Time Series Analysis

Prof. Lea Petrella

MEMOTEF Department
Sapienza, University of Rome, Italy

Class 10

Building our model

1. Find a class of models that might have generated the observed series: ARMA, ARIMA, ARCH, GARCH
2. Identify the model in a parsimonious way, that is among the models that are good fit for the observed series, the one with few number of covariates (use ACF, PACF, AIC).
3. Estimate the parameters.
4. Diagnostic: in this step, boh the goodness of fit of the selected model for the oserved series and/or the adequacy of the hypothesis about the distribution of the shock are evaluated.
5. Use the selected model for forecasting.

Model identification

- Heuristic method: compares the path of the theoretical and empirical model (estimated with data) using ACF and PACF.
- $AR(p)$ has ACF that decays slowly towards zero and vanishing PACF for lags greater than p.
- $MA(q)$ has vanishing ACF for lags greater than q and PACF that decays slowly towards zero.
- $ARMA(p, q)$ has ACF that behaves as that of an $AR(p)$ after the first q lags and PACF that behaves as that of an $MA(q)$ after the first q lags.
- In general, it is difficult to identify $ARMA$ models.

Parameters estimation for ARMA processes

- There are three ways to estimate *ARMA* models.

1. Least squares.
2. Maximum likelihood.
3. Method of moments.

Least squares

- The class of AR is generally not difficult to estimate.
- For this class we can rely on the ordinary least squares estimator (OLS)
- Consider

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \epsilon_t.$$

- The OLS estimator solves:

$$min_{\varphi_1, \varphi_2, \ldots, \varphi_p} \sum_{t=p+1}^{n} (X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \ldots - \varphi_p X_{t-p})^2$$

- For instance, consider the $AR(1)$ process:

$$X_t = \varphi X_{t-1} + \epsilon_t$$

- The OLS estimator is

$$\hat{\varphi} = \frac{\sum_{t=1}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_{t-1}^2}.$$

Maximum likelihood

- In order to compute the maximum likelihood estimator we need to make assumptions on the distribution of the shocks.
- Generally, it is assumed $\epsilon_t \sim i.i.d.N(0, \sigma^2)$.
- Maximum likelihood for AR processes.
- For the $AR(1)$, the exact maximum likelihood function is

$$L(\varphi, \sigma^2) = f(x_1, x_2, \ldots, x_n | \varphi, \sigma^2) =$$

$$= f(x_1 | \varphi, \sigma^2) \times f(x_2 | x_1, \varphi, \sigma^2) \times f(x_3 | x_2, x_1, \varphi, \sigma^2) \times$$

$$\times \ldots \times f(x_n | x_{n-1}, \ldots, x_1, \varphi, \sigma^2).$$

- This quantity is function of ϕ and σ^2.

- In order to compute the maximum likelihood estimator, it is easier to work with the logarithm of $L(\varphi, \sigma^2)$
- The maximum is found by taking partial derivatives:

$$\frac{\partial l(\varphi, \sigma^2)}{\partial \varphi} = 0$$

$$\frac{\partial l(\varphi, \sigma^2)}{\partial \sigma^2} = 0.$$

- This procedure allows to obtain estimators $\hat{\varphi}$ and $\hat{\sigma^2}$ which are the values that maximize the function $L(\varphi, \sigma^2)$.
- Such estimators do not always have closed form solution and numerical procedures are sometimes used to compute them.

AR estimation

- In the case of the $AR(1)$, what is the distribution $f(x_1|\varphi, \sigma^2)$?
- Recall that we assumed $\epsilon_t \sim i.i.d.N(0, \sigma^2)$.
- We know that

$$\mathbb{E}(X_1) = 0 \quad \text{and} \quad \mathbb{V}ar(X_1) = \frac{\sigma^2}{(1-\varphi^2)},$$

- Also, we know that the $AR(1)$ can be written as an $MA(\infty)$.
- Therefore, $\epsilon_t \sim i.i.d.N(0, \sigma^2)$ implies $X_1 \sim N(0, \sigma^2/(1-\varphi^2))$. That is,

$$f(x_1|\varphi, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\varphi^2)}} \exp\left\{-\frac{(1-\varphi^2)}{2\sigma^2}x_1^2\right\}.$$

- What is the distribution of $f(x_2|x_1, \varphi, \sigma^2)$?
- From the $AR(1)$ equation we see that

$$X_2 = \varphi X_1 + \epsilon_2.$$

- Conditioning on $X_1 = x_1$ we have

$$(X_2|X_1 = x_1, \varphi, \sigma^2) \sim N(\varphi x_1, \sigma^2),$$

- From which

$$f(x_2|x_1, \varphi, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x_2 - \varphi x_1)^2\right\}.$$

- Similarly, we can proceed for X_3. Notice, however, that for the $AR(1)$ we have

$$f(x_3|x_2, x_1, \varphi, \sigma^2) = f(x_3|x_2, \varphi, \sigma^2).$$

- $$f(x_3|x_2,\varphi,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x_3-\varphi x_2)^2\right\}$$
- Considering all the variables, the likelihood writes

$$L(\varphi,\sigma^2) = \left(1-\varphi^2\right)^{\frac{1}{2}} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{(1-\varphi^2)}{2\sigma^2}x_1^2\right\} \times$$

$$\times \prod_{t=2}^{n} \exp\left\{-\frac{1}{2\sigma^2}(x_t-\varphi x_{t-1})^2\right\} =$$

$$= \left(1-\varphi^2\right)^{\frac{1}{2}} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{(1-\varphi^2)}{2\sigma^2}x_1^2\right\} \times$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}(x_t-\varphi x_{t-1})^2\right\}.$$

- In order to compute the maximum likelihood estimator, the logarithm of the likelihood function is considered:

$$l(\varphi, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(1-\varphi^2) - \frac{1}{2}\sum_{t=2}^{n}(x_t - \varphi x_{t-1})^2.$$

- Then, the maximum is computed via partial derivatives:

$$\frac{\partial l(\varphi, \sigma^2)}{\partial \varphi} = 0$$

$$\frac{\partial l(\varphi, \sigma^2)}{\partial \sigma^2} = 0.$$

- From this we derive $\hat{\varphi}$ and $\hat{\sigma^2}$ that have no closed form solution and should be computed numerically.
- Alternatively, we can consider the conditional likelihood, where the observation x_1 is seen as deterministic.

- The conditional likelihood function is given by

$$L(\varphi, \sigma^2) = \prod_{t=2}^{n} f(x_t | x_{t-1}, \varphi, \sigma^2) =$$

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n-1}{2}} \prod_{t=2}^{n} \exp\left\{-\frac{1}{2\sigma^2}(x_t - \varphi x_{t-1})^2\right\}.$$

- Log likelihood becomes

$$l(\varphi, \sigma^2) = -\frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(x_t - \varphi x_t)^2.$$

- It can be shown that minimizing that function with respect to the parameters is equivalent to finding the least squares estimators.

- It can be shown that if the number n of observation of the time series is high, the contribution of x_1 becomes null. Then, the procedures are asymptotically equivalent.

MA estimation

- For MA processes OLS estimation does not apply because the shocks cannot be observed
- Thus, the technique of the conditional maximum likelihood is considered.
- Consider an $MA(1)$ process and assume to know $\epsilon_0 = known$.
- The likelihood function writes

$$L(\theta, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{n} (x_t - \theta\epsilon_{t-1})^2 \right\}.$$

- Differently form x_t, ϵ_t are not observed.
- However, assuming to know ϵ_0, $\epsilon_1, \epsilon_2, \ldots$ can be obtained recursively.
- Unfortunately, the resulting likelihood is non linear in θ, so it can be evaluated only by means of numerical procedures.
- The exact likelihood is even more complicated and does not allow to obtain the estimators in closed form.
- In order to write the likelihood function of an *ARMA* process it takes to put together the two parts (*AR* and *MA*).
- Also in this case, we need numerical procedures to obtain estimates.

Model selection

- The randomness associated with the fact that the relevant quantities of the generating process are estimates, does not allow to know the choice of the orders on p and q.
- For each analyzed series, we obtain a set of models (pairs of p and q) among which the best one is chosen.
- The choice is based on a parsimonious criteria, the preferred model will be that one with the smallest number of parameters.
- Empirical rule: If the value of the estimates $\hat{\xi} = (\hat{\theta}, \hat{\varphi})$ lies inside the interval $[-2se(\hat{\xi}), 2se(\hat{\xi})]$ we claim that $\hat{\xi}$ is not significant and set $\hat{\xi} = 0$. Otherwise, we claim $\hat{\xi}$ is significant and use the model.

- The same conclusion holds when considering the p-value, P.
- P is a number $\in [0, 1]$ that measure the evidence in the data in favor of the null hypothesis H_0. Small values for P indicate evidence against H_0.
- If P is high (> 0.5) accept the null hypothesis H_0: $H_0 : \hat{\xi} = 0$. Otherwise, for $P < 0.5$, reject the null hypothesis $H_0 : \hat{\xi} = 0$, so $\hat{\xi}$ is significant and can be used in the model.

- A well known automatic selection criteria is the AIC (Akaike Information Criterion)

$$AIC(p,q) = n\log(\hat{\sigma}^2(p,q)) + 2(p+q).$$

- $\hat{\sigma}^2(p,q)$ represents the variance of the residuals and $2(p+q)$ is a penalizing factor.
- Compute the AIC value for different values of p and q in the model and choose those that minimize the AIC.

- The higher the orders of p and q the better the fit (increasing the number of covariates better explains the phenomenon) and the less is the variance of residuals $\hat{\sigma}^2(p, q)$.
- However, high orders of p and q increases the number of parameters that need to be estimated, thus the randomness in the final outcome.
- In particular, when using the model for forecasting, the forecast error will depend both on the variance of the residuals and on the errors in the parameters estimation.
- For this reason, AIC depends on a penalizing factor that increases as the number of parameters increases.

Diagnostic

- After having identified the model (chosen the order) and estimated the parameters, diagnostic operations are aimed at checking the goodness of the fit based on the observed residuals

$$\hat{\epsilon} = x_t - \hat{\varphi}_1 x_{t-1} - \ldots \hat{\varphi}_p x_{t-p} - \hat{\theta}_1 \epsilon_{t-1} - \ldots - \hat{\theta}_q \epsilon_{t-q}.$$

- The goal is to check if the observed residuals satisfy the underlying hypothesis of the model, very important is the uncorrelation (WN).
- Recall that the goal of time series models is to explain the serial autocorrelation of the phenomena.
- The chosen model should capture and explain "all" the existing dependence.
- If the chosen model does not explain the phenomenon, it will not capture some part of the correlation that will remain in the residuals that will appear correlated (no WN)!

- The residuals time series can be analyzed as we studied.
- ACF, $\hat{\rho}_\epsilon(h)$, and PACF, $\hat{\phi}_\epsilon(kk)$, can be computed. If their values lie outside the interval

$$\left[-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}\right]$$

  then, $\hat{\rho}_\epsilon(h)$ and $\hat{\phi}_\epsilon(kk)$ are significantly different from zero and we can conclude that the model is inadequate.
- Indeed, under the Gaussianity assumption

$$\left[-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}\right]$$

  is the critical region for accepting a test on $\rho_\epsilon(h)$ at 5% significance level.

- The behaviour of the ACF observed on the residuals can be checked as h varies, by plotting two lines parallel to the x-axis in $[-\frac{1.96}{\sqrt{n}}, +\frac{1.96}{\sqrt{n}}]$.
- We can also use statistical test procedures to see if the residuals are uncorrelated, that is if $H_0 : \hat{\rho}_\epsilon(h) \approx 0$ for each h.
- The Box-Pierce test is based in the statistic

$$Q_M = n \sum_{h=1}^{M} \hat{\rho}_\epsilon^2(h).$$

- M is the number of autocorrelation (typically $n/2$). We expect Q_M small for a correct choice of the model.
- A modified version is the Ljung-Box test

$$Q_M^* = n(n+2) \sum_{h=1}^{M} \frac{\hat{\rho}_\epsilon^2(h)}{(n-h)}.$$

- In order to evaluate Q_M we can rely on its associated p-value.
- Empirical rule: If the p-value is less than 0.05 (0.01) then the null hypothesis H_0 that the residuals are uncorrelated should be rejected and the model should be re-considered. Otherwise, if the p-value is greater than 0.05 (0.01) we can accept the model.
- The analysis of residuals also helps to understand if the Gaussianity hypothesis is correct.
- Besides a graphical check (QQ-plot), statistical tests may be conducted (χ^2 test or tests based on residuals third and fourth moments).