# An introduction to ARMA models

Susan Thomas

19 August 2009

# Goals

Introducing the simplest **linear** time series models:

- White noise processes
- Moving Average (MA) models
- Auto Regressive (AR) models
- Mixed AR-MA models

# A white noise process

$$x_t = \epsilon_t; \qquad \epsilon_t \sim D(\alpha_0, \sigma^2)$$

- Stationary process:
  1. $E(x_t) = E\epsilon_t = \alpha_0, \quad \forall t$
  2. $E(x_t)^2 = E(\epsilon_t^2) = \sigma^2, \quad \forall t$
  3. $E(x_t)(x_{t-j}) = E(\epsilon_t)(\epsilon_{t-j}) = 0, \quad \forall j$
- Not a good DGP for most economic time series variables, which tend to change very slowly.

# Introducing AR models

- An AR process $y_t$ of order $p$ is written as $AR(p)$ and is modelled:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \epsilon_t$$

  where $\epsilon_t \sim$w.n with $\mu_\epsilon = 0$.

- It can be written as:

$$\begin{aligned} y_t &= (\alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_p L^p) y_t + \epsilon_t \\ \alpha(L) y_t &= \epsilon_t \\ \text{where } \alpha(L) &= (1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p) \end{aligned}$$

# Introducing MA models

- An MA process $y_t$ of order $q$ is written as $MA(q)$ and is modelled:

$$y_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$$

  where $\epsilon_t \sim$w.n. with $\mu_\epsilon = 0$.

- It can also be written as:

$$\begin{aligned} y_t &= (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^p)\epsilon_t \\ y_t &= \theta(L)\epsilon_t \\ \text{where } \theta(L) &= (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^p) \end{aligned}$$

# Example of AR model: AR(1)

- $y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$
- $\mathrm{E}(y_t | y_{t-1}, \ldots, y_0) = \alpha_0 + \alpha_1 y_{t-1}$

# Example of MA models: MA(1)

- $y_t = \alpha_0 + \theta_1 \epsilon_{t-1} + \epsilon_t$
- $\mathrm{E}(y_t | y_{t-1}, \ldots, y_0) = \alpha_0$

# Understanding stationarity of AR(1), MA(1)

# Recapitulating stationarity

- Strict stationarity: All the moments of the series are constants and are not functions of time.
- This implies that the distribution of innovations is bounded and not explosive.
- Weak/Covariance stationarity: The first two moments of the series is bounded:
  - $E(y_t) = C$
  - $E(y_t)^2 = M$
  - $E(y_t y_{t-s}) = f(s) \qquad \forall s = 1, 2, \ldots.$

# Stationarity of AR(1)

- $y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$
- $\mathrm{E}(y_t) = \alpha_0 + \alpha_1 E(y_{t-1})$
- If $y_t$ is **stationary**, then $E(y_t) = E(Y_{t-1})$, then

$$E(y_t) = \frac{\alpha_0}{1 - \alpha_1}$$

- This works only if $\alpha_1 < 1$

# Stationarity of AR(1), contd.

- $y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$
- $\mathrm{var}(y_t) = \alpha_1^2 \mathrm{var}(y_{t-1}) + \sigma_\epsilon^2$
- If $y_t$ is **stationary**, then $\mathrm{var}(y_t) = \mathrm{var}(Y_{t-1})$, and

$$\mathrm{var}(y_t) = \frac{\sigma^2}{1 - \alpha_1^2}$$

- This works only if $|\alpha_1| < 1$

# Stationarity of AR(1), contd.

- $y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$
- Define $\hat{y}_t = y_t - \frac{\alpha_0}{(1-\alpha_1)}$
- We can show that:

$$\hat{y}_t = \alpha_1 \hat{y}_{t-1} + \epsilon_t$$

- Then, covariance at lag 1 is: (assuming **stationarity**)

$$\begin{aligned} \mathrm{cov}(y_t, y_{t-1}) &= E(\hat{y}_t)(\hat{y}_{t-1}) = \alpha_1 E(\hat{y}_{t-1}^2) + E(\epsilon_t \hat{y}_{t-1}) \\ &= \alpha_1 \mathrm{var}(y_{t-1}) = \frac{\alpha_1 \sigma^2}{(1-\alpha_1^2)} \end{aligned}$$

- In terms of correlations, corr$(y_t, y_{t-1}) = \alpha_1$.

# To work out: covariance and correlation at lag 2 for AR(1)

- What is the autocovariance at lag 2 for an AR(1) model?
- Solution:

$$
\begin{aligned}
\hat{y}_t &= \alpha_1 \hat{y}_{t-1} + \epsilon_t \\
\text{cov}(y_t, y_{t-2}) &= \alpha_1 E(\hat{y}_{t-1})(\hat{y}_{t-2}) = \alpha_1 E(\hat{y}_{t-1}\hat{y}_{t-2}) + E(\epsilon_t \hat{y}_{t-2}) \\
&= \alpha_1 cov(y_t y_{t-1}) \text{ assuming stationarity} \\
&= \frac{\alpha_1^2 \sigma^2}{(1-\alpha_1^2)}
\end{aligned}
$$

- $\text{Corr}(y_t, y_{t-2}) = \alpha_1^2$.
- Generally, $\text{Corr}(y_t, y_{t-s}) = \alpha_1^s$.

# Stationarity of MA(1)

- $y_t = \alpha_0 + \theta_1 \epsilon_{t-1} + \epsilon_t$ where $\epsilon \sim$ w.n. with mean zero.
- Mean of MA(1): E($y_t$)

$$\begin{aligned} E(y_t) &= E(\alpha_0 + \theta_1 \epsilon_{t-1} + \epsilon_t) \\ &= \alpha_0 + \theta_1 E(\epsilon_{t-1}) + E(\epsilon_t) \\ &= \alpha_0 \end{aligned}$$

Mean $= \alpha$

# Stationarity of MA(1)

- $y_t = \alpha_0 + \theta_1 \epsilon_{t-1} + \epsilon_t$
- Variance: $\mathrm{E}(y_t - E(y_t))^2$

$$\begin{aligned}
\mathrm{var}(y_t) = E(y_t - \alpha_0)^2 &= E(\theta_1 \epsilon_{t-1} + \epsilon_t)^2 \\
&= \theta_1^2 E(\epsilon_{t-1})^2 + 2\theta_1 E(\epsilon_{t-1}\epsilon_t) + E(\epsilon_t)^2 \\
&= (1 + \theta_1^2)\sigma_\epsilon^2
\end{aligned}$$

Variance $= (1 + \theta_1^2)\sigma_\epsilon^2$.

- Both the mean and the variance are constants.

# Stationarity of MA(1), contd.

- Covariance at lag 1: $E((y_t - \mu)(y_{t-1} - \mu))$

$$
\begin{aligned}
E(y_t - \alpha_0)(y_{t-1} - \alpha_0) &= E(\theta_1 \epsilon_{t-1} + \epsilon_t)(\theta_1 \epsilon_{t-2} + \epsilon_{t-1}) \\
&= \theta_1^2 E(\epsilon_{t-1}\epsilon_{t-2}) + \theta_1 E(\epsilon_{t-1})^2 \\
&\quad + \theta_1 E(\epsilon_t \epsilon_{t-2}) + E(\epsilon_t \epsilon_{t-1}) \\
&= \theta_1 \sigma_\epsilon^2
\end{aligned}
$$

# To work out: covariance at lag 2 of MA(1)

- Covariance at lag 2: $E((y_t - \mu)(y_{t-2} - \mu))$

$$
\begin{aligned}
E(y_t - \alpha_0)(y_{t-1} - \alpha_0) &= E(\theta_1 \epsilon_{t-1} + \epsilon_t)(\theta_1 \epsilon_{t-3} + \epsilon_{t-2}) \\
&= \theta_1^2 E(\epsilon_{t-1} \epsilon_{t-3}) + \theta_1 E(\epsilon_t \epsilon_{t-3}) + \theta_1 E(\epsilon_{t-} \\
&= 0
\end{aligned}
$$

- **HW:** Work out that autocovariances at lag 3 and all further lags are also 0.

# Stationarity properties for AR(1), MA(1)

- AR1: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t; \;\; \epsilon \sim N(0, \sigma^2)$
  - Mean: $\frac{\alpha_0}{1-\alpha_1}$
  - Variance: $\frac{\alpha_0 \sigma^2}{1-\alpha_1}$
  - Correlation (at lag $s$): $\alpha_1^s$
- MA1: $y_t = \alpha_0 + \theta_1 \epsilon_{t-1} + \epsilon_t; \;\; \epsilon \sim N(0, \sigma^2)$
  - Mean: $\alpha_0$
  - Variance: $(1 - \theta_1^2)\sigma^2$
  - Correlation (at lag $s = 1$): $\theta_1$
    Correlation (at lag $s > 1$): 0
- Summary: All the conditions are met for both the AR(1) and MA(1) models to be stationary.

# Theoretical analysis of the information content of MA models and AR models

# Defining information sets in the simple AR-MA models

- Given a time series, $y_t, y_{t-1}, y_{t-2}, \ldots, y_0$.
- In both the models above, each observation at time $t$, $y_t$ has two components:
  1. The innovation, $\epsilon_t$, which is revealed in period $t$.
  2. The permanent component of the information set $I_t | y_{t-1}, y_{t-2}, \ldots, y_0$
- Understanding the time behaviour of a stochastic process is to identify which part of $y_t$ derives from the permanent component, $I_t$ and which from the innovation $\epsilon_t$.

# Information sources in the AR-MA framework

- Under the AR(1) model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$$

  The structure of information appears to be from the permanent components.

- Under the MA(1) model:

$$y_t = \alpha_0 + \theta_1 \epsilon_{t-1} + \epsilon_t$$

  The structure of information appears to be from the innovations.

- However, the base information is still the same – the innovation series:

$$\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}, \cdots, \epsilon_0$$

# MA models have limited past information

- MA(1):

$$\begin{aligned} y_t &= \alpha_0 + \theta_1 \epsilon_{t-1} + \epsilon_t \\ y_{t-1} &= \alpha_0 + \theta_1 \epsilon_{t-2} + \epsilon_{t-1} \\ y_{t-2} &= \alpha_0 + \theta_1 \epsilon_{t-3} + \epsilon_{t-2} \end{aligned}$$

- There is a link between $y_t$ and periods $t, t-1$.
  This is the nature of the link between every $y_{t-s}$. It is linked to information from $t-s, t-s-1$.

# AR models have much more dependence on past information

- AR(1), where we set $\alpha_0 = 0$:

$$
\begin{aligned}
y_t &= \alpha_1 y_{t-1} + \epsilon_t \\
y_{t-1} &= \alpha_1 y_{t-2} + \epsilon_{t-1} \\
\rightarrow y_t &= \alpha_1(\alpha_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \alpha_1^2 y_{t-2} + \alpha_1 \epsilon_{t-1} + \epsilon_t \\
y_{t-2} &= \alpha_1 y_{t-3} + \epsilon_{t-2} \\
\rightarrow y_t &= \alpha_1^2(\alpha_1 y_{t-3} + \epsilon_{t-2}) + \alpha_1 \epsilon_{t-1} + \epsilon_t \\
&= \alpha_1^3 y_{t-3} + \alpha_1^2 \epsilon_{t-2} + \alpha_1 \epsilon_{t-1} + \epsilon_t \\
y_{t-3} &= \alpha_1 y_{t-4} + \epsilon_{t-3} \\
\rightarrow y_t &= \alpha_1^3(\alpha_1 y_{t-4} + \epsilon_{t-3}) + \alpha_1^2 \epsilon_{t-2} + \alpha_1 \epsilon_{t-1} + \epsilon_t
\end{aligned}
$$

# AR models have much more dependence on past information

- If we unravel this to the starting point in the series, $y_0, \epsilon_0$:

$$y_t = \epsilon_t + \sum_{i=1}^{T} \alpha_1^i \epsilon_{t-i} + \alpha_1^T y_0$$

- There is a link between $y_t$ and every single period since the start of the series at $t = 0$, $y_0$.
- How much is the dependence between $y_t$ and old information depends upon the sigze of $\alpha_1$.
  Larger $\alpha_1$ (closer to 1, or -1), the stronger the dependence.
  Smaller $\alpha_1$ (closer to 0), the weaker the dependence.
- **Note:** Read, Hamilton on "Difference Equations" (Chapter 1, Section 1.1).

# The link between AR and MA models

- We can re-write the AR(1) model from:

$$y_t = \epsilon_t + \sum_{i=1}^{T} \alpha_1^i \epsilon_{t-i} + \alpha_1^T y_0$$

as a function of lag operators as:

$$\begin{aligned} y_t &= \alpha_1 L(y_t) + \epsilon_t \\ (1 - \alpha_1 L) y_t &= \epsilon_t \\ \text{or, } y_t &= \epsilon_t \frac{1}{(1 - \alpha_1 L)} \\ &= (1 + \alpha_1 L + \alpha_1 L^2 + \alpha_1 L^3 + \ldots \ldots) \epsilon_t \end{aligned}$$

This becomes a geometric series of lagged operators on $\epsilon_t$.
- Thus, the AR(1) model becomes an MA($\infty$) model.

# Interpreting stationarity conditions for AR/MA models

- For example, with the constraints, we can re-write the AR(1) as:

$$
\begin{aligned}
\text{AR}(1), y_t &= \alpha_1 y_{t-1} + \epsilon_t \\
\rightarrow y_t &= \alpha_1 L y_t + \epsilon_t \\
\rightarrow (1 - \alpha_1 L) y_t &= \epsilon_t \\
\rightarrow y_t &= \frac{\epsilon_t}{(1 - \alpha_1 L)}
\end{aligned}
$$

- Generally, an invertible AR(p) process yields a specific MA($\infty$) one:

$$
\begin{aligned}
\alpha(L) y_t = (1 - \sum_{i=1}^{p}) y_t &= \alpha_0 + \epsilon_t \\
y_t &= \alpha^{-1}(L)(\alpha_0 + \epsilon_t)
\end{aligned}
$$

# Interpreting stationarity conditions for AR/MA models

- For an AR(1) process to be stationary, the constraints on the coefficients are that each coefficient should fall within the bounds of $(-1, 1)$.
  With these constraints, an AR(1) process becomes "invertible."
- More generally, for an AR(p) process written as:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \epsilon_t$$

  the necessary condition on the coefficients becomes:

$$-1 < \sum_{i=1}^{p} \alpha_i < 1$$

# Interpreting the coefficients of an MA processes

- An MA process is *always* bounded as long as the *number* of the lagged terms are limited.
- Therefore, even if the values of the MA coefficients are greater than one, the process will be stationary.

# MA($\infty$) processes

$$y_t = \alpha + \epsilon_t + \sum_{j=1}^{\infty} \theta_j \epsilon_{t-j} = \alpha + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \ldots$$

- This process is auto-covariance stationary as long as the values of the coefficients are bounded.
- The condition that we check for is:

$$\sum_{j=0}^{\infty} \theta_j \; < \; \infty$$

# Analysing sample behaviour of data from an AR(1)/MA(1) process

# Sample vs. theoretical autocorrelations

- If the autocovariance is denoted by $C_{t,t-s}$, then we can define the autocorrelation as:

$$
\begin{aligned}
\rho_s &= C_{t,t-s}/\sqrt{\sigma_t^2 \sigma_{t-s}^2} \\
\sigma_t^2 &= \sigma_{t-s}^2 \\
&= C_{t,t-s}/\sigma_t^2 \\
&= \theta \text{ at s = 1 for MA(1) process} \\
&= 0 \text{ at s } > 1 \text{ for MA(1)}
\end{aligned}
$$

- Given either AR or MA models, we know the theoretical form of the autocorrelations (as given above).
- However, we only work with sample estimators of the same, $\hat{\rho}_s$.
  Thus we need inference on the estimators to identify what type of linear stochastic process most likely fits the sample.

# Sample vs. theoretical autocorrelations

- The tool most often is the *Autocorrelation function* (ACF).
- This is a graph of the sample autocorrelations (on the y-axis) against the lag (on the x-axis).

# Inference for the autocorrelation estimates

- $H_0 : \rho_s = 0$.
- What is the standard error of the autocorrelation coefficient estimators (AC)?

# Bartlett's approximation for $\sigma(\rho_k)$

- Bartlett 1928: approximation of the variance of an estimated AC of a stationary, normal process is:

$$var(\hat{\rho}_k) \sim \frac{1}{T} \sum_{i=-\infty}^{\infty} \rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_i\rho_{i+k}\rho_{i-k} + 2\rho_i^2\rho_k^2$$

For any given model, we can calculate what the variance of $\rho_i$ ought to be approximately.

# Bartlett's approximation for $\sigma(\rho_k)$

- For example, for an AR(1) model, $\text{var}(\hat{\rho}_1) = \frac{1}{T}(1 - \phi^2)$.
- For example, for an MA(q) model, where for $k > q$, $\rho_k = 0$, Bartlett's approximation works out to be:

$$\text{var}[\hat{\rho}_k] \sim \frac{1}{T}(1 + \sum_{i=1}^{q} \rho_i^2)$$

- For an AR(k) model, where $k \to \infty$ and if $\phi$ is different from 1, then

$$\text{var}[\hat{\rho}_k] \sim \frac{1}{T}(1 + \sum_{i=1}^{q} \rho_i^2)$$

- Most often used generalised form:

$$var(\hat{\rho}_k) = \frac{1}{T}$$