ARMA: Representation and Estimation

This Lecture

- What are ARMA models and how are they used
- Multiple equivalent representations of ARMA models
- Estimation of ARMA models

Later lectures:

1. Rules of thumb for model fitting
2. Properties of stationary and nonstationary ARMA processes
3. Forecasting using ARMA

ARMA Models

- (AR)(I)(MA): Autoregressive Integrated Moving Average
- Autoregressive: control for lags of the series
- Integrated: any unit roots are removed
- Moving Average: weighted lags of the residuals
- For now, focus on ARMA
  - Temporarily assume stationary, no integrated/unit root components

ARMA(1,1) vs ARMA(p,q)

Let a_t be white noise and

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t + \theta_1 a_{t-1}$$

or

$$r_t = \phi_0 + \underbrace{\phi_1 r_{t-1} + \ldots + \phi_p r_{t-p}}_{AR(p)} + a_t + \underbrace{\theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}}_{MA(q)}$$

for any combination of p and q that “fits” the data.

Uses of ARMA Models

- A way to parsimoniously represent linear dependence
- Univariate forecasting from past values
- Controlling for autocorrelation in regression residuals, e.g.

$$y_t = \beta_0 + \beta_1 x_t + e_t$$

$$e_t = \phi_0 + \phi_1 e_{t-1} + ... + \phi_p e_{t-p} + a_t + \theta_1 a_{t-1} + ... + \theta_q a_{t-q}$$

- shocks/innovations $a_t \sim iid(0, \sigma_a^2)$, e.g., white noise

ARMA and Linear Time Series

- We said any linear time series can be written

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

- $\psi_0 = 1$, ψ_i are "psi-weights".
- MA(q) is simple:

$$r_t = \phi_0 + a_t + \underbrace{\theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}}_{MA(q)}$$

- $\psi_1, \ldots, \psi_q \neq 0$ and $\psi_i = 0$ for $i > q$.
- What about AR(p)?

ARMA Representations

Claims: For a stationary process:

- $AR(1)$ has an equivalent $MA(\infty)$ representation (Linear Time Series)
- $MA(1)$ has an equivalent $AR(\infty)$ representation
- Likewise, $AR(p)$ has $MA(\infty)$ and $MA(q)$ has $AR(\infty)$
- $ARMA(1,1)$ and $ARMA(p,q)$ can be written *either* as $MA(\infty)$ or $AR(\infty)$.

$AR(1)$ as $MA(\infty)$

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t, \quad a_t \sim white\ noise(0, \sigma_a^2)$$

$$E(r_t | r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \quad Var(r_t | r_{t-1}) = \sigma_a^2$$

- weak stationarity requires
  $E(r_t) = \mu,\ Var(r_t) = \gamma_0,\ Cov(r_t, r_{t-l}) = \gamma_l$ are all constants not dependent on t.

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}) \Longrightarrow \mu = \frac{\phi_0}{1 - \phi_1},\ or\ \phi_0 = (1 - \phi_1)\mu$$

- This implies $\phi_1 \neq 1$, AND $\mu = 0$ if and only if $\phi_0 = 0$

$AR(1)$ as $MA(\infty)$

- Using $\phi_0 = (1 - \phi_1)\mu$, we can "demean", or rewrite the AR model in deviations from its mean:

$$r_t = (1 - \phi_1)\mu + \phi_1 r_{t-1} + a_t$$

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t$$

$$r_t - \mu = \phi_1(\phi_1(r_{t-2} - \mu) + a_{t-1}) + a_t = \phi_1^2(r_{t-2} - \mu) + \phi_1 a_{t-1} + a_t$$

$$= \phi_1^m(r_{t-m} - \mu) + \sum_{i=0}^{m-1} \phi_1^i a_{t-i} \approx \sum_{i=0}^{\infty} \phi_1^i a_{t-i}$$

- This is a linear time series or MA(q) with $\psi_i = \theta_i = \phi_1^i$
- Now you try it with AR(2).

$MA(1)$ as $AR(\infty)$

$$r_t = \mu + a_t + \theta a_{t-1}, \quad a_t \sim white\ noise(0, \sigma_a^2)$$

$$E(r_t|r_{t-1}) = E(r_t) = E(\mu) + E(a_t) + \theta E(a_{t-1}) = \mu,$$

$$\begin{aligned} Var(r_t|r_{t-1}) = Var(r_t) \quad &= E(r_t - \mu)^2 \\ &= E(a_t^2 + \theta^2 a_{t-1}^2 + 2\theta a_t a_{t-1}) \\ &= \sigma_a^2 + \theta^2 \sigma_a^2 + 0 \\ &= (1+\theta^2)\sigma_a^2 \end{aligned}$$

$MA(1)$ as $AR(\infty)$

$$
\begin{aligned}
a_t &= (r_t - \mu) - \theta a_{t-1} \\
a_{t-1} &= (r_{t-1} - \mu) - \theta a_{t-2} \\
&\ldots \text{etc.} \\
\Longrightarrow a_t &= (r_t - \mu) - \theta(r_{t-1} - \mu) + \theta^2(r_{t-2} - \mu) - \ldots
\end{aligned}
$$

- which can be rearranged with r_t on the left, and infinite lags on the right plus a_t.
- Now you try it with MA(q), or ARMA(1,1)

Simulation examples

Simulate an AR(1)

$$y_t = 0.8y_{t-1} + a_t$$

Estimate a large MA(q)

```
y1 <- arima.sim(model=list(ar=c(0.8)),1000)
# simulates 1000 obs with ph1 =0.8
# True model is ARMA(1,0), suppose we fit MA(12)
y1ma = arima(y1,order=c(0,0,12))
summary(y1ma)
```

Simulation examples

True model:

$$y_t = 0.8y_{t-1} + a_t$$

Estimated model has many significant MA terms:

```
##
## Call:
## arima(x = y1, order = c(0, 0, 12))
##
## Coefficients:
##          ma1     ma2     ma3     ma4     ma5     ma6     ma7     ma8
##       0.7493  0.5478  0.3876  0.3581  0.2470  0.1487  0.1609  0.1853
## s.e.  0.0316  0.0393  0.0429  0.0444  0.0454  0.0452  0.0450  0.0441
##         ma10    ma11    ma12  intercept
##       0.0999  0.1007  0.0342    -0.0254
## s.e.  0.0428  0.0388  0.0319     0.1258
##
## sigma^2 estimated as 0.9239:  log likelihood = -1379.82,  aic = 2787
```

Simulation examples

True model:

$$y_t = 0.8 y_{t-1} + a_t$$

Estimated "true" model:

```
##
## Call:
## arima(x = y1, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.8082    -0.0208
## s.e.  0.0186     0.1643
##
## sigma^2 estimated as 1.001:  log likelihood = -1420.17,  aic = 2846.
```

Simulation examples

Simulate an MA(1)

$$y_t = a_t + 0.8a_{t-1}$$

Estimate a large AR(p)

```r
y1 <- arima.sim(model=list(ma=c(0.8)),1000)
# simulates 1000 obs with tht1 =0.8
# True model is ARMA(0,1), suppose we fit AR(12)
y1ar = arima(y1,order=c(12,0,0))
summary(y1ar)
```

Simulation examples

True model:

$$y_t = a_t + 0.8a_{t-1}$$

Estimated model has many significant AR terms:

```
##
## Call:
## arima(x = y1, order = c(12, 0, 0))
##
## Coefficients:
##          ar1      ar2     ar3      ar4     ar5      ar6     ar7
##       0.7470  -0.5836  0.4612  -0.3633  0.2855  -0.2286  0.1179  -0.
## s.e.  0.0317   0.0395  0.0435   0.0458  0.0472   0.0479  0.0479   0.
##          ar9     ar10    ar11    ar12  intercept
##       0.0963  -0.1078  0.0364  0.0127    -0.0033
## s.e.  0.0458   0.0436  0.0396  0.0317     0.0504
##
## sigma^2 estimated as 0.9587:  log likelihood = -1398.32,  aic = 2824
```

Simulation examples

True model:

$$y_t = a_t + 0.8a_{t-1}$$

Estimated "true" model:

```
##
## Call:
## arima(x = y1, order = c(0, 0, 1))
##
## Coefficients:
##          ma1  intercept
##       0.7976    -0.0377
## s.e.  0.0185     0.0588
##
## sigma^2 estimated as 1.07:  log likelihood = -1453.14,  aic = 2912.2
```

Takeaway on Representations

- You can always fit an AR(p) with large p or MA(q) with large q
- An ARMA(p,q) with short p and q is usually better.
- Representations are not unique, so don't waste a lot of time.
  - Find an adequate model that fits well and removes autocorrelation in the residuals.

One more point: Companion Form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t$$

- Define $\mathbf{y_t} = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}$$

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

- Call this the “companion form”.
- Any scalar AR(p) can be written as a vector AR(1)
- We’ll come back to this

One more point: Companion Form

- $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1} + \ldots + \phi_p y_{t-p} + w_t$

Companion form:

$$\mathbf{y_t} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & & & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

$$\mathbf{w_t} = \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

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

Estimation of ARMA(p,0)

$$r_t = \phi_0 + \underbrace{\phi_1 r_{t-1} + ... + \phi_p r_{t-p}}_{AR(p)} + a_t$$

- AR(p) can be estimated by OLS
- "Burn" first p observations.

$$min_{\phi_i} \sum_{t=p+1}^{T} (r_t - \phi_0 - \phi_1 r_{t-1} - \cdots - \phi_p r_{t-p})^2$$

Estimation with MA terms

$$r_t = \mu + a_t + \underbrace{\theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}}_{MA(q)}$$

- Assume $a_t \sim N(0, \sigma_a^2)$, use MLE for $f(a_t)$
- Assume the first q values of a_t are at their mean of zero.
- Calculate $a_{q+1} = r_t - \mu - \theta_1 a_q + \cdots + \theta_q a_1$.
- Plug into log likelihood $lnf(a_{q+1})$
- Iterate

$$max \sum_{q+1}^{T} lnf(r_t - \mu - a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} | \underline{\theta})$$