# Forecasting Economic and Financial Time Series
# Week4: Cycles: AR, MA and ARIMA Models


Ron Smith
MSc/PGCE Option, EMS Birkbeck


Spring 2019, Revised version. New slides 14,18-20,55-57,60-63.

# Cycles

- Cycles are thought of as the dynamics not captured by the trend and seasonal component, which can take on the form of a irregular up and down type movement, more varied and less rigid in form.
- Less regular than seasonal patterns.
- All we require is that they have some form of dynamics, some persistence, which in some way links the present with the past and future the present - where cycles are present in most series
- However, unlike the type of trend and seasonality considered, cycles are more complicated, because of their wide variety of cyclical patterns: so as outlined in Diebold, we need break down the approach into tools/methods for modelling time series, specific models which characterise univariate representations and then how we might use these models to forecast

# Stationary Time Series

- A **realisation** of a time series is an ordered set, $\{...., y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}, ...\}$.
- In theory a time series realisation goes into the infinite future and past, but in practice is defined by a sample - but nonetheless the theoretical notion proves useful
- For forecasting models you need to assume that the probabilistic structure of the series is constant over time, otherwise we could not predict the future on the basis of the past, since the future process would differ
- A strongly stationary series has $f_t(y_t) = f(y_t)$ : distribution is constant through time (after removing any deterministic elements like trend and seasonal).
- A weakly (covariance/2nd order) stationary series has constant means, variances and covariances between current and past values, again after removing deterministic elements.

# Covariance Stationary Time Series

- **Mean** of series is covariance stationary if it is constant over time, so without any need for a time subscript:

$$E(y_t) = \mu$$

- The covariance structure needs to be stable, where we use the **autocovariance function,** which is just the covariance between $y_t$ and $y_{t-\tau}$, which may depend on both $t$ and $\tau$:

$$\gamma(t,\tau) = cov(y_t, y_{t-\tau}) = E(y_t - \mu)(y_{t-\tau} - \mu)$$

- If the structure is stationary over time, then the autocovariances depend only on displacement, $\tau$, not on time and therefore:

$$\gamma(t,\tau) = \gamma(\tau)$$

  for all $t$ - symmetric, so $\gamma(\tau) = \gamma(-\tau)$
- We examine the autocovariance structure to examine the dynamic behaviour. We also require that the variance of the series - the autocovariance at displacement 0, $\gamma(0)$, be finite.

# Autocorrelation function 1

- Note we typically use the **autocorrelation function,** which acts like a correlation coefficient in that it normalises by dividing $\gamma(\tau)$ by the variance: $\gamma(0)$ :

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}, \qquad \tau = 0, 1, 2, ....$$

- So $\rho(0) = 1$, where we focus on autocorrelations beyond displacement 0.
- Also look at the partial autocorrelation function, $p(\tau)$, which is just the coefficient of $y_{t-\tau}$ in the regression of $y_t$ on $y_{t-1}, ...., y_{t-\tau}$
- Autocorrelation are just regular correlations between $y_t$ and $y_{t-\tau}$ whereas the partial autocorrelation is between $y_t$ and $y_{t-\tau}$ after controlling for earlier lags

# Autocorrelation function 2

- We graph autocorrelations and partial autocorrelations (like regression coefficients) against $\tau$ and examine their qualitative shape, in order that we learn something about what model best represent our data, which in turn, given covariance stationarity will form the best forecasting model.
- Many series are not covariance stationary, but is often possible to remove nonstationary components such as trend and seasonality, so that the cyclical component left over is likely to be covariance stationary
- Alternatively, simple transformations can transform a non-stationary series to covariance stationarity. Most usual here being a series non-stationary in levels may well be stationary in first differences.
- Below EViews AC and PAC for Columbian Coca production, (original and after removing quadratic trend?) and stock market volume.

# Coca Series

Date: 10/24/15  Time: 17:52
Sample: 1994Q1 2005Q3
Included observations: 47

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Pro |
|---|---|---|---|---|---|---|
| | | 1 | 0.615 | 0.615 | 18.939 | 0.00 |
| | | 2 | 0.289 | -0.14... | 23.217 | 0.00 |
| | | 3 | 0.562 | 0.734 | 39.744 | 0.00 |
| | | 4 | 0.775 | 0.213 | 71.906 | 0.00 |
| | | 5 | 0.394 | -0.42... | 80.403 | 0.00 |
| | | 6 | 0.070 | -0.35... | 80.681 | 0.00 |
| | | 7 | 0.273 | -0.20... | 84.969 | 0.00 |
| | | 8 | 0.428 | 0.016 | 95.778 | 0.00 |
| | | 9 | 0.102 | 0.099 | 96.402 | 0.00 |
| | | 1... | -0.19... | -0.06... | 98.741 | 0.00 |

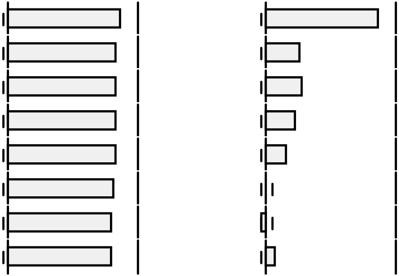
# Detrended Coca series

Date: 01/18/16 Time: 11:18
Sample: 1994Q1 2005Q3
Included observations: 47

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.217 | 0.217 | 2.3617 | 0.12 |
| | | 2 | -0.53... | -0.61... | 17.091 | 0.00 |
| | | 3 | 0.079 | 0.672 | 17.418 | 0.00 |
| | | 4 | 0.735 | 0.228 | 46.382 | 0.00 |
| | | 5 | 0.031 | -0.47... | 46.436 | 0.00 |
| | | 6 | -0.64... | 0.035 | 69.593 | 0.00 |
| | | 7 | -0.08... | -0.21... | 70.019 | 0.00 |
| | | 8 | 0.492 | -0.10... | 84.310 | 0.00 |
| | | 9 | -0.07... | 0.048 | 84.632 | 0.00 |
| | | 1... | -0.65... | -0.24... | 111.33 | 0.00 |

# Volume Series

Date: 11/24/15 Time: 16:34
Sample: 1/01/1947 7/01/2013
Included observations: 1747



| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.910 | 0.910 | 1448.1 | 0.00 |
| | | 2 | 0.873 | 0.266 | 2783.6 | 0.00 |
| | | 3 | 0.870 | 0.282 | 4110.8 | 0.00 |
| | | 4 | 0.874 | 0.232 | 5451.3 | 0.00 |
| | | 5 | 0.872 | 0.159 | 6786.6 | 0.00 |
| | | 6 | 0.851 | -0.00... | 8058.8 | 0.00 |
| | | 7 | 0.830 | -0.04... | 9267.3 | 0.00 |
| | | 8 | 0.830 | 0.068 | 10479. | 0.00 |

# White Noise Processes 1

- The simplest process is a **white noise** process, which forms the basis of the other processes. Suppose that:

$$\begin{aligned} y_t &= \varepsilon_t \\ E(\varepsilon_t) &= 0, \\ E(\varepsilon_t^2) &= \sigma^2 \\ E(\varepsilon_t \varepsilon_{t-i}) &= 0; \ i \neq 0. \end{aligned}$$

  so the "shock", $\varepsilon_t$, is uncorrelated overtime. Therefore $\varepsilon_t$ and hence $y_t$ are **serially uncorrelated.**

- Assuming $\sigma^2 < \infty$, such a process - with zero mean, constant variance and no serial correlation - is called **zero-mean white noise**, or simply **white noise,** such that we write:

$$\begin{aligned} \varepsilon_t &\sim WN(0, \sigma^2) \\ y_t &\sim WN(0, \sigma^2) \end{aligned}$$

# White Noise Processes 2

- Serially uncorrelated, does not necessarily imply serially independent, unless they are normally distributed. If they are serially independent and normally distributed, then we say that $y$ is **normal/ Gaussian white noise:**

$$y_t \overset{iid}{\sim} N(0, \sigma^2)$$

  $y$ is independently and identically distributed as normal, with zero mean and constant variance.

- Since there are no patterns to this type of data due to the independence over time - you cannot forecast the future of the process. Understanding and identifying white noise is important:
  - Richer processes are constructed from white noise
  - 1-step-ahead forecast errors from good models under some loss functions should be white noise - if they are not then they are serially correlated which means they are predictable, which may mean the model can be improved. Therefore we need to be able to recognise white noise.

# White Noise Processes 3

- White noise has unconditional mean and variance:

$$\begin{aligned} y_t &= \varepsilon_t \\ E(y_t) &= 0 \\ var(y_t) &= \sigma^2 \end{aligned}$$

- These are constant, as a requirement of covariance stationary processes, where constant unconditional variance imply autocovariances are a function of displacement only and not time
- The variance is the autocovariance at displacement 0, but because white noise is uncorrelated overtime, then the autocovariances and hence autocorrelations are zero beyond displacement zero. Same for the partial autocorrelation function.

# Unconditional/conditional 1

- We distnguish conditional and unconditional means (expected values) and variances.
- Means and variances may be unconditional $E(y_t)$, $V(y_t)$, or conditional on some information for instance
  - $E(y|X) = X\beta$, $V(y|X) = \sigma^2 I$ in traditional linear regression or
  - $E(y_t|\Omega_{t-1})$ and $V(y_t|\Omega_{t-1})$ in forecasting, where $\Omega_{t-1}$ is information at time $t-1$.
- For white noise which are unpredictable.the conditional and unconditional means and variances are the same 0 and $\sigma^2$

$$\begin{aligned} E(y_t) &= E(y_t|\Omega_{t-1}) = 0 \\ V(y_t) &= V(y_t|\Omega_{t-1}) = E((y_t - E(y_t|\Omega_{t-1}))^2|\Omega_{t-1}) \\ &= \sigma^2 \end{aligned}$$

# Unconditional/conditional 2

- For stationary AR1

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

Conditional moments

$$\begin{aligned} E(y_t|\Omega_{t-1}) &= \alpha + \rho y_{t-1} \\ V(y_t|\Omega_{t-1}) &= \sigma^2 \end{aligned}$$

Unconditional moments

$$\begin{aligned} E(y_t) &= \frac{\alpha}{1-\rho} \\ V(y_t) &= \frac{\sigma^2}{1-\rho^2} \end{aligned}$$

# The Lag Operator

- We use lag operator extensively, defined by:

$$Ly_t = y_{t-1}, \qquad L^2 y_t = L(L(y_t)) = Ly_{t-1} = y_{t-2}$$

- Typically operate on a series with a finite polynomial lag operator:

$$B(L) = b_0 + b_1 L + b_2 L^2 + ..... + b_m L^m$$

- Or infinite

$$B(L) = b_0 + b_1 L + b_2 L^2 + ..... = \sum_{i=0}^{\infty} b_i L^i$$

- Simple examples:

$$L^m = y_{t-m}, \qquad \Delta y_t = (1 - L) y_t = y_t - y_{t-1}$$

- Infinite distributed lag of current and past shocks:

$$B(L)\varepsilon_t = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + ..... = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

# Wold Representation E&T135

- Call any deterministic part (e.g. intercept and trend), $\mu_t$. The Wold representation theorem says any covariance stationary process, $\{y_t\}$, can be represented as a linear combination of serially uncorrelated white noise terms and a linearly deterministic component.
- Let $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$ then :

$$\begin{aligned} y_t &= B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} + \mu_t \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

- Any covariance stationary series can be modelled as some infinite distributed lag of white noise errors.

# Wold Representation, continued

- The innovations, $\varepsilon_t$, correspond to 1-step-ahead forecast errors: the part of the evolution that is linearly unpredictable on the basis of the past and will be uncorrelated but not necessarily independent - only if Gaussian
- It is a general linear process: general because any covariance stationary series can be written this way and linear because it is expressed as a linear function of $\varepsilon_t$.

In practice we approximate the infinite MA by a parsimonious combination of finite AR and MA terms, after differencing enough to make it stationary

$$
\begin{aligned}
y_t &= B(L)\varepsilon_t \approx \frac{C(L)}{A(L)}\varepsilon_t \\
A(L)y_t &= C(L)\varepsilon_t.
\end{aligned}
$$

# Wold Representation ARMA(1,1)

Consider the infinite moving average associated with an ARMA(1,1)

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \varepsilon_t + \mu \varepsilon_{t-1} \\
A(L) y_t &= C(L) \varepsilon_t \\
(1 - \rho L) y_t &= (1 + \mu L) \varepsilon_t \\
y_t &= \frac{(1 + \mu L)}{(1 - \rho L)} \varepsilon_t \\
y_t &= \frac{C(L)}{A(L)} \varepsilon_t \\
y(t) &= B(L) \varepsilon_t
\end{aligned}
$$

# B(L) for ARMA(1,1)

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \varepsilon_t + \mu\varepsilon_{t-1} \\
&= \rho(\rho y_{t-2} + \varepsilon_{t-1} + \mu\varepsilon_{t-2}) + \varepsilon_t + \mu\varepsilon_{t-1} \\
&= \rho^2 y_{t-2} + \rho\mu\varepsilon_{t-2} + \varepsilon_t + (\rho+\mu)\varepsilon_{t-1} \\
&= \rho^2 \left(\rho y_{t-3} + \varepsilon_{t-2} + \mu\varepsilon_{t-3}\right) + \rho\mu\varepsilon_{t-2} + \varepsilon_t + (\rho+\mu)\varepsilon_{t-1} \\
&= \rho^3 y_{t-3} + \rho^2\mu\varepsilon_{t-3} + \varepsilon_t + (\rho+\mu)\varepsilon_{t-1} + \rho(\rho+\mu)\varepsilon_{t-2} \\
&= \rho^3 \left(\rho y_{t-4} + \varepsilon_{t-3} + \mu\varepsilon_{t-4}\right) \\
&\quad +\rho^2\mu\varepsilon_{t-3} + \varepsilon_t + (\rho+\mu)\varepsilon_{t-1} + \rho(\rho+\mu)\varepsilon_{t-2} \\
&= \rho^4 y_{t-4} + \rho^3\mu\varepsilon_{t-4} + \varepsilon_t + (\rho+\mu)\varepsilon_{t-1} + \rho(\rho+\mu)\varepsilon_{t-2} \\
&\quad +\rho^2(\rho+\mu)\varepsilon_{t-3}
\end{aligned}
$$

# Continuing the sequence

We see how a simple low order ARMA can give a a complicated Wold representation as an infinite moving average:

$$
\begin{aligned}
y_t &= \varepsilon_t + (\rho+\mu)\varepsilon_{t-1} + \rho(\rho+\mu)\varepsilon_{t-2} + \rho^2(\rho+\mu)\varepsilon_{t-3} \\
&\quad + \rho^3(\rho+\mu)\varepsilon_{t-4.} + \ldots\ldots \\
y_t &= (1 + (\rho+\mu)L + \rho(\rho+\mu)L^2 + \rho^2(\rho+\mu)L^3 + \rho^3(\rho+\mu)L^4 + \ldots)\varepsilon_t \\
y_t &= \sum_{i=0}^{\infty} b_i L^i \varepsilon_t \\
b_0 &= 1; b_i = \rho^{i-1}(\rho+\mu),\ \ i \geqslant 1
\end{aligned}
$$

# General Linear Process: Unconditional Moments

For the general linear structure, taking the means and variances we obtain the unconditional moments:

$$
\begin{aligned}
E(y_t) &= E\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} b_i E(\varepsilon_{t-i}) = \sum_{i=0}^{\infty} b_i \times 0 = 0 \\
var(y_t) &= var\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} b_i^2 var(\varepsilon_{t-i}) \\
&= \sum_{i=0}^{\infty} b_i^2 \sigma^2 = \sigma^2 \sum_{i=0}^{\infty} b_i^2
\end{aligned}
$$

# General Linear Process: conditional moments

For information set $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$, the conditional mean moves in response to the evolving information set:

$$
\begin{aligned}
E(y_t|\Omega_{t-1}) &= E(\varepsilon_t|\Omega_{t-1}) + b_1 E(\varepsilon_{t-1}|\Omega_{t-1}) \\
&\quad + b_2 E(\varepsilon_{t-2}|\Omega_{t-1}) + ... \\
&= 0 + b_1\varepsilon_{t-1} + b_1\varepsilon_{t-2} + ... \\
&= \sum_{i=1}^{\infty} b_i \varepsilon_{t-i}
\end{aligned}
$$

The conditional variance:like the unconditional variance, is constant overtime.

$$
\begin{aligned}
Var(y_t|\Omega_{t-1}) &= E((y_t - E(y_t|\Omega_{t-1}))^2|\Omega_{t-1}) \\
&= E(\varepsilon_t^2|\Omega_{t-1}) = E(\varepsilon_t^2) = \sigma^2
\end{aligned}
$$

# Moving Average (MA) Models 1

- A finite-order moving average process is a natural and obvious approximation to the Wold representation, which is an infinite moving average process
- The fact that all time series are, in various ways, driven by shocks of various sorts suggest the possibility of modeling time series as distributed lags of current and past shocks.
- The first-order moving average process, or **MA(1) process**, is:

$$
\begin{aligned}
y_t &= \varepsilon_t + \theta\varepsilon_{t-1} = (1 + \theta L)\varepsilon_t \\
\varepsilon_t &\sim WN(0, \sigma^2)
\end{aligned}
$$

which defines $y_t$ as a function of unobservable shocks

# Moving Average (MA) Models 2

- The unconditional mean and variance are:

$$\begin{aligned} E(y_t) &= E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) = 0 \\ Var(y_t) &= Var(\varepsilon_t) + \theta^2 Var(\varepsilon_{t-1}) = \sigma^2 + \theta^2\sigma^2 = \sigma^2(1+\theta^2) \end{aligned}$$

- The conditional mean and variance, given information set $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$ are

$$\begin{aligned} E(y_t|\Omega_{t-1}) &= E(\varepsilon_t + \theta\varepsilon_{t-1}|\Omega_{t-1}) \\ &= E(\varepsilon_t|\Omega_{t-1}) + \theta E(\varepsilon_{t-1}|\Omega_{t-1}) = \theta\varepsilon_{t-1} \\ Var(y_t|\Omega_{t-1}) &= E((y_t - E(y_t|\Omega_{t-1}))^2|\Omega_{t-1}) \\ &= E(\varepsilon_t^2|\Omega_{t-1}) = E(\varepsilon_t^2) = \sigma^2 \end{aligned}$$

- The conditional mean adapts to the evolving information set, unlike the constant unconditional mean. In an MA(1) only the first lag of the shock enters the conditional mean - more distant shocks have no effect: a short memory.

# Moving Average (MA) Models 3

- The autocovariance function is:

$$\begin{aligned}\gamma(\tau) &= E(y_t y_{t-\tau}) = E((\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-\tau} + \theta\varepsilon_{t-\tau-1})) \\ &= \begin{cases} \theta\sigma^2, & \tau = 1 \\ 0, & \tau > 1 \end{cases}\end{aligned}$$

- The autocorrelation function, is the autocovariance function scaled by the variance:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \begin{cases} \frac{\theta}{1+\theta^2}, & \tau = 1 \\ 0, & \tau > 1 \end{cases}$$

- We could estimate $\theta$ from $\rho(1)$.

# Moving Average (MA) Models 4

- To compute the autoregressive representation rewrite the MA(1) as:

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

- By lagging successively:

$$\begin{aligned}\varepsilon_{t-1} &= y_{t-1} - \theta\varepsilon_{t-2} \\ \varepsilon_{t-2} &= y_{t-2} - \theta\varepsilon_{t-3} \\ \varepsilon_{t-3} &= y_{t-3} - \theta\varepsilon_{t-4}\end{aligned}$$

- Substituting back into the MA(1) expression $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ yields:

$$y_t = \varepsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - ....$$

- In lag operator notation:

$$\frac{1}{1+\theta L} y_t = \varepsilon_t$$

# Moving Average (MA) Models

- To represent the MA process as an AR process, it must be invertible, the roots must lie outside the unit circle..
- Polynomials of degree $m$ has $m$ roots, Therefore, the MA(1) lag operator polynomial has one root, which is the solution to $(1 + \theta L) = 0$, which is $L = -1/\theta$, so the inverse will be less than 1 in absolute value if $|\theta < 1|$
- Autoregressive representations are real world in the sense that they link present observables to the past history of observables based on present and past observables. Moving averages have this form if they can be inverted, which is why if we start with a MA process we restrict ourselves to invertible processes.
- Partial autocorrelation function of an MA(1) oscillates, from positive to negative, and declines to zero - this follows directly from the autoregressive representation derived.

# The MA(q) Process

- General finite-order process of order $q$ or MA(q):

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + .... + \theta_q \varepsilon_{t-q} = \Theta(L)\varepsilon_t \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

  where $\Theta(L) = 1 + \theta_1 L + ..... + \theta_q L^q$, is a $q$-th order lag operator polynomial.
- $MA(q)$ is covariance stationary for any value of its parameters.
- In an MA(q), only q shocks enter the conditional mean.and autocorrelations greater than $q$ are zero.

# AR(1)

- A stationary AR1 takes the form:

$$\begin{aligned} y_t &= \alpha + \rho y_{t-1} + \varepsilon_t, \\ y_t(1-\rho L) &= \varepsilon_t, \end{aligned}$$

  with $\mid \rho \mid < 1$ and $E(y_t) = \alpha/(1-\rho)$.,
- If it is stationary and repeated substitution, gives an $MA(\infty)$ :

$$\begin{aligned} y_t &= \alpha/(1-\rho) + \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \rho^3\varepsilon_{t-3}.... \qquad (1) \\ y_t &= \alpha/(1-\rho) + (1-\rho L)^{-1}\varepsilon_t, \end{aligned}$$

- $\text{Var}(y_t) = E(y_t - E(y_t))^2 = E(y_t - \alpha/(1-\rho))^2 = \sigma^2/(1-\rho^2)$ and the correlations between $y_t$ and $y_{t-i}$, $r_i = \rho^i$, so decline exponentially.
- If stationary, the parameters can be estimated consistently by OLS, though the estimates will not be unbiased ($y_{t-1}$ is uncorrelated with $\varepsilon_t$ but not independent since it is correlated with $\varepsilon_{t-1}$); the estimate of $\rho$ will be biased downwards.

# AR(p)

- An AR(p) takes the form:

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + ... + \rho_p y_{t-p} + \varepsilon_t$$

$$y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} - ... - \rho_p y_{t-p} = \varepsilon_t$$

$$\begin{aligned}(1 - \rho_1 L - \rho_2 L^2 - ... - \rho_p L^p) y_t &= \varepsilon_t; \\ A^p(L) y_t &= \varepsilon_t.\end{aligned}$$

- $A^p(L)$ is a polynomial in the lag operator.

# Stationarity

- The process is stationary if all the roots (solutions), $z_i$, of $1 - \rho_1 z - \rho_2 z^2 - ... - \rho_p z^p = 0$ lie outside the unit circle (are greater than one in absolute value).
- The condition is sometimes expressed in terms of the inverse roots, which must lie inside the unit circle.
- Usually we just check that $-1 < \sum \rho_i < 1$ for stationarity.
- For an AR1 process, for stability, the solution to $(1 - \rho z) = 0$, must be greater than unity in absolute value, since this implies $z = 1/\rho$ this requires $-1 < \rho < 1$.
- For an $AR2$ the real parts of solution to the two solutions to the quadratic $(1 - \rho_1 z - \rho_2 z^2)$ must be greater than unity.

# Random Walk and I(1)

- If a root lies on the unit circle, some $z_i = 1$, the process is said to exhibit a unit root.; stationary after being differenced once I(1)
- A special case of an I(1) variable is a random walk (with drift) where $\rho = 1$ in the AR1:

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

- Substituting back shocks have permanent effects and there is both a deterministic and stochastic trend

$$y_t = \alpha t + \sum_{i=0}^{t-1} \varepsilon_{t-i} + y_0;$$

- $\Delta y_t$ is stationary, I(0), but $y_t$ is non-stationary, I(1). If there is no drift the expected value of $y_t$ will be constant at zero, if $y_0 = 0$, but the variance will increase with t. If there is a drift term the expected value of $y_t$, as well as the variance, will increase with t.

# ARIMA

- Combining the AR and MA processes, gives the ARMA process. The first order ARMA(1,1) with intercept is

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t + \mu \varepsilon_{t-1}$$

- To make them stationary the data are differenced d times, then modelled as an ARMA process of order p and q. This gives the Autoregressive Integrated Moving Average, ARIMA(p,d,q) process:

$$A^p(L)\Delta^d y_t = \alpha + B^q(L)\varepsilon_t.$$

- The ARIMA(1,1,1) process is

$$\Delta y_t = \alpha + \rho \Delta y_{t-1} + \varepsilon_t + \mu \varepsilon_{t-1}$$

# ML Estimation 1

- For the LRM with normal (Gaussian) errors the ML estimator has a closed form solution. For MA errors, this is not the case and an iterative procedure is required.
- Consider maximising a quadratic function of a vector of $k$ parrameters $\boldsymbol{\theta}$, where $\mathbf{C}$ is a positive definite matrix

$$F(\boldsymbol{\theta}) = a + \mathbf{b}'\boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}'\mathbf{C}\boldsymbol{\theta}$$

the first order conditions for a maximum and the closed form solution are

$$\begin{aligned} \frac{\partial F(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \mathbf{b} - \mathbf{C}\boldsymbol{\theta} = \mathbf{0} \\ \boldsymbol{\theta} &= \mathbf{C}^{-1}\mathbf{b}. \end{aligned}$$

# ML Estimation 2

- If $F(\boldsymbol{\theta})$ is the likelihood function, or GMM minimand, for a non-linear model, estimation is usually done using an interative algorithm, where starting from some initial guesses, $\boldsymbol{\theta}_0$ the estimates are updated as

$$\begin{aligned} \boldsymbol{\theta}_1 &= \boldsymbol{\theta}_0 + \lambda_0 \boldsymbol{\Delta}_0 && (2) \\ \boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta} t + \lambda_t \boldsymbol{\Delta}_t && (3) \end{aligned}$$

  $\lambda_t$ is the step size and $\boldsymbol{\Delta}_t$ the direction. This continues until it stops once it has converged to a maximum.

- The most commonly used algorithms are gradient methods. Define the gradient and Hessian

$$\mathbf{g} = g(\boldsymbol{\theta}) = \frac{\partial F(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}; \ \mathbf{H} = \frac{\partial^2 F(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

# Gradient Methods

- The simplest gradient method is Newton's method based on a linear Taylor series expansion around $\boldsymbol{\theta}_0$

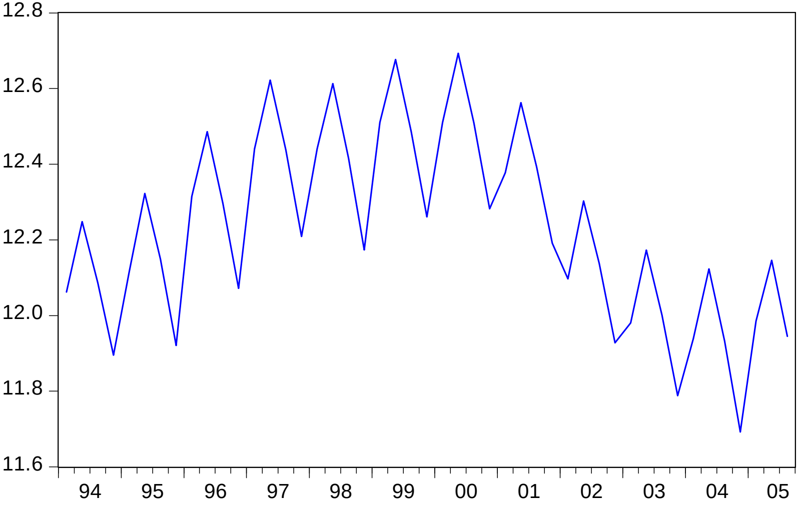
$$\begin{aligned}\frac{\partial F(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &\simeq \mathbf{g}_0 + \mathbf{H}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = 0 \\ \boldsymbol{\theta} &\simeq \boldsymbol{\theta}_0 - \mathbf{H}_0^{-1}\mathbf{g}_0.\end{aligned}$$

  In (2) this sets $\lambda_t = 1$ and $\boldsymbol{\Delta}_t = \mathbf{H}_t^{-1}\mathbf{g}_t$.
- Newton Method often works well, but may be improved by adjusting $\lambda_t$.
- It may be difficult to calculate $\mathbf{H}_t^{-1}$ and it may not be positive definite.
- In ML examples the outer product gradient, OPG, method uses $\left[\sum_{t=1}^{T} \mathbf{g}_t \mathbf{g}_t'\right]^{-1}$ instead of $(-H)^{-1}$. This is always positive definite and only requires calculating first derivatives. It is the basis of BHHH, Berndt, Hall, Hall & Hausman.

# Issues: where to start, how to climb, when to stop

- Start: Try to choose sensible initial values, $\boldsymbol{\theta}_0$, e.g. based on linear approximations and try different values to check for local maxima.
- Climb depends on choice of $\lambda_t$ and $\boldsymbol{\Delta}_t$, programs will often switch between procedures or give you a choice.
- Stop: convergence criteria: $\mathbf{g}_t < \varepsilon$, .and $F(\theta)_t - F(\theta_{t-1}) < \varepsilon$ are sensitive to scaling, the units the variables are measured in. $\mathbf{g}'\mathbf{H}^{-1}\mathbf{g}$ is less sensitive. Programs give you a choice over tolerances, $\varepsilon$, for the coefficients, log likelihood and Hessian scaled gradient.
- If in the linear case $X'X$ is singular, it will be obvious, you get no estimates. This may not be so obvious in the non-linear case and the program may provide estimates even if the likelihood is quite flat.
- For less well behaved functions there are algorithms like simulated annealing and genetic algorithms.

# Coca example



# EViews estimation for Coca

- For estimation of an equation with quadratic trend plus ARMA(1,1) enter : LD C @trend @trend^2 AR(1) MA(1)
- In options box, you have a choice of
  - Method: ML, Conditional LS, GLS
  - Information Matrix: OPG, Hessian observed
  - Starting ARMA coefficient values
  - Optimisation method
  - Step method
  - Maximum iterations
  - Convergence tolerance

EViews - [Equation: EQ02 Workfile: COLOMBIAATTACKS::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

Command

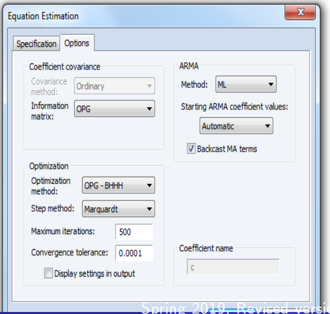
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LD
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 01/21/16 Time: 11:54
Sample: 1995Q2 2005Q3
Included observations: 42
Convergence achieved after 17 iterations
Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 12.02174 | 0.255329 | 47.08330 | 0.0000 |
| @TREND | 0.037550 | 0.022483 | 1.670184 | 0.1041 |
| @TREND^2 | -0.000903 | 0.000412 | -2.190490 | 0.0354 |
| @SEAS(2) | 0.181753 | 0.043377 | 4.190050 | 0.0002 |
| @SEAS(4) | -0.217444 | 0.028539 | -7.619200 | 0.0000 |
| AR(1) | 0.794253 | 0.219806 | 3.613428 | 0.0010 |
| MA(1) | -0.140253 | 0.503474 | -0.278571 | 0.7823 |
| SIGMASQ | 0.004387 | 0.001087 | 4.036928 | 0.0003 |

| | | | |
|---|---|---|---|
| R-squared | 0.929183 | Mean dependent var | 12.25090 |
| Adjusted R-squared | 0.914603 | S.D. dependent var | 0.251919 |
| S.E. of regression | 0.073617 | Akaike info criterion | -2.191651 |
| Sum squared resid | 0.184264 | Schwarz criterion | -1.860666 |
| Log likelihood | 54.02467 | Hannan-Quinn criter. | -2.070332 |
| F-statistic | 63.73046 | Durbin-Watson stat | 1.926742 |
| Prob(F-statistic) | 0.000000 | | |

| | |
|---|---|
| Inverted AR Roots | .79 |
| Inverted MA Roots | .14 |



# Coca Production

Going to consider a number of models on a common sample 1995Q2-2005Q3. Not fixing sample is a common mistake, all models should have 42 observations, to allow for LD(-4). Could have 47 observations for first model.

- A. Seasonals 2 and 4 plus quadratic trend (as in week 2)
- B. ARMA(1,1) with quadratic trend
- C. ARMA(1,1) with seasonals and quadratic trend (C1 by ML and C2 Conditional Least Squares)
- D. ARIMA(4,1,0) with seasonals S12=S1+S2
- E. ARMA(1-4,0,0) with constant and 3 seasonals

# A: S2+S4 + t+ t^2 OLS

Dependent Variable: LD
Method: Least Squares
Date: 10/30/15 Time: 16:01
Sample: 1995Q2 2005Q3
Included observations: 42

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 12.00834 | 0.070160 | 171.1572 | 0.0000 |
| @SEAS(2) | 0.179869 | 0.036646 | 4.908346 | 0.0000 |
| @SEAS(4) | -0.226602 | 0.037832 | -5.989694 | 0.0000 |
| @TREND | 0.042356 | 0.006041 | 7.011784 | 0.0000 |
| @TREND^2 | -0.001042 | 0.000116 | -8.993149 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.862558 | Mean dependent var | 12.25090 |
| Adjusted R-squared | 0.847699 | S.D. dependent var | 0.251919 |
| S.E. of regression | 0.098313 | Akaike info criterion | -1.689976 |
| Sum squared resid | 0.357622 | Schwarz criterion | -1.483111 |
| Log likelihood | 40.48950 | Hannan-Quinn criter. | -1.614152 |
| F-statistic | 58.05115 | Durbin-Watson stat | 0.599440 |
| Prob(F-statistic) | 0.000000 | | |

# B: ARIMA(1,0,1) + t+ t^2 ML

Dependent Variable: LD
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/30/15 Time: 15:45
Sample: 1995Q2 2005Q3
Included observations: 42
Convergence achieved after 17 iterations
Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 12.04827 | 0.163253 | 73.80139 | 0.0000 |
| @TREND | 0.037515 | 0.014034 | 2.673241 | 0.0112 |
| @TREND^2 | -0.000945 | 0.000265 | -3.562633 | 0.0011 |
| AR(1) | -0.076041 | 0.372282 | -0.204256 | 0.8393 |
| MA(1) | 0.712942 | 0.254481 | 2.801550 | 0.0081 |
| SIGMASQ | 0.020372 | 0.008504 | 2.395497 | 0.0219 |

| | | | |
|---|---|---|---|
| R-squared | 0.671157 | Mean dependent var | 12.25090 |
| Adjusted R-squared | 0.625485 | S.D. dependent var | 0.251919 |
| S.E. of regression | 0.154168 | Akaike info criterion | -0.755594 |
| Sum squared resid | 0.855644 | Schwarz criterion | -0.507356 |
| Log likelihood | 21.86748 | Hannan-Quinn criter. | -0.664605 |
| F-statistic | 14.69498 | Durbin-Watson stat | 2.029000 |
| Prob(F-statistic) | 0.000000 | | |

| | |
|---|---|
| Inverted AR Roots | -.08 |
| Inverted MA Roots | -.71 |

Dependent Variable: LD
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 01/21/16 Time: 11:54
Sample: 1995Q2 2005Q3
Included observations: 42
Convergence achieved after 17 iterations
Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 12.02174 | 0.255329 | 47.08330 | 0.0000 |
| @TREND | 0.037550 | 0.022483 | 1.670184 | 0.1041 |
| @TREND^2 | -0.000903 | 0.000412 | -2.190490 | 0.0354 |
| @SEAS(2) | 0.181753 | 0.043377 | 4.190050 | 0.0002 |
| @SEAS(4) | -0.217444 | 0.028539 | -7.619200 | 0.0000 |
| AR(1) | 0.794253 | 0.219806 | 3.613428 | 0.0010 |
| MA(1) | -0.140253 | 0.503474 | -0.278571 | 0.7823 |
| SIGMASQ | 0.004387 | 0.001087 | 4.036928 | 0.0003 |

| | | | |
|---|---|---|---|
| R-squared | 0.929183 | Mean dependent var | 12.25090 |
| Adjusted R-squared | 0.914603 | S.D. dependent var | 0.251919 |
| S.E. of regression | 0.073617 | Akaike info criterion | -2.191651 |
| Sum squared resid | 0.184264 | Schwarz criterion | -1.860666 |
| Log likelihood | 54.02467 | Hannan-Quinn criter. | -2.070332 |
| F-statistic | 63.73046 | Durbin-Watson stat | 1.926742 |
| Prob(F-statistic) | 0.000000 | | |

| | |
|---|---|
| Inverted AR Roots | .79 |
| Inverted MA Roots | .14 |

# C2: ARIMA(1,0,1) + t+ t^2+S2+S4, CLS, rho close to one

Dependent Variable: LD
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 01/21/16 Time: 11:52
Sample: 1995Q2 2005Q3
Included observations: 42
Failure to improve likelihood (non-zero gradients) after 128 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1995Q1

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 57.79258 | 796.3761 | 0.072569 | 0.9426 |
| @TREND | -0.813540 | 9.717920 | -0.083715 | 0.9338 |
| @TREND^2 | 0.004762 | 0.039408 | 0.120830 | 0.9045 |
| @SEAS(2) | 0.182742 | 0.019753 | 9.251273 | 0.0000 |
| @SEAS(4) | -0.218083 | 0.020607 | -10.58272 | 0.0000 |
| AR(1) | 0.980376 | 0.129864 | 7.549227 | 0.0000 |
| MA(1) | -0.462778 | 0.221021 | -2.093822 | 0.0436 |

| | | | |
|---|---|---|---|
| R-squared | 0.932986 | Mean dependent var | 12.25090 |
| Adjusted R-squared | 0.921498 | S.D. dependent var | 0.251919 |
| S.E. of regression | 0.070583 | Akaike info criterion | -2.313037 |
| Sum squared resid | 0.174370 | Schwarz criterion | -2.023426 |
| Log likelihood | 55.57378 | Hannan-Quinn criter. | -2.206883 |
| F-statistic | 81.21308 | Durbin-Watson stat | 1.830749 |
| Prob(F-statistic) | 0.000000 | | |

# D: ARMA(4,0,0) with constant and S1+S2+ S3

Dependent Variable: LD
Method: Least Squares
Date: 10/30/15 Time: 16:50
Sample (adjusted): 1995Q2 2005Q3
Included observations: 42 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.798403 | 0.668797 | 1.193789 | 0.2411 |
| LD(-1) | 0.889572 | 0.140214 | 6.344368 | 0.0000 |
| LD(-2) | 0.035800 | 0.184509 | 0.194027 | 0.8473 |
| LD(-3) | 0.050633 | 0.186741 | 0.271138 | 0.7880 |
| LD(-4) | 0.568246 | 0.189046 | 3.005863 | 0.0050 |
| LD(-5) | -0.617848 | 0.145874 | -4.235490 | 0.0002 |
| @SEAS(1) | 0.154994 | 0.081184 | 1.909171 | 0.0650 |
| @SEAS(2) | 0.187649 | 0.108110 | 1.735729 | 0.0919 |
| @SEAS(3) | 0.061987 | 0.080263 | 0.772299 | 0.4454 |

| | | | |
|---|---|---|---|
| R-squared | 0.946172 | Mean dependent var | 12.25090 |
| Adjusted R-squared | 0.933122 | S.D. dependent var | 0.251919 |
| S.E. of regression | 0.065148 | Akaike info criterion | -2.436900 |
| Sum squared resid | 0.140061 | Schwarz criterion | -2.064542 |
| Log likelihood | 60.17489 | Hannan-Quinn criter. | -2.300416 |
| F-statistic | 72.50737 | Durbin-Watson stat | 2.264424 |
| Prob(F-statistic) | 0.000000 | | |

# E: ARIMA(4,1,0) with S12=S1+S2

Dependent Variable: D(LD)
Method: Least Squares
Date: 10/30/15 Time: 15:53
Sample (adjusted): 1995Q2 2005Q3
Included observations: 42 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -0.082782 | 0.029583 | -2.798301 | 0.0079 |
| D(LD(-4)) | 0.604159 | 0.132136 | 4.572242 | 0.0000 |
| S12 | 0.161225 | 0.055151 | 2.923362 | 0.0057 |

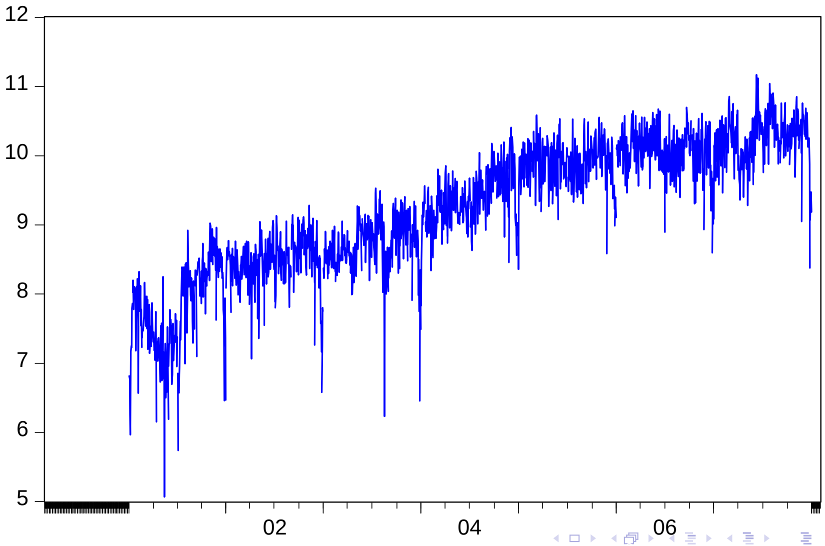
| | | | |
|---|---|---|---|
| R-squared | 0.919085 | Mean dependent var | -0.004035 |
| Adjusted R-squared | 0.914936 | S.D. dependent var | 0.214891 |
| S.E. of regression | 0.062675 | Akaike info criterion | -2.632968 |
| Sum squared resid | 0.153197 | Schwarz criterion | -2.508849 |
| Log likelihood | 58.29233 | Hannan-Quinn criter. | -2.587474 |
| F-statistic | 221.4939 | Durbin-Watson stat | 2.372920 |
| Prob(F-statistic) | 0.000000 | | |

# Schwarz criterion

- A:-1.48: Seasonals plus quadratic trend as before
- B:-0.51: ARIMA(1,0,1) with quadratic trend
- C:-1.86: ARIMA(1,0,1) with seasonals and quadratic trend
- D:-2.06: ARMA(1-4,0,0) with constant and 3 seasonals
- E:-**2.51**: ARIMA(4,1,0) with seasonals S12=S1+S2

Specification E developed from looking at D.

# Log Volume



# Volume ARMA(1,1)+trend

Dependent Variable: LV
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 01/18/16 Time: 15:55
Sample: 1/02/2001 12/31/2007
Included observations: 1747
Convergence achieved after 23 iterations
Coefficient covariance computed using observed Hessian

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 7.376531 | 0.100814 | 73.16980 | 0.0000 |
| @TREND | 0.001679 | 8.38E-05 | 20.02995 | 0.0000 |
| AR(1) | 0.921293 | 0.015534 | 59.30896 | 0.0000 |
| MA(1) | -0.596586 | 0.034126 | -17.48181 | 0.0000 |
| SIGMASQ | 0.123055 | 0.004164 | 29.55499 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.867028 | Mean dependent var | 9.219225 |
| Adjusted R-squared | 0.866723 | S.D. dependent var | 0.962264 |
| S.E. of regression | 0.351295 | Akaike info criterion | 0.748900 |
| Sum squared resid | 214.9776 | Schwarz criterion | 0.764543 |
| Log likelihood | -649.1638 | Hannan-Quinn criter. | 0.754683 |
| F-statistic | 2839.623 | Durbin-Watson stat | 1.802580 |
| Prob(F-statistic) | 0.000000 | | |

| | |
|---|---|
| Inverted AR Roots | .92 |
| Inverted MA Roots | .60 |

# ARMA(2,2) +trend and LDV

Dependent Variable: D(LV)
Method: ARMA Maximum Likelihood (BFGS)
Date: 01/18/16 Time: 16:33
Sample: 1/03/2001 12/31/2007
Included observations: 1746
Convergence achieved after 18 iterations
Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.985202 | 0.128571 | 7.662697 | 0.0000 |
| @TREND | 0.000219 | 3.17E-05 | 6.897615 | 0.0000 |
| LV(-1) | -0.132540 | 0.017548 | -7.553067 | 0.0000 |
| AR(1) | 0.510996 | 0.092592 | 5.518810 | 0.0000 |
| AR(2) | -0.463155 | 0.041198 | -11.24208 | 0.0000 |
| MA(1) | -0.928566 | 0.087923 | -10.56117 | 0.0000 |
| MA(2) | 0.466813 | 0.072240 | 6.461988 | 0.0000 |
| SIGMASQ | 0.116660 | 0.002564 | 45.49260 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.288629 | Mean dependent var | 0.001356 |
| Adjusted R-squared | 0.285764 | S.D. dependent var | 0.405076 |
| S.E. of regression | 0.342340 | Akaike info criterion | 0.698805 |
| Sum squared resid | 203.6878 | Schwarz criterion | 0.723846 |
| Log likelihood | -602.0571 | Hannan-Quinn criter. | 0.708063 |
| F-statistic | 100.7387 | Durbin-Watson stat | 2.044408 |
| Prob(F-statistic) | 0.000000 | | |

| | | |
|---|---|---|
| Inverted AR Roots | .26+.63i | .26-.63i |
| Inverted MA Roots | .46-.50i | .46+.50i |

# ARIMA(1,1,1) worse

Dependent Variable: D(LV)
Method: ARMA Maximum Likelihood (BFGS)
Date: 01/18/16 Time: 16:09
Sample: 1/03/2001 12/31/2007
Included observations: 1746
Convergence achieved after 8 iterations
Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.004666 | 0.003118 | 1.496414 | 0.1347 |
| @TREND | -2.94E-06 | 2.59E-06 | -1.134790 | 0.2566 |
| AR(1) | 0.326724 | 0.024708 | 13.22366 | 0.0000 |
| MA(1) | -0.892794 | 0.012254 | -72.85916 | 0.0000 |
| SIGMASQ | 0.120751 | 0.002581 | 46.78916 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.263682 | Mean dependent var | 0.001356 |
| Adjusted R-squared | 0.261990 | S.D. dependent var | 0.405076 |
| S.E. of regression | 0.347991 | Akaike info criterion | 0.730162 |
| Sum squared resid | 210.8310 | Schwarz criterion | 0.745812 |
| Log likelihood | -632.4313 | Hannan-Quinn criter. | 0.735948 |
| F-statistic | 155.8669 | Durbin-Watson stat | 1.973592 |
| Prob(F-statistic) | 0.000000 | | |

| | |
|---|---|
| Inverted AR Roots | .33 |
| Inverted MA Roots | .89 |

# ARIMA(1,1,2)+LDV+GARCH + t distribution for errors

Dependent Variable: D(LV)
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 01/18/16 Time: 16:15
Sample (adjusted): 1/03/2001 12/31/2007
Included observations: 1746 after adjustments
Convergence achieved after 81 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(7) + C(8)*RESID(-1)^2 + C(9)*GARCH(-1)

| Variable | Coefficient | Std. Error | z-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.317928 | 0.066564 | 4.776276 | 0.0000 |
| @TREND | 6.39E-05 | 1.52E-05 | 4.206212 | 0.0000 |
| LV(-1) | -0.041417 | 0.008910 | -4.648071 | 0.0000 |
| AR(1) | -0.038071 | 0.119159 | -0.319495 | 0.7494 |
| MA(1) | -0.519126 | 0.118646 | -4.375435 | 0.0000 |
| MA(2) | -0.234902 | 0.082894 | -2.833761 | 0.0046 |
| Variance Equation | | | | |
| C | 0.028912 | 0.010230 | 2.826276 | 0.0047 |
| RESID(-1)^2 | 0.099170 | 0.030172 | 3.286862 | 0.0010 |
| GARCH(-1) | 0.648830 | 0.106388 | 6.098719 | 0.0000 |
| T-DIST. DOF | 6.650946 | 0.820755 | 8.103446 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.274348 | Mean dependent var | 0.001356 |
| Adjusted R-squared | 0.272263 | S.D. dependent var | 0.405076 |
| S.E. of regression | 0.345561 | Akaike info criterion | 0.611231 |
| Sum squared resid | 207.7770 | Schwarz criterion | 0.642532 |
| Log likelihood | -523.6051 | Hannan-Quinn criter. | 0.622803 |
| Durbin-Watson stat | 1.928213 | | |

| | | |
|---|---|---|
| Inverted AR Roots | -.04 | |
| Inverted MA Roots | .81 | -.29 |

# ARIMA(2,1,2)+GARCH+ t Distrib

Dependent Variable: D(LV)
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 01/18/16 Time: 16:39
Sample (adjusted): 1/03/2001 12/31/2007
Included observations: 1746 after adjustments
Convergence achieved after 73 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(8) + C(9)*RESID(-1)^2 + C(10)*GARCH(-1)

| Variable | Coefficient | Std. Error | z-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.048226 | 0.128224 | 8.174972 | 0.0000 |
| @TREND | 0.000220 | 2.95E-05 | 7.453931 | 0.0000 |
| LV(-1) | -0.138628 | 0.017165 | -8.075996 | 0.0000 |
| AR(1) | 0.505295 | 0.087280 | 5.789364 | 0.0000 |
| AR(2) | -0.457146 | 0.040520 | -11.28203 | 0.0000 |
| MA(1) | -0.928034 | 0.085403 | -10.86657 | 0.0000 |
| MA(2) | 0.439411 | 0.068911 | 6.376516 | 0.0000 |
| Variance Equation | | | | |
| C | 0.040666 | 0.011335 | 3.587547 | 0.0003 |
| RESID(-1)^2 | 0.133809 | 0.035273 | 3.793468 | 0.0001 |
| GARCH(-1) | 0.508729 | 0.115038 | 4.422250 | 0.0000 |
| T-DIST. DOF | 6.732139 | 0.842945 | 7.986450 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.285635 | Mean dependent var | 0.001356 |
| Adjusted R-squared | 0.283170 | S.D. dependent var | 0.405076 |
| S.E. of regression | 0.342961 | Akaike info criterion | 0.596938 |
| Sum squared resid | 204.5452 | Schwarz criterion | 0.631369 |
| Log likelihood | -510.1270 | Hannan-Quinn criter. | 0.609667 |
| Durbin-Watson stat | 2.015783 | | |

| | | |
|---|---|---|
| Inverted AR Roots | .25+.63i | .25-.63i |
| Inverted MA Roots | .46+.47i | .46-.47i |

# AR(1,4,5)+t GARCH

Dependent Variable: LV
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 02/14/19 Time: 14:49
Sample (adjusted): 1/09/2001 12/31/2007
Included observations: 1742 after adjustments
Convergence achieved after 61 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(6) + C(7)*RESID(-1)^2 + C(8)*GARCH(-1)

| Variable | Coefficient | Std. Error | z-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.797635 | 0.165345 | 10.87200 | 0.0000 |
| @TREND | 0.000371 | 3.97E-05 | 9.329573 | 0.0000 |
| LV(-1) | 0.401197 | 0.021946 | 18.28101 | 0.0000 |
| LV(-4) | 0.201800 | 0.021437 | 9.413672 | 0.0000 |
| LV(-5) | 0.160326 | 0.022225 | 7.213834 | 0.0000 |
| | | Variance Equation | | |
| C | 0.031763 | 0.008950 | 3.548874 | 0.0004 |
| RESID(-1)^2 | 0.131672 | 0.033413 | 3.940768 | 0.0001 |
| GARCH(-1) | 0.589996 | 0.094990 | 6.211125 | 0.0000 |
| T-DIST. DOF | 6.196534 | 0.748680 | 8.276607 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.870484 | Mean dependent var | 9.226750 |
| Adjusted R-squared | 0.870186 | S.D. dependent var | 0.953041 |
| S.E. of regression | 0.343378 | Akaike info criterion | 0.576999 |
| Sum squared resid | 204.8071 | Schwarz criterion | 0.605222 |
| Log likelihood | -493.5662 | Hannan-Quinn criter. | 0.587435 |
| Durbin-Watson stat | 1.944553 | | |

# AR(1,4,5)+t OLS

Dependent Variable: LV
Method: Least Squares
Date: 02/14/19 Time: 14:51
Sample (adjusted): 1/09/2001 12/31/2007
Included observations: 1742 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.672537 | 0.169893 | 9.844664 | 0.0000 |
| @TREND | 0.000366 | 4.20E-05 | 8.714920 | 0.0000 |
| LV(-1) | 0.445218 | 0.021235 | 20.96574 | 0.0000 |
| LV(-4) | 0.188124 | 0.023269 | 8.084621 | 0.0000 |
| LV(-5) | 0.142385 | 0.023822 | 5.976934 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.871143 | Mean dependent var | 9.226750 |
| Adjusted R-squared | 0.870846 | S.D. dependent var | 0.953041 |
| S.E. of regression | 0.342503 | Akaike info criterion | 0.697796 |
| Sum squared resid | 203.7649 | Schwarz criterion | 0.713475 |
| Log likelihood | -602.7799 | Hannan-Quinn criter. | 0.703593 |
| F-statistic | 2935.768 | Durbin-Watson stat | 2.040717 |
| Prob(F-statistic) | 0.000000 | | |

# Schwarz criterion Volume

- A. 0.7645: ARMA(1,1)+trend
- B. 0.7238: ARMA(2,2) +trend and LDV
- C. 0.7458: ARIMA(1,1,1)
- D. 0.6425: ARIMA(1,1,2)+LDV+GARCH + t distribution for errors
- E. 0.6314: ARIMA(2,1,2)+GARCH+ t Distrib
- F. 0.60522: AR(1,4,5) +t+GARCH t distribution of errors
- G. 0.70359: AR(1,4,5) + t OLS

G fits a lot worse than F, but the coefficients in the mean equation are very similar so their point forecasts are likely to be very similar.

# Comments

- Identification of ARIMA(p,d,q) models more art than science. Note identification is choice of pdq, not like econometric identification.
- There are a large number of computational issues and different programs or different estimation methods will give different results.
- The ultimate test is practical forecasting, but a well established process may change when you come to forecast.
- Models that look very different may give very similar forecasts.
- May be difficult to beat very "naive" models out of sample, so in-sample significance should not be give too much weight.

# Revision & Exercises

- Hamilton alternative to HP: a regression of $y_{t+h}$ on the four (for quarterly) most recent values as of date $t$ provides a robust alternative to detrending that achieves the objectives sourght by HP users with none of its drawbacks. Corresponds to a Beveridge Nelson, long horizon, trend in some cases. Linear projection.

$$y_{t+h} = \alpha + \beta_0 y_t + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + u_t$$

- Unbiased forecast, e.g. inthe case of the log transform. A prediction is unbiased if

$$\begin{aligned} E(y_{T+1}) &= y_{T+1,T} \\ y_{T+1} &= E(y_{T+1}) + u_{T+1} = y_{T+1,T} + u_{T+1} \end{aligned}$$

for white noise $u_t$, so, we require $\alpha = 0$, and $\beta = 1$ in a regression of actual on forecast

$$y_t = \alpha + \beta y_{T+1,T} + u_t$$

# Holt Winters (Exponential) Smoothing, estimates a local mean

- Initialise at time $t = 1 : \bar{y}_1 = y_1$
- Update: $\bar{y}_t = \alpha y_t + (1 - \alpha)\bar{y}_{t-1}$
- Forecast: $\bar{y}_{T+h,T} = \bar{y}_T$
- Note
$$\bar{y}_t = \sum_{j=0}^{t-1} \alpha(1 - \alpha)^j y_{t-j}$$
- Choose $0 < \alpha < 1$ to reflect the relative importance of signal in $y_t$ (large $\alpha$) relative to noise (small $\alpha$).

# Holt Winters smoothing allowing for trend

- Initialise at time $t = 2 : \bar{y}_2 = y_2; \;\; F_2 = y_2 - y_1$
- Update

$$\begin{aligned}
\bar{y}_t &= \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + F_{t-1}) \\
F_t &= \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)F_{t-1}, \\
t &= 3, 4, ..., T
\end{aligned}$$

- Forecast: $y_{T+h,T} = \bar{y}_T + hF_t$

# Holt Winters smoothing allowing for seasonal

- Initialise at $t = s: \bar{y}_s = \sum_{t=1}^{s} y_t / s;\ F_s = 0;\ G_j = y_j - \bar{y}_s,$ $j = 1, 2, .., s$
- Update for $t = s + 1, ..., T$

$$
\begin{aligned}
\bar{y}_t &= \alpha(y_t - G_{t-s}) + (1-\alpha)(\bar{y}_{t-1} + F_{t-1}) \\
F_t &= \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)F_{t-1}, \\
G_t &= \gamma(y_t - \bar{y}_t) + (1-\gamma)G_{t-s}
\end{aligned}
$$

- Forecast

$$
\begin{aligned}
y_{T+h,T} &= \bar{y}_T + hF_T + G_{T+h-s};\ h = 1, 2, ..., s \\
y_{T+h,T} &= \bar{y}_T + hF_T + G_{T+h-2s};\ h = s+1, s+2, ..., 2s
\end{aligned}
$$

# Exam Question

In Moodle there are a number of documents relevant to the Autumn 2018 Bank of England forecasts and Monetary Policy Committee decision and the OBR forecasts for the Budget decisions.
Some questions on the exam will require you to use information from these forecasting and decision processes to answer the questions.
Start by asking the basic questions for forecasters for each of the organisations.