Problem Set III

Econometrics I -

Instructor: Raul Guarini Riva

TA: Taric Latif Padovani

Problem 1 – (points: 2)

Consider an AR(p) process:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$. This questions will explore the dynamics of conditional moments of y_{t+h} given $\mathcal{I}_t$, where $\mathcal{I}_t = \{y_t, y_{t-1}, \ldots\}$ represents the information set available at time t.

a) Consider $p \times 1$ vector $Y_t = (y_t - \mu, y_{t-1} - \mu, \ldots, y_{t-p+1} - \mu)'$. Show that there exists a $p \times p$ matrix A and a $p \times 1$ vector U_t such that:

$$Y_t = AY_{t-1} + U_t, \quad \forall t$$

Additionally, show that $\Omega \equiv \mathbb{E}[U_t U_t']$ is a $p \times p$ matrix with all elements equal to zero, except for the first element of the main diagonal, which is equal to σ^2.

Hint: Matrix A will only have 0's and 1's, except for the first row.

b) Show that $\mathbb{E}[Y_{t+h}|\mathcal{I}_t] = A^h Y_t$.

c) Find an expression for $\mathrm{Var}[Y_{t+h}|\mathcal{I}_t]$ that depends only on A, Ω and h;

d) Show that the eigenvalues of A are the roots of the polynomial

$$\Phi(z) = (-1)^p \left(z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \cdots - \phi_p \right),$$

i.e., this is its characteristic polynomial;

Hint 1: Recall that the determinant of a triangular matrix is equal to the product of its main diagonal elements.

Hint 2: Recall that if we multiply a column of a matrix by a constant and add the result to another column, the determinant does not change. Try applying operations on $(A - \lambda I)$ to make it triangular – this is a good refresher in Linear Algebra, isn't it?

e) Even if you have not completed the previous item, argue that the eigenvalues of A are all smaller than one in absolute value if the AR(p) process is stationary;

f) Find the limits of $\mathbb{E}[Y_{t+h}|\mathcal{I}_t]$ and $\mathrm{Var}[Y_{t+h}|\mathcal{I}_t]$ as $h \to \infty$ if the process is stationary. What is the intuition for this result?

Problem 2 – (points: 2)

In this question, we will explore one example of a stationary process that is not an ARMA process. Let $\psi_j = \frac{1}{j^2}$ for $j \neq 0$ and $\psi_0 = 1$. Consider the process y_t defined in the following way:

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \varepsilon_t \sim \text{i.i.d. } (0, \sigma^2).$$

1. Compute the mean and variance of y_t;

2. Compute the autocovariance function $\gamma(h) = Cov(y_t, y_{t-h})$ for $h = 1, 2, 3, \ldots$; Is this process covariance-stationary? Why?

3. Show that there are positive constants c_1 and c_2 such that $c_1 \leq h^2 \cdot \gamma(h) \leq c_2$. Conclude that $\gamma(h) = O(1/h^2)$;

4. Now, assume by contradiction that y_t is an ARMA(p, q) process for some finite p and q. Let $\tilde{\gamma}(h)$ be the h-th autocovariance implied by the coefficients of this ARMA process. Show that

$$\lim_{h \to \infty} \frac{\gamma(h)}{|\tilde{\gamma}(h)|} = +\infty$$

Conclue that y_t can never be an ARMA(p, q) process.

Problem 3 – (points: 6)

On the last problem set, you analyzed Brazilian inflation data. Now, you we will create a full pipeline for the estimation of an ARMA model for inflation and we will use it for forecasting. Please use the same dataset `ipca_igpm.csv` as last time. We will concentrate on the IPCA time series for this problem set. We will model it as an ARMA(p, q) process with unknown p and q:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{2}$$

where $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$. We will assume that the process is stationary and invertible, so that the parameters are identified.

a) Create a plot showing the time series evolution of the IPCA series;

b) For given values of p and q (and the associated parameters), write a function that returns the conditional log-likelihood function of an ARMA(p, q) model. You can assume Gaussian errors with constant variance σ^2;

c) Consider $p, q \in \{0, 1, 2\}$. For each of the 8 possible combinations of p and q (4 proper ARMA models, 2 AR models, and 2 MA models), estimate parameters by conditional maximum likelihood. Create a table in which you report, for each model, the estimated parameters, 95% confidence intervals, and the AIC and BIC values;

d) Based on the AIC and BIC values, which model would you choose? Why? Do they disagree in terms of the best model? If the model chosen by BIC is different than the one chosen by AIC, pick your favorite and justify your choice.

e) For the model with the best possible fit, create two plots: (i) the fitted values against the actual series; (ii) the residuals of the model against time. Comment on these plots.

f) Create a plot showing the autocorrelation function of residuals. Comment on this plot;

g) Now we will do some forecasting. For the model with the best fit, simulate 24 realizations of a random variable distributed as $N(0, \hat{\sigma}^2)$. Given the estimated parameters and these simulated shocks, iterate y_t forward following (**??**) to get a path of simulated future realizations of y_t. Plot the original series followed by your forecast.

h) Repeat the item above 1000 times and keep track of the forecasts. For each point in time, compute the realizations in the 5% and 95% percentiles of the forecasts. Plot your forecast with these bounds around it.

i) Intuitively, how similar is this process to the non-parametric boostrap we saw in the beginning of the class?