Autoregressive moving average models

Kevin Kotzé

Contents

Univariate models for persistent data

- Dominant feature of many time series is that today's values are close to tomorrow's values
- Observations are not independent, but autocorrelated
- Need to account for this behaviour in the explained part of the model, otherwise it will be captured by the error, which violates the assumptions of the model
- Example of stochastic process:

$$y_t = 0.7y_{t-1} + \varepsilon_t$$

- This could represent an example of a linear stochastic difference equation, that includes discrete information
- Descriptive information should be used to populate the coefficient and random noise should be contained in the error

Moving average models

- Linear combination of white noise (i.e. ε_t), such that the $MA(1)$ may take the form,

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

- where μ is a constant, while ε_t and ε_{t-1} are independent and identically distributed white noise, $\varepsilon_t \sim \mathsf{i.i.d.}\ \mathcal{N}(0, \sigma^2)$
- To determine whether the $MA(1)$ process is stationary, we calculate the different moments

MA models - Expected Mean

- Note that $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = \sigma^2$,

$$\begin{aligned}
\mathbb{E}\left[y_t\right] &= \mathbb{E}[\mu + \varepsilon_t + \theta\varepsilon_{t-1}] \\
&= \mu + \mathbb{E}[\varepsilon_t] + \theta\mathbb{E}\left[\varepsilon_{t-1}\right] \\
&= \mu
\end{aligned}$$

- Since error terms are **i. i. d.** and their expected mean value is zero

- Hence, the mean for this process is μ, which is constant and does not depend on time

MA models - Variance

$$
\begin{aligned}
\text{var}[y_t] &= \mathbb{E}\left[y_t - \mathbb{E}[y_t]\right]^2 \\
&= \mathbb{E}\left[\,(\mu + \varepsilon_t + \theta\varepsilon_{t-1}) - \mu\right]^2 \\
&= \mathbb{E}[\varepsilon_t]^2 + 2\theta\mathbb{E}[\varepsilon_t\varepsilon_{t-1}] + \mathbb{E}[\theta\varepsilon_{t-1}]^2 \\
&= \sigma^2 + 0 + \theta\sigma^2 \\
&= \left(1 + \theta^2\right)\sigma^2
\end{aligned}
$$

- which is constant and does not depend on time

MA models - Covariance

- For the first lag,

$$
\begin{aligned}
\textbf{cov}[y_t, y_{t-1}] &= \mathbb{E}\Big[\big(y_t - \mathbb{E}\left[y_t\right]\big)\big(y_{t-1} - \mathbb{E}\left[y_{t-1}\right]\big)\Big] \\
&= \mathbb{E}\big[\,(\varepsilon_t + \theta\varepsilon_{t-1})\,(\varepsilon_{t-1} + \theta\varepsilon_{t-2})\,\big] \\
&= \mathbb{E}\,\left[\varepsilon_t\varepsilon_{t-1}\right] + \theta\mathbb{E}[\varepsilon_{t-1}^2] + \mathbb{E}[\theta\varepsilon_t\varepsilon_{t-2}] + \mathbb{E}[\theta^2\varepsilon_{t-1}\varepsilon_{t-2}] \\
&= 0 + \theta\sigma^2 + 0 + 0 \\
&= \theta\sigma^2
\end{aligned}
$$

- which is constant and does not depend on time

MA models - Covariance

- For the general case of j lags,

$$
\begin{aligned}
\textbf{cov}[y_t, y_{t-j}] &= \mathbb{E}\Big[\big(y_t - \mathbb{E}\left[y_t\right]\big)\big(y_{t-j} - \mathbb{E}\left[y_{t-j}\right]\big)\Big] \\
&= \mathbb{E}\left[\left(\varepsilon_t + \theta\varepsilon_{t-1}\right)\left(\varepsilon_{t-j} + \theta\varepsilon_{t-j}\right)\right] \\
&= 0 \quad \text{for} \;\; j > 1
\end{aligned}
$$

- which is constant and does not depend on time

MA models - Stationarity

- Neither the mean, variance nor covariances depend on time
- Hence the $MA(1)$ process is covariance stationary
- Such a $MA(1)$ process is stationary regardless of the value θ

MA models - ACFs

- ACF for a $MA(1)$ may then be derived from the expression,

$$\rho\left(j\right) \equiv \frac{\gamma\left(j\right)}{\gamma\left(0\right)} = \frac{\mathsf{cov}[y_t, y_{t-j}]}{\mathsf{var}[y_t]}$$

- Hence,

$$\begin{aligned}\rho\left(1\right) &= \frac{\theta}{(1+\theta^2)} \\ \rho\left(j\right) &= 0 \quad \text{for} \quad j > 1\end{aligned}$$

- for lag orders $j > 1$, the autocorrelations are zero

Figure 1: Simulated $MA(1)$: $\varepsilon_t - 0.5\varepsilon_{t-1}$

Figure 2: Autocorrelation Functions for $MA(1)$: $\varepsilon_t - 0.5\varepsilon_{t-1}$

MA models - Higher Order

- Finite order $MA(q)$ process may be,

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

- Infinite-order moving average process, $MA(\infty)$,

$$y_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} = \mu + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots$$

- With $\theta_0 = 1$

MA models - Higher Order

- After excluding extreme cases,

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

- which implies that the coefficients are absolute summable
- Moreover, the process is covariance-stationary when,

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty$$

MA models - Identifying the order

- With a $MA(1)$ process the effect of the shock ε_{t-1} affects the value of y_t
- Hence, the value for the first autocorrelation, $\rho(1)$ should differ from zero but the others would not
- With a $MA(2)$ process the effect of the shocks ε_{t-1} and ε_{t-2} affect the value of y_t
- Hence, the value for the first two autocorrelations, $\rho(1)$ and $\rho(2)$ should differ from zero but the others would not
- This would allow us to use the ACF to identify the order of an $MA(q)$ process

Figure 3: Identifying the order - $MA(1)$, $MA(2)$ & $MA(3)$ process

AR models - Solutions

- Given the $AR(1)$,

$$y_t = \phi y_{t-1} + \varepsilon_t$$

- Relates the value of a variable y at time t, to its previous value at time $(t-1)$ and a random disturbances ε, also at time t

- Assuming that ε_t is independent and identically distributed white noise, $\varepsilon_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2)$

- We showed that if $|\phi| < 1$, the $AR(1)$ is covariance-stationary,

$$\begin{aligned} \mathbb{E}\left[y_t\right] &= 0 \\ \text{var}[y_t] &= \frac{\sigma^2}{1-\phi^2} \\ \text{cov}[y_t, y_{t-j}] &= \phi^j \text{var}[y_t] \end{aligned}$$

- To prove this we use recursive substitution, method of undetermined coefficients, or lag operators

AR models - Recursive Substitution

- Starting at some period of time, j

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t \\ &= \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi^2(\phi y_{t-3} + \varepsilon_{t-2}) + \phi\varepsilon_{t-1} + \varepsilon_t \\ &= \vdots \\ &= \phi^{j+1} y_{t-(j+1)} + \phi^j \varepsilon_{t-j} + \ldots + \phi^2 \varepsilon_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

- Explains y as a linear function of the initial value $y_{t-(j+1)}$ and the historical values of ε_t
- If $|\phi| < 1$ and j becomes large, $\phi^{j+1} y_{t-(j+1)} \to 0$
- Thus, the $AR(1)$ can be expressed as an $MA(\infty)$
- Note that if $|\phi| > 1$ and j becomes large, $\phi^j \to \infty$
- Hence, the equivalent of an autoregressive random walk is an moving average with coefficients that are not summable

AR models - Lag operators

- Lag operators are particularly useful when dealing with more complex model structures

- The straightforward $AR(1)$ model can be written as,

$$(1 - \phi L)\, y_t = \varepsilon_t$$

- Such a sequence $\{y_t\}_{t=-\infty}^{\infty}$ is bounded if there exists a finite number k, such that $|y_t| < k$ for all t

- Provided $|\phi| < 1$ and we restrict ourselves to bounded sequences, we can multiply by $(1 - \phi L)^{-1}$ on both sides of the equality (the process is invertible),

$$\begin{aligned}(1 - \phi L)^{-1} (1 - \phi L)\, y_t &= (1 - \phi L)^{-1} \varepsilon_t \\ y_t &= (1 - \phi L)^{-1} \varepsilon_t\end{aligned}$$

AR models - Lag operators

- Under the assumption that $|\phi| < 1$, we can apply the geometric rule,

$$(1-\phi L)^{-1} = \lim_{j \to \infty} \left(1 + \phi L + (\phi L)^2 + \ldots + (\phi L)^j\right)$$

- This is based on the expression, $(1-z)^{-1} = 1 + z + z^2 + z^3 + \ldots$, which holds if $|z| < 1$

- Using this we can solve for,

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

AR models - Lag operators

- This expression could be written as a $MA(\infty)$,

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$$

- Therefore, when $|\phi| < 1$,

$$\sum_{j=0}^{\infty} |\theta_j| = \sum_{j=0}^{\infty} |\phi^j|$$

AR models - Unconditional Moments

- The unconditional first-and second-order moments of a stable $AR(1)$ process may be represented by an $MA(\infty)$,
- Where for $y_t = \phi y_{t-1} + \varepsilon_t$,

$$\mathbb{E}\left[y_t\right] = \mathbb{E}\left[\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots\right] = 0$$

- The variance is then,

$$\begin{aligned}
\gamma\left[0\right] &= \textbf{var}\left[y_t\right] = \mathbb{E}\Big[y_t - \mathbb{E}\left[y_t\right]\Big]^2 \\
&= \mathbb{E}\Big[\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots\Big]^2 \\
&= \textbf{var}\left[\varepsilon_t\right] + \phi^2\textbf{var}\left[\varepsilon_{t-1}\right] + \phi^4\textbf{var}\left[\varepsilon_{t-2}\right] + \phi^6\textbf{var}\left[\varepsilon_{t-3}\right] + \ldots \\
&= \Big(1 + \phi^2 + \phi^4 + \phi^6 + \ldots\Big)\,\sigma^2 \\
&= \frac{1}{1-\phi^2}\sigma^2
\end{aligned}$$

AR models - Unconditional Moments

- The first order covariance is then,

$$\begin{aligned}
\gamma(1) &= \mathbb{E}\Big[\big(y_t - \mathbb{E}[y_t]\big)\big(y_{t-1} - \mathbb{E}[y_{t-1}]\big)\Big] \\
&= \mathbb{E}\left[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots) \times (\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \ldots)\right] \\
&= \left(\phi + \phi^3 + \phi^5 + \ldots\right)\sigma^2 = \phi\left(1 + \phi^2 + \phi^4 + \ldots\right)\sigma^2 \\
&= \phi\frac{1}{1-\phi^2}\sigma^2 \\
&= \phi \textsf{var}[y_t]
\end{aligned}$$

- While for $j > 1$ we have,

$$\gamma(j) = \mathbb{E}\Big[\big(y_t - \mathbb{E}[y_t]\big)\big(y_{t-j} - E[y-j]\big)\Big] = \phi^j \textsf{var}[y_t]$$

- which proves the result relating to the stationarity of the $AR(1)$ model when $|\phi| < 1$

AR models - Unconditional Moments

- As noted previously the ACF for an $AR(1)$ process coincides with its impulse response function
- where for the ACF of an $AR(1)$ for $j = 1, \ldots, J$

$$\rho\left(0\right) = \frac{\gamma\left(0\right)}{\gamma\left(0\right)} = 1,\ \rho\left(1\right) = \frac{\gamma\left(1\right)}{\gamma\left(0\right)} = \phi, \ldots, \rho\left(j\right) = \frac{\gamma\left(j\right)}{\gamma\left(0\right)} = \phi^{j}$$

- Which equals the dynamic multipliers that may be summarised by the impulse response function

$$\frac{\partial y_t}{\partial \varepsilon_t} = 1, \frac{\partial y_t}{\partial \varepsilon_{t-1}} = \phi, \ldots, \frac{\partial y_t}{\partial \varepsilon_{t-j}} = \phi^{j}$$

Figure 4: Autocorrelation functions for $AR(1)$ processes

AR models - Adding a constant

- To ascertain how the results change after adding a constant,

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t$$

- We can define $v_t = \mu + \varepsilon_t$, such that,

$$\begin{aligned} y_t &= \phi y_{t-1} + v_t \\ y_t &= (1 - \phi L)^{-1} v_t \\ &= \left(\frac{1}{1-\phi} \right) \mu + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \ldots \end{aligned}$$

- with unconditional first moment,

$$\mathbb{E}\left[y_t \right] = \left(\frac{1}{1-\phi} \right) \mu$$

- which does not depend on time

AR models - Higher order processes

- For higher-order autoregressive processes, things become a bit more complicated, where

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

- No longer able to consider the value of ϕ_1 alone to determine whether it is stationary
- To complete the process we have to rewrite the $AR(2)$ expression as a first order difference equation

AR models - Higher order processes

- Using a vector, Z_t, which is of dimension (2×1),

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

- With a vector for the errors,

$$\upsilon_t = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

- And the (2×1) matrix for the coefficients,

$$\Gamma = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$$

AR models - Higher order processes

- The first-order vector difference equation can be written,

$$Z_t = \Gamma Z_{t-1} + v_t$$

- The matrix Γ is termed the *companion form* matrix of the $AR(2)$ process
- To check for stationarity we can compute the eigenvalues of this matrix
- Moreover, the eigenvalues of Γ are two solutions of x polynomial that satisfy the characteristic equation:

$$x^2 - \phi_1 x - \phi_2 = 0$$

AR models - Higher order processes

- These eigenvalues $(m_1$ and $m_2)$ must then satisfy $(x - m_1)(x - m_2)$, and can be found from the formula:

$$m_1, m_2 = \frac{\left(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}\right)}{2}$$

- Stationarity requires that the eigenvalues are less than one in absolute value

- In the $AR(2)$ case, one can show that this will be the case if,

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ -\phi_1 + \phi_2 &< 1 \\ \phi_2 &> -1 \end{aligned}$$

Figure 5: Eigenvalues for difference equation $x^2 - 0.6x - 0.2 = 0$

AR models - Higher order processes

- The $AR(p)$ can then be written as,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

- Checking for stationarity involves similar calculations

- In this case the Γ matrix will be of the form:

$$\Gamma = \begin{bmatrix} \phi_1 & \phi_2 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

- Provided the eigenvalues are less than one in absolute value, (i.e. they lie within the unit circle), the p^{th} order autoregression is stable

AR models - Identify the order of AR(p)

- As in the case of the $MA(q)$ processes one could use the ACF coefficients to identify the order of the $AR(p)$ process

- However, as the $AR(p)$ process passes on the persistence to successive lags so the ACF would not be useful

- As the PACF removes the effects of the persistence that is passed on from intervening lags of the $AR(p)$ process it may be used to identify the order of an $AR(p)$ process

ARMA models

- We can specify an $ARMA(1,1)$ process as,

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

- Or using the lag polynomials, a general form of an ARMA model is,

$$\phi(L) y_t = \theta(L) \varepsilon_t$$

- Note that the number of lags, (p) and (q), can differ

- For instance, an $ARMA(2,1)$ combines an $AR(2)$ with an $MA(1)$:

$$\begin{aligned} \left(1 - \phi_1 L - \phi_2 L^2\right) y_t &= (1 + \theta_1 L) \varepsilon_t \\ y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \end{aligned}$$

ARMA processes

- Whether an $ARMA(p, q)$ process is stationary depends solely on its autoregressive past
- Assume an $ARMA(1, 1)$ and using the lag operator,

$$(1 - \phi L)\, y_t = (1 + \theta L)\, \varepsilon_t$$

- Multiplying by $(1 - \phi L)^{-1}$ on both sides,

$$\begin{aligned} y_t &= \frac{(1 + \theta L)}{(1 - \phi L)} \varepsilon_t \\ &= (1 - \phi L)^{-1} \varepsilon_t + (1 - \phi L)^{-1} \theta_1 \varepsilon_{t-1} \end{aligned}$$

ARMA processes

- When $|\phi| < 1$, this can be written as the geometric process,

$$
\begin{aligned}
y_t &= \sum_{j=0}^{\infty} (\phi L)^j \varepsilon_t + \theta \sum_{j=0}^{\infty} (\phi L)^j \varepsilon_{t-1} \\
&= \varepsilon_t + \sum_{j=1}^{\infty} \phi^j \varepsilon_{t-j} + \theta \sum_{j=1}^{\infty} \phi^{j-1} \varepsilon_{t-j} \\
&= \varepsilon_t + \sum_{j=1}^{\infty} \left(\phi^j + \theta \phi^{j-1} \right) \varepsilon_{t-1}
\end{aligned}
$$

Figure 6: ACF and PACF for $AR(1)$ with $\phi = 0.5$

Figure 7: ACF and PACF for $MA(1)$ with $\theta = 0.6$

Figure 8: ACF and PACF for $ARMA(1, 1)$ with $\phi = 0.5$ and $\theta = 0.6$

Autocorrelation patterns

- When combining the AR and MA correlation functions the results may be somewhat unclear
- Possibly $ARMA(2, 2)$, $ARMA(1, 2)$, $ARMA(2, 1)$, $ARMA(1, 1)$, $MA(2)$ or $AR(2)$

Seasonal ARMA Models

- In several cases the dependence on the past occurs with a seasonal lag s
- With monthly economic data the behaviour in Jan 2010 may be related to Jan 2011
- Could introduce autoregressive and moving average terms that arise at a seasonal interval
- For example, $ARMA(p,q)_s$ model that takes the form $ARMA(1,1)_{12}$ would be written as,

$$y_t = \phi y_{t-12} + \varepsilon_t + \theta \varepsilon_{t-12}$$

- Estimation is relatively straightforward

Seasonal ARMA Models - Identification

- The $MA(1)$ with a seasonal $(s = 12)$, which could be written as, $y_t = \varepsilon_t + \theta\varepsilon_{t-12}$
- It is easy to verify that

$$
\begin{aligned}
\gamma(0) &= (1+\theta^2)\sigma^2 \\
\gamma(12) &= \theta\sigma^2 \\
\gamma(j) &= 0, \quad \text{for values where } j \neq 12
\end{aligned}
$$

- The only non-zero autocorrelation, aside from lag zero is, $\rho(12) = \theta/(1+\theta^2)$

Seasonal ARMA Models - Identification

- Similarly, for the $AR(1)$ model with seasonal $(s = 12)$, we could calculate,

$$\begin{aligned} \gamma(0) &= \sigma^2/(1-\phi^2) \\ \gamma(12) &= \sigma^2\phi^k/(1-\phi^2) \;\text{ for } k = 1, 2, \ldots \\ \gamma(j) &= 0, \;\text{ for values where } j \neq 12 \end{aligned}$$

- Results suggest the PACF from non-seasonal are analogous to the seasonal models

Seasonal ARMA Models - Identification

- Could allow for mixed seasonal models in the general $ARMA(p, q)_s$ framework,

$$y_t = \phi y_{t-12} + \varepsilon_t + \theta\varepsilon_{t-1}$$

- While estimation would be straightforward, the identification of the structural form may be problematic

Box-Jenkins methodology

- In a real world application we would not know the functional form of the underlying data generating process
- The respective parameters in these models would then need to be estimated
- Thereafter, we could assess the model fit
- This procedure is encapsulated in the Box & Jenkins (1979) methodology
  - Identification, Estimation, Diagnostic testing

Box-Jenkins - Identification

- Examine the time plot of the data to
  - detect and correct for outliers, missing values, structural breaks (if possible)
  - detect nonstationary by a pronounced trend or prolonged meander (possibly correct)
- If you are uncertain about the degree of stationarity then perform unit root tests
  - plot ACF and PACF to consider the persistence in the data
  - when ACF quickly returns to zero then there will be no unit root
- Alternatively, if you think that the data represents white noise then use Q-statistic

Box-Jenkins - Identification

- Calculate Q-statistic to test whether a group of autocorrelations is different from zero
- Originally developed by Box-Pierce (1970) better small-sample performance reported by Ljung and Box (1978)

$$Q = T(T+2)\sum_{k=1}^{s} \rho_j^2/(T-j)$$

- If sample value of Q exceeds the critical value χ^2 with s degrees of freedom then at least one value of ρ_j is statistically different from zero at specified significance level

Box-Jenkins - Identification

- Examine the ACF and PACF functions more closely to try to identify the order of a potential $ARMA(p, q)$
- For the ACF and PACF functions that were provided previously we would consider an $ARMA(2, 2)$, an $ARMA(1, 2)$, an $ARMA(2, 1)$ or an $ARMA(1, 1)$
- Would also think about using a $MA(2)$ or an $AR(2)$, but not a $MA(1)$ or an $AR(1)$

Box-Jenkins - Estimation Stage

Fit each of the candidate models and examine the various ϕ_i and θ_i coefficients according to:

- Parsimony:
  - Additional coefficients increase fit but reduce degrees of freedom
  - Parsimonious models often produce better out-of-sample fit
- Stationarity and Invertibility:
  - Distribution theory underlying the use of sample ACF and PACF as approximations for the true DGP assume that y_t is stationary
  - t-statistics and Q-statistics presume that the data is stationary
- Be suspicious if the estimated value of $|\phi_1|$ is close to unity
- Model must be invertible since the ACF and PACF assume that y_t can be approximated by an $AR(1)$ model where $|\phi_1| < 1$

Box-Jenkins - Estimation Stage

- To evaluate the different candidate models consider the goodness-of-fit measures:
  - Look at R^2 and average of the residual sum of squares
  - AIC and BIC are more suitable criteria since they weigh-up parsimony and "goodness-of-fit"
  - Smaller values of a AIC are better (or where AIC < 0, choose the model with the most negative statistic)

Box-Jenkins estimation - AIC & BIC

- Adding additional lags will reduce the sum of squares of the estimated residuals (and will lead to a higher R^2)
- But you will also loose degrees of freedom (which may be essential)
- Akaike and Bayesian Information Criteria test for goodness of fit, while prizing parsimony

$$AIC = \log \hat{\sigma}_k^2 + \frac{T + 2k}{T}$$
$$BIC = \log \hat{\sigma}_k^2 + \frac{k \log T}{T}$$

- where $k =$ number of estimated parameters and $T =$ nobs
- $\hat{\sigma}_k^2 = \frac{SSR_k}{T}$ is the variance of the residual sum of squares

Box-Jenkins - Estimation Stage

- Make sure T is fixed, when comparing $AR(1)$ & $AR(2)$
- Including a parameter must decrease SSR_k if AIC or BIC is to decrease
- Since $\log T$ is greater than 2, BIC likes more parsimonious models

Box-Jenkins - Diagnostic checking

- Plot the residuals to look for outliers or periods where the model does not fit the data
- Construct ACF and PACF of the residuals
- Serial correlation in the residuals implies that a systematic movement in the y_t sequence is not accounted for by the $ARMA(p, q)$ coefficients
    - Those models should be eliminated and re-estimated
    - Use Q-statistic to determine whether any or all of the ACF or PACF coefficients are significant
- When applying the Q-statistic to the residuals of an $ARMA(p, q)$ model use χ^2 with $s - p - q$ degrees of freedom
    - Ensure that the standard errors for the coefficient estimates are appropriate, if not re-estimate model

Box-Jenkins - Diagnostics & Forecasts

- If possible fit $ARMA(p,q)$ models to subsamples - stability of DGP

$$F = \frac{(RSS - RSS_1 - RSS_2)/n}{(RSS_1 + RSS2)/(T-2k)}$$

- where k is the number of parameters, i.e. $p+q+1$ (with constant)
- If all the coefficients are equal $(RSS_1 + RSS_2)$ should equal RSS and $F=0$
- You could then use the model for forecasting $y_{T+1}, y_{T+1}, \ldots$ for out-of-sample comparison

Structural Breaks - Chow's Breakpoint

- Model two sub-samples of the data and see whether there are significant differences in the parameters
- Test whether the null hypothesis of "no structural change" holds after constructing a F test statistic for the parameters
- Could construct a model for changes at date τ

$$y_t = x_t^\top \beta_t + \varepsilon_t$$

- where

$$\beta_t = \begin{cases} \beta & t \leq \tau \\ \beta + \delta & t > \tau \end{cases}$$

- Or alternatively we could test for change in all the model parameters with an F test

Structural Breaks - Chow's Breakpoint

- Major drawback is that the change point must be known *a priori*
- Must ensure that each sub-sample has at least as many observations as the number of estimated parameters

Structural Breaks - Quandt LR Test

- Extension of the Chow test where an F test statistic is calculated for all potential breakpoints within an interval $[\underline{i}, \bar{i}]$
- Reject the null hypothesis of no structural change if the absolute value of any of the test statistics are too large
- Takes the form of a sup F test
- Asymptotic properties of this statistic are non-standard so use those that are referenced in the notes

Figure 9: Quandt Likelihood Ratio Test - Breakpoint at observation 100 with $n = 200$ and $p = 0.00$

Structural Breaks - CUSUM Test

- CUSUM test is based on the cumulative sum of the recursive residuals
- Plot the cumulative sum together with the 5% critical boundaries
- If the cumulative sum breaks either of the two boundaries there is parameter instability and a possible structural break
- Need to specify the model *a priori* to obtain the residuals

Figure 10: CUSUM Test - Breakpoints for change in coeffic

Conclusion

- Relatively simple ARMA models can be used to describe stationary univariate time series
- Easy to estimate and use of the straightforward Box & Jenkins method can identify possible functional forms for the underlying data generating process
- It is possible to test the data generating process for structural breaks