Lezione 3 – 30/09/2019

venerdì 1 novembre 2019 19:41

Lezione 3 – 30/09/2019

We have an introduction of what is time series (it's a special object made by one observation o dimension t). TO make inference we will need to make some restrictions for heterogeneity and in depend. Stationarity and ergodicity play a key in this task.

I can use a time series, why I would use that? impulse response and forecasting.

The response of today by what happened in the past.

1) Impulse

We saw that for impulse response the way to represent it is with two operator Yt / eps – j. When I have representation of these kinds we have this condition here and when we saw these representations we can define the whole composition. When we a stationary process we could obtain a linear representation for this and it's very easy to handle.

2) Forecasting

We began whit the assumption that we are in stationary data. We begin with assuming stationarity. Mu is actually 0 and I'ts necessary for the arguments that he's giving. The computation is much easier. If the meaning are not 0 we considering Tt – mu.

The forecast is describing of linear projection. No power and strange function linking y and x. We will use the same structure here and it's easier because there is not interception. So mu is equal to 0. Mu in this case is the 'a' value of y = a+ bx.

Linear function are easier to handle and that's why I want a linear projection. Basically, I would want a function that x is the value of today and y is the forecast for tomorrow. In this notation Yt is the value of today and the forecast it denoted with a Y with cap and given t. Yt +1 given the information of Yt. What do I need? I need to find out the parameter alpha. Instead of having a Yt with a value of temperature of today we have today and yesterday. Typically, the forecast will take temperature of today and yesterday. To specify this forecast we use Yt+1 given t and t-1.

We will have another variable to find like ➔ y = a + bx + cz.

Alpha1 refers to the first and alph2 refers to the second element. (1) mean only using one bit of information. (2) meaning that forecast is using 2 bit of information. The parameter refers to Yt will be different. We not solving the problem but he's providing us the problem. If we consider m days:

Yt+1 | t … , t-m-1 = alpha1 (m) + …. + AlphaM (m) Yt -m +1

How do I choose these guys (the alpha values) ?

Let me call Xt = (Yt, .., Yt-m+1) '

And alpha all the varius way of alpha. Alpha = (alpha1 (m) … , alphaM (m))'

My forecast Yt+1 | t … , t-m-1 will be in matrix notation = alpha' Xt. Where alpha and X are vectors.

How do I going to get alpha'?

If we remember regression it's orthogonal to the variables. I will choose the alpha so the error that I will observe tomorrow will be (Yt+1 – alpha' Xt). These errors is not correlated with the value of today that is X't. These is a mathematical condition. We will show that these condition delivers the best linear forecast in MSE. Before I do these proofs let's think why these conditions has sense: with the information that we have we expect to have 0 errors. If you have a error correlated with the information we could use these information to reduce the error. This is a mathematical formula with a very precise logical significate.

Regression is exactly this model : using X to explain Y like the linear model. How does alpha look like given these conditions? It's look like a matrix. I'm going to derive these and then comment it.

E = …

Alpha goes out because it's not random (??). This Is a row vector but I like col vectors so I translate all the matrix.

Xt = (Yt , Yt-1 …) '

Xt' = (Yt)
(Yt-1)
..

Here Xt Xt' we have to multiplicate the values

| (Yt) (Yt, Y-1 …) | | Yt Yt Yt Yt-1
| (Yt-1) * | = | Yt-1 Yt Yt-1Tt-1
| … | |

Expectation of Yt is the covariance of Yt and Yt-1 that is gamma1 and so on. So our problem is a linear projection.

These alpha really is the one that I want and that gives the best linear forecast in MSE. Given another sequence of weight that is g.

This is a square and this is the expected value of a square so will be a non-negative number (in fact >= 0).

The middle term will be 0. So alpha and g will be 0. We will prefer alpha to any other g.

In some case we will use non-linear forecast. Linear model is easy to solve, just invert the matrix and we got the answer. We saw before that a stationary process, we can invert these composition I can always get a linear projection even if the starting process wasn't linear.

Example.

Yt as a stationary process and I'm interested in the best forecast when we obtain gamma1.

Example2.

Is the same but I pretend to know a bit more: temperature of today and yesterday. So, the value in these forecast changes because we got more information. The second one is better because we take more information. If we want to compare the two forecasts putting a way of 0 here and we know that the optimal weight is not 0 but is 0.09375. So, these is better forecast.

Example3.

If we have 3 bit of information I use the same algorithm but apart from that we obtain a better forecast. The weight of the very far past doesn't matter for today so I give for these 3rd observation a less weight.

NO I DON'T NEED TO KNOW HOW TO INVERT THE MATRIX.

These technic works and we know understood more things. If I have 100 observation and we try to use that formula I can use a hundred point, right? NO! in real life there is another problem. We pretend that we know the value of gamma matrix!. In reality, the data don't with the tags of please use these autocovariance. We need to estimate it.

We have to estimate with the value of rad of T. That is like choosing a bandwidth for a kernel. If only use the 10th autocovariances will be a mass? The answer is probably not!. If you have a stationary process the value of autocovariances go down very close to 0. Not a big mistake.

Rad of T as maximal dimension we have to use.

Parametric model with AR(1). We will able to get gamma for any j as long as we know the value of phi. We will proof these but these is not just the fact that these value depends only on phi.

What works best? The advantage of using parametric model is that you use all the autocovariances that we could possible want to use. This Is a very small advantage. What matter's more is when we use Parametrics the further away we go the better is the estimation. In the parametric we need just the value of phi. All the autocovariance will be so estimated with more precision (so estimates will be more precise). Because we are using more information and phi it's easy to compute. If I get the model wrong then all my forecast model is wrong because of the autocovariances that we specified to be wrong.

Which one I choose? Nonparametric for the prof but for the majority they give more importance for the parametric.

Additional definition for stationary process: partial autocorrelation function.

Autocorrelation is the function of autocorrelation of how Yt is correlated with Yj happened in the past. How Is Yt correlated with Yt -j. What is the new correlation with Yt-j?. We consider these values using the autocorrelation will be alpha1 (1), aplha2 (2) ecc. In examples 2 will be the secondo element, in the third will be 0.014778. So, this Is the new value.

Capitolo 4, ARMA Models

White noise, MA(1) model, MA(q) model, MA(infinite) model, AR(1) model, AR(2) model, AR(p) model, ARMA(1,1) model, ARMA(p,q) model.

Sum of ARMA processes.

Required that we know all the gamma, so when make estimations we need to know the constrain of the time series. Non parametric → use gamma. Instead of gamma I will estimated the parameter that characterized this model and then obtain gamma. So required that I estimated extra things. I will reduce the large of estimation of autocovariance to low range of numbers. The goal for today and next 5 weeks will be to come up with models that will have a small number of parameter I will manage to characterized all the gamma to get the forecasting.

What kind of model works for this? The ARMA model (autoregression moving average). It doesn't use many parameter and making estimations is very simple. We could use a self-contain space to characterized the values of gamma.

The first and most simple model is the

White Noise

$\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is white noise if

$$E(\varepsilon_t) = 0 \ \forall t$$
$$E(\varepsilon_t^2) = \sigma^2 \ \forall t$$
$$E(\varepsilon_t\varepsilon_\tau) = 0 \ \forall t, \tau \text{ such that } \tau \neq t$$

So, if $Y_t = \varepsilon_t$,

$$\psi_j = 0 \ \forall j \neq 0$$
$$\gamma_j = 0 \ \forall j \neq 0$$
$$\alpha_j^{(j)} = 0 \ \forall j \neq 0$$

i.e. the process has no memory.

White noise, this proprieties hold for the error in the whole composition. We can transform correlation with density…

If I have a series that is white noise (no correlation with the past) follows the definition that we used before. The gamma will be 0 and the correlation will be 0 again and the dependence of the past will be 0 because it doesn't depend of something in the past.

If the process is white noise is maybe independent but needs not be. Y is covariance stationary. W.n(0, o^2) ➔ white noise. We will think a model of this kind will be stationary for two condition:

If ε_t is w.n.$(0,\sigma^2)$, and $Y_t = \varepsilon_t$,

- Y_t may be independent, but needs not be;
- Y_t may be strictly stationary, but needs not be;
- Y_t is covariance stationary;

If ε_t is w.n.$(0,\sigma^2)$, and
$Y_t = \mu + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$,

- Y_t is stationary if $\sum_{j=0}^{\infty}\psi_j^2 < \infty$
- Y_t is stationary and ergodic (for the mean) if $\sum_{j=0}^{\infty}|\psi_j| < \infty$

MA(1)

MA(1)

Let ε_t w.n.$(0,\sigma^2)$, then

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

is MA(1).

We can check stationarity noticing that $\psi_0 = 1$, $\psi_1 = \theta$, so $\sum_{j=0}^{\infty}\psi_j^2 = 1 + \theta^2 < \infty$.

Otherwise, we can check that the first two moments do not depend on time.

Mean:

$$\begin{aligned} E(Y_t) &= E(\mu + \varepsilon_t + \theta\varepsilon_{t-1}) \\ &= E(\mu) + E(\varepsilon_t) + E(\theta\varepsilon_{t-1}) \\ &= \mu \quad + 0 \quad + 0 = \mu \end{aligned}$$

Aggregating μ at different time. Coming average of μ's today .. with μ's yesterday. All I need to do is to check the two condition of white noise. The sum of all the psi squared will be 1^2 + teta^2 ecc. This is finite so this guarantee that the process is stationary.

The E of Yt is the E of μ (that is constant), epsilon is the new shock and teta will be 0 so μ +0 + 0 = mu. So mean it doesn't not depend on time.

Gamma0 is (1 + teta^2) o^2

Gamma1 is teta o^2

Autocovariances:

$$\begin{aligned} \gamma_0 &= E\big[(Y_t - \mu)^2\big] \\ &= E\big[(\varepsilon_t + \theta\varepsilon_{t-1})^2\big] \\ &= E(\varepsilon_t^2 + \theta^2\varepsilon_{t-1}^2 + 2\theta\varepsilon_t\varepsilon_{t-1}) \\ &= E(\varepsilon_t^2) + \theta^2E(\varepsilon_{t-1}^2) + 2\theta E(\varepsilon_t\varepsilon_{t-1}) \\ &= \sigma^2 \quad + \theta^2\sigma^2 \quad + 0 = (1+\theta^2)\sigma^2 \\ \gamma_1 &= E[(Y_t-\mu)(Y_{t-1}-\mu)] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \\ &= E(\varepsilon_t\varepsilon_{t-1} + \theta\varepsilon_{t-1}^2 + \theta\varepsilon_t\varepsilon_{t-2} + \theta^2\varepsilon_{t-1}\varepsilon_{t-2}) \\ &= 0 \quad + \theta\sigma^2 \quad + 0 \quad + 0 = \theta\sigma^2 \end{aligned}$$

$$\gamma_{j\geq 2} = 0$$

★ So, if we want to check for stationarity by checking the moments, we verify

$$E(Y_t) = \mu \text{ for any } t$$
$$Cov(Y_t, Y_{t+j}) = \gamma_j \text{ for any } t$$

and, in particular,

$$\gamma_0 = (1+\theta^2)\sigma^2,\ \gamma_1 = \theta\sigma^2,\ \gamma_j = 0 \text{ for } j \geq 2.$$

E(eps and eps t-1) is 0 because they are not correlated. If we push for longer autocovariances we will have 0 (if j >= 2 → if we push autocovariance more than y1).

They depend on how distant in time this point are. So this is stationary. In particular y0 is this quantity (slide) and gamma j with j >= 2 is 0.

Sum for j = -inf to + inf yj = sum j = -inf to -2 y1+ y-1+ y0 +y1 + sum J= 2 inf yj. The second one will be 0 so teta sigma square + (1 + teta) sigma^2, so this Is like teta sigma square (like the other) and we have (teta + 1 + teta^2 + teta) sigma^2 = (1+ teta) ^2 sigma^2.

So this quantity is the sum of (phi j) ^2

[FOTO]

$\sum_{-\infty}^{\infty}\gamma_s = \sum_{-\infty}^{-2}\gamma_s + \gamma_{-1} + \gamma_0 + \gamma_1 + \sum_{2}^{\infty}\gamma_s = (\sum\psi_s)^2$

$0 \quad \theta\sigma^2 + (1+\theta^2)\sigma^2 + \theta\sigma^2 \quad 0$

$(\theta + 1 + \theta^2 + \theta)\sigma^2$

$(1+\theta)^2\sigma^2$

$Y_t = \psi(L)\varepsilon_t$

There is $\psi(L)^{-1}$

$\psi(L)^{-1}Y_t = \varepsilon_t$

So I can have $(Y_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} \dots) = \varepsilon_t$

$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \dots + \varepsilon_t$

Autocorrelation

Autocorrelations:

$$\rho_1 = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$$

$$\rho_{j\geq 2} = 0$$

Partial autocorrelations: using the definition it is possible to compute

$$\alpha_j^{(j)} = -\frac{(-\theta)^j}{(1+\theta^2+\ldots+\theta^{2j})} = -\frac{(-\theta)^j}{\sum_{i=0}^{j}\theta^{2i}}$$

Given autocovariances we can compute autocorrelation if correlation j>= 2 will be equal to 0. It's interesting to look at the function and tell that declines to 0 to the power of j.

★ Note: the same autocorrelation structure is generated by two values of θ. Consider

$$\theta = \theta_1 \text{ and } \theta = \theta_2 = 1/\theta_1:$$

$$\text{when } \theta = \theta_1,\ \rho_1|_{\theta=\theta_1} = \frac{\theta_1}{1+\theta_1^2};$$

$$\text{when } \theta = \theta_2,\ \rho_1|_{\theta=\theta_2} = \frac{\theta_2}{1+\theta_2^2}:$$

Replacing $\theta_2 = 1/\theta_1$ in $\rho_1|_{\theta=\theta_2}$,

$$\begin{aligned} \rho_1|_{\theta=\theta_2} &= \frac{1/\theta_1}{1+1/\theta_1^2} = \frac{\theta_1^2}{\theta_1^2}\frac{1/\theta_1}{1+1/\theta_1^2} \\ &= \frac{\theta_1}{\theta_1^2+1} = \rho_1|_{\theta=\theta_1} \end{aligned}$$

i.e. θ_1 and θ_2 ($\theta_2 = 1/\theta_1$) generate two equally valid representations of the same process.

The same autocorrelation structure is generate by two values of teta. Why is interesting this result? When we look at the data we will not know what is the data generating process. We don't know the value of teta. If we have a moving average with teta

+-* = 2 we will not be ablet to distingue this model to another one just looking at the data because the data will be exactly the same way. Two different value of data will generate data exactly the same way so the same autocorrelation. If teta = teta1 nd we got teta = that is the invert of teta1.

The order of teta1 is the function on the slide ecc.

Pretend that teta2 is the invert of teta1. We substitute values and then this is the autocorrelation that we started with. So, the autocorrelation cannot be distinguishing by one from each other. They give us the same number.

Invertibility

Assume $\theta = \theta_1$, $|\theta_1| < 1$, and set $\mu = 0$, then rewrite $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ as

$$\varepsilon_t = Y_t - \theta\varepsilon_{t-1}$$

and notice that $\varepsilon_{t-1} = Y_{t-1} - \theta\varepsilon_{t-2}$ so, replacing in the formula for ε_t,

$$\begin{aligned} \varepsilon_t &= Y_t - \theta(Y_{t-1} - \theta\varepsilon_{t-2}) \\ &= Y_t - \theta Y_{t-1} + \theta^2\varepsilon_{t-2} \end{aligned}$$

In the same way, $\varepsilon_{t-2} = Y_{t-2} - \theta\varepsilon_{t-3}$ so

$$\begin{aligned} \varepsilon_t &= Y_t - \theta Y_{t-1} + \theta^2(Y_{t-2} - \theta\varepsilon_{t-3}) \\ &= Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} - \theta^3\varepsilon_{t-3} \end{aligned}$$

Iterating n times,

$$\varepsilon_t = \sum_{j=0}^{n}(-\theta)^j Y_{t-j} + (-\theta)^{n+1}\varepsilon_{t-(n+1)}$$

and, for $n \to \infty$, since $|\theta| < 1$, then $(-\theta)^{n+1} \to 0$, so

$$\varepsilon_t = \sum_{j=0}^{\infty}(-\theta)^j Y_{t-j}, \text{ i.e. } Y_t = \sum_{j=1}^{\infty} -(-\theta)^j Y_{t-j} + \varepsilon_t$$

So for an invertible MA(1) process, we can compute ε_t provided that we know $Y_t, \ldots, Y_{-\infty}$ and θ.

Invertibility is the follow up. We can describe Yt = phi (L) eps t.

There is phi (L) ^-1 so I can write phi (L)^-1 Yt = epst

So I can have (Yt- pi1 Yt2 = eps t

Yt = pi1 Yt1+ pi2 Yt2+ .. + eps t.

This will be a linear model in wise. So, if I want to make a forecast in Yt and this model is invertible, I can use a linear model and also my model works as a linear model! That great, we can make the best estimates!

I can give it the linear representation. This is the best way to obtain forecast. But how do I show it? Start with the model here and put tet to the other side so we highlight epsilon t and we do the same and highlight epsilon t-1.

In n time I get eps t = sum J = 0 to n (- teta) ^j Yt -j + (- teta) ^n+1 eps t-(n+1).

Teta of power of n this quantity is equal to 0. If n going to infinity I can write the next formula and I can write the model as a linear model! These guys will be the benchmark we use to do the benchmark. This will be a good model to make forecast and I will be able to get an easy formula. The catches I need teta to be < 1 in absolute value because I need that – teta ^ n+1 to go to 0. Because esp will explode otherwise.

If teta is bigger that 1 the model will be observationally equivalent if I get the invert of teta. SO I can use the propriety to make the forecast.

The lag operator will be more elegant way yto obtain the value. (1 + phi L) is the polimonial of Phi of L.

An alternative way to obtain this representation:

Rewrite $\varepsilon_t = Y_t - \theta\varepsilon_{t-1}$ as

$$\varepsilon_t = Y_t - \theta L\varepsilon_t$$

using the lag operator. Then,

$$\varepsilon_t + \theta L\varepsilon_t = Y_t$$
$$(1+\theta L)\varepsilon_t = Y_t$$

so, for $|\theta| < 1$, then $\varepsilon_t = (1+\theta L)^{-1}Y_t$, i.e.

$$\varepsilon_t = \frac{1}{(1+\theta L)}Y_t.$$

Since

$$\frac{1}{(1+\theta L)} = \sum_{j=0}^{\infty}(-\theta)^j L^j,$$

then $$\varepsilon_t = \sum_{j=0}^{\infty}(-\theta)^j Y_{t-j},$$

i.e. $$Y_t = \sum_{j=1}^{\infty} -(-\theta)^j Y_{t-j} + \varepsilon_t$$

Phi(L) could be (1+ tetaL) eps t . And the inverse will be 1 / (1+tetaL Yt). The inverse is the limit of geometric sum. So, apply this straight away here and we get the formula. This show how I derive Phi(L) and invert it. Using this polynomial will be easier when we became familiar.

However, if $\theta = \theta_2$, $|\theta_2| \geq 1$, then $(-\theta)^{n+1} \not\to 0$ as $n \to \infty$, so the representation is not invertible if $\theta = \theta_2$.

Finally, if $\theta_1 = 1$, then $1/\theta_1 = 1$, and in both the cases $|\theta| < 1$ is not met. So, for $\theta = 1$ no invertible representation is avalailable.