

## CHAPTER 5

# VECTOR AUTOREGRESSIVE AND VECTOR ERROR CORRECTION MODELS

1

# MOTIVATION

- Univariate models are very useful instruments in the economic analysis
  - To understand the properties of the series as a first stage before specifying a more complex econometric model. For example, generally a stationary model cannot be specified when all series in the model have trend (we'll go back later on to this point).
  - ARIMA models are very useful for forecasting purposes as they are parsimonious and easily adaptable.
- However, ARIMA models are not useful when the purpose of the analysis is to study the interaction among a group of economic variables. Frequently, this is the purpose of an economist and econometric models are required for this purpose.

# MOTIVATION

- The first step to specify an econometric model is to specify the variables to included. This is determined by economic theory. For example:
  - Reaction functions in the production decisions of two rival firms.
  - The effect of some macroeconomic policies on sales.
  - Aggregate demand.

# MOTIVATION

- The analysis of the relation among several variables usually is not static but dynamic.
  - This is similar to what happens when a person who is having a shower presses the button of the hot water. The temperature normally does not adjust instantaneously but only after a few seconds.
  - The same happen with the relation among different economic variables. For example, a movement in the official interest rate by the monetary authority only has effect on inflation after a long period of delay (sometimes more than a year).

# MOTIVATION

- Besides, when we study the relation among a group of economic variables it is not clear whether the causal relationship goes in one direction or all the variables cause each other.
  - For example, to study the relation between interest rates and business cycle.
  - An equation that has a measure of business cycle as dependent variable and short run interest rates as the explanatory variable is known as the aggregate demand function.
  - However, we can have the inverse relation when interest rate is the dependent variable and measures of business cycle and inflation the explanatory variables. This is known as the monetary policy rule (Taylor rule).

# STRUCTURE

- 4.1. Stationary VAR(p). Formulation. Time dependency. Granger causality. Contemporaneous dependency.
- 4.2. Economic analysis with VAR models.
- 4.3. Definition of Cointegration and tests.
- 4.4. VAR models and vector equilibrium correction (VeqC) models.
- 4.5. Cointegration test and specification of VeqC models.
- 4.6. Estimation
- 4.7. Testing long run relations
- 4.8. Testing short run relations.
- 4.9. Two examples.

# STATIONARY VAR MODELS

- We denote by  $X_t$  a vector of dimension  $n$  and we define the following VAR( $p$ ) model

$$X_t = A_0 + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + e_t$$

where  $A_0$  is a  $n \times 1$  vector and  $A_1, \dots, A_p$  are  $n \times n$  matrices of parameters and  $e_t$  is a  $(n \times 1)$  vector of white noise variables.

# STATIONARY VAR MODELS

- For example, a VAR(1) for two variables  $y_t$   $z_t$

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t}$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t}$$

# STATIONARY VAR MODELS.

$$E \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$E(e_{1t}, e_{1t+i}) = 0 \quad \forall i \neq 0$$

$$E(e_{2t}, e_{2t+i}) = 0 \quad \forall i \neq 0$$

$$\text{Var} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

# STATIONARY VAR MODELS.

- ◉ It is easy to see that VAR models are a multivariate extension of AR models.
- ◉ In the same way, ARMA models can be generalized to VARMA(p,q) models for the univariate case. However, in practice most papers only use the autoregressive part (VAR models) for the following 3 reasons:
  - As we have seen, every finite moving average model can be approximated by an autoregressive model that includes a sufficient number of lags.
  - Using a moving average component complicate the estimation a lot while the estimation of autoregressive models by conditional maximum likelihood coincides with the ordinary least square procedure.
  - The economic interpretation of moving average components in an econometric model is not intuitive.

# STATIONARY VAR MODELS.

- In the AR(1) process,

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

the stability condition is that the absolute value of  $a_1$  is less than unity. The same happens for VAR models. If we iterate a VAR(1) model, we obtain:

# STATIONARY VAR MODELS.

$$X_t = A_0 + A_1(A_0 + A_1X_{t-2} + e_{t-1}) + e_t = (I + A_1)A_0 + A_1^2X_{t-2} + A_1e_{t-1} + e_t$$

After n iterations

$$X_t = (I + A_1 + \dots + A_1^n)A_0 + \sum_{i=0}^n A_1^i e_{t-i} + A_1^{n+1} X_{t-n-1}$$

# STATIONARY VAR MODELS.

- It is clear that convergence requires

$$\lim_{n \rightarrow \infty} A_1^n = 0$$

Assuming process is stationary, the solution of  $x_t$  can be written as

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \mathbf{A}_1^i \mathbf{e}_{t-i}$$

# STATIONARY VAR MODELS.

Generally, a VAR(p) model

$$X_t = A_0 + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + e_t$$

is stationary if the solution of

$$\left| I - A_1 x - A_2 x^2 - \dots - A_p x^p \right| = 0$$

is outside the unit circle.

# STATIONARY VAR MODELS.

Or alternatively, if the solution of

$$\left| \lambda^p - A_1 \lambda^{p-1} - A_2 \lambda^{p-2} - \dots - A_p x^p \right| = 0$$

is inside of the unit circle.

# Stationary VAR models.

## EXAMPLE

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 0.3 & 0.6 \\ 0.1 & 0.5 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

$$\left| \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} - \begin{pmatrix} 0.3 & 0.6 \\ 0.1 & 0.5 \end{pmatrix} \right| = 0$$

$$(\lambda - 0.3)(\lambda - 0.5) - 0.06 = 0$$

$$\lambda \approx \begin{cases} 0.66 \\ 0.14 \end{cases}$$

# Stationary VAR models.

## VAR MODELS WITH EXOGENOUS VARIABLES

- The aim of the exogeneity analysis is to reduce the number of equations in the econometric system.
- Therefore, knowing if a variable is exogenous is a fundamental issue for the specification and estimation of the model.
- There are three types of exogeneity.

# Stationary VAR models.

- ◉ Generally, we consider a vector of variables  $w_t$  that can be split into two different groups  $(y_t, z_t)$
- ◉ Then:

$$f(Y_t, Z_t / Y_1^{t-1}, Z_1^{t-1}, \lambda) = f(Y_t / Y_1^{t-1}, Z_1^t, \lambda_1) f(Z_t / Y_1^{t-1}, Z_1^{t-1}, \lambda_2)$$

- Weak exogeneity: when parameters  $\lambda_1$  y  $\lambda_2$  are independent and do not have common elements. Besides, parameters to be estimated only depend on  $\lambda_1$ .

# Stationary VAR models.

- For strong exogeneity, besides the 2 former conditions, variables should not Granger cause each other. In this case

$$f(\mathbf{Y}_t, \mathbf{Z}_t / \mathbf{Y}_1^{t-1}, \mathbf{Z}_1^{t-1}, \lambda) = f(\mathbf{Y}_t / \mathbf{Y}_1^{t-1}, \mathbf{Z}_1^t, \lambda_1) f(\mathbf{Z}_t / \mathbf{Z}_1^{t-1}, \lambda_2)$$

This topic is discussed in detail by Engle, Hendry y Richard (1983).

# Stationary VAR models.

## EXOGENEITY

- When we have an econometric system, the interest is knowing which variables are exogenous, weakly exogenous and endogenous.

# Stationary VAR models.

## GRANGER CAUSALITY

- In a bivariate system, variable  $w_{1t}$  does not Granger cause  $w_{2t}$  if for all  $s > 0$  the mean square forecast error (MSFE) of  $w_{2t+s}$  given  $(w_{21}, w_{22}, \dots, w_{2t})$  is the same as the mean square forecast error (ECM) of  $w_{2t+s}$  given  $(w_{11}, w_{12}, \dots, w_{1t}, w_{21}, w_{22}, \dots, w_{2t})$ .

# Stationary VAR models.

- In order to run the Granger test, we estimate the following equation

$$w_{2t} = c + \alpha_1 w_{2t-1} + \dots + \alpha_p w_{2t-p} + \beta_1 w_{1t-1} + \dots + \beta_p w_{1t-p} + \varepsilon_t$$

and then we test

$$H_0 : \beta_1 = \dots = \beta_p = 0$$

E-views include an automatic procedure for this test.

# Stationary VAR models.

- Important!!! Granger causality does not imply economic causality. Granger causality is an indication of how useful a variable is to forecast another variable. This concept is very useful to forecast.

# Stationary VAR models.

- In order to understand this, assume that we have two variables: the rainfall in a certain region and the number of tourists who arrive to that region. If tourists have information on future whether in the region they will arrive in more number when the weather forecast is going to be good. Therefore, if we run a Granger causality test we probably reject that the number of tourists does not cause rainfall. This only means that tourists use the information available to anticipate the rainfall.

## Stationary VAR models.

- A particularly interesting case is when only one variable is endogenous and the rest are exogenous. In this case, we can efficiently focus our analysis in a single equation.

For example:

\* Commodity prices use to be excellent indicators of future inflationary pressures. In this case, it is worth to focus our analysis in a single equation that has inflation as the dependent variable and some commodity prices as explanatory variables.

# Economic analysis with VAR models.

How do you estimate the VAR?

- (It can be proved that) there are no gains to methods more complex than OLS, provided that each equation has the same set of right hand side variables
  - So, you could estimate this in Excel
  - Generally, you want to produce ancillaries
    - A specialized time series package like RATS, TSP, or E-Views is worthwhile for this

# ECONOMIC ANALYSIS WITH VAR MODELS.

What do the estimates of the VAR look like?

- You don't care
  - Personally, I rarely if ever even look at them

# ECONOMIC ANALYSIS WITH VAR MODELS

How is that justified?

- When you estimate a parameter in a regression, you estimate two things
  - The parameter itself
  - The standard error of the parameter
- Omitting a relevant variable from the regression biases the parameter and standard error estimates
  - You can't easily predict which way
- Adding an irrelevant variable from the regression biases the standard error estimate (upward)
  - But....the parameter estimate is fine



# HOW IS THAT JUSTIFIED (CONT'D)?

- With a VAR, when in doubt, you add extra lags to the right hand side
  - This make sure that you don't omit anything
    - So, your parameter estimates are fine
  - However, you almost certainly included too much
    - So, your standard errors go through the roof
    - As a result, your  $t$ -statistics are likely to indicate that your parameters are insignificant



# IF YOU'RE NOT INTERESTED IN THE SIGNIFICANCE OF THE PARAMETERS, WHAT IS THE POINT OF ESTIMATING A VAR?

- VAR's can be re-expressed as ancillaries
  - Impulse response functions
  - (Forecast error) variance decompositions



# WHY DO WE NEED VAR ANCILLARIES?

- There is a lot more going on in a simple VAR system than meets the eye
  - $x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \text{error}$
  - $y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 y_{t-1} + \text{error}$
- Suppose  $y$  changes at  $t-1$ 
  - Then  $x$  and  $y$  change at  $t$
  - Both of which will cause  $x$  and  $y$  to change again at  $t+1$
- This process could continue forever, so you need a way to sort those effects out and organize them



# THE MATH – PAGE 1

- Write the system more specifically
  - $x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \varepsilon_t$
  - $y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 y_{t-1} + \eta_t$
- Note that you can “backshift” the equations
  - $x_{t-1} = \alpha_0 + \alpha_1 x_{t-2} + \alpha_2 y_{t-2} + \varepsilon_{t-1}$
  - $y_{t-1} = \beta_0 + \beta_1 x_{t-2} + \beta_2 y_{t-2} + \eta_{t-1}$



## THE MATH – PAGE 2

- Now substitute the right hand sides of the backshifted equations for the right hand side variables in the original equations to get:
  - $x_t = \alpha_0 + \alpha_1[\alpha_0 + \alpha_1 x_{t-2} + \alpha_2 y_{t-2} + \varepsilon_{t-1}] + \alpha_2[\beta_0 + \beta_1 x_{t-2} + \beta_2 y_{t-2} + \eta_{t-1}] + \varepsilon_t$
  - $y_t = \beta_0 + \beta_1[\alpha_0 + \alpha_1 x_{t-2} + \alpha_2 y_{t-2} + \varepsilon_{t-1}] + \beta_2[\beta_0 + \beta_1 x_{t-2} + \beta_2 y_{t-2} + \eta_{t-1}] + \eta_t$



## THE MATH – PAGE 3

- These equations are a mess, but we can gather terms to get:
  - $x_t = [\alpha_0 + \alpha_1\alpha_0 + \alpha_2\beta_0] + [(\alpha_1)^2 + \alpha_2\beta_1]x_{t-2} + [\alpha_1\alpha_2 + \alpha_2\beta_2]y_{t-2} + [\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\eta_{t-1}]$
  - $y_t = [\beta_0 + \beta_1\alpha_0 + \beta_2\beta_0] + [(\beta_1\alpha_1 + \beta_1\beta_2)]x_{t-2} + [\beta_1\alpha_2 + (\beta_2)^2]y_{t-2} + [\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\eta_{t-1}]$
- This is still a mess, but the essential point is that each variable still depends on lags of both variables, and a more complex set of errors



## THE MATH – PAGE 4

- If we kept backshifting each equation and substituting back in, we'd ultimately get equations that looked like this:
  - $x_t = \text{constant} + \gamma_x x_{t-n} + \gamma_y y_{t-n} + \text{lots of errors}$
  - $y_t = \text{constant} + \delta_x x_{t-n} + \delta_y y_{t-n} + \text{lots of errors}$
- Note that the  $\gamma$ 's and  $\delta$ 's, as well as the errors would be big functions of all of the  $\alpha$ 's and  $\beta$ 's from the original equations



# HOW DO WE SORT OUT WHAT'S GOING ON HERE?

- One result that you can count on is that most of the  $\alpha$ 's and  $\beta$ 's will be less than one in absolute value
  - Only unstable processes will have a lot of  $\alpha$ 's and  $\beta$ 's that are outside of this range – and we don't usually think of our world as unstable
- This is important because:
  - The  $\gamma$ 's and  $\delta$ 's are composed of products of  $\alpha$ 's and  $\beta$ 's – which go to zero the more we backshift
  - The “lots of errors” are composed of sums of  $\alpha$ 's and  $\beta$ 's weighting the errors – which don't go to zero



# THE SIGNIFICANCE OF THE MATH

- If we backshift enough, each series can be shown to be equal to
  - A constant
    - Which is the mean of the variable
  - A (weighted) sum of past errors
    - These come from all variables
    - These are the shocks that buffet the variables



# WHAT DO WE DO WITH THIS RESULT?

- We construct two VAR ancillaries to summarize how and why a variable gets away from its mean
  - Impulse response functions
    - These trace out how typical shocks will affect a variable through time
  - Variance decompositions
    - Show which shocks are most important in explaining a variable through time



# WHAT'S AN IMPULSE RESPONSE FUNCTION?

- Recall the error term obtained for  $x_t$  on slide 34 (after one backshift and substitution had been made)
  - $\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \eta_{t-1}$
- The impulse response function is the pattern of how a shock affects  $x$  – it can be read off the coefficients
  - A shock to  $x$  (an  $\varepsilon$ ) affects  $x$  immediately, and continues to affect  $x$  next period (the weight,  $\alpha_1$  may amplify or diminish the shock), and stops affecting  $x$  after that
  - A shock to  $y$  (an  $\eta$ ), does not affect  $x$  at all right away, affects it with a weight of  $\alpha_2$  the next period, and stops affecting  $x$  after that



# WHAT'S A VARIANCE DECOMPOSITION?

- Once we're done backshifting and substituting, what's left is a constant plus errors
  - Any variance of the variable must come from those errors
    - But....the errors have a variance that we already know because it gets estimated when we run the regression
- Again, for  $x$  (after one backshift and substitution):
  - $\text{Var}(x) = E[(\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\eta_{t-1})(\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\eta_{t-1})]$
  - $\text{Var}(x) = (\sigma_\varepsilon)^2 + (\alpha_1)^2(\sigma_\varepsilon)^2 + (\alpha_2)^2(\sigma_\eta)^2$
- Note that the first term is from  $t$ , and the last two are not
  - So, 100% of the variance of  $x$  at  $t$  comes from shocks to  $x$  ( $\varepsilon$ 's)
  - However, the variance of  $x$  at  $t+1$  comes from 2 sources
    - $\{(\alpha_1)^2(\sigma_\varepsilon)^2/[(\alpha_1)^2(\sigma_\varepsilon)^2 + (\alpha_2)^2(\sigma_\eta)^2]\}$  from  $x$
    - $\{(\alpha_2)^2(\sigma_\eta)^2/[(\alpha_1)^2(\sigma_\varepsilon)^2 + (\alpha_2)^2(\sigma_\eta)^2]\}$  from  $y$



# REPORTING VAR ANCILLARIES

- Typically, the software produces a ton of numbers in tabular form when you ask for these
  - The numbers are rarely reported
- Generally, authors provide plots of both
  - An impulse response function graph shows you whether a shock to one variable has:
    - A positive or negative affect on another variable (or both)
    - An effect the strengthens or diminishes through time
  - A variance decomposition graph shows you how the sources of variation underlying a variables movements wax and wane through time



# WHAT'S THE BIGGEST PROBLEM WITH VAR ANCILLARIES IN PUBLISHED RESEARCH?

- The ancillaries are non-linear combinations of a large number of underlying parameter estimates
  - Unfortunately, parameters estimates are *point estimates*
    - They are correct with probability zero
  - So, all VAR ancillaries are also point estimates
- How do we get around this?
  - It isn't very hard, and most programs can produce confidence intervals for VAR ancillaries
- So ....
  - Many articles don't include these confidence intervals because they are very wide – indicating a lot of uncertainty in the results



# WHAT'S THE CATCH?

- At first glance, it seems like applying a VAR is nothing more than applying some (time consuming) arithmetic to plain old OLS regressions
- This isn't the case. All multi-variable estimation problems require the researcher to address something called the identification problem
  - Prior to VAR's (and still with other methods) this required solving a sophisticated linear algebra problem
    - The difficulty of this problem goes up geometrically with the size of the model you're working with
  - VAR's still require that the identification issue be addressed, but the question is couched in a form that is easier to tackle
    - The difficulty of this problem need not go up too quickly



# WHAT'S THE IDENTIFICATION PROBLEM?

- Consider a basic microeconomic situation
  - We don't observe demand and supply
  - What we do observe is a quantity sold and a price
    - This is just one point on the standard microeconomics graph
- At some other time, we may observe a different quantity sold at a different price
  - This again is just another point on the graph
- How did we get to that new point?
  - Did supply shift?
  - Did demand shift?
  - Did both shift?
- This is the identification problem



# HOW DO WE (CONCEPTUALLY) IDENTIFY A SUPPLY OR A DEMAND?

- This is actually pretty easy
  - If only one of the curves shifts, the equilibrium will move along the other curve – tracing it out
- In order to get only one curve to shift, it must be pushed by some variable that only affects that curve, and not the other one.  
For example:
  - Changes in personal income will cause demand to shift, but are often irrelevant to the firms supply decisions
  - Changes in input prices will cause supply to shift but are often irrelevant to the households demand decisions



# HOW DO WE (MATHEMATICALLY) IDENTIFY A SUPPLY AND A DEMAND?

- Write out an equation for each one. I assume that they each relates prices and quantities, along with two other (shift) variables R and S. For now, it is important to include both of those variables in both equations
  - D:  $P = a_0 + a_1Q + a_2R + a_3S + \text{demand error}$
  - S:  $P = b_0 + b_1Q + b_2R + b_3S + \text{supply error}$
- Identification amounts to saying that only one of R or S affects demand, and the other one affects supply. This amounts to the following restrictions:
  - $a_2 = b_3 = 0$ , or alternatively
  - $b_2 = a_3 = 0$
- Justifying restricting a whole bunch of parameters to zero before you even start running regressions makes this tough



# HOW DOES IDENTIFICATION DIFFER IN VAR'S? PART 1

- Suppose you are trying to get information about how 2 variables, Y and Z, behave. First, you would write down a system of 2 structural equations:
  - $Y_t = c_0 + c_1 Z_t + c_2 Y_{t-1} + c_3 Z_{t-1} + \mu_t$
  - $Z_t = d_0 + d_1 Y_t + d_2 Y_{t-1} + d_3 Z_{t-1} + v_t$
- These equations are similar to those on the previous slide – I just replaced R and S with past values of Y and Z
- These equations are structural in the sense that they contain contemporaneous values of both variables of interest *in each equation*
- Also, because we are claiming that these represent some underlying structure, we assume that the two errors are uncorrelated



# HOW DOES IDENTIFICATION DIFFER IN VAR'S? PART 2

- All multi-variable estimations require that the structural equations be estimated by first obtaining and estimating the systems reduced form equations
  - Reduced forms are what is meant in algebra when you solve equations – two equations can be solved for two variables, in this case  $y_t$  and  $z_t$ , in each case by eliminating the other variable from the right hand side to get:
    - $Y_t = e_0 + e_2 Y_{t-1} + e_3 Z_{t-1} + \text{a function of both errors}$
    - $Z_t = f_0 + f_2 Y_{t-1} + f_3 Z_{t-1} + \text{another function of both errors}$
  - The  $e$ 's and  $f$ 's will be some messy combination of the underlying  $c$ 's and  $d$ 's from the structural equations



# HOW DOES IDENTIFICATION DIFFER IN VAR'S? PART 3

- We now have the original structural system:
  - $Y_t = c_0 + c_1 Z_t + c_2 Y_{t-1} + c_3 Z_{t-1} + \mu_t$
  - $Z_t = d_0 + d_1 Y_t + d_2 Y_{t-1} + d_3 Z_{t-1} + v_t$ 
    - 10 things need to be estimated here: four c's, four d's and the variances of the two errors (recall that their covariance is zero)
- We also have the equivalent reduced form system:
  - $Y_t = e_0 + e_2 Y_{t-1} + e_3 Z_{t-1} + \text{a function of both errors}$
  - $Z_t = f_0 + f_2 Y_{t-1} + f_3 Z_{t-1} + \text{another function of both errors}$ 
    - When we estimate this we get 9 pieces of information about the 10 that we are trying to estimate above (three 3's, three f's, variances of two errors, and one covariance between the - now related - errors)



# HOW DOES IDENTIFICATION DIFFER IN VAR'S? PART 4

- An alternative way of thinking about identification is that we can only estimate as many structural parameters as we have pieces of information from the reduced forms
  - Thus, we have to eliminate one thing of interest in the structural system
    - This may seem somewhat egregious, but recall that in the economic example I gave that we had to restrict two parameters to zero – so we are already better off here!



# HOW DOES IDENTIFICATION DIFFER IN VAR'S? PART 5

- We can safely eliminate any of the ten parameters in the structural system – but we must eliminate some of them to achieve identification
- Here's where a VAR makes your life easier
  - Rather than constraining a parameter on two of the lags to zero, we constrain one of the parameters on the contemporaneous terms to zero
    - The former is tantamount to saying that particular variables from the past do not cause other variables today
    - The latter is saying something less egregious – that certain variables don't affect other ones right away. This is an easier thing to explain and justify.



# HOW DOES VAR IDENTIFICATION WORK IN PRACTICE?

- Identifying a VAR amounts to choosing an “ordering” for your variables
  - If you have  $n$  dependent variables, they can be rearranged into  $n!$  orders
  - The researchers job is to pick one of those orders
- What makes a good order?
  - An argument that one variable (say  $X$ ) is likely to affect some other variable (say  $Y$ ) before  $Y$  can feed back and affect  $X$



# AN EXAMPLE OF VAR IDENTIFICATION

- A common set of variables in a macroeconomic VAR includes output, money, prices, and interest rates (Y, M, P, and r)
  - There are 24 possible orderings
    - YMP<sub>r</sub>, YMrP, YPM<sub>r</sub>, YPrM, rPMY, and so on
- A plausible ordering would be M, r, Y, P
  - The Federal Reserve controls M, and isn't likely to respond quickly to the other variables
  - The Federal Reserve is trying to influence r
  - By influencing r, the Federal Reserve hopes to influence Y and P
    - Most first adjust quantities faster than prices, so I put Y before P



# HOW SENSITIVE ARE VAR'S TO ORDERING?

- This question doesn't have a good answer
  - There are big differences across the set of possible orderings, but a good researcher knows that most of those orderings aren't justifiable
- A good convention to go by is that if you have trouble figuring out which variable should precede and which should follow, it probably won't make much difference to the VAR ancillaries either



## 4.4. VAR AND VEQC MODELS.

- An unrestricted VAR model is specified as

$$X_t = A_0 + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + e_t \quad (4.1)$$

where  $A_0, A_1, \dots, A_p$  are  $n \times n$  matrices and  $e_t$  is a  $n \times 1$  vector of white noise processes.

$$e_t \sim (0, \Sigma_e)$$

**The stability condition of the process is**

$$\det(I_n - A_1 z - \dots - A_p z^p) = 0 \text{ for } |z| > 1$$

# VAR and VecqC models.

- If there is a unit root in the polynomial, we know for sure that not all the variables in the system are stationary.
- If some of the variables have a common trend, then a combination of  $I(1)$  variables generate a  $I(0)$  system.

# VAR and VecqC models.

## COINTEGRATION. DEFINITION

Let's consider the following VAR(1) model

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

The model is not stationary when  $|I - \Phi| = 0$  has a unit root. Let's assume that the two variables have the same integration order.

## TWO CASES (EXPLAINED WITH EXAMPLES)

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

$$\begin{vmatrix} 1 - 0.5x & -0.1x \\ 0 & 1 - x \end{vmatrix} = 0$$

$$(1 - 0.5x)(1 - x) = 0$$

$$0.5x^2 - 1.5x + 1 = 0$$

$$x = \frac{1.5 \pm \sqrt{1.5^2 - 2}}{1} = \begin{cases} 2 \\ 1 \end{cases}$$

# VAR and VecqC models.

- Both series have a unit root

$$w_{1t} = 0.5w_{1t-1} + 0.1w_{2t-1} + a_{1t}$$

$$w_{2t} = w_{2t-1} + a_{2t}$$

- However, there is only a unit root in the system.

# VAR and VecqC models.

## SECOND CASE

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} 0.5 & 0 \\ 0.1 & 1 \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

$$\begin{vmatrix} 1 - 0.5x & 0 \\ -0.1x & 1 - x \end{vmatrix} = 0$$

$$(1 - 0.5x)(1 - x) = 0$$

$$0.5x^2 - 1.5x + 1 = 0$$

$$x = \frac{1.5 \pm \sqrt{1.5^2 - 2}}{1} = \begin{cases} 2 \\ 1 \end{cases}$$

# VAR and VecqC models.

- In this case, only one of the two processes is nonstationary.

$$w_{1t} = 0.5w_{1t-1} + a_{1t}$$

$$w_{2t} = 0.1w_{t-1} + w_{2t-1} + a_{2t}$$

# VAR and VecqC models.

## QUESTION

- Why in the first example each series have a unit root however the whole system has a unit root?

# VAR and VecqC models.

## ANSWER

- Why in the first example each series have a unit root however the whole system has a unit root?
- Because the two processes share the same unit root.

# VAR and VecC models.

## COINTEGRATION

- Variables in vector  $(x_{1t}, x_{2t}, \dots, x_{nt})'$  are cointegrated of order  $d, b$ , denoted by  $CI(d, b)$  if
  - 1) Each variable is integrated of order  $d$ .
  - 2) Exist a vector  $\beta = (\beta_1, \beta_2, \dots, \beta_n)'$  such that the linear combination  $\beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt}$  is integrated of order  $(d-b)$ , where  $\beta$  is called the cointegration vector.

# VAR and VecqC models.

## FOUR IMPORTANT POINTS TO KEEP IN MIND:

- Cointegration refers to linear combination of variables. Non linear combination of variables that reduce the integration order may exist but they cannot be found with the econometric tests available.
- Cointegration vector is not unique. If  $(\beta_1, \beta_2, \dots, \beta_n)'$  is a cointegration vector then  $(\lambda\beta_1, \lambda\beta_2, \dots, \lambda\beta_n)'$  is also a cointegration vector. Generally, the coefficient of one of the variables is normalized to unity, ( $\lambda=1/\beta_1$ ).
- All variables must be integrated of the same order. If variables are integrated of different orders, they cannot be cointegrated.
- If  $x_t$  have  $n$  components it can have at most  $n-1$  cointegration vectors. Clearly, if  $x_t$  have 2 variables, it could be at most 1 cointegration vector.

# VAR and VecqC models.

An important branch of the literature focuses on the case where each variable has one unit root. The reason is that this is the most relevant case in reality. However, cointegration among  $I(2)$  series may exist or even seasonal cointegration.

# VAR and VecqC models.

- An unrestricted VAR model (model (4.1)) is general enough to accommodate variables with stochastic trends, however it is not the most suitable type of model if the interest centers on the cointegration relations because they do not appear explicitly. A vector equilibrium correction (VecqC) model expresses the adjustment of variables to the equilibrium (this type of model can be expressed whenever variables are cointegrated).

$$\begin{aligned}\Delta Y_t &= \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots \\ &+ \Gamma_{p-1} \Delta Y_{t-p+1} + e_t \quad (4.2)\end{aligned}$$

where  $\Pi = -(I_k - A_1 - \dots - A_p)$  and

$$\Gamma_i = -(A_{i+1} + \dots + A_p) \text{ for } i=1, \dots, p-1$$

# VAR and VecqC models.

- A VecqC model is obtained by subtracting  $Y_{t-1}$  on both sides of the unrestricted VAR model and reagruping terms. For example:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + e_t$$

We subtract  $Y_{t-1}$  on both sides

$$Y_t - Y_{t-1} = (A_1 - I)Y_{t-1} + A_2 Y_{t-2} + e_t$$

Adding and subtracting  $A_2 Y_{t-1}$  to the left,

$$\Delta Y_{t-1} = (A_2 + A_1 - I)Y_{t-1} - (A_2)\Delta Y_{t-1} + e_t$$

# VAR and VecqC models.

- Given that we assume all the variables in  $Y_{t-1}$  are stationary, the left hand side of (4.2) must be stationary. Therefore,  $\Pi Y_{t-1}$  must be also stationary.  $\Pi Y_{t-1}$  is sometimes called the long-run or long-term part. Notice that when there are unit roots,  $\Pi = -(I_n - A_1 - \dots - A_p)$  is singular. Matrix  $\Pi$  do not have complete rank but it can be written as the product of two  $(n \times r)$  matrices such that  $\Pi = \alpha \beta'$ .
- $\Gamma_i Y_{t-i}$  captures the short run dynamic of the system.

# VAR and VecqC models.

- In a VecqC model, the short run dynamic of each variable is influenced by deviations of the long run equilibrium. Therefore, if we assume that short and long run interest rates are I(1), a VecqC model would be:

$$\Delta r_{St} = \alpha_S (r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{St}, \alpha_S > 0$$

$$\Delta r_{Lt} = -\alpha_L (r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{Lt}, \alpha_L < 0$$

- In this system, interest rate variables react both to short and long run shocks.

# VAR and VecqC models.

- This insight does not change if we formulate a more general model by introducing lagged values of the different variables. In this case, we have the following specification:

$$\begin{aligned}\Delta \mathbf{r}_{St} &= \mathbf{a}_{10} + \boldsymbol{\alpha}_S (\mathbf{r}_{Lt-1} - \boldsymbol{\beta} \mathbf{r}_{St-1}) + \sum \mathbf{a}_{11}(\mathbf{i}) \Delta \mathbf{r}_{St-i} + \sum \mathbf{a}_{12}(\mathbf{i}) \Delta \mathbf{r}_{Lt-i} + \boldsymbol{\varepsilon}_{St} \\ \Delta \mathbf{r}_{Lt} &= \mathbf{a}_{20} + \boldsymbol{\alpha}_L (\mathbf{r}_{Lt-1} - \boldsymbol{\beta} \mathbf{r}_{Lt-1}) + \sum \mathbf{a}_{21}(\mathbf{i}) \Delta \mathbf{r}_{St-i} + \sum \mathbf{a}_{22}(\mathbf{i}) \Delta \mathbf{r}_{Lt-i} + \boldsymbol{\varepsilon}_{Lt}\end{aligned}$$

# VAR and VecqC models.

- $r_{Lt-1} - \beta r_{St-1}$  describe the long run equilibrium of both interest rates.
- Parameters  $\alpha_s$  and  $\alpha_L$  can be interpreted as the speed of adjustment to equilibrium. This is related to the concept of weak exogeneity.

# VAR and VecqC models.

## EXAMPLES OF COINTEGRATED ECONOMIC SERIES

- Money demand equations have inspired an important part of the literature on the cointegration analysis. An econometric specification for this equation can be:

$$m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t$$

- Notice that  $e_t$  must be stationary otherwise the specification is wrong.

# VAR and VecqC models.

Generally real GDP, monetary aggregates, price indices and interest rate series can be considered as nonstationary series. However, economic theory implies that there exist a linear combination of these variables that is stationary.

# VAR and VecqC models.

## OTHER EXAMPLES

- Consumption theory: total consumption is equal to the sum of permanent and transitory consumption. Transitory consumption is a stationary variable and permanent consumption is related in the long run to permanent income.
- Law of one price: Similar prices must be different in different geographical areas but arbitrage makes price deviations stationary.

# VAR and VecqC models.

- A VecqC model can be written in a compact way as:

$$\Delta x_t = \pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_p \Delta x_{t-p} + e_t.$$

- Given that all the expressions on the right hand side are stationary, the left hand side must also be stationary for the expression to be consistent. Two important points must be noticed:
  - If all the elements of  $\pi$  are zero, we have a model in first differences. There is not an equilibrium correction model in such a case given that variables do not respond to deviations of the long run equilibrium. (This happens when I(1) variables are not cointegrated)
  - If one or more elements of  $\pi$  are different from zero, variables react to shocks in the long run relation. In this case, it is unaccurate to estimate a model in first differences.

# VAR and VecqC models.

- We assume now an example of VecqC model with three endogenous variables and two cointegration relations

$$\Pi Y_{t-1} = \alpha \beta' Y_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix}$$

- Which are the cointegration relations in this system?
- Are the unique?
- How this system adjust to the long run equilibrium?

# VAR and VecqC models.

## DETERMINISTIC COMPONENTS

- In many cases, we need to include deterministic components in the mean, trend or seasonality. One way to treat them is to extract the deterministic components of the series and to specify the model for the adjusted series.
- An alternative way is to include these variables in the model. This is the most accurate way because it allows us to see the interaction among deterministic and stochastic elements in the model.
- Important!!! We will use a number of restrictions in the deterministic components in the specification of VecqC model (we will see this later on).

# VAR and VecqC models.

## EXOGENOUS VARIABLES

- If we have a vector of exogenous explanatory variables  $Z_t$  and a vector of deterministic variables  $D_t$ , they can be easily included in the model:

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + CD_t + Bz_t + e_t \quad (4.4)$$

where  $D_t$  is a vector that contains deterministic variables and  $z_t$  contains nondeterministic exogenous variables.  $C$  and  $B$  are matrices of parameters.

# VAR and VecqC models.

## AN EXAMPLE

- Imagine we are estimating a model for inflation and interest rates,  $(\pi_t \ y \ i_t)$ . These are two endogenous variables as there is a feedback between them.
- However, seasonal variables may have an effect on them. These are deterministic variables (that are exogenous).
- Also, weather could have an effect on inflation and it can be convenient to include this exogenous stochastic variable in the model.

## 5.3. VEC MODEL SPECIFICATION

- The two most popular procedures to test cointegration and specify VEC models are:
  - The Engle and Granger approach.
  - The Johansen approach.

# VecqC model specification.

## THE ENGLE AND GRANGER APPROACH

- This methodology can be used to test whether or not two series are cointegrated. Steps:
  - 1) Define the integration order of each of the variables ( D-F test). Both series must be I(1).
  - 2) Estimate the following equation

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

If the two variables are cointegrated, the OLS estimator of  $\beta_1$  and  $\beta_2$  is superconsistent.

Then, we test whether residuals of this equation are stationary. We do not need to include deterministic components in the DF test if they have been already included in the equation that relates  $y_t$  and  $x_t$ .

# VecqC model specification.

## THE ENGLE AND GRANGER APPROACH

3) Estimate the VecqC model:

$$\Delta y_t = \alpha_1 + \alpha_y (y_{t-1} - \beta_1 z_{t-1}) + \sum_{i=1} \alpha_{11}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{12}(i) \Delta z_{t-i} + \varepsilon_{yt}$$

$$\Delta z_t = \alpha_2 + \alpha_z (y_{t-1} - \beta_1 z_{t-1}) + \sum_{i=1} \alpha_{21}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{22}(i) \Delta z_{t-i} + \varepsilon_{zt}$$

Notice that this estimation amounts to:

$$\Delta y_t = \alpha_1 + \alpha_y \hat{e}_{t-1} + \sum_{i=1} \alpha_{11}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{12}(i) \Delta z_{t-i} + \varepsilon_{yt}$$

$$\Delta z_t = \alpha_2 + \alpha_z \hat{e}_{t-1} + \sum_{i=1} \alpha_{21}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{22}(i) \Delta z_{t-i} + \varepsilon_{zt}$$

Two important points:

- It is efficient to estimate the model equation by equation by OLS.
- Given that all variables are stationary, we can interpret t-statistics in a standard way.

# VeqC model specification.

## THE ENGLE AND GRANGER APPROACH

- Ensure that we have an accurate model. Some procedures are:
  - Ensure that residuals of the VAR model are white noise by including enough number of lags.
  - Parameters of speed of adjustment must be different from zero if there is cointegration among the variables.

# VeqC model specification.

## PROBLEMS WITH THE ENGLE AND GRANGER APPROACH

- This methodology is easily implemented, however it has important drawbacks.
  - In the estimation of the long run equation, researcher must decide which is the dependent and which the explanatory variable. This decision can influence the final result in short samples.
  - When there are more than two variables, there could be more than one cointegration relation. This test does not have a systematic procedure to distinguish among the different cointegration relations.
  - This is a two step procedure. Every mistake in the first step can influence the second step.

# VecC model specification.

## JOHANSEN PROCEDURE (1988)

- The test proposed by Johansen (1988) solved these problems by doing an estimation in a single step and coping with more than one cointegration relation. This procedure can be considered a multivariate generalization of unit root tests. The equation to be estimated is:

$$\Delta x_t = \pi x_{t-1} + \varepsilon_t$$

# VeqC model specification.

## JOHANSEN APPROACH

- As we did in the univariate case, if the rank of  $\pi$  is zero there is no cointegration relations in the system. However, if the rank of  $\pi$  is  $n$ , all variables are stationary. For intermediate cases the rank determines the number of cointegration relations in the system.

# VeqC model specification.

## JOHANSEN APPROACH

- The number of cointegration relations can be tested by checking the rank of  $\pi$ . We know that the rank of a matrix is the number of eigenvalues different than zero. If we sort the eigenvalues in the following way  $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n$ . If variables are not cointegrated, the rank of  $\pi$  is zero and  $\ln(1-\lambda_i)=0$ , but if the rank of  $\pi$  is the unity,  $1 > \lambda_1 > 0$ ,  $\ln(1-\lambda_1) \neq 0$ .
- In practice, we use the following test:

# VecqC model specification.

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i)$$

- In the trace test, we test if there are at most  $n$  cointegration relations and it is a sequential test.
- Johansen and Juselius (1990) provide critical values for these two statistics. The distribution of these two statistics depends on:
  - The number of nonstationary components under the null hypothesis ( $n-r$ ).
  - The deterministic components included in the system.

## 4.5. ESTIMATION

- We have a sample  $y_1, \dots, y_T$  and a presample  $y_{-p+1}, \dots, y_0$ .
- The  $n$  equations of an unrestricted VAR model can be estimated separately by OLS and this estimator is as efficient as the generalized least square estimator (GLS), Zellner (1962).
- We use the following notation  $Y = [y_1, \dots, y_T]$ ,  $A = [A_1: \dots: A_p]$ ,  $E = [e_1, \dots, e_T]$  y  $Z = [Z_0, \dots, Z_{T-1}]$ , where

$$Z_{t-1} = \begin{bmatrix} y_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ y_{t-p} \end{bmatrix}$$

# Estimation.

Then, the VAR model can be written as:

$$Y_{nxT} = A_{nx} (nxp) Z_{(nxp)xT} + E_{nxT}$$

The OLS estimator of A is

$$\hat{A} = [\hat{A}_1 : \dots : \hat{A}_p] = YZ' (ZZ')^{-1}$$

and, using standard assumptions,  $\hat{A}$  is a consistent and normal estimator

$$\sqrt{T} \text{vec}(\hat{A} - A) \rightarrow N(0, \Sigma_{\hat{A}})$$

or

$$\text{vec}\hat{A} \rightarrow N(\text{vec}A, \Sigma_{\hat{A}}/T)$$

# Estimation.

- For autoregressive models, the OLS is identical to the ML estimator conditional to the initial values.
- When the system is not stationary, the distribution does not change, although the covariance matrix is singular. In other words, coefficients of linear combination of the coefficients converge to the true value of the estimation to a higher rate than  $T^{1/2}$ .

# Estimation.

- Careful!!! generally, in a nonstationary VAR model, critical values for the t-statistics of the parameters are no longer valid. In this case:
  - When variables are cointegrated, a better option is to specify a VeqC model.
  - When variables are not cointegrated and the model is not stationary we should specify the model for first differences of the variables to interpret the statistics.

See Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996).

## ESTIMATION OF VEQC MODELS

- We assume that our VAR models is correct description of the data. We also assume that all variables are stationary or  $I(1)$ , that is, there are not  $I(2)$  variables or with seasonal roots.
- Before starting, we need to define the number of cointegration relations using standard tests (either Granger or Johansen procedure).
- This is a two step estimation. First, we estimate the long run parameters and afterwards, conditional to these coefficients, we estimate the short run dynamics.

## CONCENTRATION OF THE VAR MODEL

We assume that

$$\Pi = \alpha\beta'$$

where  $\alpha$  and  $\beta$  are  $n \times r$  matrices.

Three possible cases:

- 1) If  $r=n$ , all the elements of  $x_t$  are stationary and we can use standard inference.
- 2) If  $r=0$ , therefore  $x_t$  is not stationary and we can not obtain stationary linear combination of these variables. In this case, a VAR model in first differences must be specified.
- 3) The most interesting case is when  $n>r>0$ , then we can say that exists  $r$  stationary linear combination of variables.

# Estimation.

- We obtain the ML estimator proceeding in two steps.
  - 1) We assume that  $\beta$  is known and we derive an estimator for  $\alpha$ .
  - 2) We insert the estimation of  $\alpha$  in the likelihood function in a way that the problem is a function of  $\beta$ .

# Estimation.

## INTUITIVELY

- Cointegration relations are canonical correlations between series in levels and differences.
- Once the cointegration vector is set, the rest of parameters can be estimated. This has sense because the cointegration vector is superconsistent and converge to the true value to a faster rate than short run parameters.

## ASYMPTOTIC DISTRIBUTIONS

Under general assumptions

$$\sqrt{T}vec([\hat{\Gamma}_1 : \dots : \hat{\Gamma}_{p-1}] - [\Gamma_1 : \dots : \Gamma_{p-1}]) \rightarrow N(0, \Sigma_{\hat{\Gamma}})$$

$$\sqrt{T}vec(\hat{\Pi} - \Pi) \rightarrow N(0, \Sigma_{\hat{\Pi}})$$

Distribution of  $\hat{\Gamma}$  is nonsingular and therefore we can use standard inference for short run parameters. However, the covariance matrix  $\Sigma_{\hat{\Pi}}$  with dimension  $(n^2 \times n^2)$  has rank  $nr$  and it is singular. This happens because we have restricted the cointegration space and because cointegration estimations have specific properties.

## IDENTIFICATION

- The described procedure only identify the cointegration space but not the parameters of the cointegration relation.
- The most usual identification restriction consists of assuming that the first part of  $\beta$  is the identity matrix such that
$$B' = [I_r : B'_{(n-r)}].$$
- For  $r=1$ , this amounts to normalize the coefficient of the first variable to 1.
- Very important!!!, this normalization requires lot of care in choosing the order of variables in the system. Old versions of E-views, such as E-views 3.0 do this normalization automatically.
- The order of variables is not a trivial question, one must set coefficient one to the dependent variable in the cointegration relation.

## 4.7. RESTRICTIONS IN THE COINTEGRATION VECTOR

There are 2 ways to impose restrictions in the cointegration vector: 1) specifying  $s_i$  free parameters; or 2) specifying  $m_i$  restrictions in the cointegration vector.

1) The first specification can be written as

$$\beta^c = (\beta_1^c, \dots, \beta_r^c) = (H_1 \varphi_1, \dots, H_r \varphi_r),$$

where  $\varphi_i$  is a  $(s_i \times 1)$  matrix of coefficients,  $H_i$  is a  $(p \times s_i)$  matrix where  $p$  is the dimension of the cointegration vector (it can be different from  $n$  if there are deterministic components)

2) For the second case, we specify restrictions in matrices  $R_i$  that have dimension  $(p \times m_i)$

$$\begin{aligned} R_1' \beta_1 &= 0 \\ &\vdots \\ &\vdots \\ R_r' \beta_r &= 0 \end{aligned}$$

# Restrictions in the cointegration vector

## EXAMPLE

We assume the following system

$$\beta_1' \bar{x}_t = m_t^r - y_t^r - b_1(R_{m,t} - R_{b,t}) - b_2Ds83_t$$

$$\beta_2' \bar{x}_t = y_t^r - b_3(\Delta p_t - R_{b,t})$$

$$\beta_3' \bar{x}_t = (R_{m,t} - R_{b,t}) + b_4Ds83_t$$

Questions:

How many free parameters we have in and how many restrictions without taking into account the normalization restriction?

# Restrictions in the cointegration vector

If we sort variables in the following way

$$\bar{x}_t' = [m_t^r, y_t^r, \Delta p_t, R_{m,t}, R_{b,t}, Ds83_t]$$

We write restrictions as a function of free parameters, we have

$$\beta_1 = H_1 \varphi_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \varphi_{11} \\ \varphi_{12} \\ \varphi_{13} \end{bmatrix}$$

$$\beta_2 = H_2 \varphi_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varphi_{21} \\ \varphi_{21} \end{bmatrix}$$

# Restrictions in the cointegration vector

## HOMEWORK

- Write the restriction for the third cointegration relation.

# Restrictions in the cointegration vector

If we write restrictions using the second way, we have

$$R_1' \beta_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \\ \beta_{15} \\ \beta_{16} \end{bmatrix} = 0,$$

$$R_2' \beta_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \\ \beta_{15} \\ \beta_{16} \end{bmatrix} = 0$$

# Restrictions in the cointegration vector

## HOMEWORK

- Write the third restriction in this way.

# Restrictions in the cointegration vector

## TESTING THE SAME RESTRICTION IN ALL THE COINTEGRATION RELATIONS

- In some cases we want to test the same restriction for all cointegration vectors. For example, we can be interested in drop one variable from all the cointegration relations or if the coefficient associated to one of the variables is always the same (price homogeneity).

# Restrictions in the cointegration vector

In this case, the restricted model is specified as follows

$$H^c(r): \beta^c = (H\varphi_1, \dots, H\varphi_r) = H\varphi$$
$$\Delta x_t = \alpha\varphi' H' \bar{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-1} + \varepsilon_t$$

The likelihood ratio test is based on the ratio between the restricted and the unrestricted likelihood function.

The statistic is distributed as a  $\chi^2(v)$  where  $v = rm$  and  $m$  is the number of restrictions imposed in each of the equations.

Generally, restrictions are not the same in the different equations. In this case, degrees of freedom would be

$$v = \sum_{i=1}^r (m_i - r + 1) = \sum_{i=1}^r (n - r) - (s_i - 1)$$

# Restrictions in the cointegration vector

## RESTRICTIONS IN THE ADJUSTMENT COEFFICIENTS

- Adjustment coefficients indicate the speed of adjustment to shocks in the long run relation.
- If reaction is zero, we say this variable is weakly exogenous.
- The type of restriction tested in this case is whether or not the speed of reaction is equal to zero. If the whole row is equal to zero we consider this variable as weakly exogenous.
- For this test, the t-values associated to the estimated alpha parameters follow a standard distribution.
- If there are cointegration relations among the variables in a system at least one of the variables must be exogenous.

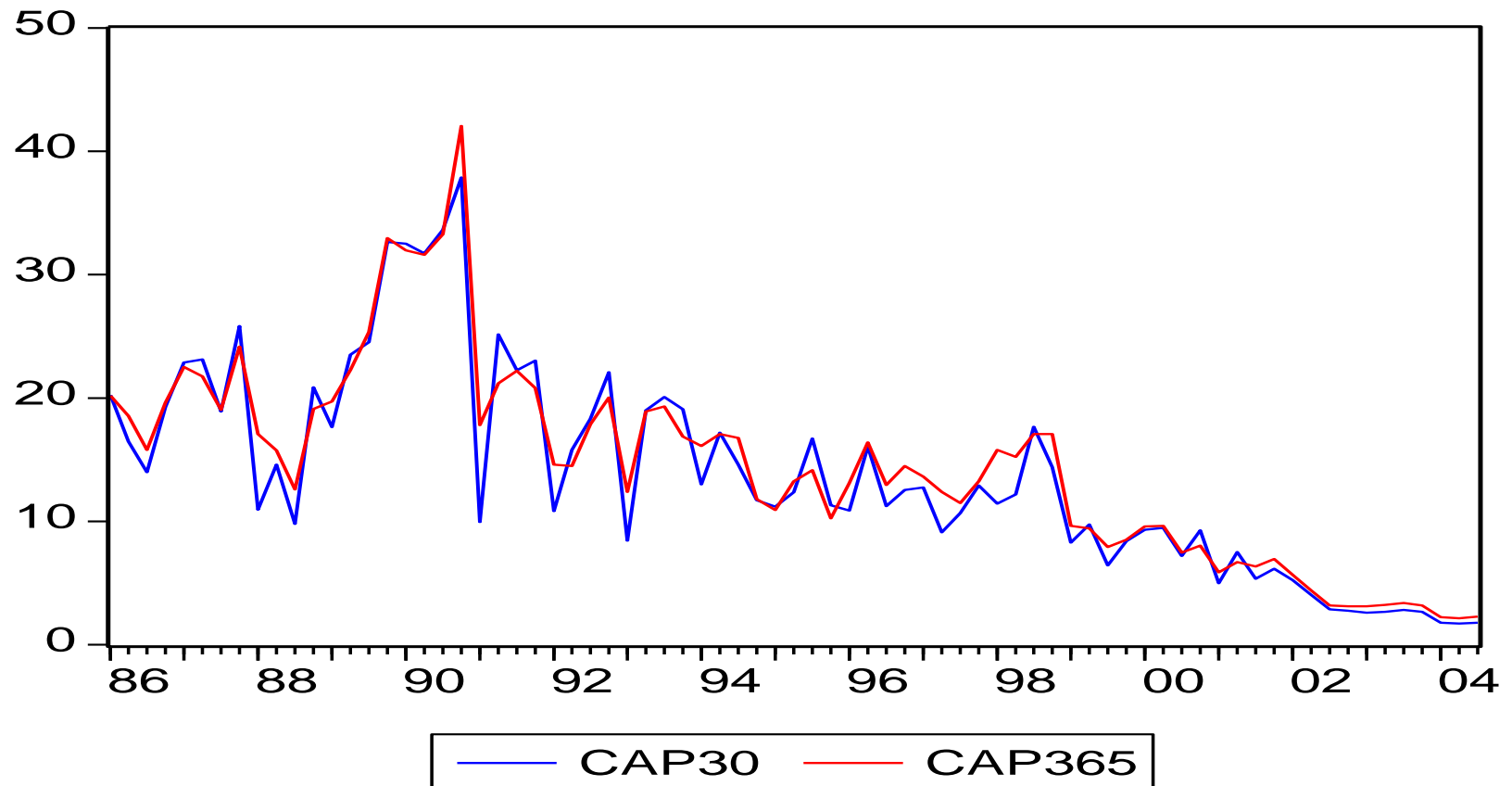
# Restrictions in the cointegration vector

## IN PRACTICE

1. We have some hypothesis about the model we want to estimate. In this model, there are equilibrium relations and there could be also parametric restrictions.
2. We pick the time series. They should have at most 1 unit root . (Cointegration of  $I(2)$  series  $I(2)$  has not been considered in this course).
3. We test the number of cointegration relations using the Johansen approach.
4. We identify the long run relations.
5. We insert restrictions in the long run relations until they have an economic meaning.
6. We test for weak exogeneity.
7. Economic conclusions are drawn.

# Two examples

## EXAMPLE 1



# Two examples

Null Hypothesis: CAP30 has a unit root  
Exogenous: Constant  
Lag Length: 7 (Automatic based on AIC, MAXLAG=11)

---

---

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.461923	0.8914
Test critical values:		
1% level	-3.531592	
5% level	-2.905519	
10% level	-2.590262	

---

---

\*MacKinnon (1996) one-sided p-values.

# Two examples

Null Hypothesis: CAP365 has a unit root

Exogenous: Constant

Lag Length: 9 (Automatic based on AIC, MAXLAG=11)

---

---

	t-Statistic	Prob.*
<b>Augmented Dickey-Fuller test statistic</b>	<b>-0.185945</b>	<b>0.9344</b>
Test critical values:		
1% level	-3.534868	
5% level	-2.906923	
10% level	-2.591006	

---

---

\*MacKinnon (1996) one-sided p-values.

# Two examples

Dependent Variable: CAP365

Method: Least Squares

Date: 11/08/06 Time: 22:55

Sample: 1986Q1 2004Q3

Included observations: 75

---

---

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.449270	0.422846	3.427415	0.0010
CAP30	0.943063	0.026003	36.26754	0.0000

---

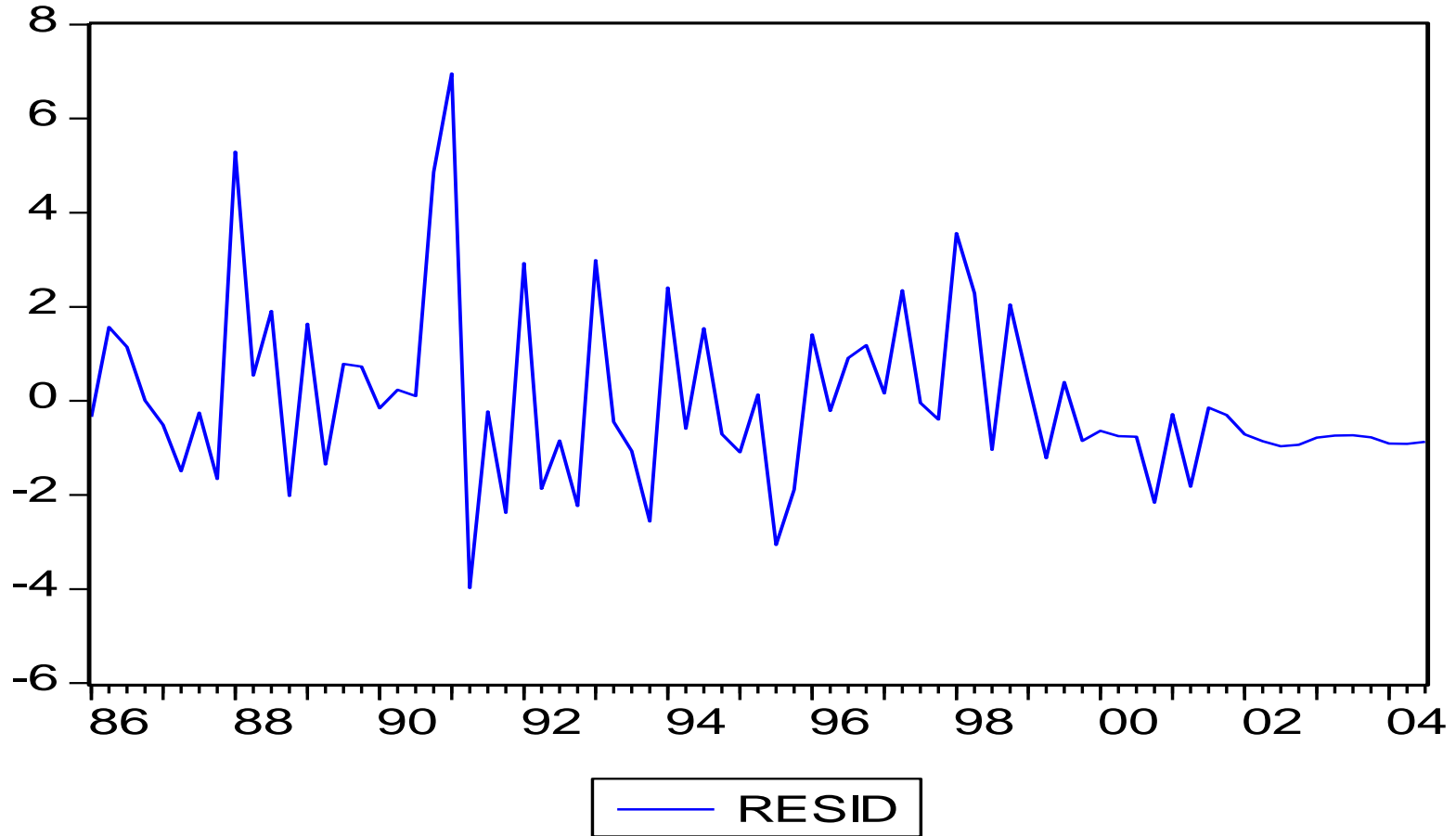
---

R-squared	0.947419	Mean dependent var	14.63253
Adjusted R-squared	0.946699	S.D. dependent var	8.103340
S.E. of regression	1.870824	Akaike info criterion	4.116939
Sum squared resid	255.4987	Schwarz criterion	4.178739
Log likelihood	-152.3852	F-statistic	1315.334
Durbin-Watson stat	2.068827	Prob(F-statistic)	0.000000

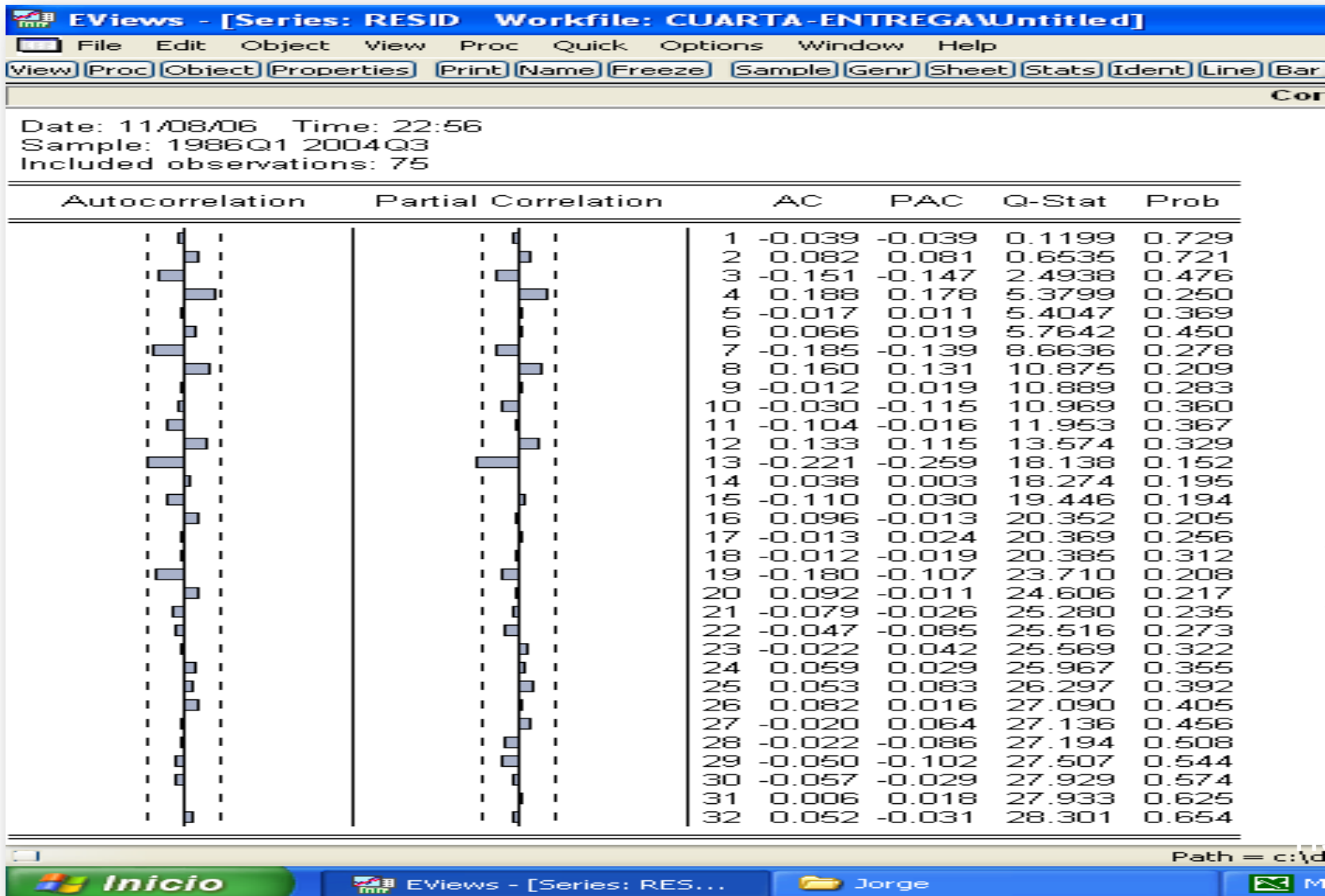
---

---

# Two examples



# Two examples



# Two examples

Null Hypothesis: RES has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on AIC, MAXLAG=11)

---

---

		t-Statistic	Prob.*
<hr/>			
<u>Augmented Dickey-Fuller test statistic</u>		-8.787314	0.0000
Test critical values:	1% level	-3.521579	
	5% level	-2.901217	
	10% level	-2.587981	

---

---

# Two examples

	D(CAP30,1)	D(CAP365,1)
D(CAP30(-1),1)	-0.145432 (0.30506) [-0.47673]	-0.026697 (0.23773) [-0.11230]
D(CAP365(-1),1)	-0.471710 (0.35067) [-1.34518]	-0.351583 (0.27327) [-1.28658]
C	-0.358505 (0.60249) [-0.59504]	-0.308582 (0.46951) [-0.65725]
RES(-1)	0.239134 (0.49589) [ 0.48223]	-0.490617 (0.38644) [-1.26958]
R-squared	0.263667	0.120998
Adj. R-squared	0.231653	0.082781
Sum sq. resids	1821.209	1105.987
S.E. equation	5.137542	4.003598
F-statistic	8.235869	3.166038
Log likelihood	-220.9956	-202.7907
Akaike AIC	6.164263	5.665500
Schwarz SC	6.289767	5.791004
Mean dependent	-0.201096	-0.223014
S.D. dependent	5.861065	4.180361
Determinant resid covariance (dof adj.)		47.89271
Determinant resid covariance		42.78799
Log likelihood		-344.2684
Akaike information criterion		9.651190
Schwarz criterion		9.902199

# Two examples

## EXAMPLE 2

- The last example considered cointegration between two series.
- Now, we develop a more complex analysis by studying the cointegration relations among a big group of variables:
  - Price of Brent,  $B_t$ .
  - Price of Copper,  $C_t$ .
  - The Dow-Jones index,  $D_t$ .
  - The exchange rate expressed as the number for Chilean pesos for one dollar,  $e_t$ .
  - The American GDP,  $y_t^{\text{US}}$ .
  - The American CPI,  $p_t^{\text{US}}$ .
  - The federal fund rate,  $i_t$ .
  - The Chilean GDP,  $y_t^{\text{Ch}}$ .
  - The Chilean CPI,  $p_t^{\text{Ch}}$ .
- All series are in quarterly basis and cover the period 1966:Q1-2006:Q3.

# Two examples

## HYPOTHESIS OF THE MODEL

$$\text{Chilean Phillips Curve: } \Delta_4 p_t^{Ch} = f(\Delta y_t^{Ch}, \Delta e_t) + \theta_{1t}$$

$$\text{Chilean aggregate demand equation: } \Delta y_t^{Ch} = f(\Delta_4 p_t^{Ch}, i_t) + \theta_{2t}$$

$$\text{Price of Brent: } B_t = \theta_{3t}$$

$$\text{Price of cooper: } C_t = \theta_{4t}$$

$$\text{Dow -Jones: } D_t = f(\Delta y_t^{US}, \Delta_4 p_t^{US}) + \theta_{5t}$$

$$\text{Determination of Exchange rate: } \Delta e_t = f(\Delta_4 p_t^{Ch} - \Delta_4 p_t^{US}) + \theta_{6t}$$

$$\text{American Phillips curve: } \Delta_4 p_t^{US} = f(\Delta y_t^{Ch}) + \theta_{7t}$$

$$\text{American aggregate demand equation: } \Delta y_t^{US} = f(\Delta_4 p_t^{US}, i_t) + \theta_{8t}$$

# Two examples

## COINTEGRATION TRACE TEST

I(1) cointegration analysis, 1966(4) - 2006(2)

H0: rank $\leq$  Trace test [ Prob]

0	336.40 [0.000] **
1	234.71 [0.000] **
2	165.03 [0.005] **
3	115.73 [0.064]
4	76.860 [0.266]
5	52.814 [0.300]
6	32.113 [0.388]
7	18.754 [0.302]
8	8.6881 [0.206]

# Two examples

- We can accept the presence of 3 cointegration relationships at the conventional level.
- We impose initial restrictions to indentify cointegration relationships. To do this, we normalize to one the three highest coefficients in the system. These are Chilean output growth, fed interest rates and American output growth.
- Then, we impose overidenfiying restrictions until cointegration relationships are irreducible.
- This can be done with E-views 5.0, 6.0, GiveWin, RATS and JMulTI.

# Two examples

**Table 6.3. Test for over identified restrictions**  
Standard error between brackets

$B_t$	$C_t$	$D_t$	$\Delta e_t$	$\Delta y_t^{US}$	$\Delta p_t^{US}$	$i_t^{US}$	$\Delta y_t^{Ch}$	$\Delta p_t^{Ch}$	Trend
-0.006 (0.002)	0.004 (0.002)	-0.001 (0.003)	0.05 (0.01)	1	0.93 (0.13)	0	0	-0.07 (0.01)	0.0001 (9.1xe <sup>-5</sup> )
0.0004 (0.002)	-0.002 (0.002)	1.9xe <sup>-5</sup> (0.003)	-0.08 (0.01)	0	-0.63 (0.13)	1	0	0.1 (0.01)	4.2xe <sup>-6</sup> (9.0xe <sup>-5</sup> )
0.01 (0.01)	0.0007 (0.02)	0.01 (0.02)	0.49 (0.07)	0	-0.22 (0.88)	0	1	-0.49 (0.09)	-0.0004 (0.0006)
-0.006 (0.002)	0.004 (0.002)	-0.002 (0.003)	0.05 (0.01)	1	0.91 (0.13)	0	0	-0.07 (0.01)	0.0001 (8.3xe <sup>-5</sup> )
0	-0.002 (0.002)	0.0005 (0.0008)	-0.08 (0.009)	0	-0.56 (0.10)	1	0	0.09 (0.01)	0
0.004 (0.005)	-0.002 (0.02)	-0.005 (0.007)	0.51 (0.07)	0	-0.46 (0.85)	0	1	-0.51 (0.09)	0
$\chi^2(\nu): 0.70(3)$					p-value=0.87				
-0.006 (0.002)	0.004 (0.002)	-0.002 (0.003)	0.05 (0.009)	1	0.93 (0.12)	0	0	-0.07 (0.01)	0.0001 (8.3xe <sup>-5</sup> )
0	-0.002 (0.001)	0.0003 (0.0006)	-0.08 (0.009)	0	-0.57 (0.07)	1	0	0.09 (0.01)	0
0.002 (0.004)	0	0	0.53 (0.07)	0	0	0	1	-0.52 (0.09)	0
$\chi^2(\nu): 1.27(6)$					p-value=0.97				
-0.005 (0.001)	0.004 (0.002)	0	0.05 (0.009)	1	0.94 (0.12)	0	0	-0.07 (0.01)	0.0001 (2.5xe <sup>-5</sup> )
0	-0.001 (0.001)	0	-0.07 (0.009)	0	-0.60 (0.06)	1	0	0.09 (0.01)	0
0	0	0	0.55 (0.07)	0	0	0	1	-0.54 (0.09)	0
$\chi^2(\nu): 1.93(9)$					p-value=0.99				
-0.005 (0.001)	0.004 (0.002)	0	0.07 (0.008)	1	0.93 (0.11)	0	0	-0.09 (0.01)	0.0001 (2.5xe <sup>-5</sup> )
0	0	0	-0.10 (0.006)	0	-0.60 (0.06)	1	0	0.12 (0.01)	0
0	0	0	1	0	0	0	1	-1	0
$\chi^2(\nu): 4.37(12)$					p-value=0.98				
-0.006 (0.001)	0	0	0.04 (0.008)	1	1.06 (0.11)	0	0	-0.06 (0.01)	0.0001 (2.4xe <sup>-5</sup> )
0	0	0	0	0	-0.62 (0.07)	1	0	0	0
0	0	0	1	0	0	0	1	-1	0
$\chi^2(\nu): 22.251(15)$					p-value=0.10				
-0.005	0	0	-0.005	1	1	0	0	-0.005 (0.001)	0.0001 (1.5xe <sup>-5</sup> )
0	0	0	0	0	-0.61 (0.07)	1	0	0	0
0	0	0	1	0	0	0	1	-1	0
$\chi^2(\nu): 23.238(18)$					p-value=0.18				

Int Production for time series (2013)