Time series analysis
Day 3.

David Pupovac

Reminder

- Let us briefly repeat what we were doing up to now:
  - First we approached time series from multiple regression perspective and we talked about FGLS and Newey-West standard errors – this was about **multivariate** analysis
  - Next, we moved to time series decomposition, signal detection, smoothers and we introduced AR(p) and MA(q) processes – this was about **univariate** analysis – today we continue in this fashion.
- Now, let us review the main concepts from the previous class:
  - White noise
  - Stationary vs. non-stationary series (random walk, deterministic trend)
  - Autoregressive process
  - Moving average process
  - Stability
  - Invertibility
  - ACF
  - PACF
- Now, lets put these concepts together.

ARMA and characteristic equation

- Using lag operators we defined ARMA as:

$$\varphi(L)y_t = \vartheta(L)w_t$$

a) The process is **stationary** when the roots of $\varphi(L)$ all exceed unity in absolute value (i.e. if roots lie outside the unit circle in the complex plane)

This is actually testing for **stability** condition. If stability condition holds, stationarity condition is satisfied

b) The process is **invertible** when the roots of $\vartheta(L)$ all exceed unity in absolute value. (The autocorrelation function only identifies a unique MA(q) if process is invertible)

- These roots are derived from **characteristic equation (polynomial)** where characteristic equation for AR(p) model is defined as:

$$\varphi(L) = 0$$
$$1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p = 0$$

and you need to solve for L

Stationarity

- Stationarity is fundamental characteristic of time series. So we will focus on this aspect.
- Stationarity of AR(1):

In general, for stationarity we require:

$$\varphi(L) = 0$$

for AR(1) process this would imply

$$y_t = \varphi y_{t-1} + w_t \Longrightarrow (1 - \varphi L) y_t = w_t$$
$$1 - \varphi L = 0$$

- The root of this equation is:

$$L = 1/\varphi$$

- Consequently:

$$|\varphi| < 1 \Longrightarrow |L| > 1$$

- Because in AR(p) this is a polynomial of arbitrary degree, the roots can be complex numbers. The extension of absolute values to complex numbers is modulus. If the moduli of roots of characteristic equation exceed 1, the process is stationary.

Examples of AR(p) and roots of characteristic equation

- Remember that AR(p) is defined as:

$$\left(1-\varphi_1 L-\varphi_2 L^2-\cdots-\varphi_p L^p\right) y_t=w_t$$

- Consequently, we are evaluating:

$$1-\varphi_1 L-\varphi_2 L^2-\cdots-\varphi_p L^p=0$$

- In characteristic equation lag operator L is formally treated as a number (real or complex).

- **The AR(1) model $\boldsymbol{y_t=0.5 y_{t-1}+w_t}$:**
- Characteristic equation is $1-0.5L=0$.
- In this case $L=2$ and as this is larger than 1, the process is stationary.

Examples of AR(p) and roots of characteristic equation

- **The AR(2) model $\boldsymbol{y_t = 0.5y_{t-1} + 0.5y_{t-2} + w_t}$:**
- Characteristic equation is: $1 - 0.5L - 0.5L^2 = 0$.

$$-0.5(-2 + L + L^2) = 0$$
$$-0.5(-1 + L)(L + 2) = 0$$

- Consequently $L = 1, -2$. Thus while one root exceeds 1 in absolute value, the other is equal to 1. This process in non-stationary and, actually, this a <u>unit root process</u> (more on this later).

- **The AR(2) model $\boldsymbol{y_t = 0.25y_{t-2} + w_t}$:**
- Characteristic equation is: $1 - 0.25L^2 = 0$.
- The solution is complex number $L = \pm 2 \times i$, where $i = \sqrt{-1}$. Absolute values are 2 and the process is stationary.

How to induce stationarity?

- There are two general ways in which non-stationary time series can be stationary**:**

  1. Stationary around deterministic trend - **a <u>trend stationary</u> process**
  2. Stationary around stochastic trend - **a <u>difference stationary</u> (unit roots) process**

- Deterministic trend is most easily subtracted from a series by estimating $y_t = \beta_0 + \beta_1 t + \varepsilon_t$. You can also address this non-parametrically using filters (e.g. moving average)

- Once the trend is estimated and removed from the data, the residual series is a stationary stochastic process and we can account for dynamics using ARMA. In this case a trend stationary AR(1) process, after removing trend, is defined as:

$$\varepsilon_t = \varphi \varepsilon_{t-1} + w_t$$

- where φ is AR coefficient and w_t is white noise term

ARIMA

- Assuming that one root of the autoregressive polynomial lies on the unit circle (i.e. has a unit root) and the remaining ones are all outside, series has to be differenced to induce stationarity. Differencing is defined as:

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

- As the differenced series needs to be aggregated (or ‘integrated’) to recover the original series, the underlying stochastic process is called autoregressive integrated moving average - **ARIMA**

- **Random walk model is prototype of the unit root process**. However, series may have unit roots without being random walk

- If the first difference filter results in stationary series, this is integrated model of order one, or the I(1)-model for short. If two roots of characteristic equation are equal to 1, series has to be differenced twice .Thus in ARIMA)(p, **d**, q), ***d*** refers to the order of integration

Differencing

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

$\boldsymbol{time}$	$\boldsymbol{y}$	$\boldsymbol{y_{t-1}}$
1	0,849468	
2	0,162461	0,849468
3	0,172593	0,162461
4	0,562669	0,172593
5	0,639223	0,562669
6	0,110537	0,639223
7	0,725579	0,110537
8	0,302856	0,725579
9	0,342341	0,302856
10	0,505434	0,342341

Unit root processes

- As we discussed earlier, we can distinguish between pure random walk defined as:

$$y_t = y_{t-1} + w_t = \boldsymbol{y_0} + \sum_{\boldsymbol{s=1}}^{\boldsymbol{t}} \boldsymbol{w_t}$$

  and random walk with drift defined as:

$$y_t = \delta + y_{t-1} + w_t = \boldsymbol{y_0} + \boldsymbol{\delta t} + \sum_{\boldsymbol{s=1}}^{\boldsymbol{t}} \boldsymbol{w_t}$$

- However, it is difficult to distinguish between a random walk process and a stable AR(1)-process in which the autoregressive coefficient is close to unity.

- Furthermore, in a finite sample, it can be shown that if a variable follows a unit-root process, there exists a stationary process that cannot be distinguished from the true unit process. Conversely, if the true process is stationary, there exists a unit root process that cannot be rejected for any finite sample.

Trend- vs. difference-stationary processes

- Differencing a series can remove **trends**, whether these trends are stochastic, as in a random walk, or deterministic.

- However:

  – Differencing a trend stationary process results in a moving average process with a unit root rather than a white noise process (so, consequently, MA component is no longer invertible)

  – Detrending unit root process creates constant mean, however variance of disturbances increases with t, so variable is still not stationary.

- Furthermore, it should be evident that the statistical discrimination between a deterministic trend contaminated with noise and a random walk with drift is not easy.

- **Unit roots test** may help us in deciding which model is appropriate (we will discuss these later)

Fitting ARIMA models - Box-Jenkins approach

- ARIMA models are very flexible, but more than one ARIMA specification may provide a good fit to an observed time series.

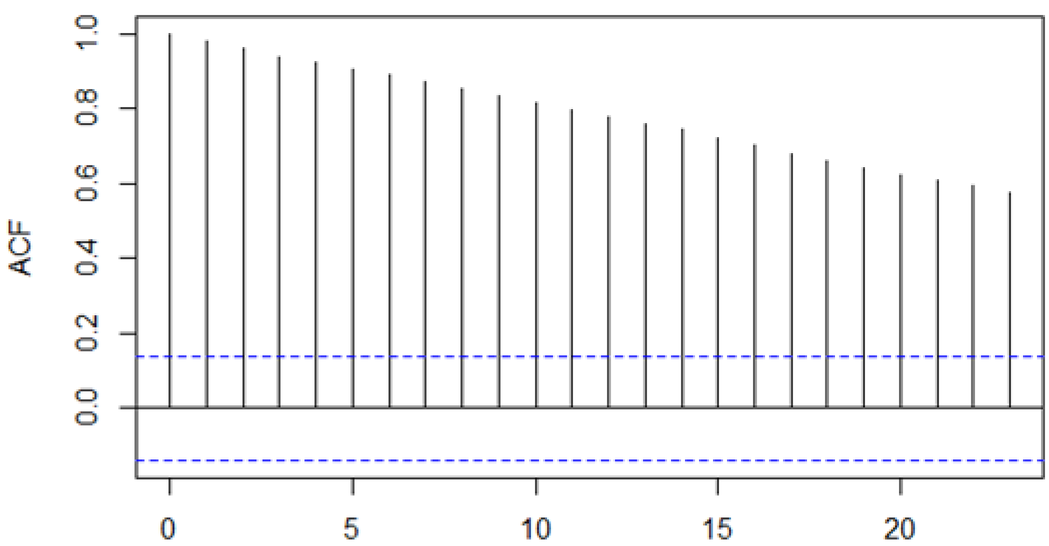
- **Box Jenkins iterative approach:**
  - Identification – the order of ARIMA model.
  - Estimation – estimation of θ, φ
  - Diagnostic checking

- **For time series variable, a model first must be fit to the data before some of the key distributional features of the variable can be measured!!!**

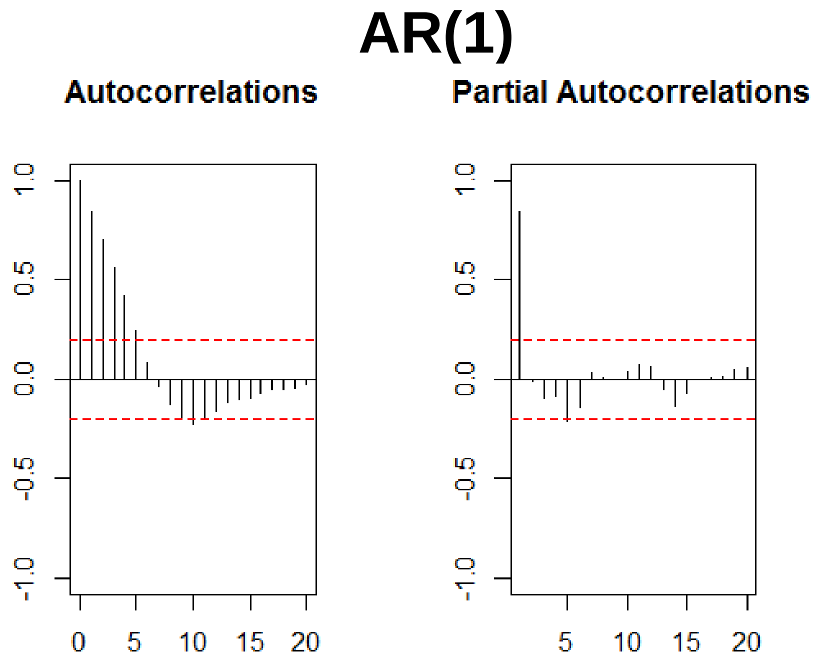
- Emphasis on parsimony

Identify order of differencing (integration)

- To determine order of differencing, ***d*** in ARIMA(p,d,q), look at autocorrelation function
- If autocorrelations collapse quickly toward zero, the time series is stationary
- If the magnitude of correlations declines approximately linearly at least one differencing is needed



Identifying order of AR(p)

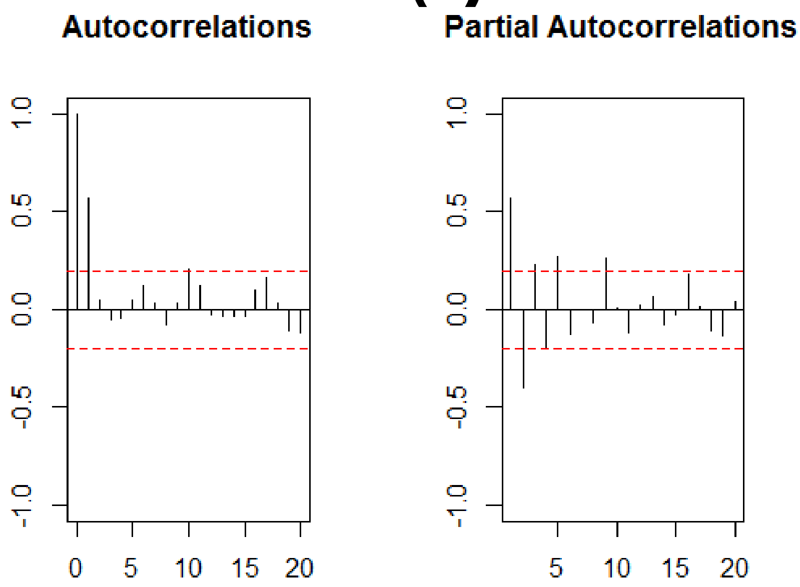


- In pure AR process **acf** dies out either exponentially (if roots are real), in damped oscillations (if roots are complex) or in combination of the two (if roots are both real and complex).

- In pure AR process **pacf** should lie outside confidence bands for first p lags of AR(p)

Identifying order of MA(q)

MA(1)



(keep in mind that in R acf function produces autocorrelation at 0 lag, which you should ignore)

- In pure MA process **acf** should lie outside confidence bands for first q lags of MA(q).

- In pure MA process **pacf** dies out in some combination of exponentially decay and damped oscillations.

Identifying p and q order of ARIMA

- Identification of the combinatory process, ARMA, is more complicated and there is no clear indication of the order using **acf** and **pacf**.

- With respect to **acf** autocorrelation should die out (collapse in towards zero in some combination of exponential decay and dumped oscillation (sinusoidal components) after first $q - p$

- With respect to **pacf** partial autocorrelation should die out after first $p - q$ lags

When fitting to data, an ARMA model will often be more parameter efficient (i.e., require fewer parameters) than a single MA or AR model

Order of AR(p) and MA(q) in ARIMA

- **The approach to identifying order of AR(p) and MA(q) can be both top-down and bottom-up.**
  - Top down would entail assessing reasonably large number of (p,q) parameters using **acf** and **pacf** graphs and successively pruning parameters on the basis of significance levels.
  - Bottom up approach would entail successively adding parameters until optimal ARIMA model is achieved.
- **The Box Jenkins emphasis on parsimony may cause tendency to fit models that are too simple**
- **Overfitting:**
  - The way to guard against too simple models it is overfitting - adding parameters to the model. AR and MA parameters should not be added simultaneously because of the risk of parameter redundancy. In the parameters are insignificant, more parsimonious specification is adequate.

Information criteria

- However, statistical significance may not be the best way to select the model.

- The modern approach to assessing the order of AR and MA orders in ARIMA is based on information criteria. Information criteria penalize models with too many parameters. Various criteria may be employed (and unfortunately they may produce different results). Three criteria are defined bellow:

$$Akaike(AIC) = \ln(\hat{\sigma}^2) + \frac{2(p+q)}{t}$$

$$Bayesian\ (BIC) = \ln(\hat{\sigma}^2) + \frac{\ln(t) \times (p+q)}{t}$$

$$Hannan\ and\ Quinn\ (HQ) = \ln(\hat{\sigma}^2) + \frac{\ln(ln(t))(p+q)}{t}$$

where $\hat{\sigma}^2$ signifies the estimated variance of an ARMA(p, q)-process

Likelihood function

- An additional approach is based on log likelihood. Similarly to information criteria, we can compare log likelihoods of two models.

- Alternatively, a likelihood-ratio test can be computed for an unrestricted and a restricted model.

- The test statistic is defined as:

$$2\left[\mathfrak{L}\left(\hat{\theta}\right) - \mathfrak{L}(\tilde{\theta})\ \right] \sim \chi^2(m)$$

where $\mathfrak{L}(\hat{\theta})$ denotes the unrestricted estimate of the log-likelihood and $\mathfrak{L}(\tilde{\theta})$ the one for the restricted log-likelihood and test statistic is distributed as χ^2 with m degrees of freedom

Test of residuals

- Measure of well specified and accurately fitted time series model is evidence that the residual w_t is white noise. So we should examine the residuals for uncorrelatedness and normality.

- The assumption of uncorrelatedness can be tested with the **Ljung-Box Portmanteau** test (we already used it).
- The hypothesis of normally distributed errors can be tested with the **Jarque-Bera** test.
- Normality also can be tested with the **Shapiro-Wilk** test.
- The assumption of normality could be visually inspected by a **normal quantiles** plot.

Interpretation

- We want to be able to interpret the results in terms general linear process: $\boldsymbol{y_t = \psi(L)w_t}$
- The estimated weights (coefficients) allow for the comparison of the stochastic process implied by the rival model specifications

Calculus for ARMA (1,2)	Calculus for AR (1)
$\psi_1 = \varphi_1 + \vartheta_1$ $\psi_2 = \varphi_1 \cdot \psi_1 + \vartheta_2$ $\psi_3 = \varphi_1 \cdot \psi_2$ $\psi_4 = \varphi_1 \cdot \psi_3$	$\psi_1 = \varphi_1$ $\psi_2 = \varphi_1 \cdot \psi_1$ $\psi_3 = \varphi_1 \cdot \psi_2$ $\psi_4 = \varphi_1 \cdot \psi_3$
Calculus for MA (2)	**Calculus for AR (2)**
$\psi_1 = \vartheta_1$ $\psi_2 = \vartheta_2$ $\psi_3 = 0$ $\psi_4 = 0$	$\psi_1 = \varphi_1$ $\psi_2 = \varphi_1 \cdot \psi_1 + \varphi_2$ $\psi_3 = \varphi_1 \cdot \psi_2 + \varphi_2 \cdot \psi_1$ $\psi_4 = \varphi_1 \cdot \psi_3 + \varphi_2 \cdot \psi_2$

(See slides on dynamic characteristics of AR(p) and MA(q) processes from previous class)

- Positive ψ is interpreted as percent of the shock from previous period persist in the following period. Negative ψ is interpreted as reversal

Unit root tests

- Now let us go back to the discussion of stationarity. The modern approach to identification of unit roots is based on testing

- **Dickey-Fuller test**

- Dickey and Fuller [1979] proposed the following test regression that is delineated from an assumed AR(1) process of y_t:

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

- Logic: If we subtract y_{t-1} from both sides of the equation, we get:

$$y_t - y_{t-1} = \varphi y_{t-1} - y_{t-1} + w_t$$
$$\Delta y_{t-1} = (\varphi - 1) y_{t-1} + w_t$$
$$\Delta y_{t-1} = \pi y_{t-1} + w_t$$

under H_0 = process contains unit root $\pi = 0$ (that is, $\varphi = 1$)

- This test statistic does not have the familiar Student t distribution or similar, so critical values have been calculated by simulation.

Augmented Dickey-Fuller (ADF) test

- However, we need to address higher order processes

- **Augmented Dickey-Fuller** includes lagged differences to soak up the serial correlation of order k

$$\Delta y_t = \pi y_{t-1} + \sum_{j=1}^{k} \boldsymbol{\gamma_j \Delta y_{t-j}} + \varepsilon_t$$

- Several methods for selecting length of lag k of Δy_{t-k} have been suggested in the literature:

  - One method starts with an a priori chosen upper bound k_{max} and then drops the last lagged regressor if it is insignificant. If no lagged regressor turns out to be significant, you choose $k = 0$, which is a simple Dickey-Fuller test.
  - Next method for selecting an appropriate order k are based on information criteria
  - The other selection method is based on $k = \sqrt[3]{t-1}$

Augmented Dickey-Fuller (ADF) test

- Furthermore the existence and the form of the deterministic component are not known – the obvious candidates are a **drift** and a **deterministic** trend (although higher polynomials are also possible)

- The extended test regression is based on addition of β_1 (drift) and $\beta_2 t$ (deterministic trend) terms.

$$\Delta y_t = \boldsymbol{\beta_1} + \boldsymbol{\beta_2 t} + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \, \Delta y_{t-j} + \varepsilon_t$$

- For testing these hypotheses we have to use particular testing strategy

ADF test procedure – step 1

- We start with full regression equation:

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \, \Delta y_{t-j} + u_{1t}$$

- start by testing $H_0\!: \pi = 0$ using the τ_t statistic. If this test is rejected, there are no unit roots and there is no need to proceed further.

- If not, the testing sequence is continued with Φ_3 statistics where $H_0\!: \beta_2 = 0$ given that $\pi = 0$. If it is rejected ($\beta_2 \neq 0$), there is a deterministic trend) then test again for a unit root. If null is rejected, series is stationary around a linear trend; if it not rejected there is unit root.

- If you cannot reject $\beta_2 = 0$ (so, there is no deterministic trend), then go to step 2.

ADF test procedure – step 2

- As we could not reject $H_0: \beta_2 = 0$ new formula is:

$$\Delta y_t = \beta_1 + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_{2t}$$

- Start by testing if $\pi = 0$ using the τ_μ. If we reject null then there is no need to go further (series is stationary around a non-zero mean).

- If it is not rejected, test for the presence of a drift (β_1) using Φ_1 statistics. If it is rejected ($\beta_1 \neq 0$, there is a drift) then test again for a unit root, $\pi = 0$. If null is rejected, series is stationary around a constant; if it is not rejected, there is unit root.

- If you cannot reject H_0: $\beta_1 = 0$, then go to step 3.

ADF test procedure – step 3

- As we could not reject $H_0\colon \beta_1 = 0$ new formula is:

$$\Delta y_t = \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \, \Delta y_{t-j} + u_{3t}$$

- Test if $H_0\colon \pi = 0$. If the null is rejected series is stationary around a zero mean; failure to reject null means that series has a unit root with a zero drift

Additional unit root tests

- There are multiple unit roots tests:

- **Phillips-Perron test** - allows for weak dependence and heterogeneity of the error process

- **Elliott-Rothenberg-Stock test** - improves the power of the unit root test, if the true data-generating process is an AR(1)-process with a coefficient close to one.

- **Schmidt-Phillips test** – accounts for the fact that the nuisance parameters (i.e., the coefficients of the deterministic regressors) are either not defined or have a different interpretation under the alternative hypothesis of stationarity.

Kwiatkowski-Phillips-Schmidt-Shin test

- Unlike all former tests, in KPSS test null hypothesis is a stationary process while the alternative is unit root.

- This is in line with conservative testing strategy; if the results of the tests above indicate a unit root but the result of the KPSS test indicates a stationary process, one should opt for the latter result

$$y_t = \xi t - r_t + \varepsilon_t$$
$$r_t = r_{t-1} + u_t$$

- where r_t is a random walk. The initial value r_0 is fixed and corresponds to the level (local mean).

Kwiatkowski-Phillips-Schmidt-Shin test

- The test is conducted as follows: regress y_t on a constant or on a constant and a trend, depending on whether one wants to test **level** (stationary around a constant mean) or **trend stationarity**. Then, calculate the sums of the residuals $\hat{\varepsilon}_{\mathrm{t}}$ from this regression. The Lagrange multiplier test statistic is then defined as:

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_\varepsilon^2}$$

- Where $\hat{\sigma}_\varepsilon^2$ is estimate of the error variance, and S_t is sums of the residuals ε_t

Unit root tests

- There are multiple unit roots tests, so what should you do?

- A combination of some of the above-mentioned tests with the inclusion of opposing null hypotheses test is a pragmatic approach in practice.