EC306 Time Series Lectures

Collected Lecture Notes

Spring Term

1. Introduction to ARMA processes

Overview

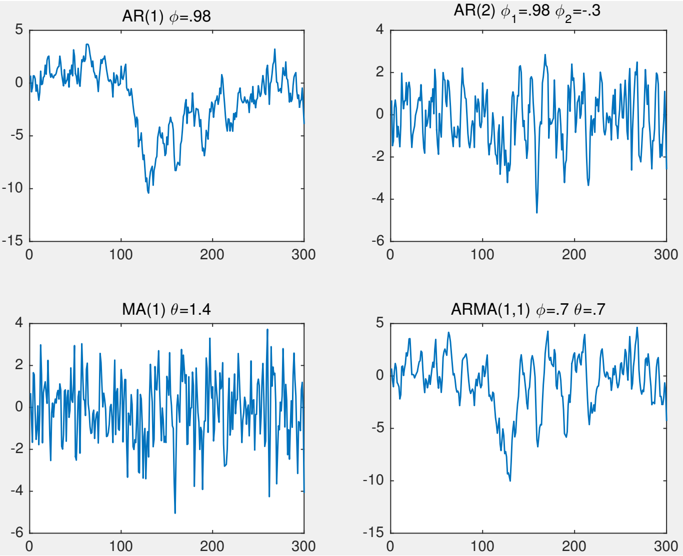
In these lectures we will cover

- MA(q) and its unconditional moments: mean, variance, covariance
- AR(p) and its unconditional moments
- ARMA(p,q) and its unconditional moments
- Aggregation, Common Factors and Misspecification

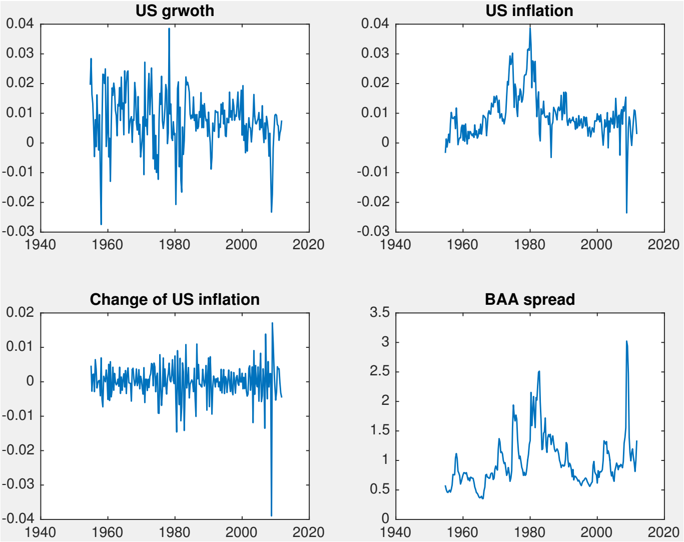
What are we doing?

- Properties of the model
- Testing properties of data
- Estimation of model parameters
- Model selection, forecasting and evaluation
- Economic inference

Some ARMA data



Some economic data



Moments and stationarity

- For a **covariance stationary** process
  - $\mathbb{E}[y_t] = \mu$
  - $\text{var}(y_t) = \gamma_0$
  - $\text{cov}(y_t, y_{t-s}) = \gamma_s$
- That is, the first two unconditional moments are all finite and time invariant

  In general the r^{th} moment of a random variable x is $\mathbb{E}[x^r]$

  The set of all moments characterises the entire probability distribution of the random variable

  Variance is the second *centred* moment $\mathbb{E}[(x-\mu)^2]$

  Covariance stationarity, a.k.a. weak stationarity, is a restriction on th first two moments of the unconditional pdf of x_t, where x_t is an observation from a process evolving over time

Properties of E[.], var(.) and cov(.)

Recall

$$\mathbb{E}[y_t] := \mu = \int y f(y) dy$$
$$\text{var}(y_t) = \mathbb{E}[(y_t - \mu)^2]$$
$$\text{cov}(y_t, y_s) = \mathbb{E}[(y_t - \mu)(y_{t-s} - \mu)]$$

Therefore

$$\mathbb{E}\Big[\sum_t \phi_t x_t\Big] = \sum_t \phi_t \mathbb{E}[x_t]$$
$$\text{var}(\sum_t^T \phi_t x_t) = \sum_t^T \phi_t^2 \,\text{var}(x_t) + 2 \sum_{t=1}^{T-1} \sum_{j=t+1}^{T} \phi_t \phi_j \,\text{cov}(x_t, x_j)$$

where the ϕ_t are *non-random*, even if time varying.

Assumptions on the error process ϵ_t

A1 $\mathbb{E}[\epsilon_t | y_{t-1}, y_{t-2} \ldots] := \mathbb{E}[\epsilon_t | y^{t-1}] = 0$

$$\Rightarrow \mathbb{E}[\epsilon_t] = \mathbb{E}[\mathbb{E}[\epsilon_t | y^{t-1}]] = \mathbb{E}[0] = 0$$

A2 (ϵ_t, ϵ_s) are independent for all $t \neq s$, and (ϵ_t, y_s) are independent for all $s < t$.

Hence $\mathbb{E}[\epsilon_t \epsilon_s] = 0$, $t \neq s$

A3 The ϵ_t are identically distributed

Hence $\mathbb{E}[\epsilon_t^2] = \mathrm{var}(\epsilon_t) = \sigma^2$ for all t

Moving Average MA(q) models

- A simple time series process
- The variable y_t is defined by a finite number of past innovations, ϵ_s
- The influence of past innovations, ϵ_s, $s \leq t$ dies off suddenly when $s < t - q$

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

- Produces a distinctive cut-off in the autorcorrelation structure of the data
- Quite powerful in practice. A good approximation to many more complex processes especially for large q
- Easy to analyse

Example MA(1)

- As a first example of time series analysis, consider an MA(1)

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

- Unconditional mean:

$$\begin{aligned} \mathbb{E}[y_t] &= \mathbb{E}[\mu + \epsilon_t + \theta_1 \epsilon_{t-1}] \\ &= \end{aligned}$$

- Unconditional variance:

$$\begin{aligned} \mathrm{var}(y_t) &= \mathrm{var}(\mu + \epsilon_t + \theta_1 \epsilon_{t-1}) \\ &= \end{aligned}$$

Example MA(1)

- First autocovariance

$$\begin{aligned}\mathbb{E}[(y_t-\mu)(y_{t-1}-\mu)] &= \mathbb{E}[(\epsilon_t+\theta_1\epsilon_{t-1})(\epsilon_{t-1}+\theta_1\epsilon_{t-2})] \\ &= \end{aligned}$$

- Second autocovariance

$$\begin{aligned}\mathbb{E}[(y_t-\mu)(y_{t-2}-\mu)] &= \mathbb{E}[(\epsilon_t+\theta_1\epsilon_{t-1})(\epsilon_{t-2}+\theta_1\epsilon_{t-3})] \\ &= \end{aligned}$$

MA(q) model

- Written in terms of the lag operator

$$\begin{aligned} y_t &= \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots \theta_q \epsilon_{t-q} \\ &= \mu + (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t \\ &= \mu + \Theta(L)\epsilon_t \end{aligned}$$

Properties of MA(q) models:

- The MA is alsways stable, i.e. y_t is a finite linear combination of innovations $y_t = \sum_{i=0}^{q} \theta_i \epsilon_{t-i}$
- Given our assumptions on ϵ_t it is covariance stationary
- It is invertible - i.e. it can be written as AR(∞) - if the roots of $\Theta(L)$ are outside the unit circle

Unconditional moments of MA(q) model

$$\begin{aligned}\mathbb{E}(y_t) &= \mathbb{E}\big(\mu + \epsilon_t + \theta_1\epsilon_{t-1} + \ldots \theta_q\epsilon_{t-q}\big)\\ &= \mu\end{aligned}$$

$$\text{var}(y_t) = \sum_{i=0}^{q} \theta_i^2 \sigma_\epsilon^2 \qquad \text{(note } \theta_0 = 1\text{)}$$

$$\text{cov}(y_t, y_{t-s}) = \sum_{j=s}^{q} \theta_j \theta_{j-s} \sigma^2$$

AR(p) model

The AutoRegression is a linear model of y_t in terms of its own past y_{t-1}, y_{t-2}... y_{t-p}, and an indpendent innovation ϵ_t.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

- We will see this is a model in which y_t can be written as a function of an infitnite number of past innovations
- Provded the AR(p) is stable
- Such that weights on past innovations decline to zero geometrically.
  (So some high oreder MA(q) would be a good approximation)

AR(p) model: stability

To find the lag polynomial $\Phi(L)$, write the AR(p)

$$(1 - \phi_1 L - \ldots - \phi_p L^p) y_t = c + \epsilon_t$$
$$\Phi(L) y_t = c + \epsilon_t$$

The AR(p) is stable, and covariance stationary under (A1)-(A3), provided:

- The roots of the lag polynomial are outside the unit circle.

  Roots, z, solve $\Phi(z) = 0$

  The cofficients of the inverse lag polynomial, $\sum_{i=0}^{\infty} \psi_i L^i$, must go to 0 as $i \to \infty$ to ensure finite mean and variance exist

  If all $|z| > 1$ then the $\psi_i \to 0$ as $i \to \infty$

AR(p) model: MA(∞) representation

Given stability we can write

$$\begin{aligned} y_t &= \Phi(L)^{-1}c + \Phi(L)^{-1}\epsilon_t \\ &= \sum_{i=0}^{\infty} \psi_i c + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \end{aligned}$$

- The **Wald Representation** expresses the realisation y_t as the sum of the mean and the effect of past shocks
- Both the mean and the inifinite sum of past shocks are finite for the stable AR(p)

Specific example: AR(1)

Mean:

$$\begin{aligned}\mathbb{E}[y_t] &= c + \phi\mathbb{E}[y_{t-1}] + \mathbb{E}[\epsilon_t] \\ &= c + \phi\mathbb{E}[y_t] \\ \mu &= \frac{c}{1-\phi}\end{aligned}$$

Variance:

$$\begin{aligned}\text{var}(y_t) &= \text{var}(c + \phi y_{t-1} + \epsilon_t) \\ &= 0 + \phi^2\,\text{var}(y_t) + \text{var}(\epsilon_t) \\ \gamma_0 &= \frac{\sigma_\epsilon^2}{1-\phi^2}\end{aligned}$$

AR(1) covariance

First autocovariance:

$$\begin{aligned}
\text{cov}(y_t, y_{t-1}) &= \mathbb{E}[(y_t - \mu)(y_{t-1} - \mu)] \\
&= \mathbb{E}[(\phi(y_{t-1} - \mu) + \epsilon_t)(y_{t-1} - \mu)] \\
&= \mathbb{E}[\phi(y_{t-1} - \mu)^2 + \epsilon_t(y_{t-1} - \mu)] \\
&= \phi\gamma_0
\end{aligned}$$

In general, by conitnued substitution

$$\begin{aligned}
\text{cov}(y_t, y_{t-s}) &= \mathbb{E}[(\phi^s(y_{t-s} - \mu) + \phi^{s-1}\epsilon_{t-s+1} + \cdots + \epsilon_t)(y_{t-s} - \mu)] \\
&= \phi^s\gamma_0
\end{aligned}$$

Note the covariance declines as $s \to \infty$, reflecting the diminishing influence of past shocks

The $cov(y_t, y_{t-s})$ is a function of s but not of t

AR(1) to MA(∞)

- To derive the coefficients of the MA(∞) representation use undetermined coefficients

$$
\begin{aligned}
1 &= \Phi(L)[\Phi(L)]^{-1} \\
&= (1 - \phi L)(1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots) \\
&=
\end{aligned}
$$

- Setting coefficients on all terms in L to 0 we find $\psi_s =$
- Notice the relationship between the autocovariance structure and the MA(∞) coefficients
- Use the representation $y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ to work out unconditional moments... (exercise)

A recursive approach

Rather than substitute backwrds s times

$$\begin{aligned}
\gamma_1 = \text{cov}(y_t, y_{t-1}) = \mathbb{E}[(\phi(y_{t-1} - \mu) + \epsilon_t)(y_{t-1} - \mu)] = \phi\gamma_0 \\
\gamma_2 = \text{cov}(y_t, y_{t-2}) = \mathbb{E}[(\phi(y_{t-1} - \mu) + \epsilon_t)(y_{t-2} - \mu)] = \phi\gamma_1 \\
\dots \gamma_k = \phi\gamma_{k-1}
\end{aligned}$$

We also have

$$\begin{aligned}
\gamma_0 &= \mathbb{E}[(\phi(y_{t-1} - \mu) + \epsilon_t)(y_t - \mu)] \\
&= \phi\gamma_1 + \sigma_\epsilon^2 \\
&= \phi(\phi\gamma_0) + \sigma_\epsilon^2 \\
\Rightarrow \gamma_0 &= \frac{\sigma_\epsilon^2}{1 - \phi^2}
\end{aligned}$$

These are the Yule-Walker equations and define the autocovariance function γ_j, $j = 0, 1, \dots$

AR(p) model: unconditional mean

- Unconditional mean: $\mu := \mathbb{E}[y_t]$,

  The steady-state value of the process: all y_{t-i} are at the same level, there are no new shocks.

  Or, if you have no information, youhave the same expectation for all t

$$
\begin{aligned}
\mathbb{E}[y_t] &= \mathbb{E}\big(c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t\big) \\
&= c + \phi_1 \mathbb{E}[y_{t-1}] + \phi_2 \mathbb{E}[y_{t-2}] + \cdots + \phi_p \mathbb{E}[y_{t-p}] + \mathbb{E}[\epsilon_t] \\
\mu &= c + \phi_1 \mu + \cdots + \phi_p \mu \\
\mu &= \frac{c}{1 - \sum_{i=1}^{p} \phi_i} \\
\text{or } \mathbb{E}[y_t] &= \mathbb{E}[\Phi(L)^{-1}(c + \epsilon_t)] = \Phi(1)^{-1} c + \Phi(L)^{-1} \mathbb{E}[\epsilon_t] \\
&= \frac{c}{\Phi(1)}
\end{aligned}
$$

AR(p) model: unconditional variance

- Unconditional variance $\text{var}(y_t)$

  Represents average variation of the process: how far away from μ you would expect some y_t to be, given no other information.

$$
\begin{aligned}
\text{var}(y_t) &= \text{var}(c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t) \\
&= \sum_{i=1}^{p} \phi_i^2 \, \text{var}(y_{t-i}) + 2 \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \phi_i \phi_j cov(y_{t-i}, y_{t-j}) + \sigma^2
\end{aligned}
$$

- Ouch!!

AR(p) model: unconditional variance

Use the MA(∞) form

$$
\begin{aligned}
y_t &= \Phi(L)^{-1}(c + \epsilon_t) \\
&= \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \\
\text{var}(y_t) &= \text{var}(\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}) \\
&= \sum_{i=0}^{\infty} \psi_i^2 \sigma^2
\end{aligned}
$$

As $\text{cov}(\epsilon_t, \epsilon_s) = 0$

AR(p) model: unconditional covariance

$$
\begin{aligned}
\operatorname{cov}(y_t, y_{t-s}) &= \mathbb{E}\big((y_t - \mu)(y_{t-s} - \mu)\big) \\
&= \mathbb{E}\big(\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \cdot \sum_{k=0}^{\infty} \psi_k \epsilon_{t-s-k}\big) \\
&= (\psi_s \mathbb{E}[\epsilon_{t-s}^2] + \psi_{s+1}\psi_1 \mathbb{E}[\epsilon_{t-s-1}^2] + \psi_{s+2}\psi_2 \mathbb{E}[\epsilon_{t-s-2}^2] + \dots) \\
&= \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+s}
\end{aligned}
$$

Specific example: AR(2) covariance

Find a function which defines $\gamma_j = \text{cov}(y_t, y_{t-j})$ for AR(2)

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

- Try repeated substitution?
- Solve the ψ_k by undetermined coefficients, then find γ_j?
- Yule Walker is actually pretty neat!

Yule Walker for AR(2)

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \epsilon_t$$

$$\begin{aligned}
\gamma_0 &= \mathbb{E}[(\phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \epsilon_t)(y_t - \mu)] \\
&= \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\epsilon^2 \\
\gamma_1 &= \mathbb{E}[(\phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \epsilon_t)(y_{t-1} - \mu)] \\
&= \phi_1\gamma_0 + \phi_2\gamma_1 \\
\gamma_2 &= \mathbb{E}[\phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \epsilon_t)(y_{t-2} - \mu)] \\
&= \phi_1\gamma_1 + \phi_2\gamma_0
\end{aligned}$$

These can be solved for $(\gamma_0, \gamma_1, \gamma_2)$

γ_0 is a bit messy

Yule Walker: Autocorrelation function AR(2)

Often is is enought to know the autorcorrelations, rather than the autocovariances

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\text{var}(y_t)} = \frac{\gamma_k}{\gamma_0}$$

$$\frac{1}{\gamma_0}\begin{bmatrix}\gamma_1\\ \gamma_2\end{bmatrix} = \frac{1}{\gamma_0}\begin{bmatrix}\gamma_0 & \gamma_1\\ \gamma_1 & \gamma_0\end{bmatrix}\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}$$
$$\begin{bmatrix}\rho_1\\ \rho_2\end{bmatrix} = \begin{bmatrix}1 & \rho_1\\ \rho_1 & 1\end{bmatrix}\begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}$$

Then $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$ for all $k = 3, 4, ...$

Mixed ARMA(p,q) model

Sometimes a mix of both models is useful

$$
\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\
\Phi(L) y_t &= \Theta(L) \epsilon_t
\end{aligned}
$$

Provided roots of $\Phi(L)$ outside unit circle,

$$
\begin{aligned}
y_t &= \Phi(L)^{-1} \Theta(L) \epsilon_t \\
&= \epsilon_t + \psi_t \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \ldots
\end{aligned}
$$

- That is, an MA(∞) representation, as seen for AR models
- Often the best way to work out the unconditional (and sometimes conditional) moments

Example: ARMA(1,1)

$$
\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} \\
&= \Phi(1)^{-1} c + \Phi(L)^{-1} \Theta(L) \epsilon_t \\
&= \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots
\end{aligned}
$$

$$
\begin{aligned}
\Psi(L) &= (1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \dots)(1 + \theta L) \\
&= (1 + (\phi + \theta) L + \phi(\phi + \theta) L^2 + \phi^2 (\phi + \theta) L^3 + \dots)
\end{aligned}
$$

$$
\begin{aligned}
y_t &= \mu + \epsilon_t + (\phi + \theta)\epsilon_{t-1} + \phi^2(\phi + \theta)\epsilon_{t-2} + \dots \\
&= \mu + \epsilon_t + \sum_{i=0}^{\infty} \phi^i (\phi + \theta) \epsilon_{t-1-i}
\end{aligned}
$$

From which you can work out $\mathrm{var}(y_t)$, $\mathrm{cov}(y_t, y_s)$ etc.

Example: ARMA(1,1)

It is occasionally useful to write an MA(q) or ARMA(p,q) as an AR(∞)

$$
\begin{aligned}
(1-\phi L)y_t &= (1+\theta L)\epsilon_t \\
\frac{(1-\phi L)}{(1+\theta L)} y_t &= \pi(L) y_t = \epsilon_t
\end{aligned}
$$

where $\pi(L) = \phi(L)/\theta(L) \Rightarrow \pi(L)\theta(L) = \phi(L)$

$$
\begin{aligned}
(1-\pi_1 L - \pi_2 L^2 - \dots)(1+\theta L) &= (1-\phi L) \\
\pi_i &= (-\theta)^{i-1}(\phi+\theta) \quad i \geq 1 \\
y_t &= \sum_{i=0}^{\infty} (-\theta)^i (\phi+\theta) y_{t-1-i} + \epsilon_t
\end{aligned}
$$

e.g. It may be possible to approximate an ARMA(p,q) with a high order AR model (which can be estimated by OLS).

Aggregation of ARMA processes

It is possible to derive a secondary process from two ARMAs

The resulting process will usually be ARMA process of different order

Aggregation of two AR(1)s:

$$
\begin{aligned}
y_t = \phi_1 y_{t-1} + \epsilon_t \qquad & z_t = \phi_2 z_{t-1} + u_t \\
x_t &= \frac{1}{1-\phi_1 L}\epsilon_t + \frac{1}{1-\phi_2 L}u_t \\
(1-\phi_1 L)(1-\phi_2 L)x_t &= (1-\phi_2 L)\epsilon_t + (1-\phi_1 L)u_t \\
x_t &= (\phi_1+\phi_2)x_{t-1} - \phi_1\phi_2 x_{t-2} + \epsilon_t - \phi_2\epsilon_{t-1} + u_t - \phi_1 u_{t-1}
\end{aligned}
$$

Which is an ARMA(2,1) model

Common Factors

Consider an ARMA(1,1) with equal and opposite coeffs

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t - \phi \epsilon_{t-1} \\ (1-\phi L) y_t &= (1-\phi L)\epsilon_t \\ y_t &= \frac{1-\phi L}{1-\phi L}\epsilon_t = \epsilon_t \end{aligned}$$

which is a white noise process.

What about the ARMA(2,2)

$$y_t = 0.2 y_{t-1} + 0.48 y_{t-2} + u_t + 0.6 u_{t-1} - 0.16 u_{t-2}$$

If you have (near) common factors you will have weak estimates

Misspecification

What if you fit the wrong model entirely?

$$
\begin{aligned}
y_t &= u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} \\
y_t &= \hat{\phi} y_{t-1} + \hat{u}_t
\end{aligned}
$$

Then

$$
\begin{aligned}
\hat{u}_t &= y_t - \hat{y}_t \\
&= u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \hat{\phi}(u_{t-1} - \theta_1 u_{t-2} - \theta_2 u_{t-3} \\
&= u_t - (\theta_1 - \hat{\phi}) u_{t-1} - (\theta_2 - \hat{\phi}\theta_1) u_{t-2} + \hat{\phi}\theta_2 u_{t-3}
\end{aligned}
$$

The error has an MA(3) structure

In general if you fit the wrong model, you will have serially correlated errors.

For what kind of mistake is this not true?

Notes/Problems

Notes/Problems

2. Unit Roots and Tests

Overview

In these lectures we will cover

- More on lag polynomial
- Moments of stationary and non-stationary processes
- Testing for unit roots

What are we doing?

- Properties of the model
- Testing properties of data
- Estimation of model parameters
- Model selection, forecasting and evaluation
- Economic inference

Stationarity

We use the covariance stationary definition

- The unconditional mean exits and is independent of time
- The autocovariances exist and depend only on the distance between observations, not the time of the observations
- In maths:

$$\mathbb{E}[y_t] = \mu$$
$$\text{cov}(y_t, y_s) = \gamma_s$$

Recall $\gamma_0 = \text{var}(y_t)$, so we require a constant, finite variance

Unit roots: AR(1)

Conisder the example

$$y_t = c + \phi y_{t-1} + \epsilon_t \tag{1}$$
$$(1 - \phi L) y_t = c + \epsilon_t$$

As $\epsilon_t \sim iid(0, \sigma^2)$, the condition for (1) to be stationary is
Roots of $\Phi(L) = 1 - \phi L$ lie outside the unit circle

$$1 - \phi z = 0$$
$$\Rightarrow z = \frac{1}{\phi}$$

Unit Roots: AR(1)

- If $z > 1 \Rightarrow \phi < 1$ therefore

$$y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$$

  is well-defined
- This intuition goes through to higher order AR(p) models
- If the roots of the lag polynomial are outside the unit circle the $\psi_i \to 0$
- However, if $\phi = 1$ then $z = 1$ and $\Phi(1) = 0$: the process has a unit root
- Question: What does this mean for $\mu = \mathbb{E}[y_t]$?

Unit roots: AR(p)

Consider a general AR(p)

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

Where $\epsilon_t \sim iid(0, \sigma^2)$, therefore v_t is stationary.

Factor the lag polyniomial

$$\begin{aligned} \Phi(L) &= (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) \\ &= (1 - b_1 L)(1 - b_2 L) \ldots (1 - b_p L) \end{aligned}$$

For any $L = \frac{1}{b_i} = z_i$, then $\Phi(L) = 0$, where z_i are the roots of $\Phi(L)$

If any $z_i = 1$ we have a unit root process: ϵ_t have permanent effects

Stationarity requires all $|z_i| > 1$

If all $|z_i| > 1$, then $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ will be finite

Order of Integration

The order of integration depends on the number of unit roots in a process

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2$$

Consider $\phi_1 = 2$, $\phi_2 = -1$

$$\begin{aligned}\Phi(L) &= 1 - 2L + L^2 \\ &= (1 - L)(1 - L)\end{aligned}$$

The the process has 2 unit roots $y_t \sim I(2)$
Consider $\phi_1 = (1 + \phi)$, $\phi_2 = -\phi$

$$\begin{aligned}\Phi(L) &= 1 - (1 + \phi)L + \phi L^2 \\ &= (1 - L)(1 - \phi L)\end{aligned}$$

has one unit root.
If $|\phi| < 1$, then $y_t \sim I(1)$, and $\Delta y_t \sim I(0)$

Order of integration...cont'd

In this case we have

$$
\begin{aligned}
\Phi(L)y_t = (1-L)(1-\phi L)y_t &= \epsilon_t \\
y_t &= (1-L)^{-1}(1-\phi L)^{-1}\epsilon_t \\
\Rightarrow (1-L)y_t &= (1-L)(1-L)^{-1}(1-\phi L)^{-1}\epsilon_t \\
\Delta y_t &= (1-\phi L)^{-1}\epsilon_t \\
&= \sum_{i=0}^{\infty}\phi^i\epsilon_{t-i}
\end{aligned}
$$

so the first difference of the process is $I(0)$, as argued above.

- $I(n)$ is the number of times a process must be differenced to become stationary
- If there is one unit root, and other roots are stable, the process is $I(1)$

Moments of an $I(0)$ variable

Consider for the AR(1), $|\phi| < 1$

$$
\begin{aligned}
y_t &= c + \phi(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \dots \\
&= c \sum_{s=0}^{t-1} \phi^s + \phi^t y_0 + \sum_{s=0}^{t-1} \phi^s \epsilon_{t-s} \\
&= \frac{c(1-\phi^t)}{1-\phi} + \phi^t y_0 + \sum_{s=0}^{t-1} \phi^s \epsilon_{t-s}
\end{aligned}
$$

$$
\text{As } t \text{ gets large, } y_t = \frac{c}{1-\phi} + \sum_{s=0}^{t-1} \phi^s \epsilon_{t-s}
$$

Moments of an $I(0)$ variable

Consider the mean of the process

$$\begin{aligned}\mathbb{E}[y_t] &= \frac{c}{1-\phi} + \sum_{s=0}^{t-1} \phi^s \mathbb{E}[\epsilon_{t-s}] \\ &= \frac{c}{1-\phi}\end{aligned}$$

which is time invariant and finite, as required.

Moments of an I(0) variable

Consider the variance

$$
\begin{aligned}
\text{var}(y_t) &= \text{var}\left(\frac{c}{1-\phi} + \sum_{s=0}^{t-1} \phi^s \epsilon_{t-s}\right) \\
&= \text{var}\left(\sum_{s=0}^{t-1} \phi^s \epsilon_{t-s}\right) = \sigma^2 \sum_{s=0}^{t-1} \phi^{2s} \\
&\to \frac{\sigma^2}{1-\phi^2} \text{ as } t \to \infty
\end{aligned}
$$

which is time invariant and finite, as required.

Exercise

- We have not shown the AR(1) is covariance stationary

  Need to show $\text{cov}(y_t, y_{t-s})$ exists and depends only on s

  Find the autocovariance funtion for an AR(1)

  Does it meet the requirements for a covariance stationary process?

- For now we move straight on to look at the contrasting behaviour of the mean and variance of an I(1) process

Moments of an $I(1)$ variable

Now consider the AR(1) with $\phi = 1$

$$\begin{aligned} y_t &= c + (c + y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \dots \\ &= ct + y_0 + \sum_{s=0}^{t-1} \epsilon_{t-s} \end{aligned}$$

This is called the random walk with drift

The value of y_t is given by the initial condition, plus a linear deterministic trend with coefficient c, plus the sum of all innovations since the start of the process

Moments of an $I(1)$ variable

Conisder the mean, (putting $y_0 = 0$)

$$\begin{aligned}\mathbb{E}[y_t] &= ct + \sum_{s=0}^{t-1} \mathbb{E}[\epsilon_{t-s}] \\ &= ct\end{aligned}$$

The mean depends on time

Unless $c = 0$, in which case the mean is 0 forever

Could a Random Walk without drift be a stationary process?

Moments of an $I(1)$ variable

Consider the variance

$$\begin{aligned}\mathrm{var}(y_t) &= \mathrm{var}\left(ct + y_0 + \sum_{s=0}^{t-1} \epsilon_{t-s}\right)\\ &= t\sigma^2\end{aligned}$$

The variance depends on time, regardless of the existence of the drift

The Random Walk is never a stationary process

It is a unit root process

- What happens to $\mathbb{E}[y_t]$ and $\mathrm{var}(y_t)$ as t gets large?

Moments of an $I(1)$ variable

Consider the covariance

$$
\begin{aligned}
\text{cov}(y_t, y_{t-s}) &= \mathbb{E}\big[(y_t - \mathbb{E}[y_t])(y_{t-s} - \mathbb{E}[y_{t-s}])\big] \\
&= \mathbb{E}\Big[\Big(\sum_{j=0}^{t-1} \epsilon_{t-j}\Big)\Big(\sum_{j=0}^{t-s-1} \epsilon_{t-s-j}\Big)\Big] \\
&= \mathbb{E}[(\epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1)(\epsilon_{t-s} + \epsilon_{t-s-1} + \cdots + \epsilon_1)] \\
&= (t-s)\sigma^2
\end{aligned}
$$

Depends on both the distance between points, s

The time at which we measure this, t

Moments of an I(1) variable

Consider the Autocorrelation Function

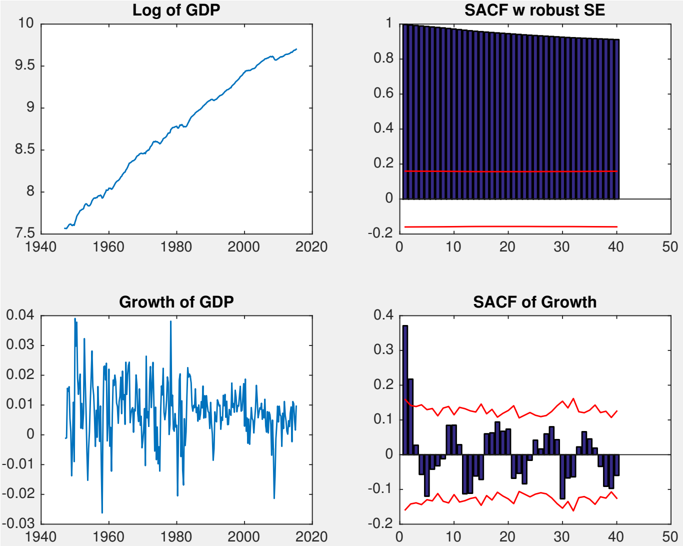
$$\rho_s = \frac{\text{cov}(y_t, y_{t-s})}{\sqrt{\text{var}(y_t)\,\text{var}(y_{t-s})}} = \frac{(t-s)\sigma^2}{\sqrt{t(t-s)}\sigma^2}$$

- The ACF tends close to 1 for $t >> s$
- We can use the Sample Autocorrelation Function, SACF, to examine the persistence of a series in more detail
- For stationary series SACF dies away quite quickly

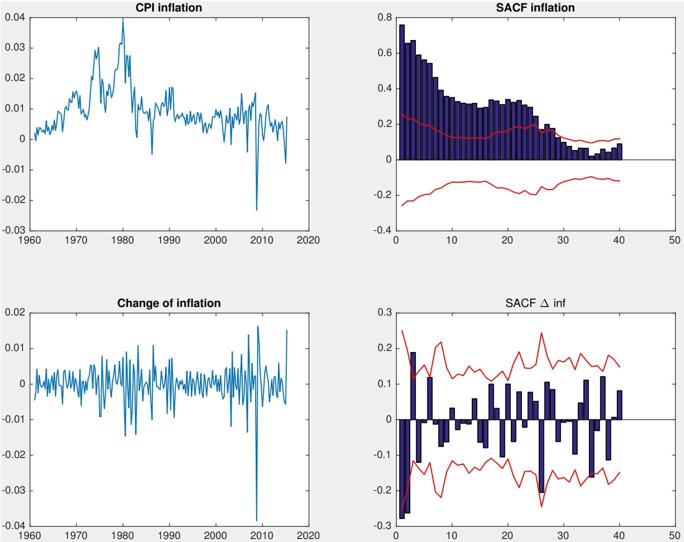
Economics and Integration

- Economic series tend to be *approximately* integrated of order
  - $I(1)$ for real variables
  - $I(1)$ or $I(2)$ for nominal variables
- Economic models often give rise to unit roots
  - The Permanent Income Hypothesis and unit root in consumption
  - The Efficient Market Hypothesis and the random walk in stock prices
  - Innovation and a unit root in output
- Great care needed when modelling relationships between such series

Some examples



Some examples



Testing for unit roots: Dickey-Fuller test

A natural test for unit root, assuming ϵ_t serially uncorrelated

$$
\begin{aligned}
y_t &= \phi y_{t-1} + \epsilon_t \\
H_0 &: \phi = 1 \\
H_A &: \phi < 1 \\
\tau &= \frac{\hat{\phi} - 1}{\widehat{\sigma_\phi}}
\end{aligned}
$$

Recall

$$
\begin{aligned}
\hat{\phi} &= \Big[\sum_{t=2}^{T} y_{t-1}^2\Big]^{-1} \Big[\sum_{t=2}^{T} y_{t-1} y_t\Big] \\
\widehat{\sigma_\phi} &= \sqrt{\frac{1}{T-1} \sum_{t=2}^{T} \hat{e}_t \Big[\sum_{t=2}^{T} y_{t-1}^2\Big]^{-1}}
\end{aligned}
$$

Testing for unit roots: Dickey-Fuller test

The crucial new fact, τ has a Dickey Fuller distrbution, $\tau \sim DF$

- The DF critical values are more negative than $N(0,1)$, or t critical values
- The critical values depend on the data generating process assumed
  - Does the regression assume a constant?
  - Does the regression assume a time trend?
- Dickey-Fuller critical values

Stat	Description	1%	5%
N(0,1)		-2.33	-1.645
τ_a	no con, no trend	-2.58	-1.95
τ_b	con, no trend	-3.43	-2.86
τ_c	con, trend	-3.96	-3.41

Testing for unit roots: Assumption on errors

For the DF test it is crucial $\mathbb{E}[\epsilon_t \epsilon_{t-s}] = 0$. Consider the DGP

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

If you run the DF regression on $y_t = \phi y_{t-1} + v_t$

$$\begin{aligned}\mathbb{E}[v_t v_{t-1}] &= \mathbb{E}[(\phi_2 y_{t-2} + \epsilon_t)(\phi_2 y_{t-3} + \epsilon_{t-1})] \\ &= \phi_2^2 \mathbb{E}[y_{t-2} y_{t-3}] \neq 0\end{aligned}$$

So the standard DF test is not suitable

Always check the errors are serially uncorrelated

Testing for unit roots: ADF test

Solution is the Augmented Dickey Fuller Test. E.g. AR(2) DGP

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \\
&= \phi_1 y_{t-1} + \phi_2 y_{t-2} - \phi_2 y_{t-1} + \phi_2 y_{t-1} + \epsilon_t \\
&= (\phi_1 + \phi_2) y_{t-1} - \phi_2 \Delta y_{t-1} + \epsilon_t \\
\Delta y_t &= (\phi_1 + \phi_2 - 1) y_{t-1} - \phi_2 \Delta y_{t-1} + \epsilon_t \\
&= \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \epsilon_t \\
\tau &= \frac{\hat{\rho}}{\widehat{se(\rho)}} \sim DF \\
H_0 &: \rho = 0 : \text{ unit root} \\
H_A &: \rho < 0 : \text{ stable process}
\end{aligned}
$$

Testing for unit roots: ADF test

The general ADF test for data from an AR(p):

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta y_{t-i} + \epsilon_t$$

$$\tau = \frac{\hat{\rho}}{\widehat{se(\rho)}} \sim DF$$

$$H_0 : \rho = 0 : \text{ unit root}$$

$$H_A : \rho < 0 : \text{ stable process}$$

Exercise: Derive expressions for ρ, γ_i, $p = 3$

Testing for unit roots: ADF test

In practice, p is unknown. How choose p?

- Begin with $p = f + 1$ lags where $f = nobs/year$ and delete insignificant lags
  - t-values are standard for $H_0 : \gamma_i = 0$
- Use an information criterion: AIC, BIC etc.
- Test OLS errors for serial correlation (BG test), and choose model with smallest no of lags
- An excessive no of lags reduces the power of the test

Testing for unit roots: ADF test: deterministic terms

- Model A: restricted intercept
  - H_0: A random walk
  - H_A: Mean-zero stationary AR(p)
- Model B: unrestricted intercept and no trend
  - H_0: A random walk with drift
  - H_A: A stationary AR(p) with non-zero mean
- Model C: unrestricted intercept, restricted trend
  - H_0: A random walk with drift
  - H_A: A trend-stationary AR(p) with non-zero intercept
- Model D: unrestricted trend and intercept
  - H_0: A random walk with drift and time-trend: a quadratic trend in t
  - H_A: A trend-stationary AR(p) with non-zero intercept

- Model B and Model C tend to be good choices for economic data

Testing for unit roots: ADF test: deterministic terms

Considerations in choosing a model

- Adding more deterministic terms than necessary reduces the power of the test

  Critical values for $\tau_d < \tau_c < \tau_b < \tau_a$
- Too few deterministic terms leads to an under-sized test

  The null rejected less than 5% of time when it is true
- Almost all econoimc variables need a constant

  Even constructed residuals may do to be on safe side: $\epsilon_0 \neq 0$
- Look at the graph: does it look like it needs a time trend?

  The test is then null of unit root vs. alternative of *trend stationary*
- A couple of examples:

  Log output: model c
  Inflation: model b

Some examples

Table: ADF Test Results

Var	Model	τ_i	p-val	Decision
GDP	C	-0.90	0.95	Unit root
ΔGDP	B	-6.44	0.00	I(0)
Inf	B	-2.46	0.13	Unit root
ΔInf	B	-7.2	0.00	I(0)

Some examples

Table: ADF Test Results

Var	Model	τ_i	p-val	Decision
GDP	C	-0.90	0.95	Unit root
ΔGDP	B	-6.44	0.00	I(0)
Inf	B	-2.46	0.13	Unit root
ΔInf	B	-7.2	0.00	I(0)

Lessons:

- From the GDP result: It's hard to tell trend stationary from unit root by eyeballing it
- From the inflation result: You can't always spot a unit root from the SACF

Monte-Carlo study of ADF test

DGP: $y_t = \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.05	0.89	0.35
C	0.05	0.68	0.21
D	0.06	0.52	0.15

Monte-Carlo study of ADF test

DGP: $y_t = \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, 1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.05	0.89	0.35
C	0.05	0.68	0.21
D	0.06	0.52	0.15

Models B,C are correctly sized

Monte-Carlo study of ADF test

DGP: $y_t = \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.05	0.89	0.35
C	0.05	0.68	0.21
D	0.06	0.52	0.15

Models B,C are correctly sized

Power declines from 89% to 35% as ϕ from 0.8 to 0.9

Typically we have low power over near unit roots

Monte-Carlo study of ADF test

DGP: $y_t = \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.05	0.89	0.35
C	0.05	0.68	0.21
D	0.06	0.52	0.15

Models B,C are correctly sized

Power declines as ϕ from 0.8 to 0.9

Typically we have low power over near unit roots

Models C and D have poor power as too many deterministic terms

Monte-Carlo study of ADF test

DGP: $y_t = 0.5 + \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.01	0.89	0.35
C	0.05	0.68	0.21
D	0.06	0.52	0.15

Model B, correctly specified, now produces an undersized test, but still has the best power

Monte-Carlo study of ADF test

DGP: $y_t = 0.5 + 0.2t + \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.00	0.00	0.00
C	0.00	0.68	0.21
D	0.06	0.53	0.15

Model B is undersized and has no power - too few deterministic terms

Monte-Carlo study of ADF test

DGP: $y_t = 0.5 + 0.2t + \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, 1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.00	0.00	0.00
C	0.00	0.68	0.21
D	0.06	0.53	0.15

Model B is undersized and has no power

Model C (correct DGP) is undersized but has comparable power to previous examples

Monte-Carlo study of ADF test

DGP: $y_t = 0.5 + 0.2t + \phi y_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$

Table: Empirical rejection frequencies

Model	$\phi = 1$	$\phi = 0.8$	$\phi = 0.9$
B	0.00	0.00	0.00
C	0.00	0.68	0.21
D	0.06	0.53	0.15

Model B is undersized and has no power

Model C (correct DGP) is undersized but has comparable power to previous examples

Model D is still oversized and lacks power - too many deterministic terms! Testing against a trend stationary alternative is difficult!

A further complication

- It is hard to reject unit root as ϕ gets towrds 1
- ADF tests have quite low power in finite samples
- They work even worse if use the wrong DGP
- Structural breaks are a further complication:

$$y_t = c_1 + \phi y_{t-1} + I(t > T_b)c_2 + \epsilon_t$$

Before $t = T_b$ process fluctuates around $c_1/(1-\phi)$

After $t = T_b$ process fluctuates around $(c_1 + c_2)/(1-\phi)$

Regression $y_t = \hat{c} + \hat{\phi} y_{t-1} + \epsilon_t$ fits best if $\hat{\phi} \to 1$

This indicates a unit root, but there isn't really one!

A further complication

This is a cause of considerable controversy

- Is output a unit root, or a trend stationary process?
- Matters - if trend stationary then losses from recessions are temproary; if $I(1)$ then losses are permanent

  Nelson and Plosser JME (1982): Almost all macroeconomic series are $I(1)$

  Perron *Econometrica* (1989): Almost all trend stationary with breaks in intercept (Great Depression) and / or trend (Oil shocks)
- Perron has tabulated critical values for unit root tests with known break dates, based on DGP like previous slide

  How do you know a break was really a break, or when it occurred?

  Treating series as $I(1)$ is still safer

Notes/Problems

Notes/Problems

3. ARMA identification, estimation and diagnostics

Overview

In these lectures we will cover

- Identification of ARMA models
- Estimation of ARMA models
- Diagnostic checking and model selection
- Forecasting
- Forecasting evaluation

What are we doing?

- Properties of the model
- Testing properties of data
- Estimation of model parameters
- Model selection, forecasting and evaluation
- Economic inference

Specification strategy for ARMA modelling

- Identify a simple ARMA structure:
  - SACF
  - SPACF
- Estimation of feasible models
  - OLS for pure autoregressions
  - MLE for models with MA terms
- Diagnostic checks on, and inspection of, residuals
  - Graphical inspection
  - Ljung-Box test
  - LM tests
- The selected model can be used for forecasting the series
  - Construct forecasts: conditional expectations
  - Evaluate forecasts: MSPE

Identification of model

- The series must be differenced until it is $I(0)$
- An ARMA model generates (amongst other things) the following theoretical moments

Quantity	Definition	Symbol
Expectation	$\mathbb{E}[y_t]$	μ
Variance	$\mathbb{E}[(y_t - \mu)^2]$	γ_0 or σ_y^2
Autocovariance	$\mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]$	γ_k or $\gamma(k)$
Autocorrelation	γ_k/γ_0	ρ_k or $\rho(k)$

- Estimate sample equivalents
- Choose plausible ARMA model to generate this behaviour

Sample estimates of moments

Quantity	Definition	Symbol
Expectation	$T^{-1}\sum_{t=1}^{T} y_t$	$\bar{\mu}$ or $\bar{y}$
Variance	$T^{-1}\sum_{t=1}^{T}[(y_t - \bar{y})^2]$	$\hat{\gamma}_0$ or s^2
Autocovariance	$\frac{1}{T-k}\sum_{t=k+1}^{T}[(y_t - \mu)(y_{t-k} - \mu)]$	$\hat{\gamma}_k$
Autocorrelation	$\hat{\gamma}_k / \hat{\gamma}_0$	$\hat{\rho}_k$

- Notice relation between $\mathbb{E}[.]$, in the population, and $T^{-1}\sum_t [.]$, in the sample
- We can compute sample moments without a model: sample averages
- Check stat. sig. of sample moments versus those implied by alternative models to build a sensible forecasting model. 'Box-Jenkins' method. Better with lots of data.

Box-Jenkins methodology

How do we choose an ARMA(p,q) model which will match our data?

Do we use an AR or MA model?

How many lags in the AR part?

How many past errors averaged in the MA part?

- We calculate the Sample Autocorrelation Function (SACF) and the Sample Partial Autocorrelation Function (SPACF)
- We choose an ARMA model with corresponding Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)
- Alternatively we start with a large ARMA(p,q) and use statistical significance to 'test down' to a more parsimonious model (Hendry)

Autocorrelation Function

We have already seen Autocorrelations.

- They are just the second moments (see slide 6):

$$\rho_k = \gamma_k / \gamma_0 = \mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)] / \operatorname{var}(y_t)$$

- Sample equivalents are defined on slide 7
- We will need to match the significant sample autocorrelations with the autocorrelations implied by our ARMA model

Partial Autocorrelation Function

The Partial Autocorrelations measure the extra information about y_t in y_{t-k} after accounting for the predictive power of $y_{t-1} \dots y_{t-k+1}$.

$$\phi_k = \text{corr}\left((y_t - \hat{\mathbb{E}}[y_t]), (y_{t-k} - \hat{\mathbb{E}}[y_{t-k}])\right)$$

where $\hat{\mathbb{E}}[x]$ stands for projection of x on $y_{t-1} \dots y_{t-k+1}$.

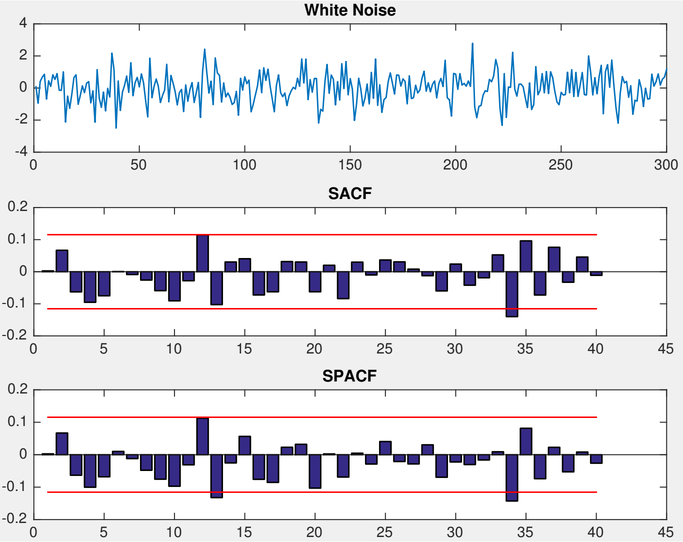
This is the last coefficient in an estimated AR(k) model

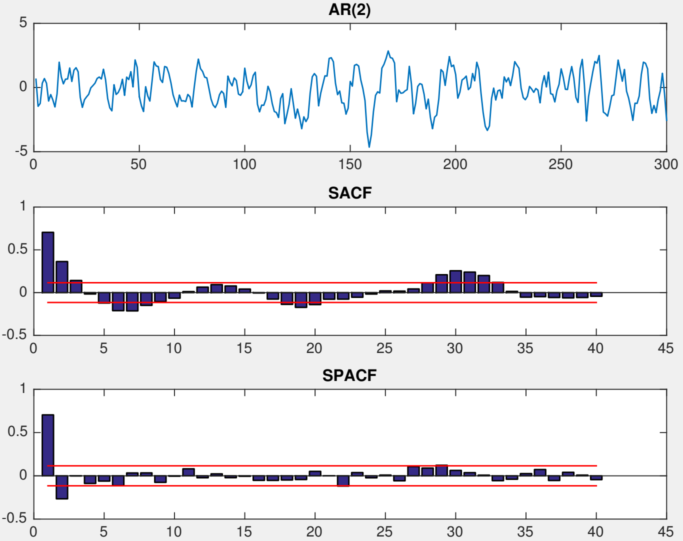
To find the PACF:

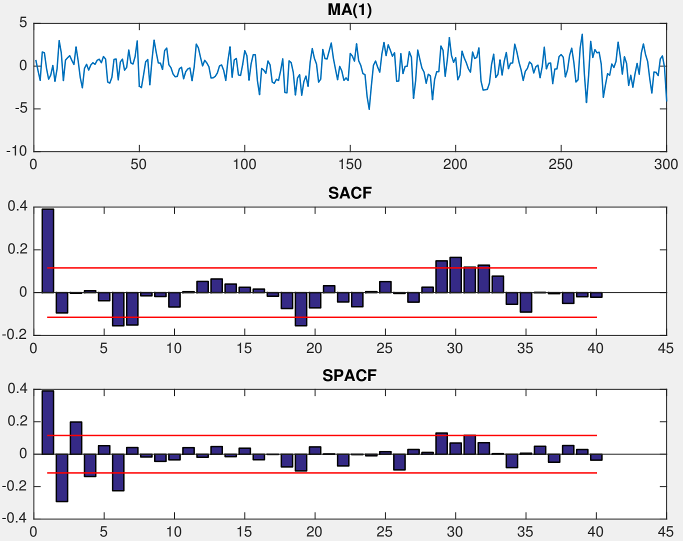
- Estimate successive AR(k) models $k = 1, 2, ...$
- $\hat{\phi}_k$ is last estimated coefficient in each model
- We need to capture the significant PACs

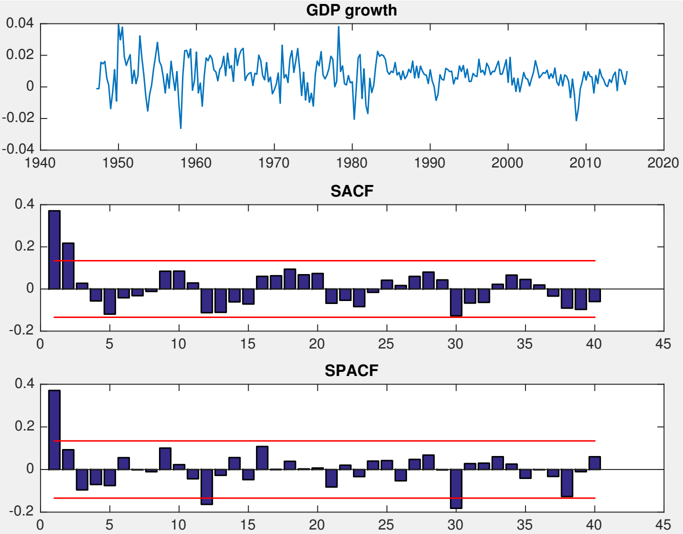
Significance

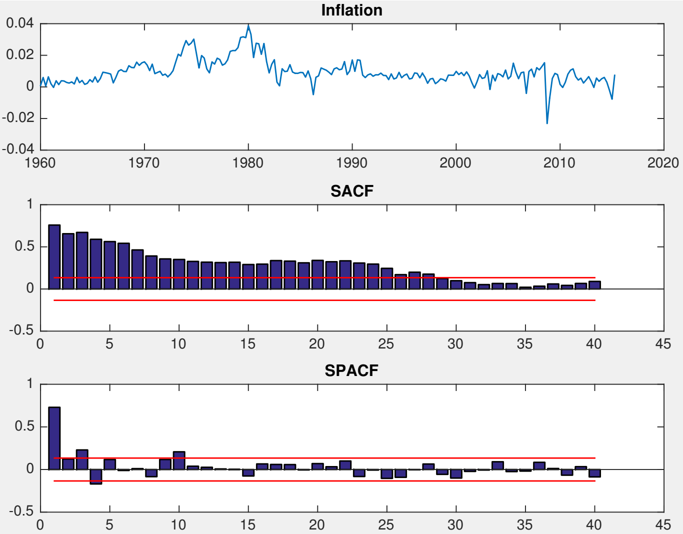
- How many SACF and SPACF terms are significant?
- Test using $H_0 : \rho(k) = 0$, the series is white noise
- Under H_0, $\hat{\rho}(k)$ and $\hat{\phi}(k)$ are approx $\sim N(0, 1/T)$
- $1/\sqrt{T}$ is the standard error of $\hat{\rho}(k)$ and $\hat{\phi}(k)$
- 95% confidence intervals are given by $\pm 2/\sqrt{T}$
- $\hat{\rho}(k)$, $\hat{\phi}(k)$ outside this interval are significantly different from zero
- Compute the SACFs up to $T/4$
- Model the significant autocorrelations and partial autocorrelations


White Noise
4
2
0
-2
-4
0
50
100
150
200
250
300
SACF
0.2
0.1
0
-0.1
-0.2
0
5
10
15
20
25
30
35
40
45
SPACF
0.2
0.1
0
-0.1
-0.2
0
5
10
15
20
25
30
35
40
45


AR(2)
5
0
-5
0
50
100
150
200
250
300
SACF
1
0.5
0
-0.5
0
5
10
15
20
25
30
35
40
45
SPACF
1
0.5
0
-0.5
0
5
10
15
20
25
30
35
40
45


MA(1)
5
0
-5
-10
0
50
100
150
200
250
300
SACF
0.4
0.2
0
-0.2
0
5
10
15
20
25
30
35
40
45
SPACF
0.4
0.2
0
-0.2
-0.4
0
5
10
15
20
25
30
35
40
45


GDP growth
0.04
0.02
0
-0.02
-0.04
1940
1950
1960
1970
1980
1990
2000
2010
2020
SACF
0.4
0.2
0
-0.2
0
5
10
15
20
25
30
35
40
45
SPACF
0.4
0.2
0
-0.2
0
5
10
15
20
25
30
35
40
45


Inflation
0.04
0.02
0
-0.02
-0.04
1960
1970
1980
1990
2000
2010
2020
SACF
1
0.5
0
-0.5
0
5
10
15
20
25
30
35
40
45
SPACF
1
0.5
0
-0.5
0
5
10
15
20
25
30
35
40
45

Box-Jenkins summary

Use SACF, SPACF as follows to suggest plausible ARMA structure

Model	(S)ACF	(S)PACF
AR(p)	Geometric decay	Dies after p lags
MA(q)	Dies after q lags	Geometric decay
ARMA(p,q)	Geometric decay	Geometric decay

Begin estimation with large plausible model and eliminate terms

Estimation

- AR(p) estimate with OLS:
  - Regress y_t on p lags of itself ($+$ constant)
  - Regressors $x_t = (y_{t-1}, \ldots, y_{t-p})'$ are known and satisfy $\mathbb{E}[x_t \epsilon_t] = 0$
  - Variance assumed constant

Estimation

- AR(p) estimate with OLS:
  - Regress y_t on p lags of itself ($+$ constant)
  - Regressors $x_t = (y_{t-1}, \ldots, y_{t-p})'$ are known and satisfy $\mathbb{E}[x_t \epsilon_t] = 0$
  - Variance assumed constant
- MA(q) models are estimated by MLE, as sequence $\{\epsilon_t\}$ is unknown before estimation
  - Likelihood contribution of y_t for MA model (normal errors)

$$f(\epsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{-\epsilon_t^2}{2\sigma^2}\right)$$

    where $\epsilon_t = y_t - \theta\epsilon_{t-1}$
  - ϵ_t are i.i.d. therefore joint likelihood of $\{\epsilon_t\}_{t=1}^T$ is

$$L(\theta, \sigma^2; \epsilon) = \Pi_{t=1}^T f(\epsilon_t)$$

  - Giving log likelihood function

$$l(\theta, \sigma^2; \epsilon) = \frac{-T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \epsilon_t^2$$

Estimation MA(1)

The likelihood can be written in terms of the data y^T for a given MA(q) model

- E.g. MA(1) with initial condition $\epsilon_0 = 0$

$$\epsilon_t = (1 + \theta L)^{-1} y_t = \sum_{j=0}^{t-1} (-\theta)^j y_{t-j}$$

- Gives the loglikelihood in terms of y^T

$$l(\theta, \sigma^2; y^T) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \Big(\sum_{j=0}^{t-1} (-\theta)^j y_{t-j} \Big)^2$$

- $\hat{\theta}$ and $\hat{\sigma}^2$ are the values of θ and σ^2 that maximise the above
  Results are conditional on $\epsilon_0 = 0$

Diagnostics: Serial Correlation 1

After estimation CHECK THE PROPERTIES OF MODEL

- Essential that residuals are serially uncorrelated.

  Check a plot of residuals

  Calculate the SACF of residuals

$$r_k = \frac{\sum_{t=k+1}^{T} \hat{\epsilon_t} \hat{\epsilon_{t-k}}}{\sum_{t=1}^{T} \hat{\epsilon}_t^2}$$

  Ljung-Box test formalises the idea of testing significance of first m r_k terms

$$Q = T(T+2) \sum_{k=1}^{m} \frac{r_k^2}{T-k} \sim \chi^2_{m-p-q}$$

  Choose m around $\sqrt{T}$. (This is a big sample thing!)

Diagnostics: Serial Correlation 2

The Breusch Godfrey test is more often used in econometrics

It is a Lagrange-multiplier test

But it is implemented via a secondary regression

$$
\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \\
\hat{\epsilon}_t &= \rho_1 \hat{\epsilon}_{t-1} + \cdots + \rho_r \hat{\epsilon}_{t-r} + \beta' x_t + u_t \\
H_0 &: \rho_1 = \cdots = \rho_r = 0
\end{aligned}
$$

I Estimate the AR(p) on y_t data, save residuales $\hat{\epsilon}_t$

II Regress $\hat{\epsilon}_t$ on r lagged residuals and the other regressors (collected in x_t) to test for r^{th} order serieal correlation

III $nR^2 \sim \chi^2_r$ where $n = T - p$. I.e. there are r restrictions under H_0
Or $R^2(T - r - p)/((1 - R^2)r) \sim F(r, T - r - p)$ for the exact test

Diagnostics: Distributional Assumptions

E.g. F-tests require distributional assumptions

Jarque-Brera test for normality of residuals

Normal distribution implies

- Skewness $\mathbb{E}[(y_t - \mu)^3/\sigma^3] = 0$
- Kurtosis $\mathbb{E}[(y_t - \mu)^4/\sigma^4] = 3$

A basis for a test of normality is whether sample equivalents, SK and K, differ from 0 and 3.

$$JB = \frac{T}{6}\Big[SK^2 + \frac{(K-3)^2}{4}\Big] \sim \chi^2_2$$

Significant JB test may be driven by large outliers or heteroskecasticity

Model Selection: Information Criteria

May choose between admissible models with Information Criteria

- AIC: $\ln(RSS/T) + 2k/T$ or in MLE setting $-2(\ln(L) - k)$
- SBIC: $\ln(RSS/T) + k\ln(T)/T$ or in MLE setting $-2\ln(L) + k\ln(T)$

  Choose the model with lower Info criterion

- SBIC is more parsimonious in model it chooses

  Penalises parameters more

  Small models often capture key feature of data

  Lower estimation variance may help forecasts

Forecasting: The Optimal Predictor

Given a model, the optimal predictor, $\hat{y}_{T+h}$ minimises

$$\mathbb{E}_T[y_{T+h} - \hat{y}_{T+h}]^2 = \mathbb{E}_T[e^2_{T+h}]$$

where $\mathbb{E}_T$ denotes expectation conditional on information up to date T

It can be shown (see later in the course) that the optimal predictor is

$$\hat{y}_{T+h} = \mathbb{E}_T[y_{T+h}] = \mathbb{E}[y_{T+h}|y_T, y_{T-1}, \ldots, y_1]$$

So forecasting is *taking conditional expectations* using the (estimated) model

Forecasting: AR(1)

Forecasts from an AR(1) model are conditional expectations

$$
\begin{aligned}
\mathbb{E}[y_{T+1}|y_T] &= \mathbb{E}[c + \phi y_T + \epsilon_{T+1}|y_T] \\
&= c + \phi y_T \\
\mathbb{E}[y_{T+2}|y_T] &= \mathbb{E}[c + \phi y_{T+1} + \epsilon_{T+2}|y_T] \\
&= c + \phi \mathbb{E}[y_{T+1}|y_T] + \mathbb{E}[\epsilon_{T+2}|y_T] \\
&= c + \phi(c + \phi y_T) \\
\ldots \mathbb{E}[y_{T+h}|y_T] &= c \sum_{k=0}^{h-1} \phi^k + \phi^h y_T
\end{aligned}
$$

Forecasting: AR(1)

Forecast error variances are a series of conditional variances

$$\begin{aligned}
\mathbb{E}[(y_{T+1} - \hat{y}_{T+1})^2 | y_T] &= \mathbb{E}[\epsilon_{T+1}^2 | y_T] = \sigma_\epsilon^2 \\
\mathbb{E}[(y_{T+2} - \hat{y}_{T+2})^2 | y_T] &= \mathbb{E}[(\epsilon_{T+2} + \phi\epsilon_{T+1})^2] \\
&= (1 + \phi^2)\sigma_\epsilon^2 \\
\mathbb{E}[(y_{T+h} - \hat{y}_{T+h})^2] &= \sigma_\epsilon^2 \sum_{k=0}^{h-1} \phi^{2k}
\end{aligned}$$

Notice as $h \to \infty$

$$\begin{aligned}
\mathbb{E}[y_{T+h} | y_T] &\to \frac{c}{1 - \phi} \\
\mathbb{E}[(y_{T+h} - \hat{y}_{T+h})^2 | y_T] &\to \frac{\sigma_\epsilon^2}{1 - \phi^2}
\end{aligned}$$

Forecasting: MA(1)

Forecasts from an MA(1) model are also conditional expectations

$$
\begin{aligned}
\mathbb{E}[y_{T+1}|y_T, .., y_1] &= \mathbb{E}[\mu + \theta\epsilon_T + \epsilon_{T+1}|y^T] \\
&= \mu + \theta\epsilon_T \\
\mathbb{E}[y_{T+2}|y^T] &= \mathbb{E}[\mu + \theta\epsilon_{T+1} + \epsilon_{T+2}|y^T] \\
&= \mu
\end{aligned}
$$

And the variances are conditional variances:

$$
\begin{aligned}
\mathbb{E}[(y_{T+1} - \hat{y}_{T+1})^2|y^T] &= \mathbb{E}[\epsilon_{T+1}^2|y^T] = \sigma_\epsilon^2 \\
\mathbb{E}[(y_{T+2} - \hat{y}_{T+2})^2|y^T] &= \mathbb{E}[(\epsilon_{T+2} + \theta\epsilon_{T+1})^2|y^T] = (1+\theta^2)\sigma_\epsilon^2
\end{aligned}
$$

And so on for $h > 2$, as there is only 1 period of memory.

Forecasting: ARMA(p,q)

We use the MA(∞) representation

$$\begin{aligned}
\Phi(L)y_t &= \Theta(L)\epsilon_t \\
y_t &= \Phi(L)^{-1}\Theta(L)\epsilon_t \\
&= \Psi(L)\epsilon_t
\end{aligned}$$

to separate y_{T+h} into realised and future shock components:

$$\begin{aligned}
y_{T+h} &= \Psi(L)\epsilon_{T+h} \\
&= \sum_{j=1}^{h} \psi_{h-j}\epsilon_{T+j} + \sum_{j=0}^{\infty} \psi_{h+j}\epsilon_{T-j} \\
\mathbb{E}_T[y_{T+h}] &= \sum_{j=0}^{\infty} \psi_{h+j}\epsilon_{T-j}
\end{aligned}$$

Forecasting: ARMA(p,q)

Consider forecast error the variance

$$\begin{aligned}
\mathbb{E}[(y_{T+h} - \hat{y}_{T+h})^2 | y^T] &= \mathbb{E}_T\Big[\Big(\sum_{j=1}^{h} \psi_{h-j}\epsilon_{T+j}\Big)^2\Big] \\
&= \sigma_\epsilon^2(1 + \psi_1^2 + \cdots + \psi_{h-1}^2) \\
&\to \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j^2 \text{ as } h \to \infty
\end{aligned}$$

which converges to the unconditional variance of the process as the forecast horizon increases

Forecasting: ARMA(1,1) example

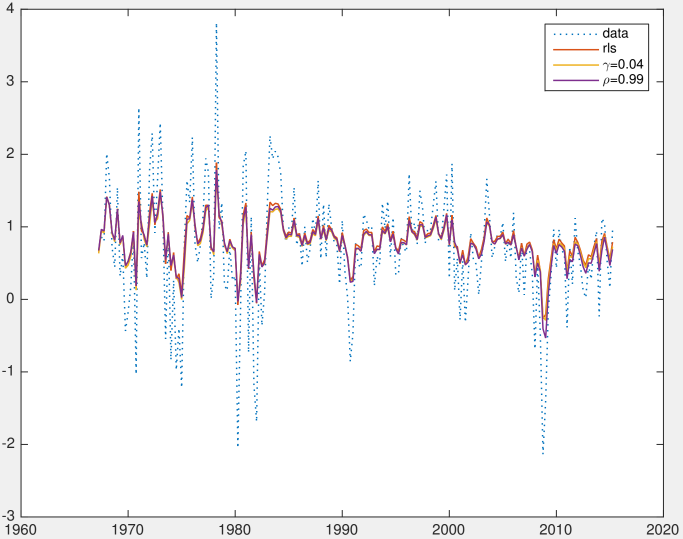
From the MA(∞)

$$\begin{aligned}
(1-\phi L)y_{T+h} &= (1+\theta L)\epsilon_{T+h} \\
y_{T+h} &= (1+\phi L+\phi^2 L^2+\dots)(1+\theta L)\epsilon_{T+h} \\
&= (1+(\phi+\theta)L+\phi(\phi+\theta)L^2+\phi^2(\phi+\theta)L^3+\dots)\epsilon_{T+h}
\end{aligned}$$

giving forecast and forecast error variance:

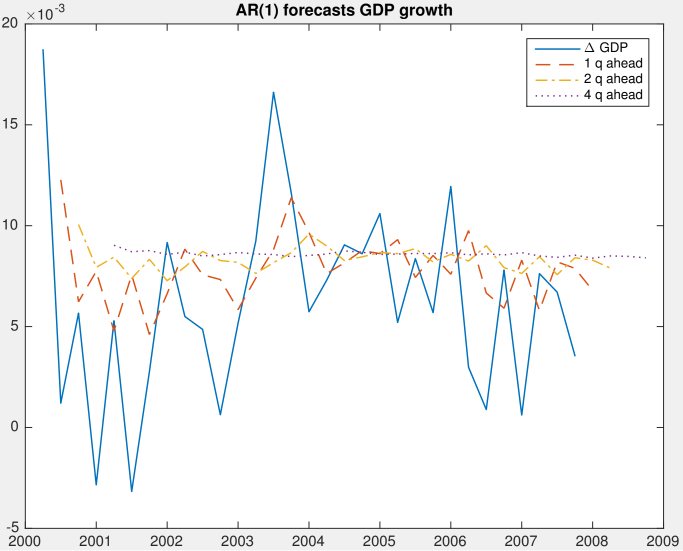
$$\begin{aligned}
\mathbb{E}_T[y_{T+h}] &= \sum_{j=0}^{\infty} \phi^{h-1+j}(\phi+\theta)\epsilon_{T-j} \\
\mathbb{E}_T[\hat{e}^2_{T+h}] &= (1+(\phi+\theta)^2+\phi^2(\phi+\theta)^2+\dots+\phi^{2(h-2)}(\phi+\theta)^2)\sigma^2_\epsilon
\end{aligned}$$

Example: GDP growth forecasts



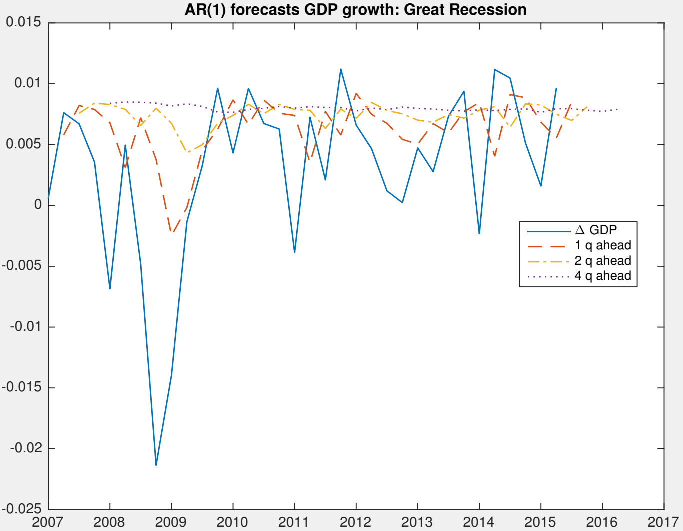
In sample model fits well: fitted values not true forecasts

Example: GDP growth forecasts



Out Of Sample error variance quite large

Example: GDP growth forecasts



Around the great recession the model does poorly

Forecast confidence intervals

How likely is the outcome to be near the forecast?

The outcome y_{T+h} has zero chance of being equal to $\hat{y}_{T+h}$

But we would like to quantify the uncertainty

We can use the Normal distribution assumption on ϵ_t

$$\Pr(\hat{y}_{T+h} - z_c \Sigma_h^{1/2} \leq y_{T+h} \leq \hat{y}_{T+h} + z_c \Sigma_h^{1/2}) = (1 - \alpha)$$

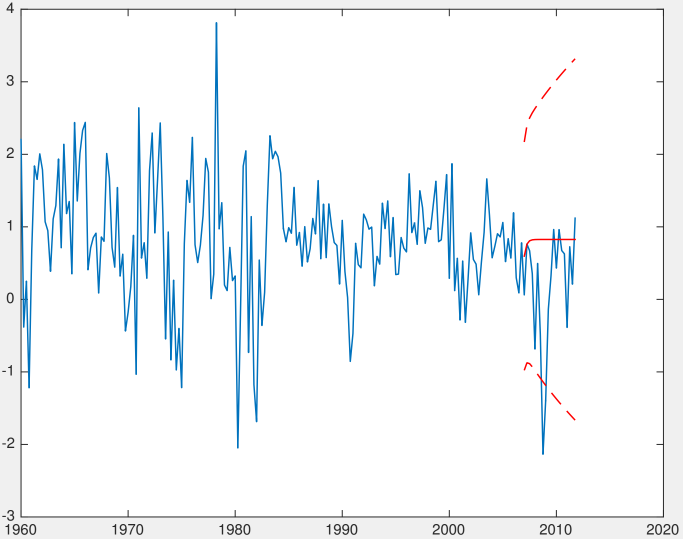
where Σ_h is the forecast error variance of the $h-$step ahead predictor

z_c is the upper $\alpha/2$ critical value for an $N(0,1)$ distribution

$z_c = 1.96$ for a 95% forecast confidence interval

Variance due to *estimation uncertainty* ignored here so these intervals overstate your real confidence in the forecast

Forecast confidence intervals



20 quarter forecast from 2006q4

Forecasting a Random Walk

You can forecast a nonstationary process e.g. $y_t = y_{t-1} + \epsilon_t$

$$\begin{aligned}
\mathbb{E}_T[y_{T+h}] &= y_T \\
\mathbb{E}_T[\hat{e}^2_{T+h}] &= \mathbb{E}_T[(\epsilon_{T+h} + \cdots + \epsilon_{T+1})^2] \\
&= h\sigma^2_\epsilon
\end{aligned}$$

but the forecast error variance grows without bound as h increases

There is no information beyond the current level of the series, sometimes called 'unforecastable'

Forecast evaluation

Let $T = n + m$

- Estimate the model(s) on first n observations
- Compute forecasts (out of sample) over the next m observations
- Compute the criterion
  - (Root) Mean Square Prediction Error (R)MSPE

$$MSPE = \frac{1}{m} \sum_{j=1}^{m} (y_{n+j} - \hat{y}_{n+j})^2$$

  - Mean Absolute Percentage Error MAPE

$$MAPE = \frac{1}{m} \sum_{j=1}^{m} \left| \frac{y_{n+j} - \hat{y}_{n+j}}{y_{n+j}} \right|$$

- Choose model with lowest error criterion

Notes/Problems

Notes/Problems

4. Conditional Heteroskedasticity: processes for second moments

Overview

In these lectures we will cover

- ARCH and GARCH models
- The ARMA representation
- Unconditional variance and expectation
- Tarch and Egarch
- Estimation of GARCH-type models

What are we doing?

- Properties of the model
- Testing properties of data
- Estimation of model parameters
- Model selection, forecasting and evaluation
- Economic inference

Notation: the information set $\boldsymbol{\Psi}_t$

We use $\boldsymbol{\Psi}_t$ to dentoe information available up to time t
It includes

- All y_j observations $j \leq t$
- All constants, seasonals and other deterministic terms
- All exogenous terms up to and including x_{t+1}, if you have such terms

We are interested in the conditional and unconditional expectations and variances of y_t

- $\mathbb{E}[y_t]$ and $\text{var}(y_t)$
- $\mathbb{E}[y_t | \boldsymbol{\Psi}_{t-1}] = \mathbb{E}_{t-1}[y_t]$ and $\text{var}(y_t | \boldsymbol{\Psi}_{t-1}) = \text{var}_{t-1}(y_t)$

Variance of y_t

Recall the AR(1) model $y_t = \phi y_{t-1} + \epsilon_t$. With usual assumptions

$$\mathsf{var}_{t-1}(y_t) = \sigma^2$$
$$\mathsf{var}_{t-1}(y_{t+1}) = (1+\phi^2)\sigma^2$$
$$\ldots \mathsf{var}_{t-1}(y_{t+h-1}) = \sigma^2 \sum_{i=0}^{h-1} (\phi^2)^i$$

Which is a function fo the *horizon*, h

But NOT of the time, t, at which you take expectations

What if there are some periods when your mean equation fits well, and others when it does not? The ARMA models do not capture this.

A model for conditional variance

Why does the e.g. AR(1) behave like this?

$$\begin{aligned}\operatorname{var}(y_t | \boldsymbol{\Psi}_{t-1}) &= \mathbb{E}[(y_t - \mathbb{E}[y_t | \boldsymbol{\Psi}_{t-1}])^2 | \boldsymbol{\Psi}_{t-1}] \\ &= \mathbb{E}[\epsilon_t^2 | \boldsymbol{\Psi}_{t-1}] \\ &= \sigma^2\end{aligned}$$

- So the constant conditional variance of y_t is a consequence of the constant variance assumption for the errors ϵ_t
- GARCH-type models allow $\mathbb{E}[\epsilon_t^2 | \boldsymbol{\Psi}_{t-1}] = h(\boldsymbol{\Psi}_{t-1})$

ARCH(p)

The ARCH(p) model assumes

$$\begin{aligned}
\epsilon_t | \mathbf{\Psi}_{t-1} &\sim N(0, h_t) \\
h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 \\
\Rightarrow \mathbb{E}_{t-1}[\epsilon_t^2] &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 \\
&= h_t
\end{aligned}$$

- $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\sum_{i=1}^{p} \alpha_i < 1$
- The conditional variance at date t, h_t is a function of date $t-1$ or earlier variables
- It will vary over time - high variance implies hight expected variance
- The ARCH process will display volatility clustering

Equivalent forumlation

You should be familiar with both representations of the ARCH model

$$\begin{aligned} \epsilon_t &= z_t \sqrt{h_t} \\ z_t | \boldsymbol{\Psi}_{t-1} &= z_t \sim^{iid} N(0,1) \\ h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 \end{aligned}$$

Conditional mean and variance of ϵ_t

$$\begin{aligned} \mathbb{E}[\epsilon_t | \boldsymbol{\Psi}_{t-1}] &= \mathbb{E}[z_t \sqrt{h_t} | \boldsymbol{\Psi}_{t-1}] \\ &= \mathbb{E}[z_t | \Psi_{t-1}] \sqrt{h_t} = 0 \\ \text{var}(\epsilon_t | \boldsymbol{\Psi}_{t-1}) &= \mathbb{E}[z_t^2 h_t | \boldsymbol{\Psi}_{t-1}] \\ &= \mathbb{E}[z_t^2 | \boldsymbol{\Psi}_{t-1}] \mathbb{E}[h_t | \boldsymbol{\Psi}_{t-1}] \\ &= h_t \end{aligned}$$

GARCH(p,q) model

GARCH is a generalisation which often works well in practice

$$\epsilon_t | \boldsymbol{\Psi}_{t-1} \sim N(0, h_t)$$
$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}$$

- More parsimonious specification compared to high-order ARCH models
- $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$

ARCH and GARCH

GARCH implies a restricted ARCH(∞). E.g. GARCH(1,1)

$$\begin{aligned}
h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \\
&= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 \epsilon_{t-2}^2 + \beta_1 h_{t-2}) \\
&= (1+\beta_1)\alpha_0 + \alpha_1(\epsilon_{t-1}^2 + \beta_1 \epsilon_{t-2}^2) + \beta_1^2 h_{t-2} \\
\ldots &= \frac{\alpha_0}{1-\beta_1} + \alpha_1(\epsilon_{t-1}^2 + \beta_1 \epsilon_{t-2}^2 + \beta_1^2 \epsilon_{t-3}^2 + \ldots)
\end{aligned}$$

- You can only do this if $\beta_1 < 1$
- Is this enough for stability of the GARCH(1,1) process?

ARMA representation

(G)ARCH models imply ARMA models for the squared residuals

- Define $v_t = \epsilon_t^2 - \mathbb{E}[\epsilon_t^2 | \mathbf{\Psi}_{t-1}]$

$$
\begin{aligned}
\epsilon_t^2 &= h_t + v_t \\
&= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} + v_t \\
&= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 (\epsilon_{t-1}^2 - v_{t-1}) + v_t \\
&= \alpha_0 + (\alpha_1 + \beta_1) \epsilon_{t-1}^2 + v_t - \beta_1 v_{t-1}
\end{aligned}
$$

- Which is ARMA(1,1) for the squared residuals
- So stability requires $(\alpha_1 + \beta_1) < 1$
- Exercise: Show $\mathbb{E}[v_t | \mathbf{\Psi}_{t-1}] = 0$ and $\mathbb{E}[v_t v_s | \mathbf{\Psi}_{t-1}] = 0$

More on persistence and staitonarity

There are two channels through which past shocks v_{t-1}, v_{t-2}, etc will affect conditional volatility h_t

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \\ &= \alpha_0 + \alpha_1 v_{t-1} + (\alpha_1 + \beta_1) h_{t-1} \end{aligned}$$

- The direct effect lasts 1 period
- But effects of past volatility shocks are persistent, through $(\alpha_1 + \beta_1) h_{t-1}$
- If $\alpha_1 + \beta_1 = 1$ then a shock to volatility in one period raises *all* future volatility forecasts
- This is called Inegrated GARCH (IGARCH)

General ARMA(m,q) for GARCH(p,q)

The general ARMA(m,q) model from a general GARCH model is

$$
\begin{aligned}
\epsilon_t^2 &= h_t + v_t \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} + v_t \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j (\epsilon_{t-j}^2 - v_{t-j}) + v_t \\
&= \alpha_0 \sum_{i=1}^{m} (\alpha_i + \beta_i) \epsilon_{t-i}^2 - \sum_{j=1}^{q} \beta_j v_{t-j} + v_t
\end{aligned}
$$

With $\alpha_i = 0$ for $i > p$; $\beta_i = 0$ for $i > q$ and $m = \max(p, q)$

Unconditional moments

It is interesting to consider the unconditional moments of the error process, $\mathbb{E}[(\epsilon_t)^r]$, up to order four. E.g. for a GARCH(1,1)

- Expectation

$$\mathbb{E}[\epsilon_t] = \mathbb{E}[z_t\sqrt{h_t}]$$
$$= \mathbb{E}[z_t]\mathbb{E}[\sqrt{h_t}] = 0$$

- Variance

$$\begin{aligned}
\sigma^2 := \mathbb{E}[\epsilon_t^2] &= \mathbb{E}[z_t^2]\mathbb{E}[h_t] = \mathbb{E}[h_t] \\
&= \alpha_0 + \alpha_1\mathbb{E}[\epsilon_{t-1}^2] + \beta_1\mathbb{E}[h_{t-1}] \\
&= \alpha_o + (\alpha_1 + \beta_1)\mathbb{E}[h_t] \\
\Rightarrow \mathbb{E}[\epsilon_t^2] &= \frac{\alpha_0}{1 - \alpha_1 - \beta_1}
\end{aligned}$$

Unconditional moments

- The third moment, which tells us about the asymmetry of the process, will be 0 by the normality of ϵ_t
- However, the fourth moment, which tells us about the probability of large shocks, will be raised compared to the usual case.
- $\mathbb{E}[\epsilon_t^4]$ depends on the variance of the surprise process v_t, as $v_t^2 = \epsilon_t^4 - 2\epsilon_t^2 h_t + h_t^2$
- E.g. for ARCH(1)
$$\frac{\mathbb{E}[\epsilon_t^4]}{\sigma^4} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} > 3$$
- GARCH processes are in general *leptokurtic*, the probability of seeing a large shock ($>$ 2 std devs) is greater than you would see under a normal distribution
- He and Tirasvirta (1997) for details

Extensions: Asymmetric GARCH process

What if good news and bad news surprises have different effects on future volatility?

- Threshold GARCH (TARCH)

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \gamma \epsilon_{t-1}^2 d_{t-1} + \sum_{j=1}^{q} \beta_j h_{t-j}$$

where $d_{t-1} = 1$ iff $\epsilon_{t-1} < 0$, else $d_{t-1} = 0$

- γ measures the asymmetry of the process: if $\gamma > 0$ bad news has a bigger effect on future volatility than good news

Extensions: Asymmetric GARCH

The EGARCH model

- This is an alternative way to model asymmetry

$$\log(h_t) = \alpha_0 + \beta \log(h_{t-1}) + \alpha \left| \frac{\epsilon_{t-1}}{h_{t-1}^{1/2}} \right| + \gamma \left(\frac{\epsilon_{t-1}}{h_{t-1}^{1/2}} \right)$$

- Effect of past shocks is exponential rather than quadratice
- Log form ensures variance estimates stay positive
- γ measures asymmetry: significantly negative γ implies bad shocks have larger effects than good shocks

Extensions: GARCH in mean

Investors may require higher returns to hold assets known to have higher variance, ie. to face conditional volatility in their protfolio

$$y_t = \beta' x_t + \delta h_{t-1} + \epsilon_t$$
$$\epsilon_t | \mathbf{\Psi}_{t-1} \sim N(0, h_t)$$

- The coefficient δ measures the relationship between risk and return
- Higher expected variance should be rewarded with higher expected returns
- Conditional variance may enter via logs, square roots etc instead of linearly

Estimation

- GARCH models are estimated via Maximum Likelihood
- E.g. assuming conditionally normal errors

$$
\begin{aligned}
y_t &= x_t'\gamma + \epsilon_t \\
\epsilon_t &= z_t\sqrt{h_t} \\
z_t &\sim^{iid} N(0,1) \\
\Rightarrow \epsilon_t | \boldsymbol{\Psi}_{t-1} &\sim N(0, h_t) \\
\Rightarrow y_t | \boldsymbol{\Psi}_{t-1} &\sim N(x_t'\gamma, h_t)
\end{aligned}
$$

- From the specification of the model, we have derived the conditional density function of an observation
- These are the individual likelihood contributions for each observation
- Note that x_t is here a set of variables which are known w.r.t $\boldsymbol{\Psi}_{t-1}$

Estimation

It is easier to work with the log-likelihood contributions

$$L_t(\theta) = \frac{1}{(2\pi h_t)^{1/2}} exp\left\{\frac{-\epsilon_t^2}{2h_t}\right\}$$

$$l_t(\theta) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln h_t - \frac{\epsilon_t^2}{2h_t}$$

The log-likelihood function for the whole sample exploits conditional independence of observations

$$l(\theta) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(h_t) - \frac{1}{2}\sum_{t=1}^{T}(\epsilon_t^2/h_t)$$

It is the sum of the individual log-likelihood contributions

Estimation

- To finish our specification of the likelihood procedure we need to give details of the GARCH model
- For GARCH(1,1) we then have specification

$$\epsilon_t = y_t - x_t'\gamma \text{ and } h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1}$$
$$\theta_{MLE} = \underset{\theta}{\operatorname{argmax}}\, l(\theta) \text{ s.t. } \alpha_0 > 0,\ \alpha_1, \beta_1 \geq 0$$

- The estimated parameter vector θ_{MLE} is the value of the vector θ which maximises $l(\theta)$; i.e. which makes the observed realisation the most likely it could be conditional on the model we estimate
- In this example the vector $\theta =$

Estimation

- Note that positive variance imposes constraints on the estimation of the α_i, β_j
- Constrianed optimisation is a hard problem and your estimates may fail to converge
- A sign of a poorly specified model for the data you consider
- Some routines also impose stability
- Often $x_t'\gamma$ is estimated beforehand be OLS, and ϵ_t must be a mean 0 series
- If the normal assumptions in the errors is wrong, we have a QMLE set up, and the estimates of the GARCH parameter are still conistent.

Forecasting: Mean deviation form

For example, consider the ARCH(1) model

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$. Provided $\alpha_1 < 1$, $\sigma^2 := \mathbb{E}[\epsilon_t^2]$ exists

$$\begin{aligned} \sigma^2 &= \frac{\alpha_0}{1-\alpha_1} \\ h_t &= \sigma^2(1-\alpha_1) + \alpha_1 \epsilon_{t-1}^2 \\ h_t - \sigma^2 &= \alpha_1(\epsilon_{t-1}^2 - \sigma^2) \end{aligned}$$

Iterate on the mean-deviation form to produce forecasts

Forecasting: ARCH(1) volatility forecasts

$$
\begin{aligned}
\mathbb{E}[h_t - \sigma^2 | \boldsymbol{\Psi}_{t-1}] &= h_t - \sigma^2 = \alpha_1(\epsilon_{t-1}^2 - \sigma^2) \\
\mathbb{E}[h_{t+1} - \sigma^2 | \boldsymbol{\Psi}_{t-1}] &= \mathbb{E}[\alpha_1(\epsilon_t^2 - \sigma^2) | \boldsymbol{\Psi}_{t-1}] \\
&= \alpha_1(\mathbb{E}[z_t^2 h_t | \boldsymbol{\Psi}_{t-1}] - \sigma^2) \\
&= \alpha_1^2(\epsilon_{t-1}^2 - \sigma^2)) \\
\mathbb{E}[h_{t+2} - \sigma^2 | \boldsymbol{\Psi}_{t-1}] &= \mathbb{E}[\alpha_1(\epsilon_{t+1}^2 - \sigma^2) | \boldsymbol{\Psi}_{t-1}] \\
&= \alpha_1(\mathbb{E}[z_{t+1}^2 h_{t+1} | \boldsymbol{\Psi}_{t-1}] - \sigma^2) \\
&= \alpha_1(\mathbb{E}[h_{t+1} | \boldsymbol{\Psi}_{t-1}] - \sigma^2) \\
&= \alpha_1^3(\epsilon_{t-1}^2 - \sigma^2) \\
\ldots \mathbb{E}[h_{t+j} - \sigma^2 | \boldsymbol{\Psi}_{t-1}] &= \alpha_1^{j+1}(\epsilon_{t-1}^2 - \sigma^2)
\end{aligned}
$$

Excercise: derive the j-period ahead vol. forecast condtional on $\boldsymbol{\Psi}_t$

Forecasting: GARCH(1,1) forecast

Conditional volatility follows

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}$$

Proceed in the mean-deviation form

$$\sigma^2 = \mathbb{E}[h_t] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

Giving

$$\begin{aligned} h_t &= \sigma^2(1 - \alpha_1 - \beta_1) + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \\ h_t - \sigma^2 &= \alpha_1(\epsilon_{t-1}^2 - \sigma^2) + \beta_1(h_{t-1} - \sigma^2) \end{aligned}$$

Plug this in to expectations of future conditional volatility

Forecasting: GARCH(1,1) forecast

$$
\begin{aligned}
\mathbb{E}[h_{t+1} - \sigma^2 | \, \boldsymbol{\Psi}_t] &= \alpha_1(\epsilon_t - \sigma^2) + \beta_1(h_t - \sigma^2) \\
\mathbb{E}[h_{t+2} - \sigma^2 | \, \boldsymbol{\Psi}_t] &= \mathbb{E}[\alpha_1(\epsilon_{t+1}^2 - \sigma^2) + \beta_1(h_{t+1} - \sigma^2) | \, \boldsymbol{\Psi}_t] \\
&= \alpha_1(\mathbb{E}[\epsilon_{t+1}^2 | \, \boldsymbol{\Psi}_t] - \sigma^2) + \beta_1 \mathbb{E}[(h_{t+1} - \sigma^2) | \, \boldsymbol{\Psi}_t] \\
&= \alpha_1(\mathbb{E}[z_{t+1}^2 h_{t+1} | \, \boldsymbol{\Psi}_t] - \sigma^2) + \beta_1 \mathbb{E}[(h_{t+1} - \sigma^2) | \, \boldsymbol{\Psi}_t] \\
&= \alpha_1(\mathbb{E}[h_{t+1} - \sigma^2 | \, \boldsymbol{\Psi}_t]) + \beta_1 \mathbb{E}[(h_{t+1} - \sigma^2) | \, \boldsymbol{\Psi}_t] \\
&= (\alpha_1 + \beta_1) \mathbb{E}[h_{t+1} - \sigma^2 | \, \boldsymbol{\Psi}_t] \\
&= (\alpha_1 + \beta_1) \big(\alpha_1(\epsilon_t^2 - \sigma^2) + \beta_1(h_t - \sigma^2) \big)
\end{aligned}
$$

$$
\ldots \mathbb{E}[h_{t+j} - \sigma^2 | \, \boldsymbol{\Psi}_t] = (\alpha_1 + \beta_1)^{j-1} \big(\alpha_1(\epsilon_t^2 - \sigma^2) + \beta_1(h_t - \sigma^2) \big)
$$

Notes/Problems

Notes/Problems

5. Cointegration: the single-equation approach

Overview

In these lectures we will cover

- Spurious regression
- Cointegration: some theory
- Cointegration: testing and static estimation
- Error correction model

What are we doing?

- Properties of the model
- Testing properties of data
- Estimation of model parameters
- Model selection, forecasting and evaluation
- Economic inference

Spurious regression: set up

Using standard testing procedures on intgrated variables suggests significant relationship where none exists

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \epsilon_t \\
x_t &= \rho x_{t-1} + v_t \\
\mathbb{E}[\epsilon_t v_s] &= 0 \quad \forall t, s
\end{aligned}
$$

Putting $\rho = 1$ gives the integrated variables

$$
\begin{aligned}
y_t &= y_{t-1} + \epsilon_t \\
x_t &= x_{t-1} + v_t
\end{aligned}
$$

Spurious regression: experiment

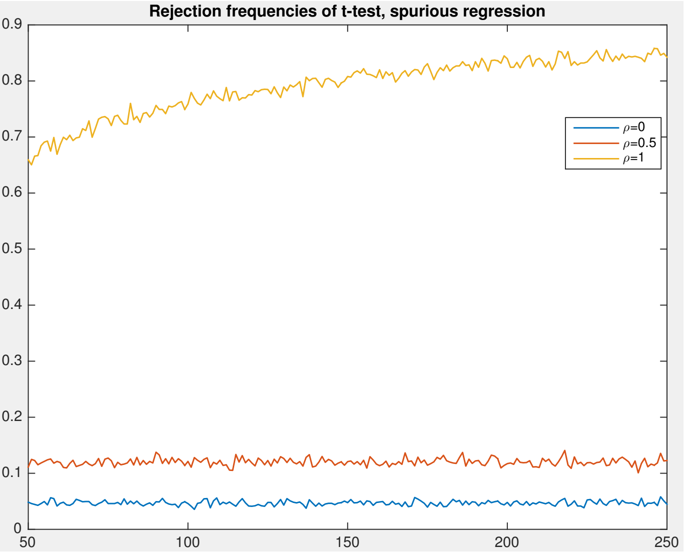
Use OLS to find out if there is a linear relationship between x an y

$$y_t = \alpha + \beta x_t + \epsilon_t \tag{2}$$
$$H_0 : \hat{\beta} = 0$$
$$\tau = \frac{\hat{\beta}}{\widehat{se(\beta)}} \sim t_{T-2}$$

The standard t-test will tend to over-reject the null as $T \to \infty$!

The closer to a unit root, the worse the problem

Spurious regression: experiment



Spurious regression: experiment

- When $\rho = 0$ the t-tes is correctly sized:
  - Rejects true null, $\beta = 0$, 5% of the time
- When $\rho = 0.5$ the test is a bit oversized
  - Rejects about 12% of the time
- When $\rho = 1$ the test is terrible, and it gets *worse* as $T \to \infty$!
  - For $T = 200$ you have $> 80\%$ chance of finding a significant correlation between two completely unrelated random walks.

Spurious regression: main problem

Under the null $\beta = 0$, the regression equation (2) is dynamically misspecified, implying

$$y_t = \alpha + \epsilon_t$$

- The variance of this process is finite while the true $\text{var}(y_t) = t\sigma^2$ grows with the sample size.
- This misleads the regression as the variance of the x process also grows with the sample size, so there is a multiple of the x data which can better approximate the y data.
- As the sample size grows, the problem gets worse, and the t-value is more likely to be significant

Spurious regression: partial fix

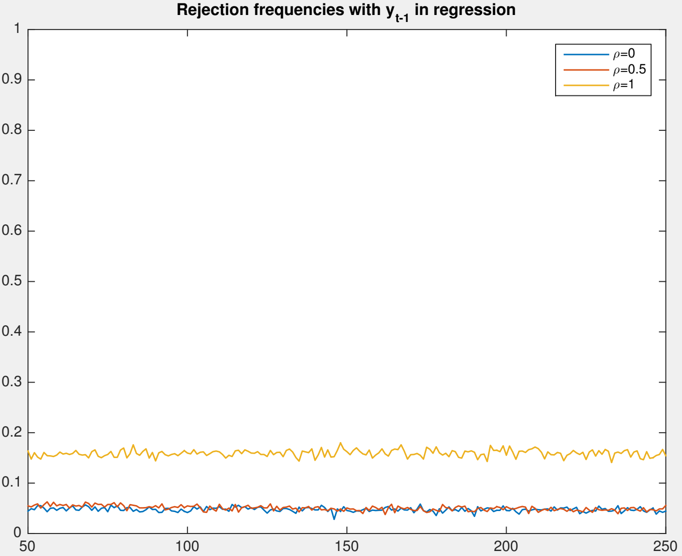
We could ameliorate the problem by embedding the true process in the regression equation under H_0. Regress

$$y_t = \alpha + \gamma y_{t-1} + \beta x_t + \epsilon_t$$
$$H_0 \Rightarrow y_t = \alpha + \gamma y_{t-1} + \epsilon_t$$

So the regression equation caputres the true Data Generating Process (DGP) with $\alpha = 0$, $\beta = 0$ and $\gamma = 1$.

The problem is less bad, but the t-test still does not have a normal distribution, even asymptotically; the test is still oversized, for (near) unit roots

Spurious regression: partial fix



Direction

- When there are unit roots standard distributions of tests do not apply
- There are two situations to distinguish
  - We have a spurious regression problem
  - There is a genuine relationship between the series: they are *cointegrated*

Rules of unit roots

Let $x_t \sim I(1)$, $y_t \sim I(1)$ $z_t \sim I(0)$ and c_1, c_2 be scalars. Then:

I $x_t + z_t \sim I(1)$: the unit root is the dominant property

II $c_1 x_t \sim I(1)$

III $c_2 z_t \sim I(0)$

IV For arbitrary c_1, c_2 then $r_t = c_1 x_t + c_2 y_t \sim I(1)$

V If there exist particular c_1, c_2, say c_1^* c_2^*, such that $r_t = c_1^* x_t + c_2^* y_t \sim I(0)$ we say x_t and y_t are cointegrated, and $[c_1^*, c_2^*]$, is the cointegrating vector

Cointegration: general definition

Two or more series which are individually integrated of order d ($\sim I(d)$) are said to be cointegrated when a linear combination of the series is integrated of order $d - b$, that is the combination is $\sim I(d - b)$, for $b > 0$.

For $I(1)$ series, they are cointegrated if there exists a linear combination which is $I(0)$. (In this case $d = 1$, $b = 1$.)

Normalising the cointegrating vector

You have to impose a restriction in order to interpret the cointegrating vector (it is only defined up to a scalar transform)

$$\begin{bmatrix} c_1^* & c_2^* \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = r_t \sim I(0)$$

If $r_t \sim I(0)$ then $nr_t \sim I(0)$, let $n = \frac{1}{c_1^*}$

$$\begin{aligned} nr_t &= \begin{bmatrix} 1 & \frac{c_2^*}{c_1^*} \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} \sim I(0) \\ &= x_t - \beta y_t \sim I(0) \end{aligned}$$

This normalises the cointegrating vector on x_t, giving relationship between unit of x_t and y_t

Cointegration: common stochastic trends

Series can only cointegrate if they have a common $I(1)$ component

$$
\begin{aligned}
y_t &= \beta X_t + \tilde{y}_t \\
x_t &= X_t + \tilde{x}_t
\end{aligned}
$$

where $X_t \sim I(1)$ and $\tilde{x}_t$, $\tilde{y}_t \sim I(0)$. $x_t \ y_t \sim I(1)$ but:

$$
\begin{aligned}
y_t - \beta x_t &= \beta X_t + \tilde{y}_t - \beta(X_t + \tilde{x}_t) \\
&= \tilde{y}_t - \beta \tilde{x}_t \sim I(0)
\end{aligned}
$$

So the linear combination $\begin{bmatrix} 1 & -\beta \end{bmatrix}$ is $I(0)$ and the series are cointegrated

The series X_t is known as the common stochastic trend $X_t = \sum_{s=1}^{t} \epsilon_s$, c.f. the deterministic trend $t = \sum_{s=1}^{t} 1$.

Uniqueness of cointegrating vector

Does the alternative linear combination $y_t - \tilde{\beta} x_t$ cointegrate?

$$\begin{aligned} y_t - \tilde{\beta} x_t &= \beta X_t + \tilde{y}_t - (\beta + \delta)(X_t + \tilde{x}_t) \\ &= \tilde{y}_t - \tilde{\beta} \tilde{x}_t - \delta X_t \sim I(1) \end{aligned}$$

So the cointegrating vector is unique(up to a scalar multiple)

Cointegration: a long-run equilibrium

Variables that ciontegrate:

- 'move together' through time
- the 'error' $r_t = y_t - \beta x_t$ is $I(0)$
- so $\mathbb{E}[r_t^2]$ is constant over time
- so y_t can not drift increasingly far away from βx_t: cointegration is a statistical notion of equilibrium

Testing for cointegration: static OLS

A simple test for cointegration is based on the regression

$$y_t = \beta_0 + \beta_1 x_t + v_t$$
$$\Rightarrow \hat{v}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$$

- The linear combination of x_t y_t defined by $\hat{v}_t$ is the linear combination most likely to be a cointegrating vector
- The test for cointegration is based on the behaviour of $\hat{v}_t$ NOT of $\hat{\beta}_1$.
- If the series are cointegrated then $\hat{v}_t$ will be stationary

OLS picks the 'most stationary' residual

OLS solves:

$$(\hat{\beta}_0 \;\; \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \, T^{-1} \sum_{t=1}^{T} \hat{v}_t^2$$

Consider the case where $v_t^* = y_t - \beta_0^* - \beta_1^* x_t \sim I(0)$ is the unique cointegrating vector, but $v_t = y_t - \beta_0 - \beta_1 x_t \sim I(1)$ for any other β

- $\text{var}(v_t^*) = M < \infty$
- $\text{var}(v_t) = \tilde{M} t$, say

OLS picks the 'most stationary' residual

The sum of square residuals depends on the estimated error sequence

$$\hat{v}_t = v_t^* \Rightarrow \mathbb{E}\left[T^{-1}\sum \hat{v}_t^2\right] = T^{-1}\sum \mathbb{E}(v_t^*)^2 = M$$

$$\hat{v}_t = v_t \Rightarrow \mathbb{E}\left[T^{-1}\sum \hat{v}_t^2\right] = T^{-1}\sum \mathbb{E}v_t^2 = T^{-1}\tilde{M}\sum t$$

$$= T^{-1}\tilde{M}(T(T+1)/2) \sim O(T)$$

- The squared error sequence is finite iff $\hat{v}_t$ is the cointegrating vector
- Therefore OLS will choose $\hat{\beta} = (\beta_0^*, \beta_1^*)$ to estimate the cointegrating linear combination of x_t and y_t, if there is one

Testing for cointegration

We use ordinary unit root tests on series $\hat{v}_t$ to see if it is stationary

$$\Delta \hat{v}_t = \phi \hat{v}_{t-1} + \epsilon_t$$
$$H_0 : \phi = 0 \Rightarrow \text{ no cointegration}$$
$$H_A : \phi < 0 \Rightarrow \text{ the series cointegrate}$$

- You may need to add lagged $\Delta \hat{v}_{t-i}$ to ensure ϵ_t is serially uncorrelated
- Critical values are from McKinnon's tables (more -ve)
- Critical values depend on the number of regressors: deterministic terms and $x_{t,i}$ variables in the regression: see Table 1
- m is the total number of variables, so $m = 2$ is relevant for case of 1 regressor

McKinnon critical values

Test stat	1%	5%	10%
m=2			
τ_t	-3.90	-3.34	-3.04
τ_{tc}	-4.32	-3.78	-3.50
m=3			
τ_t	-4.29	-3.74	-3.45
τ_{tc}	-4.66	-4.12	-3.84
m=4			
τ_c	-4.64	-4.10	-3.81
τ_{tc}	-4.97	-4.43	-4.14
m=5			
τ_c	-4.96	-4.42	-4.13
τ_{tc}	-5.25	-4.72	-4.43

See *Esimtation and Inference in Econometrics*, Davidson and McKinnon (1993) for further detail

Example

Test for cointegration between y_t and x_t allowing for a constant and no trend

$$y_t = \beta_0 + \beta_1 x_t + v_t$$
$$\Delta \hat{v}_t = \phi \hat{v}_{t-1} + \epsilon_t$$

Compare the unit root DF test statistic to the $m = 2$, τ_c critical value from the table: -3.34 for a 5% test

- If the test stat is less than cv then reject H_0, so there is evidence of cointegration
- Critical values are more negative than standard ADF tests as OLS will tend to minimise variance, so need extra confidence that the stationary behaviour of $\hat{v}_t$ is genuine

Static regressions with cointegrated variables: Advantages and disadvantages

- Simple: only requires OLS estimation
- $\hat{\beta}_1$ is consistently estimated (in fact 'superconsistent')

- But $\hat{v}_t$ will typically be autocorrelated, as any dynamics in the y_t x_t relationship are ignored $\Rightarrow$ significant small-sample biases in $\hat{\beta}_1$
- Where more than 2 variables, there could be more than 1 cointegrating relationship, but OLS only estimates a linear combination of the cointegrating equations.
- $\hat{\beta}$ is not normally distributed - cannot test inferences about β

- A VAR based approach to cointegration will fix some of these problems

Allowing for structural breaks

Run the test on

$$y_t = \alpha_1 + I(t > T_b)\alpha_2 + \lambda t + \beta_1 x_t + I(t > T_b)\beta_2 x_t + \epsilon_t$$

This allows for the possibility that x_t and y_t cointegrate with

I. A change in the intercept but no change in the slope if $\lambda = 0$, $\beta_2 = 0$
II. A change in intercept, allowing for a time trend $\beta_2 = 0$
III. A change in slope and intercept, $\lambda = 0$

- Critical values of the ADF test are more negative again
- For unkown break date compute $\hat{\epsilon}_t$ for all possible T_b and select T_b with the most negative ADF test - this is most likely date of break assuming series are really cointegrated

  Harris and Solis, pp85-86

Error correction model

Consider a dynamic model for income and consupmtion where $y_t \sim I(1)$,$c_t \sim I(1)$.

$$\begin{aligned}
c_t &= \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 c_{t-1} + \epsilon_t \\
\Delta c_t &= \beta_1 y_t + \beta_2 y_{t-1} + (\beta_3 - 1) c_{t-1} + \epsilon_t \\
&= \beta_1 \Delta y_{t-1} + (\beta_3 - 1)\Big(c_{t-1} - \frac{\beta_1 + \beta_2}{1 - \beta_3} y_{t-1}\Big) + \epsilon_t
\end{aligned}$$

Δc_t and Δy_t are I(0) as differences of $I(1)$ variables

Ther term in brackets $(c_t - \beta y_t)$ is $I(0)$ iff $(1, -\beta)$ is a cointegrating vector for c_t, y_t, where $\beta = \frac{\beta_1+\beta_2}{1-\beta_3}$

Then $z_t = \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} c_t \\ y_t \end{bmatrix} \sim I(0)$

Estimation of ECM

We want to estimate

$$\Delta c_t = \alpha_1 \Delta y_t + \alpha_2 (c_{t-1} - \beta y_{t-1}) + \epsilon_t$$

How can we distinguish α_2 and β?

How do we regress on the unobserved error correction term $z_t = c_t - \beta y_t$?

I Estimate the cointegrating equation $c_t = \hat{\gamma} y_t + \hat{v}_t$

II The residuals, $\hat{v}_t = c_t - \hat{\gamma} y_t$, estimate the error correction term

III Estimate the ECM by OLS

$$\begin{aligned}\Delta c_t &= \alpha_1 \Delta y_t + \alpha_2 \hat{v}_{t-1} + \epsilon_t \\ &= \alpha_1 \Delta y_t + \alpha_2 (c_{t-1} - \hat{\gamma} y_{t-1}) + \epsilon_t\end{aligned}$$

$\hat{\gamma} \to \beta$ as T gets large

Identification: recovering the ARDL parameters

Recall the unrestricted ARDL model was

$$c_t = \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 c_{t-1} + \epsilon_t$$

From the 2-step VECM we recover estimates of the original parameters

- $\beta_1 = \alpha_1$
- $\beta_3 = \alpha_2 + 1$
- $\beta_2 = \gamma\alpha_2 - \alpha_1$

This identification scheme imposes no restrictions on the the original model

- There are 3 parameters in the ARDL and 3 parameters in the VECM
- The model is *just identified*

Economic theory may suggest *overidentifying* restrictions

Example: overidentifying restrictions

For example, a theory suggests consumption and income move 1 for 1

$$\Delta c_t = \alpha_1 \Delta y_t + \alpha_2 (y_{t-1} - c_{t-1}) + \epsilon_t \tag{3}$$

This implies

$$\begin{aligned} c_t &= \alpha_1 y_t + (\alpha_2 - \alpha_1) y_{t-1} + (1 - \alpha_2) c_{t-1} + \epsilon_t \\ &= \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 c_{t-1} + \epsilon_t \end{aligned}$$

Identification implies $\alpha_1 = \beta_1$, $(1 - \alpha_2) = \beta_3$, $\alpha_2 - \alpha_1 = \beta 2$

The restriction is $\beta_1 + \beta_2 + \beta_3 = 1$

Such overidentifying restrictions can be tested.

Economic interpretation

Along the steady state growth path, the restricted model implies the APC (C/Y) will depend on the growth rates of variables, but not on their levels. Steady state $\Rightarrow \Delta y_t = \Delta c_t = g$. Along the ss path (3) implies:

$$g = \alpha_1 g + \alpha_2(y_{t-1} - c_{t-1})$$
$$\Rightarrow (1-\alpha_1)g = \alpha_2(\ln Y_t - \ln C_t)$$
$$\Rightarrow \frac{C}{Y} = exp\Big(g\frac{\alpha_1 - 1}{\alpha_2}\Big)$$

On the other hand

$$g = \alpha_1 g + \alpha_2(\beta y_{t-1} - c_{t-1})$$
$$\Rightarrow \frac{C}{Y} = Y^{\beta-1} exp\Big(g\frac{\alpha_1 - 1}{\alpha_2} - g(\beta - 1)\Big)$$

The restricted ARDL is only valid if there is cointegration with a unit coefficient

Notes/Problems

Notes/Problems

6. Vector Autoregressions

Overview

In these lectures we will cover

- What is a VAR
    - Writing system as a VAR
    - Expectations and variances
    - Granger causality
    - The impulse response function
    - Variance decomposition
- Why a VAR
    - Review of assumption for OLS: unbiasedness vs. consistency
    - Endogeneity in time series; finding a consistent estimator
    - Estimation and inference in a VAR

Overview

What are we doing?

- Properties of the model
- Testing properties of data
- Estimation of model parameters
- Model selection, forecasting and evaluation
- Economic inference

What is a VAR

A VAR is a set of n interdependent time series processes which have

- multiple cross-equation linkages
- correlation between equation errors

We can write this as an autoregression for the *vector* $\mathbf{y}_t$. Define the matrix lag polynomial

$$\mathbf{\Phi}(L) = (I_n - \mathbf{\Phi}_1 L - \cdots - \mathbf{\Phi}_p L^p)$$

Then the VAR satisfies

$$\begin{aligned} \mathbf{\Phi}(L)\,\mathbf{y}_t &= c + \epsilon_t \\ \mathbb{E}[\epsilon_t \epsilon_t'] &= \mathbf{\Sigma} \\ \mathbb{E}[\epsilon_t \epsilon_s] &= 0_{n.n}\ \forall t \neq s \end{aligned} \tag{4}$$

where $\mathbf{y}_t$, c and ϵ_t are $n.1$ vectors and $\mathbf{\Phi}_i$, $\mathbf{\Sigma}$ are $n.n$ matrices

What is a VAR

From (4) we have the VAR(p)

$$\underbrace{\mathbf{y}_t}_{n.1} = \underbrace{c}_{n.1} + \underbrace{\mathbf{\Phi}_1}_{n.n}\underbrace{\mathbf{y}_{t-1}}_{n.1} + \cdots + \underbrace{\mathbf{\Phi}_p}_{n.n}\underbrace{\mathbf{y}_{t-p}}_{n.1} + \underbrace{\epsilon_t}_{n.1}$$

Example $n = 2$, $p = 2$

$$\begin{aligned} y_{1,t} &= c_1 + \phi_{11}^1 y_{1,t-1} + \phi_{12}^1 y_{2,t-1} + \phi_{11}^2 y_{1,t-2} + \phi_{12}^2 y_{2,t-2} + \epsilon_{1,t} \\ y_{2,t} &= c_2 + \phi_{21}^1 y_{1,t-1} + \phi_{22}^1 y_{2,t-1} + \phi_{21}^2 y_{1,t-2} + \phi_{22}^2 y_{2,t-2} + \epsilon_{2,t} \end{aligned}$$

Gives

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}^1 & \phi_{12}^1 \\ \phi_{21}^1 & \phi_{22}^1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11}^2 & \phi_{12}^2 \\ \phi_{21}^2 & \phi_{22}^2 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

which is a VAR(2) in 2 variables.

Forecasting with VARs

Forecasting with VARs is just like forecasting with autoregressions

- We are taking conditional expectations again

Consider the VAR(1)

$$\begin{aligned}
\mathbb{E}_t[\mathbf{y}_{t+1}] &= \mathbb{E}[c + \mathbf{\Phi}\,\mathbf{y}_t + \epsilon_{t+1} | \,\mathbf{y}^t] \\
&= c + \mathbf{\Phi}\,\mathbf{y}_t \\
\mathbb{E}_t[\mathbf{y}_{t+2}] &= \mathbb{E}[c + \mathbf{\Phi}\,\mathbf{y}_{t+1} + \epsilon_{t+2} | \,\mathbf{y}^t] \\
&= (I_n + \mathbf{\Phi})c + \mathbf{\Phi}^2\,\mathbf{y}_t \\
\ldots \mathbb{E}[\mathbf{y}_{t+k}] &= (I_n + \mathbf{\Phi} + \cdots + \mathbf{\Phi}^{k-1})c + \mathbf{\Phi}^k\,\mathbf{y}_t
\end{aligned}$$

which is just like a fat autoregression

Forecasting with VARs

Consider the example of the VAR(2) in n variables

$$\begin{aligned}
\mathbb{E}_t[\mathbf{y}_{t+1}] &= \mathbb{E}[c + \mathbf{\Phi}_1\,\mathbf{y}_t + \mathbf{\Phi}_2\,\mathbf{y}_{t-1} + \epsilon_{t+1} |\, \mathbf{y}^t] \\
&= c + \mathbf{\Phi}_1\,\mathbf{y}_t + \mathbf{\Phi}_2\,\mathbf{y}_{t-1} \\
\mathbb{E}_t[\mathbf{y}_{t+2}] &= \mathbb{E}[c + \mathbf{\Phi}_1\,\mathbf{y}_{t+1} + \mathbf{\Phi}_2\,\mathbf{y}_t + \epsilon_{t+2} |\, \mathbf{y}^t] \\
&= c + \mathbf{\Phi}_1\,\mathbb{E}_t[\mathbf{y}_{t+1}] + \mathbf{\Phi}_2\,\mathbf{y}_t \\
&= (I_n + \mathbf{\Phi}_1)c + \left(\mathbf{\Phi}_1^2 + \mathbf{\Phi}_2\right)\mathbf{y}_t + \mathbf{\Phi}_1\,\mathbf{\Phi}_2\,\mathbf{y}_{t-1} \\
\mathbb{E}_t[\mathbf{y}_{t+3}] &= \dots
\end{aligned}$$

Generating forecasts recursively is a pain and I can't be bothered

Companion Form

Write the n-variable VAR(p) as a VAR(1) in dimension np:

$$\mathbf{y}_t = c + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \epsilon_t$$

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} = \begin{bmatrix} c \\ 0_{n.1} \\ \vdots \\ 0_{n.1} \end{bmatrix} + \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \dots & \mathbf{\Phi}_p \\ I_n & 0_{n.n} & \dots & 0_{n.n} \\ \vdots & \ddots & & \vdots \\ 0_{n.n} & \dots & I_n & 0_{n.n} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0_{n.1} \\ \vdots \\ 0_{n.1} \end{bmatrix}$$

Which gives the VAR(1) process:

$$\mathbf{x}_t = \mathbf{c} + \mathbf{F}\mathbf{x}_{t-1} + v_t$$

for vector $\mathbf{x}_t = \begin{bmatrix} \mathbf{y}_t' & \dots & \mathbf{y}_{t-p+1}' \end{bmatrix}'$

Companion Form Forecasts

Now forecast the 1st-order process:

$$\mathbf{x}_{t+k|t} := \mathbb{E}_t[\mathbf{x}_{t+k}] = (I_{np} + \mathbf{F} + \cdots + \mathbf{F}^{k-1})\mathbf{c} + \mathbf{F}^k \mathbf{x}_t$$
$$\mathbb{E}_t[\mathbf{y}_{t+k}] = \begin{bmatrix} I_n & 0_{n.n} & \dots & 0_{n.n} \end{bmatrix} \mathbf{x}_{t+k|t}$$

And you never have to worry about anything other than a VAR(1) again.

The same goes for working out the coefficients of the MA(∞) representation of high-order processes; so for calculating variances, impulse responses, correlations etc.

Forecasting with VARs

Do VARs forecast well - better than ARMA?

- It depends on the system
- If there is 'Granger causality' then they may do
  - 'Granger causality':=forecasting power, $\phi^k_{i,j} \neq 0$ for some $i \neq j$ and $k \in 1 \ldots p$
- But it is not guaranteed:
- VARs are large and this introduces lots of estimation variance into the forecast
- This can offset gains in prediction due to Granger causality
- Stock and Watson (2005) Inflation is an IMA(1,1): no multivariate Granger causality
- Killian and Baumeister (2013) do find a VAR which outforecasts small models for oil prices

Granger Causality

Granger causality is very simple

- If variable x helps to forecast variable y we say x Granger-causes y
- This does not mean x causes y in any physical or economic sense:
  - E.g. the weather forecast Granger-causes the weather
- Straightforward to test in a VAR

$$y_{1,t} = c_1 + \phi_{11}^1 y_{1,t-1} + \phi_{12}^1 y_{2,t-1} + \phi_{11}^2 y_{1,t-2} + \phi_{12}^2 y_{2,t-2} + \epsilon_{1,t}$$
$$y_{2,t} = c_2 + \phi_{21}^1 y_{1,t-1} + \phi_{22}^1 y_{2,t-1} + \phi_{21}^2 y_{1,t-2} + \phi_{22}^2 y_{2,t-2} + \epsilon_{2,t}$$

Test $\phi_{12}^1 = \phi_{12}^2 = 0$: approximate F-test. H_0: no Granger causailty from y_2 to y_1

Test $\phi_{21}^1 = \phi_{21}^2 = 0$: approximate F-test. H_0: no Ganger causality from y_1 to y_2

These are separate hypotheses

Conditional variance in a VAR

Condtional variances are similar to the AR model. E.g. VAR(1)

$$
\begin{aligned}
\text{var}_t(\mathbf{y}_{t+1}) &= \mathbb{E}_t\Big[\big[\,\mathbf{y}_{t+1} - \mathbb{E}_t[\mathbf{y}_{t+1}]\big]\big[\,\mathbf{y}_{t+1} - \mathbb{E}_t[\mathbf{y}_{t+1}]\big]'\Big] \\
&= \mathbb{E}_t[\epsilon_{t+1}\epsilon'_{t+1}] \\
&= \mathbf{\Sigma} \\
\text{var}_t(y_{t+2}) &= \mathbb{E}\Big[[\mathbf{\Phi}\,\epsilon_{t+1} + \epsilon_{t+2}][\mathbf{\Phi}\,\epsilon_{t+1} + \epsilon_{t+2}]'\Big] \\
&= \mathbf{\Phi}\,\mathbf{\Sigma}\,\mathbf{\Phi}' + \mathbf{\Sigma} \\
\text{var}_t(\mathbf{y}_{t+k}) &= \sum_{i=0}^{k-1} \mathbf{\Phi}^i\,\mathbf{\Sigma}\,\mathbf{\Phi}'^i
\end{aligned}
$$

with the definition $\mathbf{\Phi}^0 = I_n$

Unconditional Expectations

What happens as $k \to \infty$?

$$\begin{aligned}\lim_{k\to\infty} \mathbb{E}[\mathbf{y}_{t+k}\,|t] &= \lim_{k\to\infty} (I_n + \mathbf{\Phi} + \cdots + \mathbf{\Phi}^{k-1})c + \mathbf{\Phi}^k\,\mathbf{y}_t \\ &= (I_n - \mathbf{\Phi})^{-1}c\end{aligned}$$

where the second line requires that the VAR is stable: roots of $\mathbf{\Phi}(L)$ outside unit circle

Notice the unconditional expectation

$$\begin{aligned}\mathbb{E}[\mathbf{y}_t] &= \mathbb{E}[c + \mathbf{\Phi}\,\mathbf{y}_{t-1} + \epsilon_t] \\ &= c + \mathbf{\Phi}\,\mathbb{E}[\mathbf{y}_t] \\ \Rightarrow \mathbb{E}[\mathbf{y}_t] &= (I_n - \mathbf{\Phi})^{-1}c\end{aligned}$$

as in the stable AR model

Unconditional variance of $\mathbf{y}_t$, $\mathbf{\Sigma}_\infty$, solves $\mathbf{\Sigma}_\infty = \mathbf{\Phi}\,\mathbf{\Sigma}_\infty\,\mathbf{\Phi}' + \mathbf{\Sigma}$

The Impulse Response Function

- A shock to variable i will affect future path of all variables
- We want to know all of these effects $\frac{\partial \mathbf{y}_{i,t+k}}{\partial \epsilon_{j,t}}$ $\forall i, j$
- These objects are called 'Impulse Response Functions'
- You have already seen them for the AR model: they are coeffs of the MA(∞)

- VARs are often used to create 'stylized facts' about how one variable responds to shocks in another
- VARs 'let the data speak', so model predictions should be refinements of VAR predictions
- Smets and Wouters (2003): their DSGE predictions (IRFs) lie inside confidence intervals for VAR IRFs

The Impulse Response Function

The VMA repsresentation:

$$\begin{aligned}
\mathbf{y}_{t+s} &= \mathbf{\Phi}(L)^{-1}\epsilon_{t+s} \\
&= \epsilon_{t+s} + \mathbf{\Psi}_1\,\epsilon_{t+s-1} + \mathbf{\Psi}_2\,\epsilon_{t+s-2} + \dots + \mathbf{\Psi}_s\,\epsilon_t + \dots \\
\frac{\partial\,\mathbf{y}_{t+s}}{\partial \epsilon_t} &= \mathbf{\Psi}_s
\end{aligned}$$

Solving for the $\mathbf{\Psi}_s$ coefficients: $\mathbf{\Psi}(L)\,\mathbf{\Phi}(L) = I_n$

E.g. VAR(1)

$$\begin{gathered}
(I_n + \mathbf{\Psi}_1\,L + \mathbf{\Psi}_2\,L^2 + \dots)(I_n - \mathbf{\Phi}\,L) = I_n \\
\Rightarrow \mathbf{\Psi}_s = \mathbf{\Phi}^s
\end{gathered}$$

For the VAR(p) $\mathbf{\Psi}_s = \mathbf{\Phi}_1\,\mathbf{\Psi}_{s-1} + \mathbf{\Phi}_2\,\mathbf{\Psi}_{s-2} + \dots + \mathbf{\Phi}_p\,\mathbf{\Psi}_{s-p}$ with $\mathbf{\Psi}_0 = I_n$ and $\mathbf{\Psi}_{-s} = 0_{n.n}$.
You could verify this. Or use compaion form!

Interpreting the IRFs

What does it mean to say 'response of variable i to shock j'?

- Not a lot, yet
- ϵ_j is correlated with all other shocks $\mathbb{E}[\epsilon_t \epsilon_t'] = \mathbf{\Sigma}$, and $\sigma_{j,i} = \mathbb{E}[\epsilon_{j,t}\epsilon_{i,t}]$
- Consdier the statement 'the monetary policy shock causes a hump shaped decline in output gap'
- We need to identify the unique part of a shock attributable to monetary policy (i.e. to the short term interest rate)
- Need to isolate part of the short-rate shock uncorrelated with all other shocks
- Then we trace the effects of this unique shock through the system

Orthogonalised IRFs

We need to choose a matrix $\mathbf{H}$ such that $u_t = \mathbf{H}\epsilon_t$ and $\mathbb{E}[u_t u_t'] = \mathbf{D}$

$$\begin{aligned}\mathbb{E}[u_t u_t'] &= \mathbb{E}[\mathbf{H}\epsilon_t\epsilon_t'\mathbf{H}'] \\ &= \mathbf{H}\mathbb{E}[\epsilon_t\epsilon_t']\mathbf{H}' \\ \Rightarrow \mathbf{D} &= \mathbf{H}\mathbf{\Sigma}\mathbf{H}'\end{aligned}$$

A popular choice exploits the Choleski decomposition of $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}'$. Put $\mathbf{H} = \mathbf{P}^{-1}$

$$\begin{aligned}\mathbb{E}[u_t u_t'] &= \mathbb{E}[\mathbf{P}^{-1}\epsilon_t\left(\mathbf{P}^{-1}\epsilon_t\right)'] \\ &= \mathbf{P}^{-1}\mathbb{E}[\epsilon_t\epsilon_t'](\mathbf{P}^{-1})' \\ &= \mathbf{P}^{-1}\mathbf{P}\mathbf{P}'(\mathbf{P}^{-1})' \\ &= \mathbf{I}_n\end{aligned}$$

which is a diagonal matrix, as required.

Orthogonalised IRFs

Given our choice of $\mathbf{H}$ we have

$$\begin{aligned}
\mathbf{y}_t &= \epsilon_t + \mathbf{\Psi}_1 \, \epsilon_{t-1} + \mathbf{\Psi}_2 \, \epsilon_{t-2} + \dots \\
&= \mathbf{H}^{-1} u_t + \mathbf{\Psi}_1 \, \mathbf{H}^{-1} u_{t-1} + \mathbf{\Psi}_2 \, \mathbf{H}^{-1} u_{t-2} + \dots \\
&= \mathbf{\Psi}_0^* \, u_t + \mathbf{\Psi}_1^* \, u_{t-1} + \mathbf{\Psi}_2^* \, u_{t-2} + \dots \\
\frac{\partial \, \mathbf{y}_{t+s}}{\partial u_t} &= \mathbf{\Psi}_s^*
\end{aligned}$$

Element i, j of $\mathbf{\Psi}_s^*$ gives the effect of j^{th} identified shock on i^{th} variable, s periods in the future

The series of coefficents $\psi_{i,j}^{*(s)}$, $s = 0, 1, \dots$, is called the othogonalised impulse response of i to j

This is what your computer typically plots

Economic justification of Choleski decomposition

Choleski decomposition sets $u_t = \mathbf{P}^{-1}\epsilon_t$ where $\mathbf{P}^{-1}$ is a lower triangular matrix which imposes a causal chian

- $y_{1,t}$ responds contemporaneously only to its own shock
- $y_{2,t}$ responds to the $y_{1,t}$ shock and its own shock ...
- $y_{n,t}$ responds contemporaneously to all other variables in the system

- In fact $u_{2,t}$ is the residual from a regression of ϵ_2 on ϵ_1, u_j is the residual from a regression of ϵ_j on $\epsilon_1, ...\epsilon_{j-1}$.
- So the ordering of the variables affects the interpretation of the shocks, and the shape of the IRFs
- Think: would you order $\mathbf{y}_t$ as $[GDP_t, Stock_t]$ or $[Stock_t, GDP_t]$?
- What about inflation v.s. GDP?

- More theoretically robust identification schemes may be available: SVAR literature

The SVAR

- (5) is called the SVAR: $\mathbf{A}_0$ captures the contemporaneous correlations between the $\mathbf{y}_t$ variables
- $\mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \mathbf{D}$: these are independent *structural* shocks

$$\mathbf{A}_0 \, \mathbf{y}_t = \mathbf{a} + \mathbf{A}_1 \, \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \, \mathbf{y}_{t-p} + u_t \tag{5}$$

- This structural form of the model is not directly observed: we do not see $\mathbf{u}_t$ and $\mathbf{A}_0$ is not available by regression
- We can try to recover the structural parameters (and shocks) from observable reduced form models
- Many ways of doing this, e.g. the Choleski decomposition we have seen, sign restrictions on contemporary effects, rather than zeros, sign or zero restrictions on long-run effects, implications of a fully specified model. See Canova 'Applied Macroeconomics'.

The SVAR

- The structural form is related to the observable reduced form by the relation between the strucutal and observable errors, e.g. $\mathbf{u}_t = \mathbf{H}\epsilon_t$

$$
\begin{aligned}
\mathbf{y}_t &= c + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \epsilon_t \\
&= c + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \mathbf{H}^{-1} u_t \\
\mathbf{H}\,\mathbf{y}_t &= \mathbf{H}\Big(c + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \mathbf{H}^{-1} u_t\Big) \\
\mathbf{A}_0\,\mathbf{y}_t &= \mathbf{a} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + u_t
\end{aligned}
$$

- The SVAR the dynamic system associated with the orthogonalised impulse responses identified by a rotation of the observable errors to produce the series of uncorrelated - orthogonal - shocks.

Variance Decomposition

- The Forecast Error Variance Decomposition attributes a proportion of the unpredictable compenent of each variable to each orthogonalised shock, and can be calculated for each horizon, h, of interest
- This follows from the conditional variance calculations earlier:

$$\mathbb{E}_t\left[[\mathbf{y}_{t+h}-\hat{\mathbf{y}}_{t+h}][\mathbf{y}_{t+h}-\hat{\mathbf{y}}_{t+h}]'\right]=\sum_{i=0}^{h-1}\mathbf{\Phi}^i\,\mathbf{\Sigma}(\mathbf{\Phi}^i)'$$

$$\begin{aligned} FEV[y_{j,t}(h)] &= \left[\sum_{i=0}^{h-1}\mathbf{\Phi}^i\,\mathbf{\Sigma}(\mathbf{\Phi}^i)'\right]_{jj} \\ &= \left[\sum_{i=0}^{h-1}\mathbf{\Phi}^i\,\mathbf{P}\mathbf{P}'(\mathbf{\Phi}^i)'\right]_{jj} \\ &= \left[\sum_{i=0}^{h-1}\mathbf{\Psi}_i^*(\mathbf{\Psi}_i^*)'\right]_{jj} \end{aligned}$$

Variance Decomposition

The (j,j) element of the forecast error variance is itself a sum of components from each variable

$$FEV[y_{j,t}(h)] = \Big[\sum_{i=0}^{h-1}\sum_{k=1}^{N}(\psi_{jk}^{*i})^2\Big]$$

So the proportion of the error variance in $y_{j,t+h}$ explained by $u_{m,t}$ is simply

$$v_{j,m}(h) = \frac{\sum_{i=0}^{h-1}(\psi_{jm}^{*i})^2}{\sum_{i=0}^{h-1}\sum_{k=1}^{n}(\psi_{jk}^{*i})^2}$$

Part II: Why a VAR ?

Plan for Part II

- In e.g. RAE we like you to think about endogeneity, causality and economics as well as correlation
- Time-series can seem a bit dry and technical c.f. microeconometrics
- Let's try to think about a few issues in econometric interpretation

- Review exogeneity and unbiasedness in OLS
- Endogeneity, predeterminedness and consistency in univariate time series
- Endogeneity in multivariate time series
- VARs are consistent
- VAR as an estimation technique for simultaneous equations model

Review of OLS assumptions

y_i is the dependent variabe. Stack obs $(y_1 ... y_N)$ in $\mathbf{y}$

$\mathbf{x}_i$ is a $k.1$ vector of regressors for unit i. Stack in

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix}$$

Regression equation:

$$\mathbf{y} = \mathbf{X}\beta + u$$

OLS solves $(\min \sum_i \hat{u}_i^2)$. We can do this by setting $\mathbf{X}'\hat{u} = 0$:

$$X'(\mathbf{y} - \mathbf{X}\hat{\beta}) = 0$$

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\,\mathbf{y}$$

Unbiasedness

For OLS to be BLUE we require exogeneity of the errors w.r.t regressors

$$\mathbb{E}[u|\mathbf{X}] = 0 \qquad \text{(A1)}$$

To see this consider:

$$\begin{aligned}
\mathbb{E}[\hat{\beta}|\mathbf{X}] &= \mathbb{E}\Big[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\,\mathbf{y}\,|\mathbf{X}\Big] \\
&= \mathbb{E}\Big[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\beta + u\right)|\mathbf{X}\Big] \\
&= \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbb{E}[u|\mathbf{X}] \\
&= \beta \text{ under (A1)}
\end{aligned}$$

Then $\mathbb{E}[\hat{\beta}] = \mathbb{E}[\mathbb{E}[\hat{\beta}|\mathbf{X}]] = \beta$ by Law of Iterated Expectations

(A1) requires that $\mathbb{E}[u_i\mathbf{x}_j] = 0$ for ALL i, j

Endoegeneity in univariate time series

Time series models will never satisfy this definition of exogeneity

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

In the notation of the previous slides

$$\mathbf{y} = \begin{bmatrix} y_2 \\ \vdots \\ y_T \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} y_1 \\ \vdots \\ y_{T-1} \end{bmatrix}$$

To see why, e.g.:

$$\mathbb{E}[\epsilon_t \mathbf{x}_{t+1}] = \mathbb{E}[\epsilon_t y_t] = \sigma^2 \neq 0$$

Predetermined variables and consistency

Where can we go from here? Notice

$$\begin{aligned}
\mathbb{E}[\epsilon_t \mathbf{x}_t] &= \mathbb{E}[\epsilon_t y_{t-1}] \\
&= \mathbb{E}[\epsilon_t \sum_{i=1}^{\infty} \phi^i \epsilon_{t-i}] \\
&= \sum_{i=1}^{\infty} \phi^i \mathbb{E}[\epsilon_t \epsilon_{t-i}] \\
&= 0
\end{aligned}$$

where the last line uses $\mathbb{E}[\epsilon_t \epsilon_s] = 0$, $t \neq s$.

Predetermined variables and consistency

This is enough to use LLN and associated theory to state the consistency of OLS estimator of $\hat{\phi}$

$$\hat{\phi} = \Big(\mathbf{X}'\mathbf{X}\Big)^{-1}\mathbf{X}'\,\mathbf{y}$$
$$= \Big(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\Big)^{-1}\Big(\sum_{t=1}^{T}\mathbf{x}_t\,\mathbf{y}_t\Big)$$

Under the LLN $\frac{1}{N}\sum_i x_i \to \mathbb{E}[x]$ as $N \to \infty$, or $\operatorname{plim}\bar{x} = \mathbb{E}[x]$

$$\operatorname{plim}\hat{\phi} = \operatorname{plim}\Big\{\Big(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\Big)^{-1}\Big(\sum_{t=1}^{T}\mathbf{x}_t(\mathbf{x}_t'\phi + \epsilon_t)\Big)\Big\}$$
$$= \phi + \operatorname{plim}\Big\{\Big(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\Big)^{-1}\Big(\sum_{t=1}^{T}\mathbf{x}_t\epsilon_t\Big)\Big\}$$

Predetermined variables and consistency

The regression meets techincal assumptions such that

$$\operatorname{plim}(ab) = \operatorname{plim} a \cdot \operatorname{plim} b$$
$$\operatorname{plim}(a^{-1}) = (\operatorname{plim} a)^{-1}$$

So we have

$$\begin{aligned}
\operatorname{plim} \hat{\phi} &= \phi + \operatorname{plim} \left\{ \Big(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \Big)^{-1} \Big(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \epsilon_t \Big) \right\} \\
&= \phi + \big[\mathbb{E}[\mathbf{x}_t \mathbf{x}_t'] \big]^{-1} \mathbb{E}[\mathbf{x}_t \epsilon_t] \\
&= \phi
\end{aligned}$$

by the fact $\mathbb{E}[\mathbf{x}_t \epsilon_t] = 0$

What was that?

We just said that the OLS estimator $\hat{\phi}$, or $\hat{\phi}_T$ (to emphasise dependence of estimator on the sample size) has good large sample properties such that

$$\hat{\phi}_T \to \phi$$

as $T \to \infty$, provided that $\mathbb{E}[\epsilon_t \epsilon_{t-k}] = 0 \ \forall k$.

Those few slides were a special treat for budding econometricians and statisticians in the room

The rest of us (e.g. applied macroeconomists!) should remember

- It's crucial to get your errors serially uncorrelated
- Sample size matters in time-series, as we're relying on consistency, not unbiasedness

Endogeneity in multivariate time sereis

What about the regression

$$y_t = \phi y_{t-1} + \beta z_t + \epsilon_t$$

We need to have exogeneity of z_t w.r.t. ϵ_t, $\mathbb{E}[z_t \epsilon_t] = 0$

Two possible problems

- Simultaneous equations
- Correlated errors

Simultaneous equations

$$
\begin{aligned}
y_t &= \phi_y y_{t-1} + \beta z_t + \epsilon_{yt} \\
z_t &= \phi_z z_{t-1} + \gamma y_t + \epsilon_{zt}
\end{aligned}
$$

$$y_t = \phi_y y_{t-1} + \beta(\phi_z z_{t-1} + \gamma y_t + \epsilon_{zt}) + \epsilon_{yt}$$

But ϵ_{yt} is certainly correlated with y_t so $\mathbb{E}[z_t \epsilon_t] = \gamma \mathbb{E}[y_t \epsilon_{yt}] \neq 0$

For simultaneous equations, the standard regression framework will not do

Correlated errors

What if $\mathbb{E}[\epsilon_{yt}\epsilon_{zt}] = \sigma_{yz}$?

$$\begin{aligned} y_t &= \phi_y y_{t-1} + \beta z_t + \epsilon_{yt} \\ z_t &= \phi_z z_{t-1} + \epsilon_{zt} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[z_t \epsilon_{yt}] &= \mathbb{E}[(\phi_z z_{t-1} + \epsilon_{zt})\epsilon_{yt}] \\ &= \sigma_{yz} \neq 0 \end{aligned}$$

Once again, the z_t variable will fail the exogeneity requirement

What about macroeconomics

We want to do regressions about e.g. output, prices, interest rates

Clearly this is a simultaneous system

Clearly the date t errors will be correlated with each other

This is endemic in macroeconomics

It is pretty much impossible to call any variable (other than constant, trend, seasonals) exogeneous in macroeconomics

The VAR is one response to this

- Treat ALL variables as endogenous and model them as a single system
- $\mathbf{y}_t = c + \mathbf{F}\,\mathbf{y}_{t-1} + \epsilon_t$
- All regressors are predetermined, i.e. dated $t-1$ or earlier

Simultaneous equations models

- A fully specified model of the dynamic system would contain information about contemporary correlations between (y_{it}, y_{jt}) and the dynamic relations between the variables.
- Then we could write

$$A_0\,\mathbf{y}_t = \tilde{c} + A_1\,\mathbf{y}_{t-1} + \cdots + A_p\,\mathbf{y}_{t-p} + v_t$$

  with $\mathbb{E}[v_t v_t'] = \mathbf{D}$, a diagonal matrix
- A_0 captures the information about contemporary correlations between variables
- But this is not a regression equation: non-zero off-diagonal elements of A_0 imply endogeneity of variables in the system

Simultaneous equations models

- However, we can always estimate the *Reduced Form*

$$y_t = A_0^{-1}\tilde{c} + A_0^{-1}A_1\,\mathbf{y}_{t-1} + \cdots + A_0^{-1}A_p\,\mathbf{y}_{t-p} + A_0^{-1}v_t$$

- Which is just the VAR

$$y_t = c + \mathbf{\Phi}_1\,\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p\,\mathbf{y}_{t-p} + \epsilon_t$$

- The SVAR recovers the full simultaneous equations model by imposing $n(n+1)/2$ restrictions on the covariance matrix of the errors $v_t = \mathbf{H}\epsilon_t$, with $A_0 = \mathbf{H}$
- The VAR with the Choleski decomposition of $\mathbf{\Sigma}$ is a way of estimating the siumltaneous equations model

Inference in a VAR

Inference is conducted in the usual way for OLS estimators

- $\hat{\mathbf{\Sigma}}$ is estimated from $\frac{1}{T-p-k}\sum_{t=p}^{T}\hat{\epsilon}_t\hat{\epsilon}_t'$
- The estimated variance matrix of the data $\hat{\mathbf{Q}} = \frac{1}{T-p}\sum_{t=p}^{T}\mathbf{x}_t\mathbf{x}_t'$, where $\mathbf{x}_t = [1, \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}']'$
- The standard errors of the parameters of the i^{th} equation are sqrts of diagonal elements of $\hat{\sigma}_{ii}\hat{\mathbf{Q}}^{-1}$
- Allows us to form t and F tests in usual way - though usually use limiting χ^2 distribution for restrictions on multiple coefficients
  - E.g. the Granger causality tests discussed in Part I

Notes/Problems

Notes/Problems

7. Multivariate Cointegration: VECM models

Overview

In these lectures we will cover

- VAR to VECM
- Equilibrium and adjustment with n variables
- Estimating the number of cointegrating relations
- How do we estimate and test hypotheses in a VECM

What are we doing?

- Properties of the model
- Testing properties of data
- Estimation of model parameters
- Model selection, forecasting and evaluation
- Economic inference

VAR with I(1) data

- Let $\mathbf{y}_t$ be vector of $I(1)$ variables, $y_{i,t} \sim I(1)$
- Is it a good idea to estimate $\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{\Phi}_i \, \mathbf{y}_{t-i} + \epsilon_t$?
  - No: there may be spurious regression. Estimate VAR in differences
  - But: in Engle Granger we run a levels regression $y_t = \beta x_t + \epsilon_t$
  - Yes: if there is cointegration the VAR in levels will be fine
- The question is whether $\epsilon_t \sim I(0)$
- How do you know, is there a multivariate test of stationarity of a vector, ϵ_t? (not really)
- Estimate a version of the model which always has I(0) errors, and collapses to a VAR in differences if there is no cointegrating vector
- This representation of the model is called the Vector Error Correction Model

The Vector Error Correction Model

- The VECM always exists (Granger Representation Theorem)
- It is just a reparameterisation of the VAR process
- We can always estimate the VECM
- From the estimation of the VECM we can discover the number of cointegrating vectors in the system (Johansen Test)
- Then we either
  - Proceed with VECM estimation and identification questions
  - Estimate VAR in levels for e.g. IRF analysis
  - Estimate VAR in differences if there is no cointegration

VAR to VECM

For example, with $p = 2$ we can write

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{\Phi}_1 \, \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \, \mathbf{y}_{t-2} + \epsilon_t \\
&= \mathbf{\Phi}_1 \, \mathbf{y}_{t-1} \pm \mathbf{\Phi}_2 \, \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \, y_{t-2} + \epsilon_t \\
&= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2) \, \mathbf{y}_{t-1} - \mathbf{\Phi}_2 \, \Delta \, \mathbf{y}_{t-1} + \epsilon_t \\
\Delta \, \mathbf{y}_t &= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n) \, \mathbf{y}_{t-1} - \mathbf{\Phi}_2 \, \Delta \, \mathbf{y}_{t-1} + \epsilon_t
\end{aligned} \tag{6}
$$

VAR to VECM

For example, with $p = 2$ we can write

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{\Phi}_1\, \mathbf{y}_{t-1} + \mathbf{\Phi}_2\, \mathbf{y}_{t-2} + \epsilon_t \\
&= \mathbf{\Phi}_1\, \mathbf{y}_{t-1} \pm \mathbf{\Phi}_2\, \mathbf{y}_{t-1} + \mathbf{\Phi}_2\, y_{t-2} + \epsilon_t \\
&= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2)\, \mathbf{y}_{t-1} - \mathbf{\Phi}_2\, \Delta\, \mathbf{y}_{t-1} + \epsilon_t \\
\Delta\, \mathbf{y}_t &= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n)\, \mathbf{y}_{t-1} - \mathbf{\Phi}_2\, \Delta\, \mathbf{y}_{t-1} + \epsilon_t \qquad (6)
\end{aligned}$$

$$\underbrace{\Delta\, \mathbf{y}_t}_{I(0)} = (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n)\, \mathbf{y}_{t-1} - \mathbf{\Phi}_2\, \underbrace{\Delta\, \mathbf{y}_{t-1}}_{I(0)} + \epsilon_t$$

VAR to VECM

For example, with $p = 2$ we can write

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \epsilon_t \\
&= \mathbf{\Phi}_1 \mathbf{y}_{t-1} \pm \mathbf{\Phi}_2 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 y_{t-2} + \epsilon_t \\
&= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2) \mathbf{y}_{t-1} - \mathbf{\Phi}_2 \Delta \mathbf{y}_{t-1} + \epsilon_t \\
\Delta \mathbf{y}_t &= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n) \mathbf{y}_{t-1} - \mathbf{\Phi}_2 \Delta \mathbf{y}_{t-1} + \epsilon_t
\end{aligned} \tag{6}
$$

$$
\underbrace{\Delta \mathbf{y}_t}_{I(0)} = (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n) \mathbf{y}_{t-1} - \mathbf{\Phi}_2 \underbrace{\Delta \mathbf{y}_{t-1}}_{I(0)} + \epsilon_t
$$

Which makes sense if also have

$$
\Delta \mathbf{y}_t = \underbrace{(\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n) \mathbf{y}_{t-1}}_{I(0)} - \mathbf{\Phi}_2 \Delta \mathbf{y}_{t-1} + \underbrace{\epsilon_t}_{I(0)}
$$

VAR to VECM

There are two possibilities:

- If there is no cointegration $(\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n) = \mathbf{0}$ and so ϵ_t is stationary because $\Delta\,\mathbf{y}_i$ are stationary
- Alternatively $(\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n)\,\mathbf{y}_{t-1} \sim I(0)$ as the matrix $(\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n)$ picks out the cointegrating relationships between the $\mathbf{y}_t$ variables, producing an I(0) vector when applied to the lagged level term
- The rank of the matrix $(\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n) = r < n$ tells us the number of cointegrating vectors

Stationary representation of the data

- If the the linear combination

$$(\mathbf{\Phi}_1 + \mathbf{\Phi}_2 - \mathbf{I}_n)\,\mathbf{y}_{t-1} \sim I(0)$$

  with $r > 0$ then we have found a model of how $\Delta\,\mathbf{y}_t$ adjusts towards the long-run equilibrium relations in the system
- This is the Vector Error Correction Model

$$\Delta\,\mathbf{y}_t = \mathbf{\Pi}\,\mathbf{y}_{t-1} + \mathbf{\Gamma_1}\Delta\,\mathbf{y}_{t-1} + \epsilon_t$$

  where definitions of $\mathbf{\Pi}$ and $\mathbf{\Gamma}$ follow from (6)
- Each row of $\mathbf{\Pi}$ picks out linear combinations of $\mathbf{y}_t$ which are stationary *and* estimates how the model adjusts towards these long-run equilibria
- A linear combination of $I(1)$ variables which is $I(0)$ is called a cointegrating vector.

VECM in general

In general we can reparameterise the VAR(p)

$$\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \epsilon_t$$

$$\Delta \mathbf{y}_t = \mathbf{\Pi} \mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta \mathbf{y}_{t-i} + \epsilon_t$$

where

$$\mathbf{\Pi} = \sum_{i=1}^{p} \mathbf{\Phi}_i - \mathbf{I}_n$$

$$\mathbf{\Gamma}_i = - \sum_{j=i+1}^{p} \mathbf{\Phi}_j$$

Make sure you understand for $p > 2$

Equilibrium and adjustment with n variables

The matrix $\mathbf{\Pi}$ contains information about

- The number of cointegrating vectors
- The cointegrating vectors if any exist
- The speed of adjustment towards these long-run equilibria

The rest of this lecture is about $\mathbf{\Pi}$

Equilibrium and adjustment: 3 possibilities

The number of cointegrating vectors is given by the rank of $\mathbf{\Pi}$

- The rank of $\mathbf{\Pi} = 0$
  - There are no cointegrating relations
- The rank of $\mathbf{\Pi} = n$, the number of variables
  - The variables are all $I(0)$
- The rank of $\mathbf{\Pi} = r$, $0 < r < n$
  - There are r cointegrating relations among the n $I(1)$ variables

Rank($\mathbf{\Pi}$) = 0

This is the simplest case:

If $rk(\mathbf{\Pi}) = 0$ then $\mathbf{\Pi}$ is the null matrix $\mathbf{\Pi}\,\mathbf{x} = 0 \;\forall \mathbf{x}$

There is no conitegration

The VECM reduces to

$$\Delta\,\mathbf{y}_t = \sum_{i=1}^{p-1} \Gamma_i \Delta\,\mathbf{y}_{t-i} + \epsilon_t$$

The VAR in differences is the appropriate model

There may be a relation between the growth rates of the variables, but there is no long-run relationship between the levels.

Rank($\mathbf{\Pi}$) $= n$

- In this case the $\mathbf{y}_t$ is a vetor of $I(0)$ variables to see this, e.g. $n = 2$

$$\mathbf{\Pi}\,\mathbf{y}_t = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \sim I(0)$$

$$z_{1t} = \pi_{11} y_{1t} + \pi_{12} y_{2t}$$

$$z_{2t} = \pi_{21} y_{1t} + \pi_{22} y_{2t}$$

$$\Rightarrow \Big(-\frac{\pi_{21}\pi_{12}}{\pi_{11}} + \pi_{22} \Big) y_{2t} = z_{2t} - \frac{\pi_{21}}{\pi_{11}} z_{1t} \sim I(0)$$

$$\Rightarrow y_{2t} \sim I(0)$$

as linear combinations of I(0) variables are also I(0).

Similar argument shows $y_{1t} \sim I(0)$.

I.e. two stationary combinations of two variables $\Rightarrow$ both variables are stationary

Rank($\mathbf{\Pi}$) $= r$

When $0 < r < n$ we have cointegration

- There are r linearly independent vectors in $\mathbf{\Pi}$
- It is possible to write

$$\mathbf{\Pi} = \alpha\beta'$$

  where α is an $n.r$ matrix and β' is $r.n$

$$\begin{aligned} \mathbf{\Pi}\,\mathbf{y}_t &= \alpha\beta'\,\mathbf{y}_t \sim I(0) \\ &= \alpha z_t \end{aligned}$$

- The variables $z_t = \beta'\,\mathbf{y}_t \sim I(0)$ are the equilibrim errors, $\mathbb{E}z_t = 0$
- The r vectors in β are the cointegrating vectors
- The coefficients α tell us how quickly each variable adjusts to disequilibrium in each long run relation (i.e. to $z_{i,t-1} \neq 0$, $i = 1 \ldots r$)

Example when n=p=2, r=1

In this case we have equations for the two **y** variables as

$$\Delta y_{1t} = \alpha_1(\beta_1 y_{1,t-1} + \beta_2 y_{2,t-1}) + \gamma_{11}\Delta y_{1,t-1} + \gamma_{12}\Delta y_{2,t-1} + \epsilon_{1,t}$$
$$\Delta y_{2,t} = \alpha_2(\beta_1 y_{1,t-1} + \beta_2 y_{2,t-1}) + \gamma_{21}\Delta y_{1,t-1} + \gamma_{22}\Delta y_{2,t-1} + \epsilon_{t}$$

We need at least one of α_i to be significantly negative:

- There must be some adjustment in the system so that changes (on l.h.s.) repsond to (correct to) errors in the preivious period (on r.h.s), i.e. to $\beta_1 y_{1,t-1} \neq -\beta_2 y_{2,t-1}$

The estimation problem

We want to estimate parameters of

$$\Delta \mathbf{y}_t = \mathbf{\Pi} \mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta \mathbf{y}_{t-i} + \epsilon_t$$

subject to the constraint that $0 < rank(\mathbf{\Pi}) = r < n$

- How do we know what r is?
- How do we impose this restriction on our estimates of parameters?

The estimation problem

We want to estimate parameters of

$$\Delta \mathbf{y}_t = \mathbf{\Pi}\, \mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i\, \Delta\, \mathbf{y}_{t-i} + \epsilon_t$$

subject to the constraint that $0 < rank(\mathbf{\Pi}) = r < n$

- How do we know what r is?
- How do we impose this restriction on our estimates of parameters?

- Use MLE under the hypothesis $r = 0, 1, \ldots n$
- Compare value of test statistics to choose best r
- Estimate α, β, $\mathbf{\Gamma}_i$ under this hypothesis
- This is known as the Johansen procedure e.g. Johansen (1995) OUP

Test statistics

- r is the number of linearly independent vectors in $\mathbf{\Pi}$
- It is the number of eigenvectors in $\mathbf{\Pi}$
- Each eigenvector has associated eigenvalue $\Rightarrow$ test the number of non-zero eigenvalues

Max eigenvalue		Trace	
H_0	H_1	H_0	H_1
$r = 0$	$r = 1$	$r = 0$	$r \geq 1$
$r \leq 1$	$r = 2$	$r \leq 1$	$r \geq 2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r \leq n-1$	$r = n$	$r \leq n-1$	$r = n$

- Start from H_0: $r = 0$, and keep testing until fail to reject H_0
- If different results Trace test is generally preferred

Test statistics

Critical values depend on

- The number of deterministic terms - trends, constants etc
- How these terms enter the model
  - (i) No deterministic terms: no growth rates, no constants
  - (ii) Restricted constant: zero growth rates, but a constant in cointegrating eqns
  - (iii) Unrestricted constant: non-zero growth rates
  - (iv) Restricted trend: the variables cointegrate but this ratio changes deterministically with time
  - (v) Unrestricted trend and constant $\Rightarrow$ quadratic time trend: bad idea
- Options ii-iv are typical for economic data

Deterministic components

- No deterministic components: no growth no constants

$$\Delta y_t = \Pi y_{t-1} + \dots$$

- Restricted intercept : no growth, constant in cointegrating equations

$$\Delta y_t = \begin{bmatrix} \Pi & \mu \end{bmatrix} \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \dots$$

- Unrestricted intercept: growth rates and constant in cointegrating equations

$$\Delta y_t = \delta + \begin{bmatrix} \Pi & \mu \end{bmatrix} \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \dots$$

Deterministic components

- Restricted trend: growth rates and a trend in the cointegrating equations

$$\Delta y_t = \delta + \begin{bmatrix} \Pi & \gamma & \mu \end{bmatrix} \begin{bmatrix} y_{t-1} \\ t \\ 1 \end{bmatrix} + \dots$$

- Unrestricted trend: quadratic growth rates and a trend in cointegrating equations

$$\Delta y_t = \delta + \gamma_0 t + \begin{bmatrix} \Pi & \gamma_1 & \mu \end{bmatrix} \begin{bmatrix} y_{t-1} \\ t \\ 1 \end{bmatrix} + \dots$$

Identification problem

Re-consider the example

$$\Delta y_{1t} = \alpha_1(\beta_1 y_{1,t-1} + \beta_2 y_{2,t-1}) + \gamma_{11}\Delta y_{1,t-1} + \gamma_{12}\Delta y_{2,t-1} + \epsilon_{1,t}$$
$$\Delta y_{2,t} = \alpha_2(\beta_1 y_{1,t-1} + \beta_2 y_{2,t-1}) + \gamma_{21}\Delta y_{1,t-1} + \gamma_{22}\Delta y_{2,t-1} + \epsilon_t$$

$$\Delta y_{1t} = \frac{\alpha_1}{\lambda}(\lambda\beta_1 y_{1,t-1} + \lambda\beta_2 y_{2,t-1}) + \gamma_{11}\Delta y_{1,t-1} + \gamma_{12}\Delta y_{2,t-1} + \epsilon_{1,t}$$
$$\Delta y_{2,t} = \frac{\alpha_2}{\lambda}(\lambda\beta_1 y_{1,t-1} + \lambda\beta_2 y_{2,t-1}) + \gamma_{21}\Delta y_{1,t-1} + \gamma_{22}\Delta y_{2,t-1} + \epsilon_t$$

$$\Delta y_{1t} = \alpha_1^*(\beta_1^* y_{1,t-1} + \beta_2^* y_{2,t-1}) + \gamma_{11}\Delta y_{1,t-1} + \gamma_{12}\Delta y_{2,t-1} + \epsilon_{1,t}$$
$$\Delta y_{2,t} = \alpha_2^*(\beta_1^* y_{1,t-1} + \beta_2^* y_{2,t-1}) + \gamma_{21}\Delta y_{1,t-1} + \gamma_{22}\Delta y_{2,t-1} + \epsilon_t$$

Yet the economic interpretation of $[\beta_1, \beta_2]$ may be different from $[\beta_1^*, \beta_2^*]$

Identification problem

This arises because $\mathbf{\Pi}$ is rank deficient, therefore we can write

$$\begin{aligned}\mathbf{\Pi} &= \alpha\beta' \\ &= \alpha\Lambda\Lambda^{-1}\beta' \\ &= \alpha^*(\beta^*)'\end{aligned}$$

So the data identifies $\mathbf{\Pi}$, but NOT the decomposition into interpretable cointegrating vectors β' and associated adjustment coefficients α

- Notice Λ is an $r.r$ matrix, i.e. it has r^2 elements
- We must choose r^2 restrictions in order to identify β'
- This is r restrictions per cointegrating equation
- Or, one normalisation restriction and $r - 1$ other restrictions per cointegrating equation
- These restrictions come from economic theory

An example

Suppose $n = 4$, $r = 2$ and

$$\mathbf{y}_t = \begin{bmatrix} rm_t & i_t & r_t^d & r_t^b \end{bmatrix}'$$

where variables are the real money, income, deposit rate and bond rate.

There are 2 cointegrating equations so we have

$$\beta' = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{bmatrix}$$

We need $r = 2$ restrictions per cointegrating vector, 1 normalisation and one other restriction

Just-Identification

Theory suggests long run money demand moves 1 for 1 with income, so normalise the first cointegrating relation on *rm*; theory then suggests $\beta_{12} = -1$

$$\beta' = \begin{bmatrix} 1 & -1 & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{bmatrix}$$

Make the second equation a long run relation for the spread of deposit rate over the bond rate

$$\beta' = \begin{bmatrix} 1 & -1 & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & 1 & -1 \end{bmatrix}$$

β is now *exactly identified*

This does not change $\mathbf{\Pi}$, it does not change the likelihood, it cannot be tested

Over-identification

What if you have a theory that *also* tells you that the bond rate does not enter the long-run money demand?

$$(\beta^*)' = \begin{bmatrix} 1 & -1 & \beta_{13} & 0 \\ \beta_{21} & \beta_{22} & 1 & -1 \end{bmatrix}$$

- This is an overidentifying restriction
- It will change $\mathbf{\Pi}^*$ relative to previous $\mathbf{\Pi}$
- This is a testable restriction - likelihood ratio test: $-2(ll(r) - ll(u)) \sim \chi^2_1$ (as there was 1 overidentifying restriction in this case)

Notes/Problems

Notes/Problems