Lecture 5: Time Series Econometrics
Cointegration and Error Correction Models (ECMs): Introduction and Overview

Daniel Buncic

Institute of Mathematics & Statistics
University of St. Gallen
Switzerland

February 5, 2016 (tse5)

Table of Contents

1. Introduction

1.1. Notation

We have so far used α and β to denote the polynomial terms in the ARMA model. Unfortunately, α and β are popular Greek letters in the cointegration literature as well. Here I will follow the standard convention to denote the $k \times r$ dimensional **cointegrating vector** by β and the $k \times r$ dimensional **speed of adjustment** vector by α. For the $(k \times r)$ dimensional cointegrating vector β, I will use subscript notation to refer to elements of the vector referring to variables that cointegrate with each other. Note that I will not use bold letters to denote vectors and/or matrices, only in the case of companion form notation. Also, I will use standard y_t and x_t notation to illustrate single equation cointegration examples and will follow the vector notation X_t as used in the VAR notes in Lecture 4.

1.2. Background

We have seen from Lecture 3 on unit-root processes that there are consequence of unit-roots in the data, ie., in x_t, on the asymptotic distribution as well as the rate of convergence of the OLS estimator $\hat{\rho}$ (or MLE under normality of the error term) when the true data generating process (DGP) is a random walk model. That is, the DGP is

$$x_t = x_{t-1} + u_t \tag{1}$$

with $u_t \sim \mathsf{WN}(0, \sigma^2)$ and we estimate either of the following three models

$$\begin{aligned} x_t &= \rho x_{t-1} + u_t \\ x_t &= c + \rho x_{t-1} + u_t \\ x_t &= c + \delta t + \rho x_{t-1} + u_t. \end{aligned}$$

Variables that contain a unit-root are said to be integrated (written as $I(1)$), where the 1 signifies that the $I(1)$ process becomes stationary when differenced 1 time.

Integration is formally defined as follows:

Definition 1 (Integration): *A series x_t is said to be integrated of order d, denoted $x_t \sim I(d)$ if it becomes stationary with an ARMA representation (without any deterministic trend in it) after being differenced d times.*

From the definition of Integration above it should be clear that a stationary variable is integrated of order $I(0)$. It should also be pointed out here that when mixing time series data with different orders of integration, the one with the highest order will dominate the lower order integrated series. This means that if we linearly combine two series, one of which is $I(0)$ and the other is $I(1)$, then the new resulting series will inherit the integratedness properties of the $I(1)$ series. Similarly, for higher order integrated series, combining and $I(1)$ and an $I(2)$ series, yields an $I(2)$ series, and we can see how this generalises. Most of the time when dealing with economic and financial time series data, it is rather rarely the case that we will observe a series that is integrated of order 2, it is pretty much a zero probability event to see a series that is integrated of an order higher than 2.

1.3. Integration accounting (or keeping 'balance')

A concept that is related to the orders of integration that were discussed above and what is frequently referred to as *'integration accounting'* or keeping *'balance'* in the literature is that the left hand side of an equation has to have the same order of integration as the right hand side. That is, it has to be balanced. For instance, if we have a standard set-up of the form

$$y_t = c + \beta x_t + u_t \tag{2}$$

and y_t, x_t and u_t have different orders of integration, then, these have to balance out. As an example, if $y_t \sim I(1)$ and $x_t \sim I(0)$, then u_t has to be $I(1)$ for (2) to be balanced. An $I(0)$ variable such as x_t here can never explain the variation in an $I(1)$ variable y_t so it must be the case that all the '*integratedness*' of the $I(1)$ variable y_t gets absorbed in the '*error term*' u_t. Since the left hand side is $I(1)$ and $x_t \sim I(0)$ we have that $u_t \sim I(1)$ so that the right hand side $I(0) + I(1)$ gives an $I(1)$ variable. Now it should be clear that it is not a good idea to have the *'error term'* u_t be an $I(1)$ series as we will then obtain **spurious (invalid/non-sense) regression** results. This means that none of the standard errors or $t-$statistics of β the '*regression*' in (2) can be used for inference (we will see more on this below).

Similarly, if we have that $y \sim I(0)$ and $x \sim I(1)$ then this equation is also unbalanced and the variation in the $I(0)$ variable can never be explained by the variation in the $I(1)$ variable. What is different to the previous case is that this relation can easily be balanced by setting $\beta = 0$. What happens asymptotically is that, unlike in the spurious regression result above, the $t-$statistic still has a standard normal distribution. What exactly asymptotic here means we will see later, as it is well known now that there can be substantial small sample distortions in the distribution. Having an $I(0)$ left hand side variable and an $I(1)$ (or at least very very persistent) predictor or regressors on the right hand side is a common problem in economics and finance where a stationary variable such as returns are regressed on a very persistent predictor variable like the dividend yield.

When we have two $I(1)$ variables, ie., both y_t and x_t are $I(1)$, then we have a balanced equation again and it then depends upon what happens when these two $I(1)$ variables are linearly combined. That is, if $(y_t - c - \beta x_t) = u_t$ yields an $I(0)$ random variable, then we have the special case of the two integrated variables being cointegrated (they form a stable long-run equilibrium relation), allowing us to (more or less) conduct standard inference on $\hat{c}$ and $\hat{\beta}$. If they are not, then there will be two unrelated $I(1)$ variables which when linearly combined to not yield an error term $u_t \sim I(0)$, so that u_t will also have to be an $I(1)$ variable, leading again to spurious regression and all asymptotic inference being invalid.

1.4. Some examples of cointegrated variables

It is evident from many time series, particularly from economic and financial time series, that these contain unit-roots and are therefore integrated or $I(1)$ random variables. Nevertheless, we also know form economic theory that many of these $I(1)$ variables are bound by some relation to another $I(1)$ variable. If the linear combination of $I(1)$ variables results in an $I(0)$ variable, then the $I(1)$ variables are said to have formed a '*long-run equilibrium relation*'. This is known as **co-integration** (or **cointegration**) in the economic literature. The term '*long-run*' here means that the $I(1)$ variables drift apart from one another by only a transitory part and that there exists a (linear) attractor set (the equilibrium relation) that ensures that the movement away from equilibrium only lasts for a '*finite*' amount of time. Moreover, the movement of a cointegrated set of variables is, *on average*, bound together.

In economics and finance, some classic examples of $I(1)$ variables that form a cointegrating relation (an $I(0)$ relation) are the following:

1. Permanent income model: Consumption and Income $\sim I(0)$
2. Money demand model: Money, Income, Prices and Interest rates $\sim I(0)$
3. Models from Growth Theory: Income, Consumption and Investment $\sim I(0)$
4. PPP: Exchange rates and Price levels in two countries $\sim I(0)$
5. Fisher equation: nominal interest rates and inflation $\sim I(0)$

6. Term structure models of interest rates: longer term rates and the short $\sim I(0)$
7. Asset pricing models: Consumption, asset prices and Income $\sim I(0)$

and many many others.

Integrated series that form a stationary $I(0)$ relationship with each other often do so in a way that binds them together quite tightly. Any deviations from the long-run equilibrium relation (the attractor set) that are encountered are then adjusted to either by each variable separately, or jointly as a whole system. To illustrate this theoretical relation, consider the pair $\{x_1, x_2\}$ of $I(1)$ variables that form a (positive) cointegration relationship resulting in an $I(0)$ equilibrium error. This cointegrating relation can be expressed visually as shown below in Figure 1.

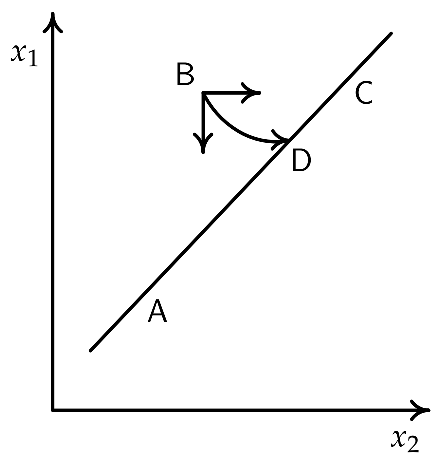


Figure 1: Equilibrium relation attractor set. The diagonal line represents the long-run equilibrium relation between x_1 and x_2.

The entire diagonal line in Figure 1 represents points on the equilibrium long-run relation between x_1 and x_2. Suppose now that there is a (transitory) movement away from this long-run relationship to point B in Figure 1 above equilibrium (equilibrium error is positive). Then there are 3 possible scenarios for this system to move back to equilibrium.

1. x_1 adjusts (decreases) in the next period to move to point A on the attractor set,
2. x_2 adjusts (increases) in the next period to move to point C on the attractor set, or
3. both adjust towards point D on the diagonal.

Under the first scenario, the error-correction mechanism for the first variable takes the from

$$\underbrace{\Delta x_{1t}}_{<0} = \alpha_1 \underbrace{(x_{1t} - c - \beta_2 x_{2t})}_{\text{equilibrium error}>0}$$

where the equilibrium error $\varepsilon_t = (x_{1t} - c - \beta_2 x_{2t})$ is positive due to point B in Figure 1 being above the equilibrium relation and the adjustment in Δx_{1t} is less than zero due to $x_{1t+1} < x_{1t}$. This implies that $\alpha_1 < 0$. Under the second scenario we have

$$\underbrace{\Delta x_{2t}}_{>0} = \alpha_2 \underbrace{(x_{1t} - c - \beta_2 x_{2t})}_{\text{equilibrium error}>0}$$

with the adjustment in Δx_{2t}being greater than 0 due to $x_{2t+1} > x_{2t}$ which implies that $\alpha_2 > 0$. Under the last scenario, both adjustment occur.

1.4.1. Permanent income example: real consumption and real income

To see how these relationships take shape in '*real empirical*' data, consider the relationship between log of real consumption and log of real income for the US over the period from 1947:Q1 to 2013:Q4. The two series are plotted below in Figure 2. From the plot in

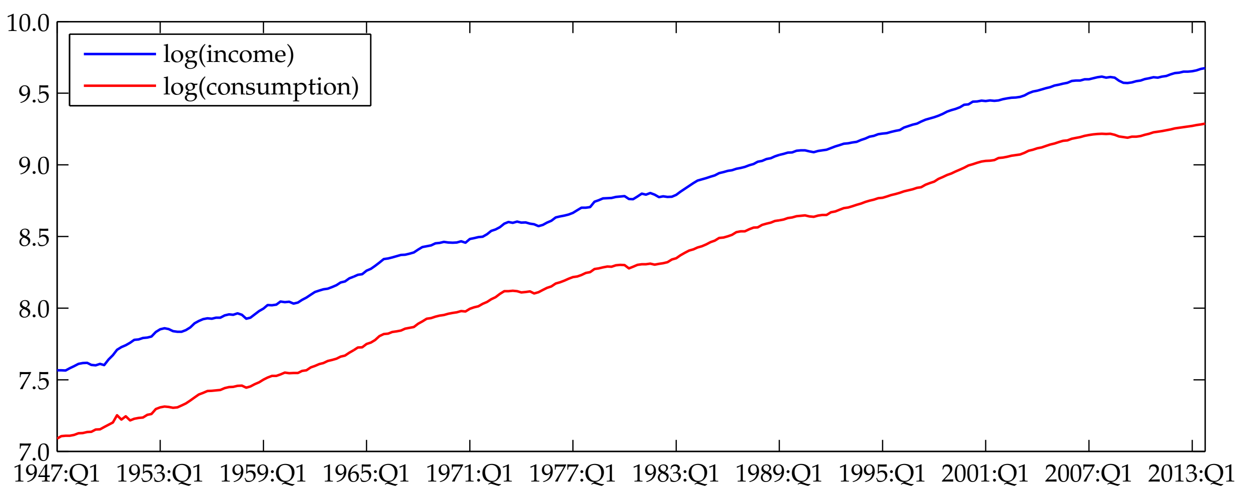


Figure 2: Time series plots of real income (GDP) and real personal consumption (in logs) from 1947:Q1 to 2013:Q4.

Figure 2 the clear upward trend in both series is obvious, as is the close '*co-movement*' of these two series. To see what type of relationship they form, we can look at a scatter plot of log income and log consumption, which is shown below in Figure 3. Comparing this scatter plot to the visual depiction of the theoretical attractor set generated by the cointegrating relation in Figure 1, we see that there is a fairly tight resemblance of the relation staying close to equilibrium for most of the time over the interval of the sample data.

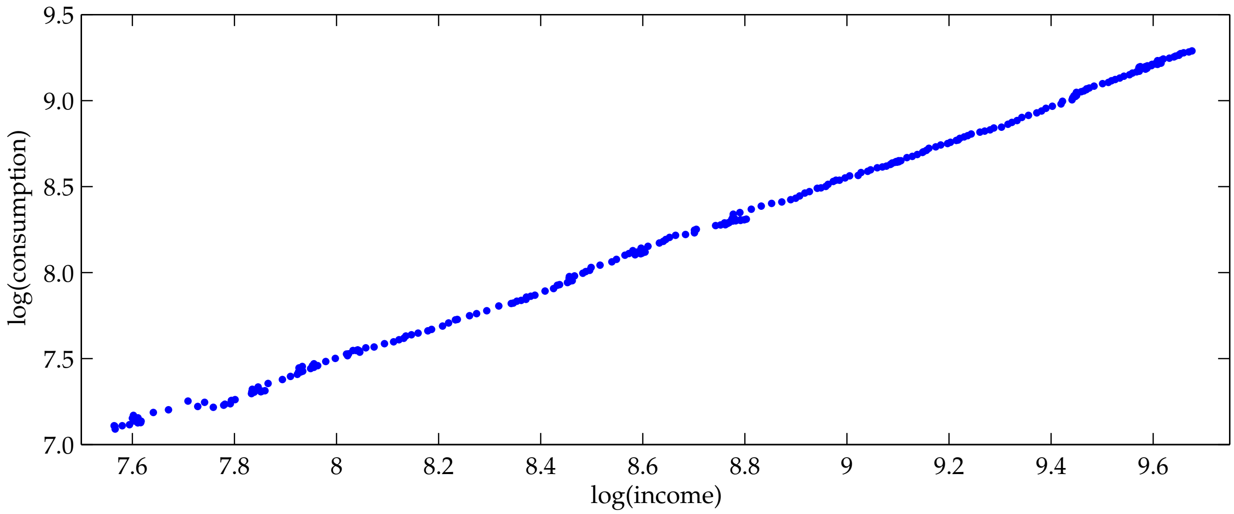


Figure 3: Scatter plot of real income (GDP) and real personal consumption (in logs) from 1947:Q1 to 2013:Q4.

1.4.2. Term structure data: 1, 5 and 10 year yields on US bonds

As another example, consider the time series evolution of three different US bonds with yields to maturity of 1, 5 and 10 years. A plot over the time period form 1962:Q1 to 2010:Q3 is shown below in Figure 4. There is no definite upward trend (not for a long period of time anyway) as was the case for real consumption and real income, nevertheless, the series are again bound together and co-move over the approximately 50 years of data. A 3D scatter plot of the three series shown in Figure 5 reveals the again fairly tight long-run relationship that is formed between the yields. In the three variable scenario the attractor set is still a straight line but now in the 3D space so that deviations from this equilibrium line can be adjusted to in a number of different ways.

2. Spurious regression and ignoring long-run relationships

An important scenario to look at now is to see what happens if we make a mistake. There are two possible scenarios again. The first one is '*running a regression*' of two independent $I(1)$ series. The second one is ignoring long-run relationships all together and using the differenced data to a regression.

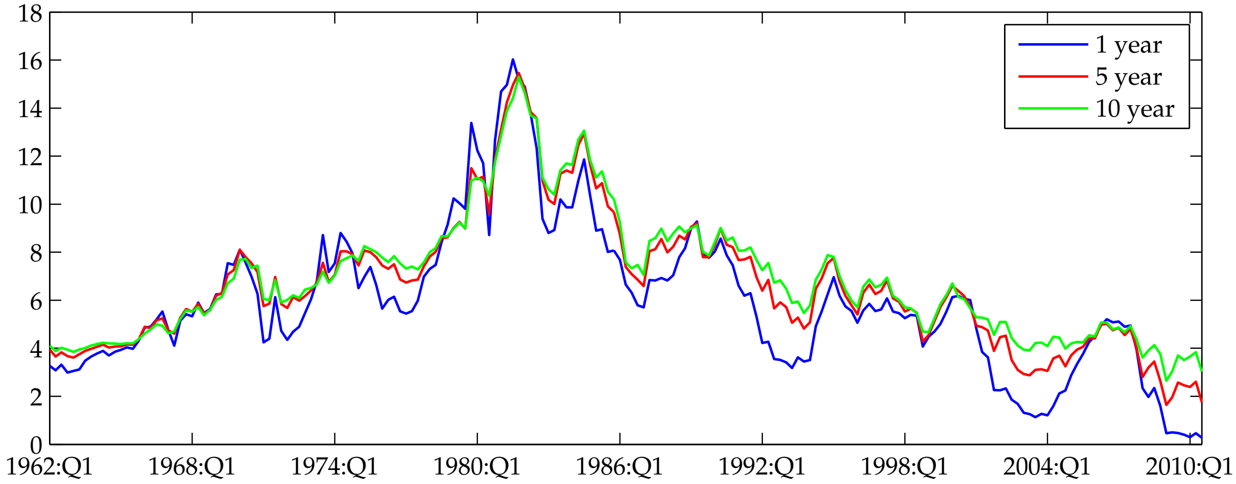


Figure 4: Time series plots of yields of 1 year, 5 year and 10 year US government bonds from 1962:Q1 to 2010:Q3.

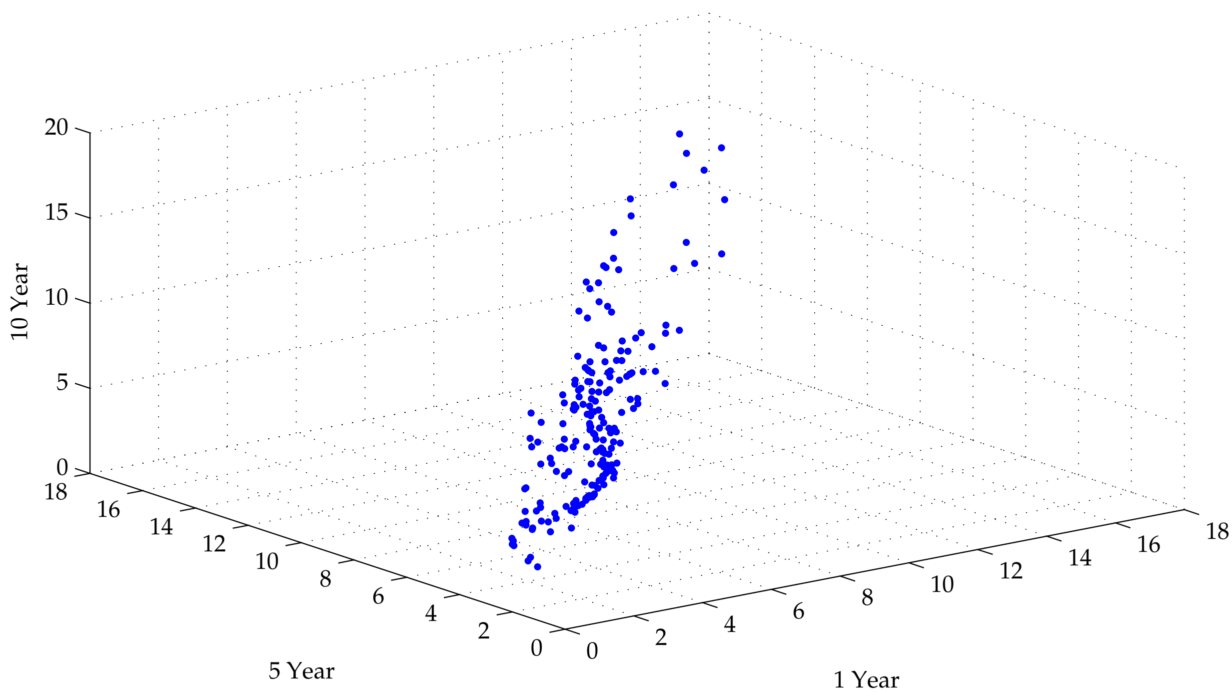


Figure 5: 3D scatter of the yields of 1 year, 5 year and 10 year US government bonds from 1962:Q1 to 2010:Q3.

2.1. Spurious regression

What happens when to the OLS estimator when two independent $I(1)$ series are regressed on each other is a question that was raised initially by Yule (1926) and later on by Granger and Newbold (1974). Theoretical results were provided later by Phillips (1986). The observed problem was that when very persistent time series data were used in *'stan-*

dard regression' analysis, one would generally find a very high R^2 (or $\bar{R}^2$) of well above 90%, highly significant $t-$statistic and yet a very low Durbin-Watson statistic, indicating strong serial correlation in the residuals of the regression model. This is the classic case when two $I(1)$ series do not form a long-run relation (thus do not cointegrate) and in the context of our '*integration accoutring*' framework discussed above this means that the residual inherits the properties of the left hand side variable.

The simplest and most intuitive way to see what happens to our usual OLS estimator and its $t-$statistic when two independent $I(1)$ series are regressed on another is via a simulation exercise. Let us set the sample size to $T = 100$ and simulate $N = 100\,000$ sets of $I(1)$ series for y_t and x_t (no constant, hence no drift) and regress then y_t on just x_t, ie, estimate the '*regression model*' $y_t = \beta x_t + u_t$. The Matlab code for this simulation is below.

```
% example spurious regression with two independent random walks.
clear all;clc;
% check matlabpool if open
if ~matlabpool('size') > 0; matlabpool; end
% sample size and number of simulations
%%
T = 1e2;
Nsim = 1e5;
seed(1234);
c = 0.0;
tic;
% level series
y = cumsum(c+randn(T,Nsim));
x = cumsum(c+randn(T,Nsim));

% differenced series
dy = y(2:T,:)-y(1:T-1,:);
dx = x(2:T,:)-x(1:T-1,:);

% storage allocation
bhat = zeros(Nsim,1);
tstat = zeros(Nsim,1);

% main loop
parfor n = 1:Nsim
  olsout   = fastols(y(:,n), x(:,n));
  bhat(n,:)  = olsout.bhat;
  tstat(n,:) = olsout.tstat;
  % based on differenced data
  olsoutd   = fastols(dy(:,n), dx(:,n));
  bhatd(n,:) = olsoutd.bhat;
  tstatd(n,:) = olsoutd.tstat;
```

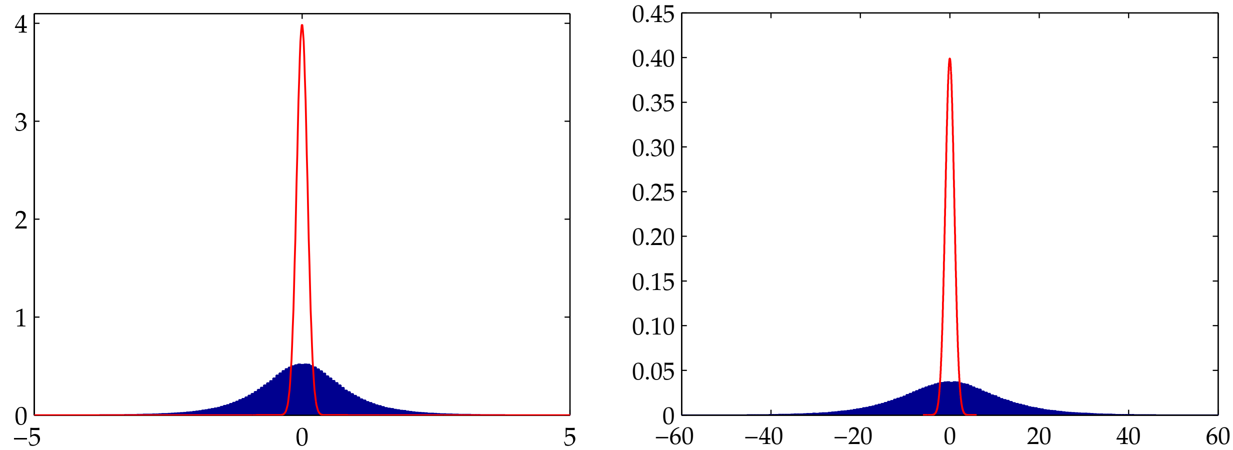
```
end;
toc;

%% level series
% p-value for 95% CI (it is not 5%)
fprintf('Pr(|t-stat|>1.96) (should be 0.05): % 2.4f \n', mean(abs(tstat)>1.96))
% plots
clf;
% bhat
dims = [.4 .4];
%subplot(1,2,1);
setplot(dims,11);
h1 = histogram(bhat,300,[0 1/T]);
ylim([0 4.1])
xlim([-5 5])
setytick(1);
setyticklabels([0:1:5])
%print2pdf('../lectures/graphics/spurious_bhat');
%
% tstat
setplot(dims,10.50);
setytick(2);
h2 = histogram(tstat,300,1);
xlim([-60 60])
setytick(2);
%print2pdf('../lectures/graphics/spurious_tstat');

%% differenced data
% p-value for 95% CI (it is not 5%)
fprintf('Pr(|t-stat|>1.96) (should be 0.05): % 2.4f \n', mean(abs(tstatd)>1.96))
% plots
clf;
% bhat
dims = [.4 .4];
%subplot(1,2,1);
setplot(dims,11);
h1 = histogram(bhatd,300,[0 1/T]);
ylim([0 4.1])
xlim([-5 5])
setytick(1);
setyticklabels([0:1:5])
%print2pdf('../lectures/graphics/spurious_bhatd');
%
% tstat
setplot(dims,10.50);
setytick(2);
h2 = histogram(tstatd,300,1);
xlim([-60 60])
setytick(2);
%print2pdf('../lectures/graphics/spurious_tstatd');
```

Matlab Code 1: `example_spurious_regression.m`

The histograms, together with the corresponding distributions under a stationary scenario are plotted below in Figure 6. As we can see from Figure 6, the $t-$statistic in Panel (b) should be standard normal distributed (as indicated by the red densities) under our usual regression assumptions. This is clearly not the case. If one counts the number of



(a) Sampling distribution of $\hat{\beta}$.

(b) Sampling distribution of $t-$statistic.

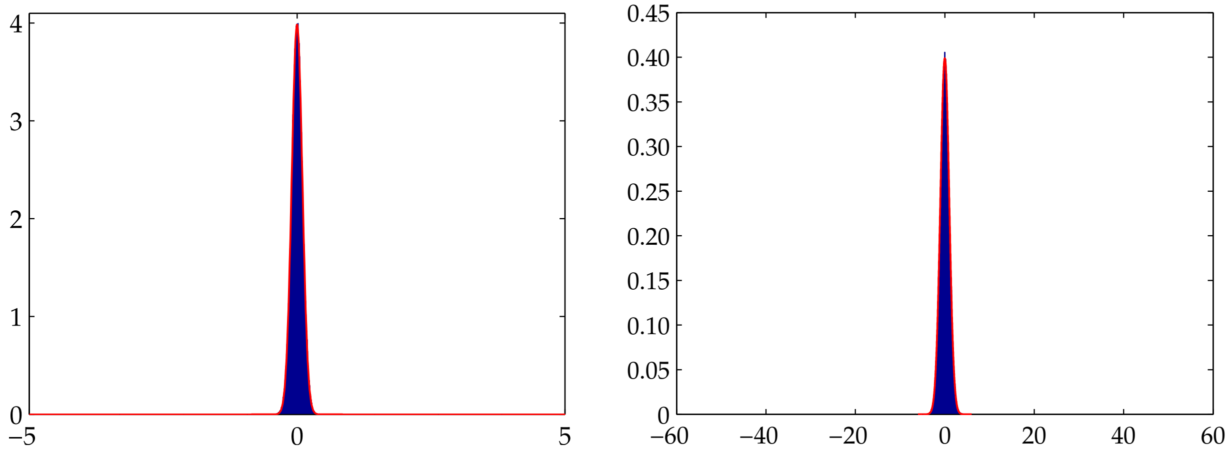
Figure 6: Sampling distribution of the OLS estimate and it's $t-$statistic.

times one gets a significant $t-$statistic when considering a 95% level of significance, this values is nearly 86%, when a sample size of $T = 100$ is used and we average over the $N = 100\,000$ simulations. This is evidently much larger than the nominal value that it should be of 5%. Thus using a standard $t-$statistic is meaningless in this scenario. Similarly, from Panel (a) we can see that the sampling distribution of our estimator is also much more dispersed in the $I(1)$ spurious regression case than what it should be if the data were $I(0)$.

It should be pointed out here that this is not a *small sample* problem hence does not disappear as the samples size increases. There is no way of remedying spurious regression problems in any other way but to difference the two $I(1)$ series and then use the difference series to look at the relationship between the variables of interest.[1]

To see what happens in the same simulation scenario that we have created when the differenced data are used in the regression, ie., $\Delta y_t = \beta \Delta x_t + u_t$, we can again plot the sampling distributions of $\hat{\beta}$ and its $t-$statistic. These distributions are shown below in Figure 7. As is evident from Figure 7, the distributions take on the expected shapes that asymptotic theory dictates. The proportion of significant $t-$statistics is 0.0525 so that the size of the $t-$test is expected.

[1] If the two series cointegrate, then things are different again and we will discuss this later.



(a) Sampling distribution of $\hat{\beta}$.

(b) Sampling distribution of $t-$statistic.

Figure 7: Sampling distribution of the OLS estimate and it's $t-$statistic.

2.2. Consequences of ignoring Cointegration

Should we always then work with differenced data? Let us now look at what happens when you have a set of $I(1)$ variables that form a cointegrating relationship but you decide to difference the data and use the differences to relate the variables. More specifically, consider the following cointegrated system with one common stochastic trend.

$$y_t = w_t + u_{yt} \tag{3a}$$

$$x_t = w_t + u_{xt} \tag{3b}$$

$$w_t = w_{t-1} + u_{wt} \tag{3c}$$

where $u_{it} \sim \text{WN}(0, \sigma^2_{u_i}), \forall i = y, x, w$ and $\text{Cov}(u_{yt}, u_{xs}) = \sigma_{u_{xy}}$ when $t = s$ and 0 otherwise and u_{wt} uncorrelated with u_{yt} and u_{xt}. That is,

$$\begin{bmatrix} u_{yt} \\ u_{xt} \\ u_{wt} \end{bmatrix} \sim WN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u_y} & \sigma_{u_{xy}} & 0 \\ \sigma_{u_{xy}} & \sigma^2_{u_x} & 0 \\ 0 & 0 & \sigma^2_{u_w} \end{bmatrix} \right) \tag{4}$$

Ignoring the cointegrating relation formed by the relations in (3) and differencing yields to

$$\begin{aligned} \Delta y_t &= \Delta w_t + \Delta u_{yt} \\ &= u_{wt} + \Delta u_{yt} \end{aligned}$$

and

$$\begin{aligned}\Delta x_t &= \Delta w_t + \Delta u_{xt} \\ &= u_{wt} + \Delta u_{xt}\end{aligned}$$

Now, estimating the relation by OLS regression

$$\Delta y_t = \gamma \Delta x_t + \epsilon_t$$

gives an estimate of γ. This estimate will have the following property:

$$\begin{aligned}\hat{\gamma} &= \frac{T^{-1}\sum_{t=1}^{T}\Delta x_t \Delta y_t}{T^{-1}\sum_{t=1}^{T}\Delta x_t^2} \\ &= \frac{T^{-1}\sum_{t=1}^{T}(u_{wt}+\Delta u_{xt})(u_{wt}+\Delta u_{yt})}{T^{-1}\sum_{t=1}^{T}(u_{wt}+\Delta u_{xt})^2} \\ &= \frac{T^{-1}\sum_{t=1}^{T}(u_{wt}^2+\Delta u_{xt}u_{wt}+u_{wt}\Delta u_{yt}+\Delta u_{xt}\Delta u_{yt})}{T^{-1}\sum_{t=1}^{T}(u_{wt}^2+2u_{wt}\Delta u_{xt}+\Delta u_{xt}{}^2)} \qquad (5)\end{aligned}$$

where the plims of the individual terms in (5) are as follows:

$$\begin{aligned}\operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} u_{wt}^2 &= \sigma_{u_w}^2 \\ \operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \Delta u_{xt}u_{wt} &= \underbrace{E(\Delta u_{xt}u_{wt})}_{=0 \text{ by assumption}} \\ \operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} u_{wt}\Delta u_{yt} &= \underbrace{E(u_{wt}\Delta u_{yt})}_{=0 \text{ by assumption}} \ .\end{aligned}$$

Further, for the other relations we obtain:

$$\begin{aligned}\operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \Delta u_{xt}\Delta u_{yt} &= E(\Delta u_{xt}\Delta u_{yt}) \\ &= 2E(u_{xt}u_{yt}) \\ &= 2\sigma_{u_x u_y}\end{aligned}$$

$$\operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\Delta u_{xt}^2 = E(\Delta u_{xt}^2)$$
$$= 2\sigma_{u_x}^2.$$

The plim of the OLS estimate $\hat{\gamma}$ is then:

$$\operatorname*{plim}_{T\to\infty} \hat{\alpha} = \frac{\sigma_{u_w}^2 + 2\sigma_{u_x u_y}}{\sigma_{u_w}^2 + 2\sigma_{u_x}^2}. \quad (6)$$

Thus, unless $\sigma_{u_x}^2 = \sigma_{u_x u_y}$, there is a wedge between the true value of $\gamma = 1$ and $\hat{\gamma}$, and the size of this wedge will depend on the magnitude and sign of $\sigma_{u_x u_y}$. From this we see that ignoring a cointegrating relationship creates a bias in the OLS regression coefficients.

3. The Cointegrated VAR model

3.1. Formal Definitions

Let us now formally define cointegrating relations and the error-correction representation, also sometimes known as equilibrium correction, in a general VAR(p) model. To motivate the set-up, consider initially the $k-$variable VAR(1) model (with zero mean here for simplicity) which we can write in its Vector Error Correction Model (VECM) form as:

$$\begin{aligned}
A(L)X_{t-1} &= U_t \\
(I_k - A_1 L)X_t &= U_t \\
X_t &= A_1 X_{t-1} + U_t \\
\Delta X_t &= (A_1 - I_k)X_{t-1} + U_t && [-X_{t-1}] \\
\Delta X_t &= -\underbrace{(I_k - A_1)}_{\Pi = -A(1)} X_{t-1} + U_t \\
\Delta X_t &= \Pi X_{t-1} + U_t. && (7)
\end{aligned}$$

The relation in (7) is the VECM form of the VAR(1). Similarly, for a VAR(2), we get

$$\begin{aligned}
A(L)X_{t-1} = U_t, \text{ with } A(L) &= (I_k - A_1 L - A_2 L^2) \\
X_t &= A_1 X_{t-1} + A_2 X_{t-2} + U_t \\
\Delta X_t &= -(I_k - A_1)X_{t-1} + A_2 X_{t-2} + U_t && [-X_{t-1}]
\end{aligned}$$

$$
\begin{aligned}
\Delta X_t &= -(I_k - A_1 - A_2)X_{t-1} + \underbrace{A_2 X_{t-2} - A_2 X_{t-1}}_{=-A_2\Delta X_{t-1}} + U_t \qquad [\pm A_2 X_{t-1}] \\
\Delta X_t &= \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + U_t. \qquad (8)
\end{aligned}
$$

where $\Pi = -A(1)$ and $\Gamma_1 = -A_2$.

For a VAR(3), we get

$$
\begin{aligned}
A(L)X_{t-1} &= U_t, \text{ with } A(L) = (I_k - A_1 L - A_2 L^2 - A_3 L^3) \\
X_t &= A_1 X_{t-1} + A_2 X_{t-2} + A_3 X_{t-3} + U_t \\
\Delta X_t &= -(I_k - A_1)X_{t-1} + A_2 X_{t-2} + A_3 X_{t-3} + U_t \qquad [-X_{t-1}] \\
\Delta X_t &= -(I_k - A_1)X_{t-1} + (A_2 + A_3)X_{t-2} - A_3 \Delta X_{t-2} + U_t \quad [\pm A_3 X_{t-2}] \\
\Delta X_t &= -(I_k - A_1 - A_2 - A_3)X_{t-1} + \underbrace{(A_2 + A_3)X_{t-2} - (A_2 + A_3)X_{t-1}}_{-(A_2+A_3)\Delta X_{t-1}} \\
&\qquad [\pm (A_2 + A_3)X_{t-1}] \\
&\quad + \Gamma_1 \Delta X_{t-1} + U_t \\
\Delta X_t &= \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + U_t. \qquad (9)
\end{aligned}
$$

where $\Pi = -A(1)$, $\Gamma_1 = -A_2 - A_3$ and $\Gamma_2 = -A_3$.

Now you can see how this generalises to the k dimensional VAR(p) model as:

$$
\begin{aligned}
A(L)X_t &= U_t \qquad (10) \\
X_t &= A_1 X_{t-1} + A_2 X_{t-2} + \ldots + A_p X_{t-p} + U_t
\end{aligned}
$$

where $A(L) = I + A_1 L + A_2 L^2 + \ldots + A_p L^p$. VAR in (10) can be written as the VECM

$$
\Gamma(L)\Delta X_t = \Pi X_{t-1} + U_t \qquad (11)
$$

where

$$
\begin{aligned}
\Pi &= -A(1) = -(I - A_1 - A_2 - \ldots A_p) \\
\Gamma(L) &= I_k - \Gamma_1 L - \Gamma_2 L^2 - \Gamma_3 L^3 - \ldots - \Gamma_{p-1} L^{p-1}
\end{aligned}
$$

and

$$
\Gamma_j = -\sum_{i=j+1}^{p} A_i.
$$

We have seen in Lecture 3 (equation 91) that we can always factor a lag polynomial as:

$$\Psi(L) = \Psi(1) + \Delta\tilde{\Psi}(L).$$

Making use of this fact for the $A(L)$ polynomial, that is, factor

$$A(L) = A(1) + \Delta\Gamma(L) \tag{12}$$

we can see that the VECM representation in (11) is nothing else than a multivariate Beveridge-Nelson decomposition, that is, (10) becomes

$$\begin{aligned} A(L)X_t &= U_t \\ A(1) + \Delta\Gamma(L)X_t &= U_t \\ \Delta\Gamma(L)X_t &= \underbrace{-A(1)}_{\Pi} + U_t \\ \Gamma(L)\Delta X_t &= \Pi X_{t-1} + U_t. \end{aligned} \tag{13}$$

From the VECM representation of the VAR in (13) we can now also see that '*running a regression in first differences*' is equivalent to setting $\Pi = 0$, so this is a restriction that is put on the whole VECM system. If this is not supported by the data, ie., the truth is $\Pi = 0$, then this has the same effect as omitting an important variable from a cross-sectional regression, that is, lead to ommitted variable bias.

Also, notice here that if there exists cointegration in the X_t vector, then Π is of reduced rank and can be factored into

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

where α and β are as before the speed of adjustment parameters and the cointegrating vectors. What does it mean for Π to be of reduced rank? When a matrix is of reduced rank, then the number of linearly independent columns is less than k, so that $\det(\Pi) = 0$ and hence inverse of Π does not exist. This is important because now there is no other way for the VAR in differences to be balanced in terms of stationarity. To see this, suppose Π was not of reduced rank. Then we could take the inverse of Π, and could write the system in (13) as:

$$\underbrace{\Pi^{-1}\Gamma(L)\Delta X_t}_{I(0)} = \underbrace{X_{t-1}}_{I(1)} + \underbrace{\Pi^{-1}U_t}_{I(0)}. \tag{14}$$

But we have assumed that ΔX_t is $I(0)$ and also that $U_t \sim \text{WN}(0, \Sigma)$ which is $I(0)$, so this equation will not balance in terms of orders of integration. Therefore, Π must be of reduced rank and hence form a cointegrating relation that makes $\beta' X_t \sim I(0)$ for the VECM to be consistent with the assumptions that are imposed.

Formally, we have the following definitions.

Definition 2 (Cointegration): *The components of the k dimensional vector X_t, are said to be cointegrated of order d, b, denoted $X_t \sim CI(d, b)$, if*

(i) all components of X_t are $I(d)$;

(ii) $\exists\, \beta\, (\neq 0)$ so that $\beta' X_t \sim I(d - b),\ b > 0$.

The vector is called the co-integrating vector.

Also,

Definition 3 (Error-correction representation): *A vector time series X_t, has an error correction representation if it can be expressed as:*

$$\Gamma(L)\Delta X_t, = \alpha\beta' X_{t-1} + U_t$$

where U_t is a stationary multivariate disturbance, with $\Gamma(L) = I_k - \Gamma_1 L - \Gamma_2 L^2 - \Gamma_3 L^3 - \ldots,$ $\Gamma(0) = I_k$, $\Gamma(1)$ is finite, and $\alpha \neq 0$. (Point 4 of the Granger Representation theorem)

Remark 1 (Some points to keep in mind:):

- *β has dimension $(k \times r)$, where the cointegrating rank r is equal to the number of linearly independent cointegrating vectors and k is the dimension of X_t.*
- *the cointegrating vectors are the columns of β.*
- *the speed of adjustment vectors are the columns of α.*
- *the number of common stochastic trends (unit-roots) that remain in system is equal to $(k - r)$.*
- *the decomposition of $\Pi = \alpha\beta'$ is not unique*

Let us now look at a few examples.

3.2. Examples of cointegrated VAR models

3.2.1. Example 1: 2 variables, 1 cointegrating vector

Consider the '*triangular system*' representation (see Phillips (1991))

$$x_{1t} = \beta_2 x_{2t} + u_{1t}$$
$$\Delta x_{2t} = u_{2t}$$

where $u_{it} \sim I(0), \forall i = 1,2$. Here we will only assume $I(0)$ or $I(1)$ for simplicity, so u_t could be white noise, but does not need to be and could also be a stationary ARMA process.

The term $x_{2t} = x_{2,0} + \sum_{j=1}^{t} u_{2j}$ is the common stochastic trend in this example, with $x_{2,0}$ being the initial condition. The cointegrating vector is $\beta = [1, -\beta_2]'$. To check this, it must be the case that, given $x_{1t} \sim I(1)$ and $x_{2t} \sim I(1)$, $\beta' X_t \sim I(0)$. To confirm this, write out $\beta' X_t$ as:

$$\begin{aligned}
\begin{bmatrix} 1 & -\beta_2 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} &= x_{1t} - \beta_2 x_{2t} \\
&= \beta_2 x_{2t} + u_{1t} - \beta_2 (x_{2t-1} + u_{2t}) \\
&= \beta_2 \Delta x_{2t} + u_{1t} - \beta_2 u_{2t} \\
&= \beta_2 u_{2t} + u_{1t} - \beta_2 u_{2t} \\
&= u_{1t} \sim I(0).
\end{aligned}$$

3.2.2. Example 2: 3 variables, 1 cointegrating vector

Consider the system

$$\begin{aligned}
x_{1t} &= \beta_2 \sum_{j=1}^{t} u_{2j} + \beta_3 \sum_{j=1}^{t} u_{3j} + u_{1t} \\
\Delta x_{2t} &= u_{2t} \\
\Delta x_{3t} &= u_{3t}
\end{aligned}$$

where $u_{it} \sim I(0), \forall i = 1,2,3$ and the intimal conditions can be set to 0 again. We can see that all elements of $X_t \sim I(1)$, but that $\beta' X_t \sim I(0)$, that is:

$$\begin{bmatrix} 1 & -\beta_2 & -\beta_3 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = x_{1t} - \beta_2 x_{2t} - \beta_3 x_{3t}$$

$$\begin{bmatrix} 1 & -\beta_2 & -\beta_3 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \beta_2 \sum_{j=1}^{t} u_{2j} + \beta_3 \sum_{j=1}^{t} u_{3j} + u_{1t} - (\beta_2 x_{2t} + \beta_3 x_{3t})$$

but $x_{it} = x_{i,0} + \sum_{j=1}^{t} u_{ij}$ for $i = 2,3$, so that

$$\begin{aligned} \begin{bmatrix} 1 & -\beta_2 & -\beta_3 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} &= \beta_2 \sum_{j=1}^{t} u_{2j} + \beta_3 \sum_{j=1}^{t} u_{3j} + u_{1t} \\ &- \left(\beta_2 \sum_{j=1}^{t} u_{2j} + \beta_3 \sum_{j=1}^{t} u_{3j} \right) \\ &= u_{1t} \sim I(0) \end{aligned}$$

3.2.3. Example 3: 3 variables, 2 cointegrating vectors

Suppose you have the system:

$$\begin{aligned} x_{1t} &= \beta_{13} x_{3t} + u_{1t} \\ x_{2t} &= \beta_{23} x_{3t} + u_{2t} \\ x_{3t} &= x_{3t-1} + u_{3t} \end{aligned}$$

where again $u_{it} \sim I(0), \forall i = 1,2,3$. x_{3t} follows a pure random walk thus $x_{3t} = x_{3,0} + \sum_{j=1}^{t} u_{3j}$ is the common stochastic trend.

What are the two cointegrating vectors? Let's look at $\beta' X_t \sim I(0)$ for $\beta = [1, 0, -\beta_{13}]'$

$$\begin{bmatrix} 1 & 0 & -\beta_{13} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = x_{1t} + 0 x_{2t} - \beta_{13} x_{3t}$$

$$\begin{bmatrix}1 & 0 & -\beta_{13}\end{bmatrix}\begin{bmatrix}x_{1t}\\ x_{2t}\\ x_{3t}\end{bmatrix} = \beta_{13}x_{3t} + u_{1t} - \beta_{13}\left(x_{3t-1} + u_{3t}\right)$$

$$= \beta_{13}\Delta x_{3t} + u_{1t} - \beta_{13}u_{3t}$$

$$= u_{1t} \sim I(0)$$

so $\beta = [1, 0, -\beta_{13}]'$ is a cointegrating vector.

What about the second cointegrating vector? $\beta = [0, 1, -\beta_{23}]'$ gives

$$\begin{bmatrix}0 & 1 & -\beta_{23}\end{bmatrix}\begin{bmatrix}x_{1t}\\ x_{2t}\\ x_{3t}\end{bmatrix} = 0\beta_{13}x_{3t} + \beta_{23}x_{3t} + u_{2t} - \beta_{23}\left(x_{3t-1} + u_{3t}\right)$$

$$= \beta_{23}\Delta x_{3t} + u_{2t} - \beta_{13}u_{3t}$$

$$= u_{2t} \sim I(0)$$

Example 1 (Cointegration in VARs: example with numbers): There is a lot more theory, but let's look at an example: Let the VAR(2) be given by

$$X_t = \underbrace{\begin{bmatrix}0.65 & 0.11 & -0.1454\\ -0.27 & 1.28 & -0.0358\\ -0.81 & 0.43 & 0.4962\end{bmatrix}}_{A_1} X_{t-1} + \underbrace{\begin{bmatrix}0.12 & 0.09 & 0.16\\ 0.21 & -0.21 & 0.02\\ 0.70 & -0.17 & 0.33\end{bmatrix}}_{A_2} X_{t-2} + U_t$$

Is this model stationary? Forming the companion form and looking at eig($\mathbf{A}$) where

$$\mathbf{A} = \begin{bmatrix}A_1 & A_2\\ I_3 & 0_3\end{bmatrix}$$

gives -0.6349, 0.0308, 0.3676, 0.7574, **1.0000**, 0.9053. So one root is equal to one and the others are less then one. There is one common stochastic trend (one unit root) and there are 2 cointegrating vectors.

Let's form the VECM from the cointegrated VAR as:

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + U_t$$

where

$$
\begin{aligned}
\Pi &= -A(1) \\
&= -\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.65 & 0.11 & -0.1454 \\ -0.27 & 1.28 & -0.0358 \\ -0.81 & 0.43 & 0.4962 \end{bmatrix} - \begin{bmatrix} 0.12 & 0.09 & 0.16 \\ 0.21 & -0.21 & 0.02 \\ 0.70 & -0.17 & 0.33 \end{bmatrix} \right)
\end{aligned}
$$

The rank of Π is 2, thus there are two cointegrating vectors. The eigenvalues of Π are 0 and $-0.1669 \pm 0.0396i$. Recall that a VAR system will be cointegrated if some eigenvalues of Π are non-zero.

We can now put Π in reduced row echelon form to find the coefficients of the cointegrating relations. In Matlab type rref(Π) to get:

$$
\mathsf{rref}(\Pi) = \begin{bmatrix} 1 & 0 & -1.02 \\ 0 & 1 & -1.1 \\ 0 & 0 & 0 \end{bmatrix}
$$

thus 2 cointegrating vectors with $\beta = \begin{bmatrix} 1 & 0 & -1.02 \\ 0 & 1 & -1.1 \end{bmatrix}'$. You can then solve for α as

$$
\begin{aligned}
\alpha\beta' &= \Pi \\
\alpha\beta'\beta'^{+} &= \Pi\beta'^{+} \\
\alpha &= \Pi\beta'^{+}
\end{aligned}
$$

where β'^{+} is the generalised or Moore-Penrose inverse of β' (in Matlab pinv)

So α is

$$
\begin{aligned}
\alpha &= \underbrace{\begin{bmatrix} -0.23 & 0.20 & 0.0146 \\ -0.06 & 0.07 & -0.0158 \\ -0.11 & 0.26 & -0.1738 \end{bmatrix}}_{\Pi} \underbrace{\begin{bmatrix} 0.6799 & -0.3452 \\ -0.3452 & 0.6277 \\ -0.3138 & -0.3384 \end{bmatrix}}_{\beta'^{+}} \\
&= \begin{bmatrix} -0.23 & 0.20 \\ -0.06 & 0.07 \\ -0.11 & 0.26 \end{bmatrix}
\end{aligned}
$$

so we get the first two columns of the Π matrix.

4. Estimation of Cointegration relations

Cointegrating relations can fundamentally be estimated by two different approaches. 1) a single equation approach and 2) a systems approach. As an alternative, one may choose not to estimate the cointegrating relation at all and impose the cointegrating restriction base upon economic theory. This evidently works only if one knows what the long-run relationship should be in terms of the parameters. For instance, from the term structure example above, we know for instance that if bond markets are efficient, then the 1 year and the 5 and 10 year rates should form a long-run relationship with two cointegrating vectors so that $\beta' X_{t-1}$ should take the from

$$\beta' X_{t-1} = \underbrace{\begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}}_{=\beta'} \underbrace{\begin{bmatrix} 1y \\ 5y \\ 10y \end{bmatrix}}_{=X_{t-1}}.$$

So here we have a pretty strong view on what β should look like based on economic/finance theory.

In the first example we had the relation between real consumption and income. In this example there is no theory that tells us what the relation should be between real income and consumption. All we know is the the income elasticity of consumption should be less than 1 in the long-run. Thus here we are interested in finding out what this long-run elasticity is form the data.

4.1. Single Equation Methods

Single equation estimation of cointegration relations was the first proposed method in the original Engle-Granger framework. It has some fairly stringent assumptions placed onto it though, so this can be a major disadvantge. These are:

- only one cointegrating vector exists.
- x_t and $y_t \sim I(1)$ but $(x_t, y_t) \sim CI(1,1)$
- we know what goes on the left hand side and what on the right (the normalisation)

Nevertheless, as we have seen that there can be potential issues with modelling real world (non-experimental) problems in a VAR framework due to the need to select the right model dimension as well as the lag structure, on top of the many unknown parameters problem, it is not that clear wether a VAR approach is uniformley better.

The Engle-Granger (EG) 2-step approach proceeds as follows:

1. Regress y_t on x_t and a constant (and possibly a time trend, dummies etc.)

$$y_t = c + \beta_2 x_t + \varepsilon_t \tag{15}$$

2. form residuals from the regression in 1) as $\hat{\varepsilon}_t = y_t - \hat{c} - \hat{\beta} x_t$
3. test for a unit-root in $\hat{\varepsilon}_t$ by running the DF regression (using tables in E

$$\Delta\hat{\varepsilon}_t = \rho\hat{\varepsilon}_{t-1} + e_t$$

(no constant in this regression, why?). NOTE that because $\hat{\varepsilon}_t$ are constructed based on estimated quantities (ie., the $\hat{c}$ and $\hat{\beta}$ term), we cannot treat it as know. This is classic constructed regressor problem. Hence, there are *special* DF critical values available that need to be use with the EG procedure.

If $\hat{\varepsilon}_t \sim I(0)$ we have an equilibrium relation with cointegrating vector $\beta = (1, -\beta_2)'$ plus constant.

Given we know our disequilibrium error in the last period, we can use $\hat{u}_{t-1}$ to form the Error correction representation

$$\begin{aligned} A_y(L)\Delta y_t &= c_y - \alpha_y \hat{\varepsilon}_{t-1} + u_{1t} \\ A_x(L)\Delta x_t &= c_x - \alpha_x \hat{\varepsilon}_{t-1} + u_{2t} \end{aligned}$$

for an arbitrary lag order VECM, where $\hat{\varepsilon}_{tt-1} = \beta'[y_{t-1},\ x_{t-1}]'$ (plus constant) is the cointegrating relation. The VECM can then be estimate by OLS. The terms α_y and α_x are "speed of adjustment" pacemakers, that tell us how quickly either of the left hand variables adjusts to deviations from long-run path. The term '*weak exogeneity*' is often used to refer to a variable that does not respond to the long-run equilibrium when one of the $\alpha_j, \forall j = y, x$ is equal to zero.

For example, when $\alpha_x = 0$, then x_t is said to weakly exogenous and in our case thus collapses to a random walk model. To see this, set the lag polynomials $A_j(L) = I, \forall j =$

y, x, so that we get (ignoring constants in the CI relation) the triangular system

$$\Delta y_t = c_y - \alpha_y \hat{\varepsilon}_{t-1} + u_{1t} \tag{16}$$

$$\Delta x_t = c_x + u_{2t} \tag{17}$$

so that

$$y_t = c_y + (1 - \alpha_y) y_{t-1} + \beta_2 x_{t-1} + u_{1t}$$

$$x_t = x_0 + c_x t + \sum_{i=1}^{t} u_{2i}$$

where the x_t relation is the common stochastic trend.

It is possible to test for cointegration in this set up also by testing $\alpha_y = 0$ in (16) directly. These tests have been shown to be more powerful than the EG 2-step procedure. However, the test hinges on the following assumptions:

- the cointegrating vector is being imposed so only possible for certain economic relations
- assumes a lower triangular structure, so an exogeneity restriction on the system

If these are not satisfied, it is not clear if the improvement in power still goes through.

There exist many other cointegration tests for single equation models. Some well known ones are the Dynamic OLS (DOLS) approach of Stock and Watson (1993) or the AutoRegressive Distributed Lag of Pesaran-Shin-Smith (1998,2001)

These approaches try to correct for the correlation between x_t and ε_t in (15) by running the OLS regression with lags and leads of Δx_t, that is, instead of

$$\begin{aligned} y_t &= c + \beta_2 x_t + \gamma_1 \Delta x_{t-1} + \gamma_2 \Delta x_{t-2} + \ldots + \gamma_p \Delta x_{t-p} \\ &\quad + \varpi_1 \Delta x_{t+1} + \varpi \Delta x_{t+2} + \ldots + \varpi_q \Delta x_{t+q} + \varepsilon_t \end{aligned}$$

and use the cointegrating vector from this relation.

4.2. System Estimators based on the VECM

4.2.1. Johansen's FIML Estimator

The VECM relation in (13) is very useful because it enables us to be precise about when cointegration doesn't hold. In these cases there is no β for which $\beta' X_t$ is $I(0)$ $(r = 0)$ and so Π is always of full rank and so $A(1)$ will be zero rank i.e $A(1) = 0$ and the model is an AR in ΔX_t i.e. there are no error correction terms. This either suggests that we test the rank of $A(1)$ or whether $\alpha = 0$. There are some distributional issues about the latter so the former has become the standard approach.

Since the rank of $A(1)$ depends on how many non-zero eigenvalues it has, Johansen focussed on these. Let the eigenvalues of $A(1)$ be ordered as $\lambda_1, ..., \lambda_k$. Then if we took the one closest to zero (generally the eigenvalues will be positive and so this will be λ_k) we could form -$T \ln(1 - \hat{\lambda}_k)$ as a test. If $\hat{\lambda}_k = 0$ this test statistic would have the value zero and the further $\hat{\lambda}_k$ is away from zero the bigger the test values. The maximum value of the true eigenvalue $\lambda_k = 1$ i.e. the k unit roots case.

One can then adapt this test to test not rank zero but rank k by testing if $\hat{\lambda}_{k-r}$ is zero etc. This is called the max *test*. The *trace test* looks at the trace of $A(1)$, which is the sum of the eigenvalues. If there is no co-integration then all eigenvalues are zero so that the trace of $A(1)$ would be zero. This is tested by something that is not strictly a trace but motivated by it viz. the statistic -$T \sum_{j=1}^{k} \ln(1 - \lambda_j)$. The scaling factor of T is used since the distribution of the eigenvalues essentially involve the normalizing factor of T. The distribution is non-standard. Later we will see that the estimators of cointegrating vectors generally have non-standard distributions and these are part of $A(1)$, so it is not surprising that this is also true of the estimated eigenvalues.

The most widely used method of estimating systems of equations with cointegrating relations is that of Johansen who assumed that the X_t followed a VAR. There are obvious analogues here with unit root testing. In the Dickey-Fuller approach it was assumed that X_t was an AR while in the Phillips and Perron test one did not need such an assumption. Johansen is essentially a vector extension of the Dickey-Fuller approach and the Phillips/Hansen work is the multivariate equivalent of the Phillips-Perron framework.

The VAR(1) with cointegration has the form

$$\Delta X_t = \alpha\beta' X_{t-1} + U_t$$

and, if the U_t are $i.i.d.MN(0,\Sigma)$, the log likelihood of this system is

$$\mathscr{L} = -\frac{(T-1)k}{2}\ln(2\pi) - \frac{(T-1)}{2}\ln(\Sigma) - \frac{1}{2}\sum_{t=2}^{T}(\Delta X_t - \alpha\beta' X_{t-1})'\Sigma^{-1}(\Delta X_t - \alpha\beta' X_{t-1})$$

Johansen then maximized $\mathscr{L}$ w.r.t. α, β and Σ. From the form of the likelihood it is clear that, if β was known, we would just obtain estimators of α and Σ by performing OLS on each equation (this will be the same as SURE since all the regressors are identical). Hence one can concentrate them out of $\mathscr{L}$ and then just solve the optimization w.r.t. β.

In fact the structure is identical to that of the LIML estimator in simultaneous equations work and the Cowles Commission solved that by finding the eigenvalues and vectors of a set of covariance matrices. Johansen does likewise. This means that there are no iterations and so convergence to the MLE's is assured. Precise details on this algorithm are available in quite a few sources. The estimator of β has a non-standard distribution just as in the discussion above, so testing hypotheses about the cointegrating vectors needs to be done carefully as any test statistics which utilize $\hat{\beta}$ will therefore have non -standard distributions.

Johansen (1991) presents an algorithm using a full-information maximum likelihood estimation to test for the number of cointegrating vectors in equation (13). To implement this, two auxiliary OLS regressions need to be performed. These are:

$$\Delta X_t = \hat{\pi}_0 + \hat{\Xi}_1 \Delta X_{t-1} + \hat{\Xi}_2 \Delta X_{t-2} + \hat{u}_t \tag{18}$$

and

$$X_{t-1} = \hat{\theta} + \hat{\aleph}_1 \Delta X_{t-1} + \aleph \Delta X_{t-2} + \hat{v}_t \tag{19}$$

where ΔX_t is the first difference of X_t, Ξ and $\aleph$ stand for an $(k \times k)$ matrix of OLS coefficient estimates and $\hat{u}_t$ and $\hat{v}_t$ denotes the $(k \times 1)$ vector of OLS residuals. k denotes the dimension of the vector X_t. Secondly, the OLS residuals $\hat{u}_t$ and $\hat{v}_t$ are used to calculate sample variance-covariance matrices

$$\hat{\Sigma}_{VV} \equiv (1/T)\sum_{t=1}^{T}\hat{v}_t\hat{v}_t' \qquad \hat{\Sigma}_{UU} \equiv (1/T)\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$$

$$\hat{\Sigma}_{UV} \equiv (1/T)\sum_{t=1}^{T}\hat{u}_t\hat{v}_t' \qquad \hat{\Sigma}_{VU} \equiv \hat{\Sigma}_{UV}' \tag{20}$$

Then, the eigenvalues of the matrix

$$\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}\hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV} \tag{21}$$

are ordered in a descending sequence $\hat{\lambda}_1 > \hat{\lambda}_3 > ... > \hat{\lambda}_k$ and can be used to determine the cointegration rank r. Precisely r eigenvalues are positive for k $I(1)$ variables with a cointegration rank r and the other $k - r$ eigenvalues are asymptotically zero. Furthermore, two different procedures exist to test the rank of r. First, the trace test assesses the null hypothesis, that there are at most r positive eigenvalues whereby the alternative is that there are more than r positive eigenvalues. The test statistic takes the form

$$Tr(r) = -T \sum_{i=r+1}^{k} \ln(1 - \hat{\lambda}_i) \tag{22}$$

The maximum eigenvalue test has the null hypothesis that there are exactly r positive eigenvalues against the alternative that there are precisely $r + 1$ positive eigenvalue. The corresponding test statistic is given by

$$\lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \tag{23}$$

Critical values are supplied in Johansen (1991) and also in Osterwald-Lenum (1992).

4.3. The ”Common Trends” Representation

In a similar way to what we have assumed earlier let the series X_t be written as

$$\Delta X_t = \Psi(L)U_t \tag{24}$$

ie., the X_t follow a VMA(∞) process where the U_t are now taken to be $i.i.d(0, V)$. From that earlier lecture we can write this as

$$\Delta X_t = \Psi(1)U_t + \Psi^*(L)\Delta U_t$$

and then it is a simple extension of the univariate analysis that the multivariate Beveridge-Nelson permanent component would be

$$X_t^P = \Psi(1) \sum_{t=1}^{T} U_t.$$

Notice that $\Psi(1)$ is a matrix so that the permanent component is potentially influenced by all the k shocks U_t unlike in the univariate case where there was a single shock affecting the permanent component. In general we could have k permanent components in the k series.

Now let us look at the cointegration case. This says that

$$\beta' X_t = \beta'(X_t^P + X_t^T) = \beta'\Psi(1)\sum_{t=1}^{T} U_t + \beta' X_t^T$$

is $I(0)$. In particular the $var(\beta' X_t)$ must be bounded. But, because the variance of $\sum_{t=1}^{T} U_t$ will rise with T, it is clear that the only way we can get $var(\beta' X_t)$ to be bounded is if $\beta'\Psi(1) = 0$. Hence a first implication of co-integration is that the matrix $\Psi(1)$ is not of full rank.

Now the restriction $\beta'\Psi(1) = 0$ will mean that $\Psi(1)$ must be of rank $(k-r)$. Hence we could factorize it as $\Psi(1) = JG$ where J is an $k \times (k-r)$ matrix of rank $(k-r)$ while G is $(k-r) \times k$ and of rank $(k-r)$. It is important to note that this decomposition is not unique i.e. $\Psi(1) = JFF^{-1}G = J^*G^*$, where F is an arbitrary non-singular $(k-r) \times (k-r)$ matrix. Using the decomposition we get

$$X_t^P = JG\sum_{t=1}^{T} U_t$$

and, defining $\tau_t = (J'J)^{-1}Jz_t^P$, we would have

$$\tau_t = G\sum_{t=1}^{T} U_t$$

which implies

$$\Delta\tau_t = Ge_t.$$

The τ_t are $(k-r)$ in number and are the $k-r$ *common permanent components* among the X_t i.e. the presence of r co-integrating vectors among the X_t means that that there are $k-r$ permanent components. When $r=0$ we have the maximum number of permanent components of k. Using Stock and Watson's description of the τ_t as "common trends" we see that this is the generalization of the Beveridge-Nelson decomposition to co-integrated series. Notice that the rank of the matrix β is r and this is the number of co-integrating vectors whereas the number of common permanent components is $k-r$. Once one knows either of these one can work out the other.

Which is the best way to think about co-integration? For some purposes the common trend approach is easier and it is often what people think about when they are describing co-integration. Suppose for example that we had the prices on a stock index in three counties. Designate these as p_{1t}, p_{2t} and p_{3t}. These should be $I(1)$ processes and, by arbitrage, we would expect to see that $p_{1t}-p_{2t}$ and $p_{1t}-p_{3t}$ would be $I(0)$. Consequently. with $X_t = \begin{bmatrix} p_{1t} \\ p_{2t} \\ p_{3t} \end{bmatrix}$ we would have

$$\beta_1' = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

But it's also true that $p_{1t}-p_{2t}$ and $p_{2t}-p_{3t}$ would be $I(0)$, since the first outcome would produce the second in that

$$p_{2t}-p_{3t} = p_{2t}-p_{1t}+p_{1t}-p_{3t}.$$

These second set of relations produces co-integrating vectors

$$\beta_2' = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

which emphasizes the non-uniqueness of co-integrating vectors. Now we also could think of these prices as having a single permanent component in them and it is that which we would see when we graph these series. In fact the idea of a *common trend* was around for a long time. Before co-integration people tended to think of it deterministically i.e. they regarded the series with the format $X_t = \mu t + u_t$ as having a common trend. Today we would say that these series *co-trend* i.e. the deterministic parts of the permanent components (i..e the μt) are related versus the situation of co-integration in which the stochastic parts of the permanent components are connected . In fact we will often need to think of X_t as being composed of a deterministic permanent component bt and a stochastic permanent component $G\sum U_{t-j}$ plus some transitory variation. In such instances we would write

$$\Delta X_t = \mu + C(L)U_t$$

and then $\beta'\mu$ may not be zero i.e. the co-integrating vector may not eliminate the deterministic trends. Indeed we may find that the co-trending vector i.e. the combination that eliminates the deterministic parts of any permanent components may be quite different from the co-integrating vectors which eliminate the stochastic parts.

The other advantage of the common trends approach is that it enables us to broaden the class of processes to which the idea of co-integration can be applied. Thus we might start with the idea that each of the series is driven by a common factor f_t and some idiosnycratic shock u_t i.e. we can postulate that

$$X_t = Jf_t + u_t$$

We can now specify the nature of f_t and u_t in many ways. For example we might have Δf_t being a TAR or MS model and we could treat u_t the same way. If f_t is $I(1)$ then this results in co-integration whenever the number of factors is less than k. This is essentially a multivariate components model. If one wanted to make it replicate the common (stochastic) trends approach above it is clear that one has to have the shocks into f_t and u_t being constructed from a common set of shocks U_t and one also needs to make some restrictions upon the autocorrelation structure of Δf_t and Δu_t. Factor models like this have become very popular in recent years. In some sense they are a bit more flexible than the approach described above which has ΔX_t being the linear process $C(L)U_t$, since one can account for some non-linear structure through an appropriate choice of a process for Δf_t.

References

Granger, Clive W. J. and Paul Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, **2**(2), 111–120.

Johansen, Sren (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, **59**(6), 1551–1580.

Osterwald-Lenum, Michael (1992): "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, **54**(3), 461–472.

Phillips, Peter C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, **33**(3), 311–340.

——— (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, **59**(2), 283–306.

Stock, James H. and Mark W. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, **61**(4), 783–820.

Yule, George Udny (1926): "Why Do We Sometimes Get Nonsense-Correlations between Time-Series? A Study in Sampling and the Nature of Time-Series," *Journal of the Royal Statistical Society*, **89**(1), 1–63.