical models

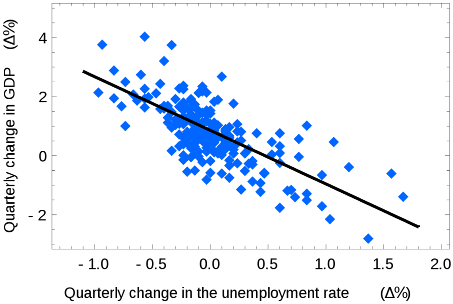
Why AR model is fundamentally different from an OLS?
Based on the model equation, AR is a special case of OLS where the explanatory variables are the lagged time series

$$AR(1): \quad y_t = \alpha + \beta \times y_{t-1} + \epsilon_t$$

$$OLS: \quad y_t = \alpha + \beta \times X_t + \epsilon_t$$

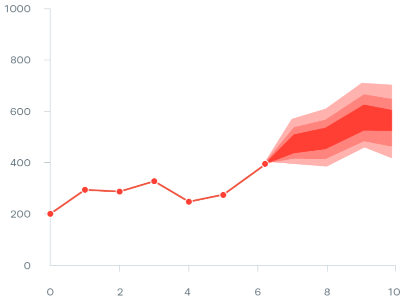
The hint to the answer is **Interpolation** vs **Extrapolation**

OLS: Interpolation



- Both X and Y are observed at the same time.
- For any potential value of X, we can compute a predicted value $\hat{Y}$ with a certain confidence

AR: Extrapolation



- Structural time dependence, $\hat{Y}_{t+1}$ depend on Y_t
- Predicting $\hat{Y}_{t+h}$ requires predicting all $\hat{Y}_{t+i} \forall i \in [1, h-1]$
- Cumulative estimation error make the results unusable

Intuitively, the farther we move in the future, the higher the chance of exogenous external shock occurring (Risk of structured breaks, innovation, macroeconomic shocks, news...) => the lower the accuracy of the model

Table of Contents

Stylized Properties

Fan and Yao (2015, the Elements of Financial Econometrics) identify 8 main "stylized facts"

1. Stationarity
2. Absence of autocorrelations
3. Heavy tails
4. Asymmetry
5. Volatility clustering
6. Aggregational Gaussianity
7. Long-range dependence
8. Leverage effect

Stationarity

Definition: Strong Stationarity

If y_t is a stationary time series, then for any period s in the future, the distribution $\{y_t, \dots, y_{t+s}\}$ doesn't depend on t

Definition: Weak Stationarity (Second Order Stationarity)

A stochastic process $X_{t \in \mathbb{Z}}$ is weakly stationary if and only if:

- $\mathbb{E}(X_t^2) < \infty \quad \forall\, t \in \mathbb{Z}$
- $\mathbb{E}(X_t) = \mu \quad \forall\, t \in \mathbb{Z}$ doesn't depend on t
- $\mathrm{Cov}(x_t, x_{t+h}) = \mathbb{E}[(x_{t+h} - m)(x_t - m)] = \gamma_h \quad \forall\, (t, h) \in \mathbb{Z}^2$ doesn't depend on t

Intuitive Characterization

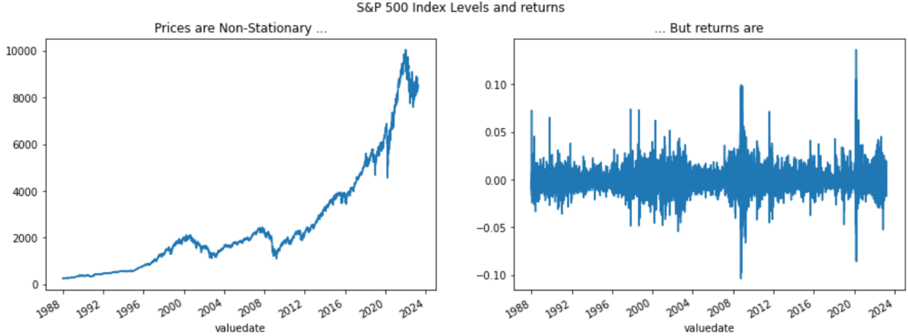
A **stationary series** is:

- Roughly horizontal.The stochastic process oscillates around a constant level
$$\mathbb{E}(X_t) = \mu \forall t$$
- Constant variance. "*covariance doesn't change when shifted in time*"
$$\mathbb{V}(X_t) = \mathbb{C}ov(X_t, X_t) = \gamma(0) \qquad \forall\ t\ \in\ \mathbb{Z}$$
- No predictable patterns in the long term

Example

Example

In general, prices are non-stationary but returns are stationary



How do you identify non-stationary series?

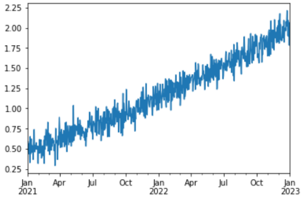
Tips: It is easier to reject the stationarity of a time series rather than confirm it

1. **Visually**:
2. **Global vs local**:
3. **Compute the ACF**:
4. **Statistical tests**

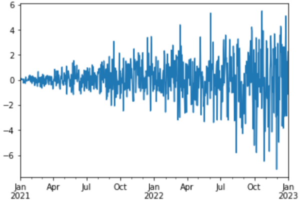
Visually

1. **Visually**: the time plot gives information on the first moments throw time

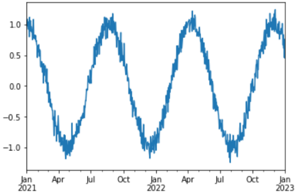
μ is not constant



σ not constant



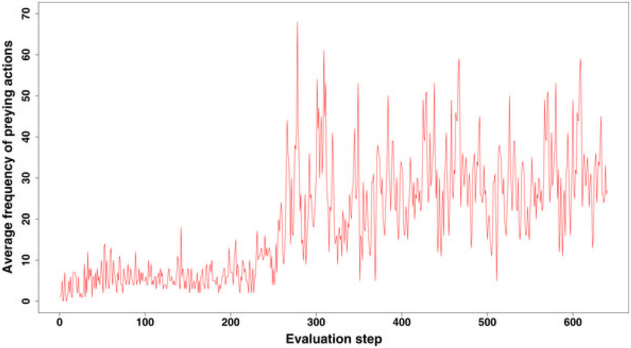
Seasonal Pattern



Global vs Local

2. **Global vs local**: compute the first moments locally and compare them with the global moments computed on the entire-time series

Typical **structural breaks** can easily be identified by this technique

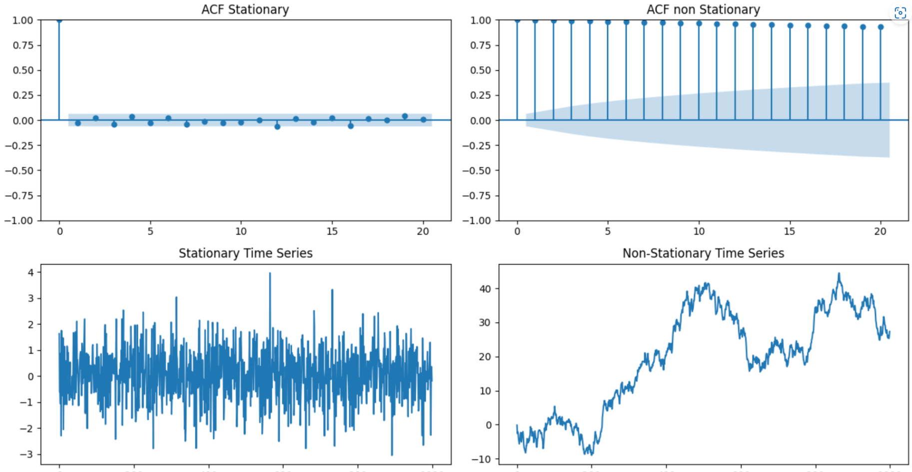


Double structural break:

- Change in Mean
- Increased Volatility

ACF Analysis

3. **Compute the ACF**:
   - The ACF of stationary data drops to zero relatively quickly
   - The ACF of non-stationary data decreases slowly
   - For non-stationary data, the value of the first coefficient is often large and positive



Unit Root Tests

4. Statistical tests for the presence of unit roots

1. **Augmented Dickey-Fuller test**: null hypothesis is that the data is **non-stationary** and non-seasonal
2. KPSS (Kwiatkowski-Phillips-Schmidt Shin) test: the null hypothesis is that the data is **stationary** and non-seasonal
3. Other tests are available for seasonal data

Absence of autocorrelations

Definition: Autocorrelation

The autocorrelation denoted $\rho(k)$ is the correlation between the values of the process at different times:

$$\rho_k \;=\; \text{Corr}(X_t, X_{t-k}) \;=\; \frac{\mathbb{E}\left[(X_t - \mu)(X_{t-k} - \mu)\right]}{\mathbb{V}(X_t)} = \frac{\gamma_k}{\sigma^2}$$

with $\mu = \mathbb{E}[X_t]$, $\sigma^2 = \mathbb{V}(X_t),\ \forall\ t$ and γ_k the autocovariance of order k

Absence of autocorrelations

Definition: Autocorrelation

The autocorrelation denoted $\rho(k)$ is the correlation between the values of the process at different times:

$$\rho_k = \text{Corr}(X_t, X_{t-k}) = \frac{\mathbb{E}\left[(X_t - \mu)(X_{t-k} - \mu)\right]}{\mathbb{V}(X_t)} = \frac{\gamma_k}{\sigma^2}$$

with $\mu = \mathbb{E}[X_t]$, $\sigma^2 = \mathbb{V}(X_t),\ \forall\ t$ and γ_k the autocovariance of order k

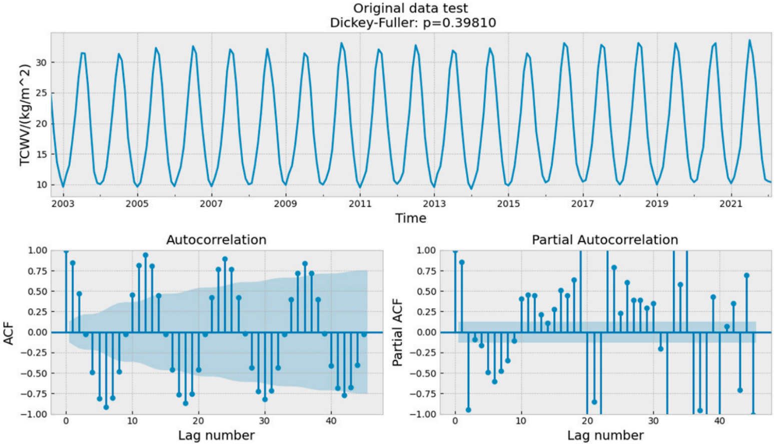
Definition: Sample Autocorrelation

$$\hat{\rho}_k = \text{corr}(X_t, X_{t-k}) = \frac{1}{(T-k)\hat{\sigma^2}} \sum_{t=k+1}^{R} (X_t - \hat{\mu})(X_{t-k} - \hat{\mu})$$

where $\hat{\sigma^2}$ and $\hat{\mu}$ are consistent estimators of the mean $\mu = \mathbb{E}(X_t)$ and the variance $\sigma^2 = \mathbb{V}(X_t)\ \forall t$

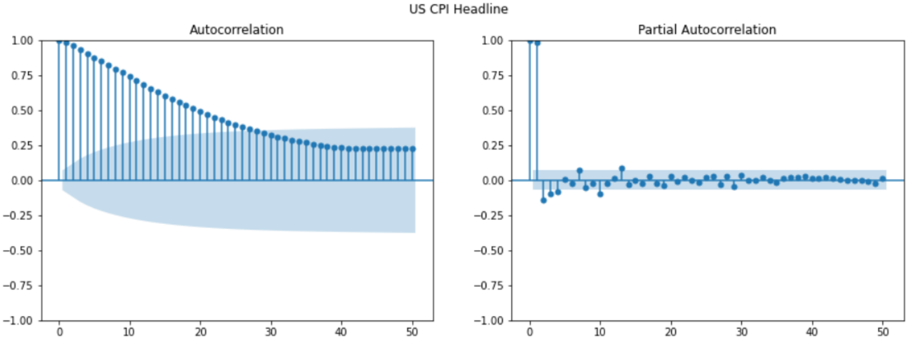
Autocorrelation and Partial Autocorrelation

- Autocorrelation: $\text{corr}(X_t, X_{t-k})$
- Partial autocorrelation at lag k is the correlation after removing the effect of the terms at shorter lags



Credit: *Shangguan, S et al. Atmosphere (2022)*

Example: Autocorrelation of the US CPI

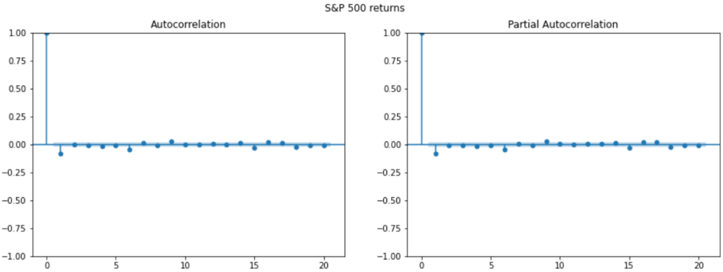


- Highly significant autocorrelation in the first 20 lags
- However, PACF display one single significant bar at the first lag
- Relation between t and t-s, s >1 goes through t-1
- AR(1) model is the best fit to the timeseries

Example: Autocorrelation of the SP500

Absence of autocorrelations in asset returns

The autocorrelations of assets returns are often insignificant, except for intraday time scales (around 20 minutes) for which the microstructure effects come into play



- The fact that returns hardly show any serial correlation does not mean that they are independent
- Financial time seires are difficult to model. That's why we need Quants!

Heavy Tails

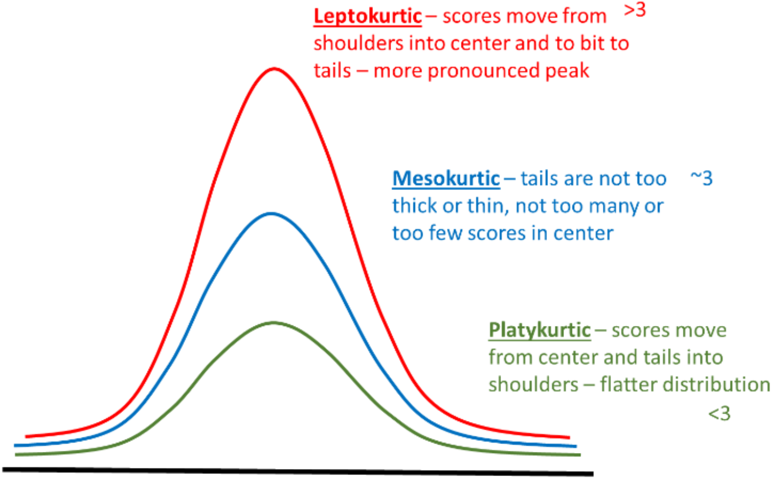
Fact: Heavy Tails

The probability distribution of many financial variables, including asset returns, often exhibit **heavier tails** than those of a normal distribution

- "Heavier tails" are rigorously defined by the kurtosis, which is the fourth-order moment (see before)
- Mandelbrot (1963) recognized the heavy-tailed, highly peaked nature of certain financial time series
- These heavy tails can be explained by risk aversion, heard behavior, market microstructure (illiquidity, asymmetric information, etc.)

Forms of Kurtosis (Fat Tails)

There are different shapes of kurtosis:



Asymmetry

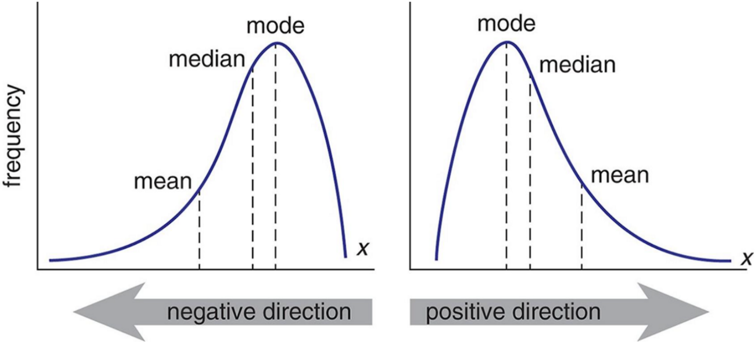
Fact: Asymmetry

The distribution of many financial variables, including asset returns, are often **asymmetric** and **negatively skewed**

- Asymmetry is defined by the skewness, which is the third-order moment
- This reflects the fact that the downturns of financial markets are often much steeper than the recoveries
  - Investors tend to react more strongly to negative news than to positive news
  - Rush toward the exit door/ flight to safety

Skewness

There are different shapes of kurtosis



Credit: *Towards Data Science*

- Skew $= 0 \Rightarrow$ Symmetric distribution $\Rightarrow$ **Mean** $=$ **Median**
- Skew $\geqslant 0 \Rightarrow$ Positive skew implies that the **Mean** is driven by a small number of high values
- Skew $\leqslant 0 \Rightarrow$ Positive skew implies that the **Mean** is driven by a small number of small values

Volatility Clustering

Fact: Volatility Clustering

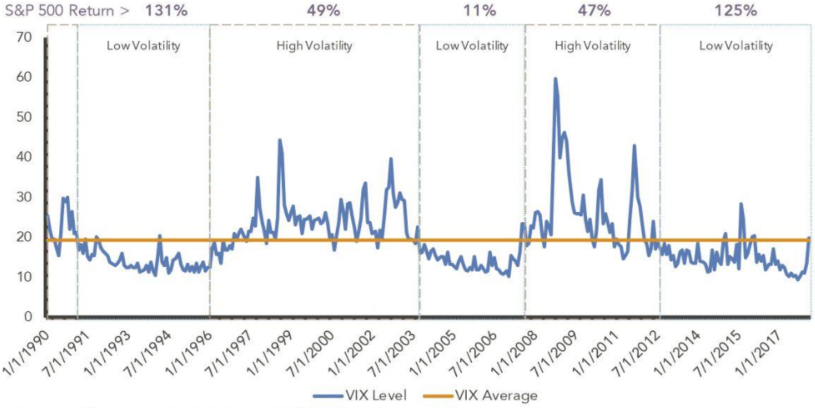
- Large price changes tend to be followed by large price changes (up and down)
- It means that returns with large absolute values or large squares occur in clusters

Note:
volatility clustering is the consequence of the autocorrelation of the squared returns

Volatility Regimes: US VIX

Periods of tranquility alternate with periods of high volatility (volatility regimes)

Figure: The VIX is the implied volatility of the US SP 500



Long Range Dependence

Definition

At the difference of returns, squared returns and absolute returns exhibit significant autocorrelations (**long-memory**)

The ARCH effect

- The autocorrelation of the squared returns is called the **ARCH effect** (auto-regressive conditional heteroskedasticity)
- ARCH effect is important in finance, because it describes patterns on the dynamic of financial volatility
- Those autocorrelations become weaker and less persistent when the sampling interval is increased to a week or a month

Long Range Dependence

SP 500 Returns (left) and squared returns (right)

Figure: ACF for the S&P500 returns

Date: 09/09/18 Time: 15:19
Sample: 8/19/2013 8/17/2018
Included observations: 1259

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.019	-0.019	0.4455	0.504
		2	-0.054	-0.054	4.1067	0.128
		3	0.024	0.022	4.8464	0.183
		4	-0.043	-0.045	7.1488	0.128
		5	-0.028	-0.027	8.1122	0.150
		6	-0.008	-0.014	8.1869	0.225
		7	0.013	0.011	8.3965	0.299
		8	-0.037	-0.039	10.165	0.254
		9	-0.054	-0.057	13.887	0.126
		10	-0.019	-0.028	14.324	0.159
		11	0.002	-0.003	14.330	0.215
		12	-0.001	-0.004	14.330	0.280
		13	-0.015	-0.022	14.617	0.332
		14	-0.020	-0.027	15.102	0.371
		15	-0.073	-0.079	21.926	0.110

Figure: ACF for the S&P500 squared returns

Date: 09/09/18 Time: 11:25
Sample: 8/19/2013 8/17/2018
Included observations: 1259

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.302	0.302	114.94	0.000
		2	0.251	0.176	194.81	0.000
		3	0.288	0.196	299.82	0.000
		4	0.251	0.116	379.50	0.000
		5	0.115	-0.052	396.30	0.000
		6	0.135	0.018	419.45	0.000
		7	0.095	-0.020	430.83	0.000
		8	0.108	0.044	445.63	0.000
		9	0.102	0.043	458.93	0.000
		10	0.123	0.060	478.14	0.000
		11	0.084	0.003	487.21	0.000
		12	0.050	-0.041	490.35	0.000
		13	0.048	-0.020	493.24	0.000
		14	0.040	-0.012	495.29	0.000
		15	0.031	0.010	496.51	0.000

Leverage Effect

Fact: the Leverage Effect

Assets returns are negatively correlated with the changes in their volatilities

Financial explanations

- An asset price declines, companies mechanically become more leveraged (debt-to-equity ratio up) and riskier: therefore, their stock prices become more volatile
- On the other hand, when stock prices become more volatile, investors demand high returns and hence stock prices go down
- Volatilities caused by price decline are typically larger than prices appreciation due to declined volatilities

Aggregational Gaussianity

Definition: Aggregational Gaussianity

- Asset returns over k days is simply the aggregation of k daily returns
- When the time horizon k increases, the central limit theory says that the distribution of returns over a long-time horizon (a few months) tends toward a **normal distribution**

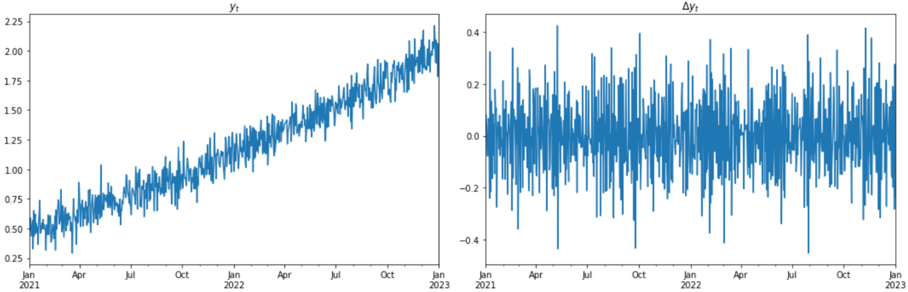
- Aggregational gaussianity implies that over long horizons, the pecularities of financial time series over short-term horizon (skewness, kurtosis, ARCH effect etc.) tend to vanish
- However, in finance, people are mostly interested in relatively short-term movements, suggesting that working under the gaussianity assumption is often not appropriate

Table of Contents

Differencing

- Differencing helps to **stabilize the mean**
- The differenced series is the *change* (or first difference) between each observation in the original series: $y'_t = y_t - y_{t-1}$
- The differenced series will have only $T - 1$ values since it is not possible to calculate a difference y'_1 for the first observation

Differencing



Suppose

$$y_t = \beta_0 + \beta_1 \times t + \epsilon_t$$

Let

$$z_t = \Delta y_t = y_t - y_{t-1}$$

$$\begin{aligned} z_t &= (\beta_0 + \beta_1 \times t + \epsilon_t) - (\beta_0 + \beta_1 \times (t-1) + \epsilon_{t-1}) \\ &= \beta_1 + (\epsilon_t - \epsilon_{t-1}) \\ &= \beta_1 + \nu_t \sim N(0, \sqrt{2\sigma^2}) \end{aligned}$$

Second-Order Differencing

Occasionally, the differenced data will not appear stationary and it may be necessary to difference the data a second time:

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

- y''_t will have $T - 2$ values
- In practice, it is almost never necessary to go beyond second-order differences

Seasonal Differencing

Definition: Seasonal Difference

A seasonal difference is the difference between an observation and the corresponding observation from the previous year

$$y'_t = y_t - y_{t-m}$$

where m = number of seasons

- For monthly data, $m = 12$
- For quarterly data, $m = 4$

Differencing in Practice

When both seasonal and first differences are applied:

- It makes no difference which one is done first - the result will be the same
- If seasonality is strong, we recommend that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference
- It is important that, if differencing is used, the differences are **interpretable**: for instance, taking lag 3 differences for yearly data is difficult to interpret

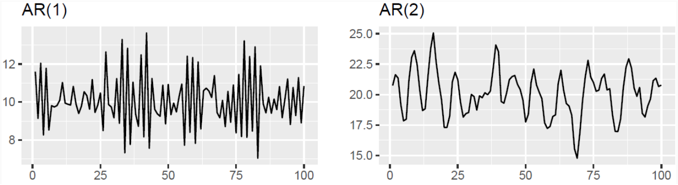
Table of Contents

Autoregressive (AR) Models

Definition

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_P y_{t-p} + \epsilon_t$$

- where ϵ_t is a white noise
- This is a multiple regression with **lagged variables**



AR(1) Model

Specification

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

- When $\phi_1 = 0$, y_t is equivalent to a **white noise**
- When $\phi_1 = 1$ and $c = 0$, y_t is equivalent to a **random walk**
- When $\phi_1 = 1$ and $c \neq 0$, y_t is equivalent to a **random walk with drift**
- When $\phi_1 < 0$ and $c = 0$, y_t tends to oscillate between positive and negative values

Stationarity Conditions

To restrict AR models to stationary data, some contraints on the coefficients are needed

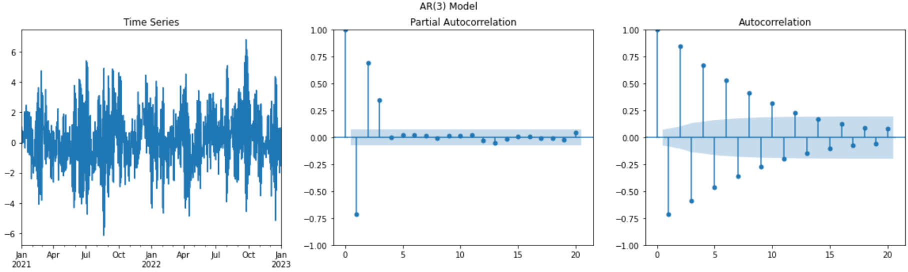
For low lags orders, the stationarity conditions are simply:

- For $p = 1$: $-1 < \phi_1 < 1$
- For $p = 2$: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$ and $\phi_2 - \phi_1 < 1$
- More complex conditions hold for $p \geqslant 3$
- Estimation software (R, Python, Eviews, etc.) takes care of this

How to identify the order of an AR process

By definition, PACF gives the direct effect of lag k on the current value of the time series

$$y_t = \alpha - 0.5 \times y_{t-1} + 0.8 \times y_{t-2} + 0.4 \times y_{t-3} + \epsilon_t$$

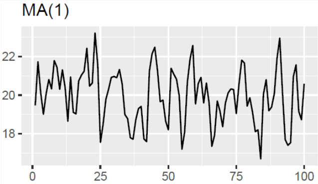


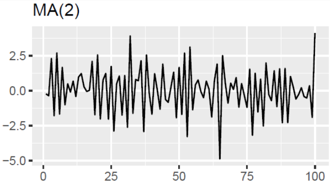
MA: Moving Average Model

Definition: Moving Average Model

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

- ϵ_t is a white noise
- This is a multiple regression with **past errors** as predictors
- Do NOT confuse this with *moving average smoothing*!





How to identify the order of an MA process

Suppose the following MA(q) model

$$MA(q): \quad X_t = \mu + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + ... + \phi_q \epsilon_{t-q} + \epsilon_t$$

The ACF computes the correlation between X_t and X_{t-k}

$$\mathbb{C}orr(X-t, X_{t-k}) = E[X_t X_{t-k}] - E[X_t]E[X_{t-k}]$$

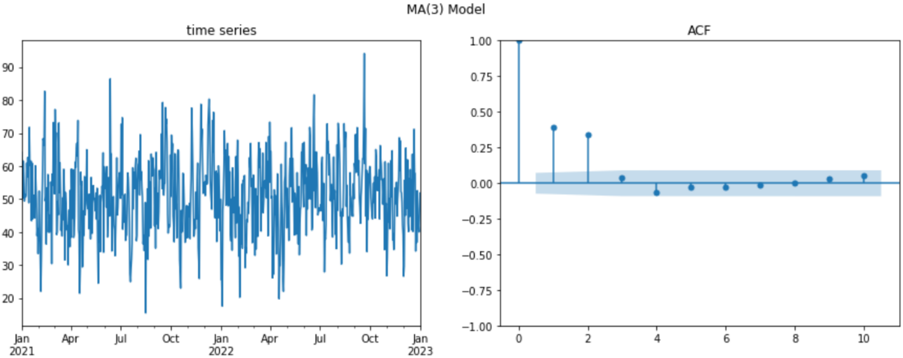
$$X_t : [\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_{t-q}]$$
$$X_{t-k} : [\epsilon_{t-k}, \epsilon_{t-k-1}, \epsilon_{t-k-2}, ..., \epsilon_{t-k-q}]$$

$$\mathbb{C}orr(X-t, X_{t-k}) \neq 0 \quad IF \quad t-q \leqslant t-k \Rightarrow k \leqslant q$$
$$= 0 \quad otherwise$$

From an ACF we can deduce the order of the MA model as the lag on which the correlation turns to 0

Example

$$MA(3): \quad X_t = 50 + 5 \times \epsilon_{t-1} + 3 \times \epsilon_{t-2} + 10 \times \epsilon_{t-3} + \epsilon_t$$



Wold Decomposition: From AR(p) to MA(∞) Model

Wold Decomposition

It is possible to write any **stationary** AR(p) model as an MA(∞)

- Intuitive: just go backward!
- $y_t = \phi_1 \underbrace{y_{t-1}}_{} + \epsilon_t$
- $y_t = \phi_1(\phi_1 y_{t-1} + \epsilon_{t-1}) + \epsilon_t = \phi_1^2 y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t$
- ...
- Providing that $1 < \phi_1 < 1$:

$$\begin{aligned} y_t &= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \phi_1^3 \epsilon_{t-3} + \dots \\ &= \epsilon_t + \sum_{i=1}^{\infty} \phi_i \epsilon_{t-i} \end{aligned}$$

Invertibility: From MA(q) to AR(∞)

- Under certain conditions, an MA(1) process can be written as an AR(∞) process
- In this case, the MA model is said to be **invertible**
- Invertible models have some mathematical properties that make them easier to use in practice
- This is intuitive: AR processes are embedding new information on the most recent lags

ARMA(p, q) Model

Specification

$$y_t = c + \underbrace{\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}}_{\text{AR}} + \underbrace{\theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}}_{\text{MA}} + \epsilon_t$$

- The predictors include both **lagged values of** y_t and **lagged errors**
- Important specification: the future value of the series depends both on the past values it took (dynamic), as well as recent random noise / error term
- This simple model is "learning" both from the dynamic of the past values and from its inherent randomness
- Conditions on AR coefficients ensure **stationarity**
- Conditions on MA coefficients ensure **invertibility**

Autoregressive Integrated Moving Average (ARIMA)

ARIMA: AR**I**MA stands for: Autoregressive **Integrated** Moving Average model

- Basically, it is a non-stationary model that can be made stationary by differencing
- $(1 - B)^d y_t$ follows an ARMA model: d is the degree of differencing
- Once differenced d times, it is stationary and behaves as an ARMA model

Generalization

- $ARIMA(p, d, q)$ where p is the autoregressive order, d the degree of differencing and q the order of the moving average part
- All linear models we discussed are special cases of the ARIMA model:
  - White noise model: $ARIMA(0, 0, 0)$
  - Random walk: $ARIMA(0, 1, 0)$ with no constant
  - Random walk with drift: $ARIMA(0, 1, 0)$ with constant
  - $AR(p) = ARIMA(p, 0, 0)$, $MA(q) = ARIMA(0, 0, q)$

Seasonal ARIMA (SARIMA)

Specification

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\text{Non-Seasonal part of the model}} \quad \underbrace{(P,D,Q)_m}_{\text{Seasonal part of the model}}$$

where m is the number of observations in a cycle

Seasonal ARIMA (SARIMA)

Specification

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\text{Non-Seasonal part of the model}} \quad \underbrace{(P, D, Q)_m}_{\text{Seasonal part of the model}}$$

where m is the number of observations in a cycle

Example: $ARIMA(1,1,1)(1,1,1)_4$, without constant

$$\underbrace{(1-\phi_1 B)}_{\text{Non-Seas. AR1}} \; \underbrace{(1-\Phi_1 B^4)}_{\text{Seas. AR1}} \; \underbrace{(1-B)}_{\text{Non-Seas. Diff.}} = \underbrace{(1+\theta_1 B)}_{\text{Non-seas. MA(1)}} \; \underbrace{(1+\Theta_1 B^4)}_{\text{Seasonal MA(1)}}$$

All factors can be multiplied to obtain the model's component form

Intuition

The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF

$ARIMA(0,0,0)(0,0,1)_{12}$ will show

- A spike at lag 12 in the ACF but no other significant spikes
- The PACF will show exponential decay in the seasonal lags; that is, at lags 12, 24, 36, etc.

$ARIMA(0,0,0)(1,0,0)_{12}$ will show

- Exponential lags in the seasonal lags of the ACF
- A single significant spike at lag 12 in the PACF

Treatment of Seasonality

- Daily data can exhibit multiple seasonal periodicities. This is a complication for all high-frequency forecasting problems: day in the month, day in the week, etc.

- This comes with additional complexity:
  - Months has different number of days
  - Leap years with different number of days
  - Weeks do not align with daily cycles (the year is not divisible in an exact number of weeks)

- Seasonality can be irregular: Ramadan and some other religious festivities for instance

- We use two approaches to deal with complex seasonality
  1. Trigonometric representation of seasonality
  2. Simplification of seasonal terms

Problems of an ARIMA with Many Binary Variables

- Often, central banks forecast currency in circulation by including a large number of binary variables ("dummies") to capture different seasonality patterns. This is called **binary seasonality**
- In general, this approach should be discouraged, because:
  - Including a large number of variables imply to estimate many more parameters, hence **adding parametric noise to the model**
  - Many parameters are not relevant/useful, **increasing noise/signal ratio**
  - Reducing the degrees of freedom spurs the **risk of overfit**
- We tested a large number of models in different countries, confirming the relatively bad performance of ARIMA with binary seasonality

Table of Contents

Empirical Strategy

The general approach of (financial) econometrics is as follows:

1. Specification of the model
2. Estimation of the parameters
3. Diagnostic tests
    - Significance tests
    - Specification tests
    - Backtesting tests
    - etc.
4. Interpretation and use of the model (forecasting, historical studies, etc.)

How to Specify an Appropriate Time Series Model

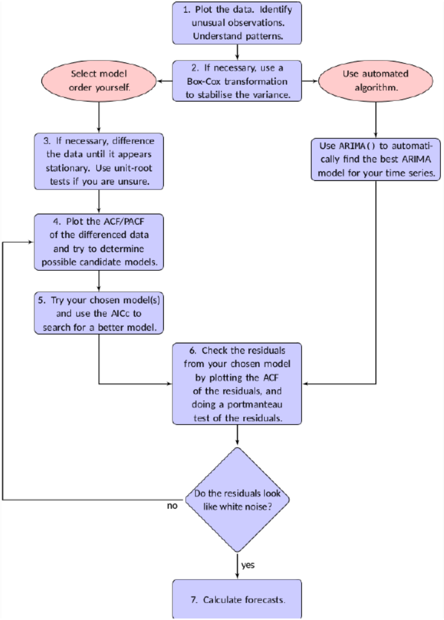
1. Study some **statistical properties** of the observed data $\{x_t\}$, for instance, the **stationarity**, the patterns of the autocorrelation function **ACF**, or the **partial autocorrelation function**, etc.

2. Compare these properties to the theoretical properties of some **typical time series models**, such as AR, MA, ARIMA, SARIMA, etc.

3. Choose the most appropriate model and **estimate its parameters**

4. Use this model for forecasting

Modeling Process



General Approach

- **Plot the data**. Identify any unusual observations
- If necessary, **transform the data** (using a Box-Cox transformation) to stabilize the variance
- If the data are non-stationary, **first-difference** it until stationarity
- **Examine the ACF/PACF**: is an $AR(p)$ or $MA(q)$ reasonable assumptions
- Try your chosen models: use AIC to compare with other models
- Plot the residuals, look at the residuals ACF. Residuals should look like a **white noise**. As long as the residuals of the model have structure, the model is not capturing the whole signal but only part of it
- Once the residuals look like a white noise, **compute the forecasts**

Thank you for you attention !