Lecture 3: ARMA models and Forecast

Dr. Joao B. Duarte[1]

[1]Nova School of Business and Economics
University of Cambridge

Masters, Economics: Macroeconometrics
Lisbon
Spring 2017

Lecture Objectives:

- Introduce the MA models and their properties.
- How to identify MA models.
- Describe the MA(∞) representation and the Wold Decomposition.
- ARMA models and their properties.
- Identification and estimation of ARMA models.
- Forecast of univariate series using ARMA.

Secondary Readings:

- Chapter 2, Applied Econometric Time Series, Enders, Walter, Fourth Edition
- Chapter 3, Time Series Analysis, Hamilton, James, first edition

Moving Average Models (MA models)

- The general form of a MA model of order 1 , MA(1), is the following:

$$y_t = c_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{1}$$

where ε_t is a white noise process

- Similarly, a MA(2) model is in the form

$$y_t = c_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \tag{2}$$

- and a MA(q) model is

$$y_t = c_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \tag{3}$$

MA(q) Properties

- MA models are always weakly stationary because they are finite linear combinations of a white noise process.
- The unconditional mean is:

$$E(y_t) = c_0 \tag{4}$$

- And the Variance:

$$Var(y_t) = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2 \tag{5}$$

- The autocovariance and autocorrelation is zero after q lags.

MA(1) Autocovariance and Autocorrelation

- Assume for simplicity $c_0 = 0$ for a MA(1). If we multiply the model by y_{t-j} we have

$$y_{t-j}y_t = y_{t-j}\varepsilon_t - \theta_1 y_{t-j}\varepsilon_{t-1}$$

- Taking the expectation we obtain

$$\begin{aligned} \gamma_1 &= -\theta_1\sigma^2 && (6) \\ \gamma_j &= 0 \qquad \forall \quad j > 1 && (7) \end{aligned}$$

- And the autocorrelation (just divide by γ_0) is:

$$\begin{aligned} \rho_0 &= 1 && (8) \\ \rho_1 &= \frac{-\theta_1}{1+\theta_1^2} && (9) \\ \rho_j &= 0 \qquad \forall \quad j > 2 && (10) \end{aligned}$$

MA(1) Invertibility

- Rewriting the zero-mean MA(1) as $\varepsilon_t = y_t + \theta_1 \varepsilon_{t-1}$, one can use the method of iteration to find:

$$\varepsilon_t = y_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \ldots$$

- Intuitively, θ_1^j should go to zero as j increases because the remote y_{t-j} should have very little impact on y_t.
- Consequently, for a MA(1) model to be plausible, we require $|\theta_1| < 1$
- Such a MA(1) model is said to be invertible. Note: When a MA(1) is invertible, we have a stationary AR(∞) representation.

MA(q) Identification

- The same is true for the MA(q) model. The autocorrelation coefficients are non-zero for q lags and then they are all zero.
- Hence, we can use the ACF to determine the order of the MA model.
- Example: Here we choose lag-12, MA(12)

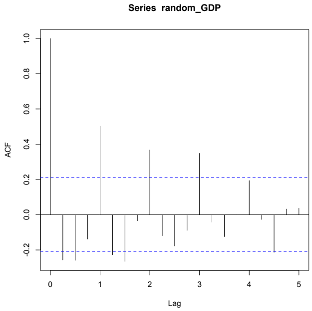

Figure: Portugal's GDP business cycle ACF

MA(q) Identification

- In fact the ACF provides exact information on which specific lags to inclue.
- This is in contrast to the PACF for AR processes.
- To see why. Consider a simple MA(2) model with $\theta_1 = 0$. The model is $y_t = c_0 + \varepsilon_t - \theta_2 \varepsilon_{t-2}$. The ACF of the model is:

$$\rho_0 = 1, \quad \rho_1 = 0, \quad \rho_2 = \frac{-\theta_2}{1+\theta_2^2}, \quad \text{and } \rho_j = 0 \quad \forall j > 2$$

- Hence, in the last example we would select a MA(12) with positive coefficients at lags 1, 2, 4, 5, 6, 8, 12.

MA(q) Estimation

- The commonly used method is maximum likelihood method.
- Example: Lets estimate the MA(12) model of the Portugal GDP business cycle (For presentation sake, we estimate all lags coefficients):

```
Call:
arima(x = random_GDP, order = c(0, 0, 12))

Coefficients:
          ma1      ma2      ma3     ma4      ma5      ma6     ma7     ma8
      -0.4681  -0.0888  -0.0106  0.2836  -0.4785  -0.1090  0.2843  0.2857
s.e.   0.1154   0.1289   0.1187  0.1206   0.1439   0.1208  0.1270  0.1154
          ma9     ma10    ma11    ma12  intercept
      -0.1623  -0.0430  0.0607  0.2668    -6.4700
s.e.   0.1176   0.1232  0.1254  0.1375    12.5716

sigma^2 estimated as 22133:  log likelihood = -562.48,  aic = 1152.95
```

Figure: Portugal's GDP business cycle MA(12) Estimation

MA(q) Estimation

- Moreover, the model is adequate as we have the following ACF of the residuals:

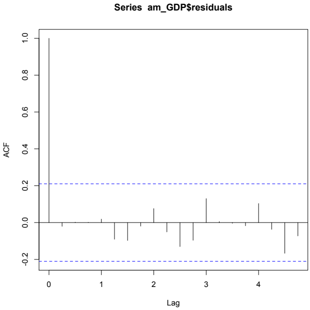

Figure: ACF of MA(12) residuals

- and the Ljung-Box statistic is: 12.818. Hence, we do not reject the null hypothesis of no serial correlation.

Summary

- for AR model, PACF is useful for selecting the lags because the PACF cuts off at lag p.
- for MA models, ACF is useful in determining the lags because it cuts off at lag q
- a MA series is always stationary, but an AR series to be stationary, all of its characteristic roots must be less than 1 in modulus
- AR models can be estimated by LS while the MA models are generally estimated by MLE.

The MA(∞) process and Wold Decomposition

- Wold (1938) showed that the autoregressive and moving average processes are specific cases of a general representation of stationary processes.

- Wold Decomposition: any weakly stationary stochastic process, z_t, with finite mean, μ, that does not contain deterministic components, can be written as a linear function of uncorrelated random variables, a_t , as:

$$z_t = \mu + \sum_{i=0}^{\infty} \phi_i a_{t-i} \tag{11}$$

where a_t is a white noise process and $\phi_0 = 1$

The MA(∞) process and Wold Decomposition

- The Wold Decomposition is extremely important because it shows that any stationary process has a linear MA(∞) representation.
- However, in practice we cannot estimate an infinite number of coefficients. Hence, we need to impose some restrictions.
- The AR admit an MA(∞) structure, but they impose restrictions on the decay patterns of the coefficients ϕ_i.
- The MA require a number of finite terms, however, they do not impose restrictions on the coefficients.
- From the point of view of the autocorrelation structure, the AR processes allow many coefficients different from zero, but with a fixed decay pattern, whereas the MA permit a few coefficients different from zero with arbitrary values.

ARMA models

- ARMA models combine the features of both AR and MA models and allow us to represent in a reduced form (using few parameters) those processes whose first q coefficients can be any, whereas the following ones decay according to simple rules.
- This allows for a very flexible model of linear time series.
- An ARMA(1,1) model satisfies:

$$y_t - \phi_1 y_{t-1} = \phi_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{12}$$

- The left-hand side of (47) is the AR component and the right-hand side gives the MA component. For the model to be meaningful we need $\phi_1 \neq \theta_1$ (why?)

ARMA(1,1) Properties

- Properties of the ARMA(1,1) are generalizations of those of AR(1) with modifications to deal with the MA(1) component.
- We start again assuming that the model is stationary and in the process we reach the conditions under which the process is indeed stationary.
- Lets take expectation of (47):

$$E(y_t) - \phi_1 E(y_{t-1}) = \phi_0 + E(\varepsilon_t) - \theta_1 E(\varepsilon_{t-1}) \Rightarrow$$

$$\boxed{E(y_t) = \mu = \frac{\phi_0}{1 - \phi_1}} \tag{13}$$

Which is exactly that same as that of the AR(1) in (13).

ARMA(1,1) Properties

- Next we assume again for simplicity that $\phi_0 = 0$ and consider the autocovariance function of y_t. First multiplying the model by ε_t and taking expectation:

$$E(y_t\varepsilon_t) = E(\varepsilon_t^2) - \theta_1 E(\varepsilon_t\varepsilon_{t-1}) + \phi_1 E(y_{t-1}\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2 \tag{14}$$

- Next, we rewrite the model as:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

- Taking the Variance of the previous equation we find that:

$$\boxed{Var(y_t) = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2)\sigma^2}{1 - \phi_1^2}} \tag{15}$$

ARMA(1,1) Properties

- Because the variance is positive, we need $\phi_1^2 < 1$ (i.e., $|\phi_1| < 1$)
- Again, this is the same starionarity condition of the AR(1) process.
- To obtain the autocovariance we just multiply the model by y_{t-j} to obtain:

$$y_t y_{t-j} - \phi_1 y_{t-1} y_{t-j} = \varepsilon_t y_{t-j} - \theta_1 \varepsilon_{t-1} y_{t-j}$$

- Take expectation and use (49) for $t-1$ we can find that for $j = 1$:

$$\gamma_1 - \phi_1 \gamma_0 = -\theta_1 \sigma^2$$

- This is different from the AR(1) process where $\gamma_1 - \phi_1 \gamma_0 = 0$

ARMA(1,1) Properties

- However, for $j = 2$ we have:

$$\gamma_2 - \phi_1 \gamma_1 = 0$$

- In fact, that is also true $\forall \quad j > 2$
- Hence, for an ARMA(1,1), the ACF is going to be given by:

$$\rho_1 = \phi_1 - \frac{\theta_1 \sigma^2}{\gamma_0}, \quad \rho_j = \phi_1 \rho_{j-1} \quad \text{for } j > 1 \tag{16}$$

- Thus, the ACF of an ARMA(1,1) behaves very similar to that of an AR(1) model except that the exponential decay starts at lag-2.

ARMA(1,1) Properties

- Notice that the ACF does not cut off at any finite lag.
- Turning to the PACF, one can show that it does not cut off at any lag either.
- In summary, the stationarity condition of an ARMA(1,1) is the same of AR(1)
- The ACF of ARMA(1,1) behaves similarly to AR(1) after lag-2
- The PACF of ARMA(1,1) similarly to MA(1) after lag-2

General ARMA(p, q) models

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{17}$$

- Note the AR and MA models are special cases of ARMA models. Using the Lag operator we have:

$$(1 - \phi_1 L - ... - \phi_p L^p) y_t = \phi_0 + (1 - \theta_1 - ... - \theta_q L^q)\varepsilon_t \tag{18}$$

- The AR component on the left-hand side introduces the characteristic equation. The ARMA(p,q) model is stationary if all the characteristics roots are less than 1 in modulus.

Identification and Estimation of General ARMA(p, q) models

- We cannot use the ACF and the PACF to identify the order of an ARMA model.
- The information criteria are the commonly used methods to select the order of an ARMA model.
- In practice, we compute the AIC for all different combinations of lags p and q and select theone that gave the minimum AIC.
- ARMA models are typically estimated with ML methods.

Forecast

- ARMA models are particularly useful to make predictions of a univariate time series.
- First, consider the forecasts from the AR(1) - ARMA(1,0) model $y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$. Updating one period, we obtain

$$y_{t+1} = \phi_0 + \phi_1 y_t + \varepsilon_{t+1}$$

- If you know the coefficients ϕ_0 and ϕ_1, you can forecast y_{t+1} conditional on the information available at period t as

$$E_t(y_{t+1}) = \phi_0 + \phi_1 y_t$$

Forecast

- In the same way, since $y_{t+2} = \phi_0 + \phi_1 y_{t+1} + \varepsilon_{t+2}$, the conditional expectation of y_{t+2} at time t, i.e. $E_t(y_{t+2}|y_{t+1}, y_t)$ is

$$E_t(y_{t+2}) = \phi_0 + \phi_1 E_t(y_{t+1})$$

- Hence, we can use the one-step ahead forecast to compute the two-step ahead forecast.

$$E_t(y_{t+2}) = \phi_0 + \phi_1(\phi_0 + \phi_1 y_t)$$

- Using forward iteration we can get the entire sequence of forecasts

$$E_t(y_{t+j}) = \phi_0(1 + \phi_1 + \phi_1^2 + ... + \phi_1^{j-1}) + a_1^j y_t \tag{19}$$

Forecast

- (54) is called the forecast equation.
- Note that if the AR model is stationary, $|\phi_1| < 1$. Hence, as j goes to infinity we have

$$E_t(y_{t+j}) \to \frac{\phi_0}{1-\phi_1} \tag{20}$$

- Which is the unconditional expectation of the AR(1)!
- Actually this is a very general result: *For any stationary ARMA model, the conditional forecast of y_{t+j} converges to the unconditional mean as $j \to \infty$*

Forecast

- Now lets take a look at MA(1) or ARMA(0,1).

$$y_{t+1} = c_0 + \varepsilon_{t+1} - \theta_1 \varepsilon_t$$

- Thus, the one-step ahead forecast is

$$E_t(y_{t+1}) = c_0 - \theta_1 \varepsilon_t$$

- The two-step ahead forecast from the equation

$$y_{t+2} = c_0 + \varepsilon_{t+2} - \theta_1 \varepsilon_{t+1}$$

- is

$$E_t(y_{t+2}) = c_0$$

- Point: It quickly reverts to the unconditional mean!!

Forecast

- The ARMA(p,q) combines both and the forecast function of the j-step ahead is given by:

$$E_t(y_{t+j}) = \phi_0 + \sum_{i=1}^{p} \phi_i E_t(y_{t+j-i}) - \sum_{i=1}^{q} \theta_i \varepsilon_{t+j-i} \qquad (21)$$

- Notice that the AR component dominates the forecast in $j > q$.

Forecast Error

- Forecasting from time period t, we can denote the j-step-ahead forecast error, called $e_t(j)$:

$$e_t(j) \equiv y_{t+j} - E_t(y_{t+j}) \tag{22}$$

- Since the one-step-ahead forecast error is equivalent to $e_t(1) = y_{t+1} - E_t(y_{t+1}) = \varepsilon_{t+1}$
- Hence, $e_t(1)$ is precisely the "unforecastable" portion of y_{t+1}, given the information available in t.

Forecast Error

- Lets introduce the MA Representation of the ARMA(p,q) model, i.e. this should not come as a surprise given the Wold decomposition.

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + ... = \mu + \psi(L)\varepsilon_t \qquad (23)$$

- The coefficients $\{\psi_i\}$ are referred to as the *impulse response function*. For a weakly stationary series, these coefficients decay exponentially as i increases. Under this representation it is easy to compute the forecast error:

$$e_t(j) = \varepsilon_{t+j} + \psi_1 \varepsilon_{t+j-1} + ... + \psi_{j-1} \varepsilon_{t+1} \qquad (24)$$

Forecast Error

- Hence, the variance of the forecast error in j period is given by:

$$Var[e_t(j)] = (1 + \psi_1^2 + ... + \psi_{j-1}^2)\sigma^2 \tag{25}$$

- Note that as j increases, the variance of the forecast error increases. But if the series is stationary, at some point it converges to the unconditional variance of y_t
- In practice, we do not observe ψ. We need to estimate them and the forecast error is compounded by the parameter uncertainty. This is one of the reasons many time series practitioners and theorist advocate for small models.
- This is probably one of the reasons simple ARIMA models out-performed the large scale macro models used in the Cowles Commission.

Combining Forecasts

- What if there are many plausible models to explain the data? Should we discard them? Or use them to make predictions?
- It turns out that it can be quite beneficial to combine forecasts of different plausible models.
- Let f_{it} be the one-step-ahead forecast. Then the combined forecast is:

$$f_{ct} = w_1 f_{1t} + w_2 f_{2t} + ... + w_n f_{nt} \tag{26}$$

where $\sum_{i=1}^{n} w_i = 1$

Combining Forecasts

- If all the forecasts are unbiased, so will be the combined forecast:

$$E_{t-1}[f_{ct}] = w_1 E_{t-1}[f_{1t}] + w_2 E_{t-1}[f_{2t}] + \ldots + w_n E_{t-1}[f_{nt}] = y_t \quad (27)$$

- Moreover, lets take a look at the variance of the combined forecast error. For simplicity, lets assume we have two plausible models:

$$e_{ct} = w_1 e_{1t} + (1 - w_1) e_{2t} \quad (28)$$

So

$$var(e_{ct}) = w_1^2 var(e_{1t}) + (1 - w_1)^2 var(e_{2t}) + 2 w_1 (1 - w_1) cov(e_{1t}, e_{2t}) \quad (29)$$

Hence, the combined variance can be lower than the variance of either individual forecast.

Forecast - Example

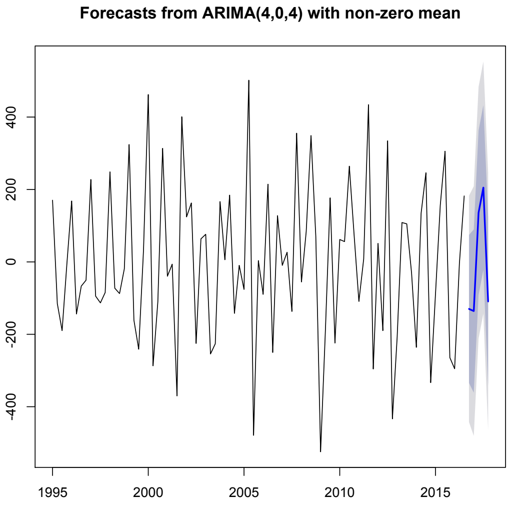

Figure: Forecast 5-step ahead of Portugal GDP Business Cycle

Forecast Seasonal- Example

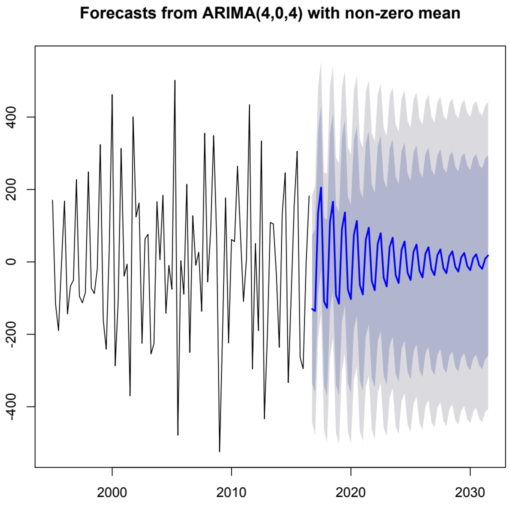

Figure: Forecast 60-step ahead of Portugal GDP Business Cycle

Forecast - Example

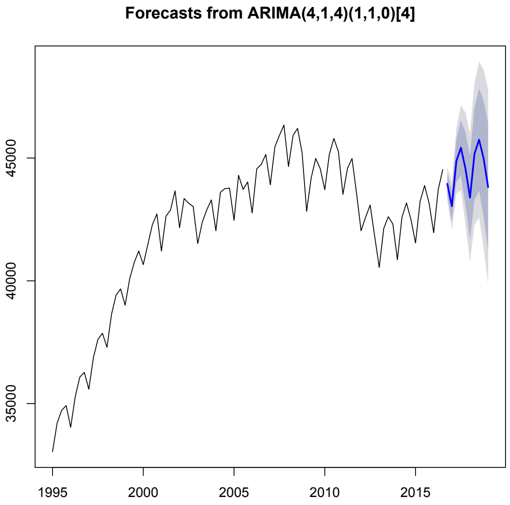

Figure: Forecast 10-step Portugal GDP

Summary

- Any stationary process can be represented by an MA(∞).
- The properties of an ARMA model follow closely the properties of an AR process woth some modifications to account for the MA component.
- For an ARMA model to be stationary, the characteristic roots of the difference equation must lie inside the unit circle.
- The ARMA models are particularly good at forecasting few steps ahead.

Questions to think about

- What is the most appropriate data transformation?
- What should be done about seemingly significant coefficients at reasonably long lags?
- How to deal with seasonality? How to deal with non-stationarity? Lecture 5.
- What if I am interested in the relationship between economic variables? Next lecture.