universität innsbruck

Time Series Analysis

Cointegration

Contents

Contents

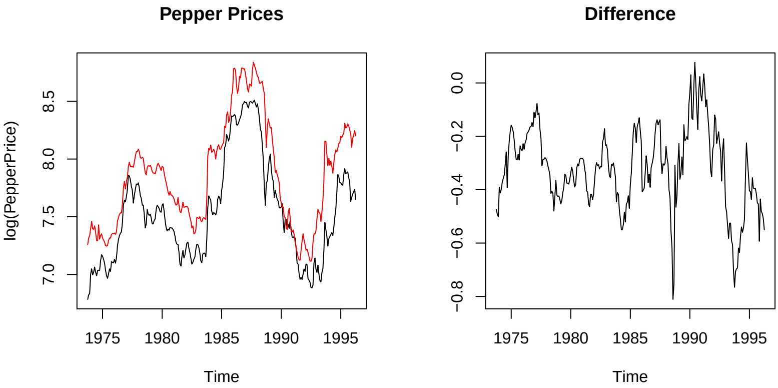
Cointegration

Motivation

Motivation

- Common features, such as
  - common trends,
  - common seasonality,
  - common volatility dynamics,
  - ...
- Cointegration: Economic time series often nonstationary but with common stochastic trends.
- Seminal paper: Engle RF, Granger CWJ (1987). "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*.
- Earlier work by Granger 1981, and in statistical literature: Box GEP, Tiao GC (1977). "A Canonical Analysis of Multiple Time Series," *Biometrika*.
- 2003 Nobel prize to CWJ Granger.

Motivation


Pepper Prices
log(PepperPrice)
7.0
7.5
8.0
8.5
1975
1980
1985
1990
1995
Time
Difference
−0.8
−0.6
−0.4
−0.2
0.0
1975
1980
1985
1990
1995
Time

Motivation

- Starting point: Several time series appear to be individually nonstationary, but certain linear combinations appear to be stationary.
- Cointegration studies effects of such combinations and relationships among components.
- Linear combinations may be known in advance or may have to be estimated.
- Cointegration methodology motivated by examples from macroeconomics and finance (empirical and theoretical).

Two approaches:

- Single-equation methods (Engle and Granger 1987).

  Problem: When does this make sense?
- System methods, using VARs (Johansen 1988, 1991).

Motivation

Examples: Cointegrating relations (Zivot & Wang 2006, p. 436).

- Term structure of interest rates: Nominal interest rates at different maturities. *Example:* y_t pair of interest rates. Then $\beta = (1, -1)^\top$ specifies that the spread is stationary.
- Permanent income: Consumption and income.
- Money demand: Money, income, prices, interest rates.
- Purchasing power parity (PPP): Nominal exchange rate, foreign, domestic prices.
- Fisher equation: Nominal interest rates, inflation.
- Commodities: Prices of close substitutes (e.g., black and white pepper).

Economic implications: Cointegration defines long-run equilibrium. Models also describe adjustment after deviations from equilibrium.

Integrated regressors: Spurious

Caution: Regression with integrated variables is different from regression with stationary variables.

Consider:

$$\begin{aligned} x_t &= x_{t-1} + \epsilon_t, \qquad \epsilon_t \sim WN(0, \sigma_\epsilon^2) \\ y_t &= y_{t-1} + \nu_t, \qquad \nu_t \sim WN(0, \sigma_\nu^2) \end{aligned}$$

where processes $\{\epsilon_t\}$ and $\{\nu_t\}$ are independent for all t.

Observation: In the meaningless regression

$$y_t = \alpha + \beta x_t + u_t$$

$\hat{\beta}$ is often significant, and R^2 is large.

Phenomenon: Known as *spurious regression*, first described by statistician Yule in 1926. Influential simulation study in econometrics by Granger and Newbold (*J Econometrics* 1974). Technical explanation by Phillips (*J Econometrics* 1986).

Integrated regressors: Spurious

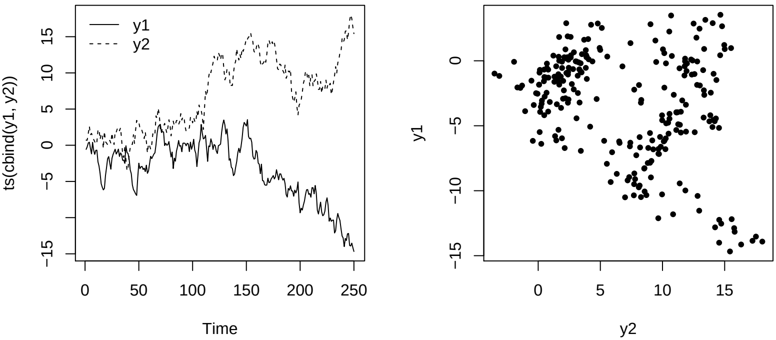
```
R> set.seed(4002)
R> e1 <- rnorm(250)
R> e2 <- rnorm(250)
R> y1 <- cumsum(e1)
R> y2 <- cumsum(e2)
R> summary(lm(y1 ~ y2))
Call:
lm(formula = y1 ~ y2)

Residuals:
   Min     1Q Median     3Q    Max
-8.991 -2.607  0.413  2.657  9.015

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.5274     0.4137   -3.69  0.00027
y2           -0.2693     0.0479   -5.62  5.1e-08

Residual standard error: 3.89 on 248 degrees of freedom
Multiple R-squared:  0.113,        Adjusted R-squared:  0.109
F-statistic: 31.6 on 1 and 248 DF,  p-value: 5.07e-08
```

Integrated regressors: Spurious



Integrated regressors: Cointegration

In general: Regressions with integrated processes are not meaningful.

Exception: A *common* nonstationary component.

Definition: Two processes $\{x_t\}$ and $\{y_t\}$ are cointegrated if

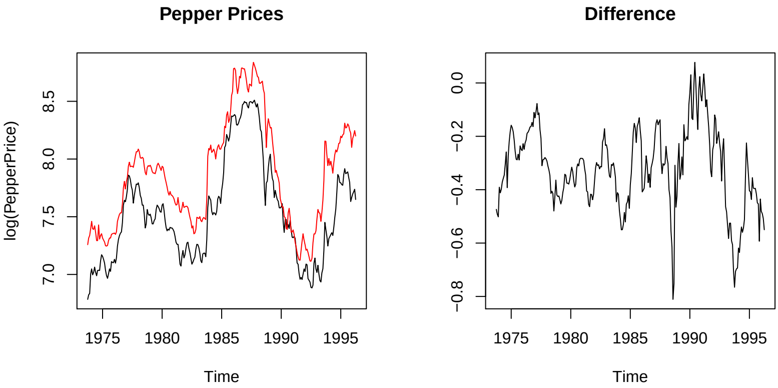
- both are I(1) processes,
- there exists a constant $b \neq 0$ such that $\{x_t - by_t\}$ is stationary.

Problem: How to test for cointegration?

Two approaches:

1. b is known: Compute $x_t - by_t$ and test using standard unit root test.
2. b is unknown. Regress y_t on x_t and test residuals for unit roots. Known as *residual-based test for cointegration*, or *Engle-Granger two-step method*.

Integrated regressors: Cointegration



Integrated regressors: Cointegration

Approach (1): b is 'known'.

- Compute $y_t - x_t$.
- Use unit root test with difference.

```
R> library("CADFtest")
R> CADFtest(log(PepperPrice[, "white"]) - log(PepperPrice[, "black"]),
+    type = "drift", max.lag.y = 5, criterion = "AIC")

        ADF test

data:  log(PepperPrice[, "white"]) - log(PepperPrice[, "black"])
ADF(0) = -3.3, p-value = 0.01
alternative hypothesis: true delta is less than 0
sample estimates:
   delta
-0.08483
```

Could also use Phillips-Perron test.

Integrated regressors: Cointegration

Approach (2): Residual-based test.

- Regress y_t on x_t.
- Use unit root test (usually ADF or Phillips-Perron) with residuals.

  Small detail: Here, ADF *without* constant makes sense, because OLS residuals in regression with a constant have mean zero.
- *Caution:* The usual critical values for the ADF are no longer valid!
  Reason: New test employs residuals, not raw data.
  Theory developed by Phillips and Ouliaris 1991 $\rightarrow$ new tables needed . . .

Integrated regressors: Cointegration

```
R> po.test(log(PepperPrice))
        Phillips-Ouliaris Cointegration Test

data:  log(PepperPrice)
Phillips-Ouliaris demeaned = -24, Truncation lag
parameter = 2, p-value = 0.02
```

This function uses the Phillips-Perron test with residuals.

Integrated regressors: Cointegration

Problems of residual-based tests:

- Method is asymmetric, although definition of cointegration is symmetric in x and y. Test result may depend on ordering of variables!

```
R> po.test(log(PepperPrice[, 2:1]))
        Phillips-Ouliaris Cointegration Test

data:  log(PepperPrice[, 2:1])
Phillips-Ouliaris demeaned = -23, Truncation lag
parameter = 2, p-value = 0.03
```

- Unit root in residuals means that we assume cointegration for the regression, then test whether this is *not* true. A stationarity test would be more consistent methodologically.
- Only suitable for cointegration involving two series. More general method: Cointegration in a multiple time series setting.

Cointegration in Multivariate Time Series Models

Cointegration in multivariate models

Example: Consider bivariate process $\{y_t\}$ with

$$
\begin{aligned}
y_{1t} &= \sum_{j=1}^{t} e_j, \quad \{e_t\} \sim WN(0, \sigma_e^2) \\
y_{2t} &= y_{1t} + v_t, \quad \{v_t\} \sim WN(0, \sigma_v^2)
\end{aligned}
$$

So $\{y_t\}$ is nonstationary. By construction, $\{y_{1t} - y_{2t}\} =: \{\beta^\top y_t\}$ is stationary, hence process is cointegrated with $\beta = (1, -1)^\top$.

First differences:

$$
\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_t \\ e_t + v_t \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} e_{t-1} \\ e_{t-1} + v_{t-1} \end{pmatrix}
$$

Cointegration in multivariate models

Observation: This has MA matrix polynomial

$$\Theta(B) \;=\; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} B$$

with $\det(\Theta(z)) = 0$ for $z = 1$. Hence there is no AR(∞) representation.

Conclusion: No differencing is too little, but differencing is too much ...

Cointegrated VARs

Consider: VAR(p) with deterministic component $\delta_0 + \delta_1 t$.

$$y_t \;=\; \delta_0 + \delta_1 t + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + e_t, \quad \{e_t\} \sim WN(0, \Sigma_e)$$

Subtracting y_{t-1} (compare DF trafo) yields

$$\Delta y_t \;=\; \delta_0 + \delta_1 t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* \Delta y_{t-j} + e_t$$

where $\Pi = -\Phi(1) = -(I - \Phi_1 - \cdots - \Phi_p)$ and
$\Phi_j^* = -\sum_{\ell=j+1}^{p} \Phi_\ell$.

This is the *error correction representation* of a cointegrated VAR.

Granger representation theorem: A VAR(p) $\{y_t\}$ with $\det(\Phi(z)) \neq 0$ for all $|z| < 1$ is cointegrated with $\beta \in \mathbb{R}^{k \times r}$ if and only if $rank(\Pi) = r$ and $\Pi = \alpha\beta^\top$, where $\alpha \in \mathbb{R}^{k \times r}$ with $rank(\alpha) = r$.

Cointegrated VARs

Remark: Note that β is not identified without normalization.

- If β defines a cointegration relation, so does $c\beta$ for any $c \neq 0$.
- When $r = 1$, usually set one element of β equal to 1, compare $\beta = (1, b)^\top$ above.
- For $r > 1$, there exist several different identification schemes.

Cointegrated VARs

Interpretation and terminology for VECM:

- $\mathsf{E}(\beta^\top y_t) =: \mu$ defines long-run equilibrium or steady state.
- $\beta^\top y_t - \mu$ defines deviations from equilibrium (‘error’).
- α describes readjustment to equilibrium (‘correction’). Hence α is matrix of adjustment coefficients.
- Π is long-run impact matrix.
- The number of linearly independent cointegration vectors is the cointegration rank.
- The space spanned by the cointegration vectors is the cointegration space.

Remark: It can be shown that in a system with r cointegration relations there are $k - r$ stochastic trends or unit roots. There exists a ‘dual’ relation of the VECM in levels, the *common trends representation* of a cointegrated VAR.

Cointegration

The Johansen Test

The Johansen test

Hypotheses: On number of cointegration relations. Usually, a sequence of tests is considered. Null hypotheses are

$$H(r): \quad \text{rank}(\Pi) \leq r, \quad r = 0, 1, \ldots, k.$$

Two types of tests:

- Trace test.

$$H(r), \text{ i.e., rank}(\Pi) \leq r, \; vs. \, H(k), \text{i.e., rank}(\Pi) = k$$

- Lambda max (λ_{max}) test.

$$H(r), \text{ i.e., rank}(\Pi) \leq r, \; vs. \, H(r+1), \text{ i.e., rank}(\Pi) = r+1$$

The Johansen test

Technical issues:

- Both tests use eigenvalues of certain matrices.
- Distributions are non-standard and involve integrals wrt. Brownian motion (compare Dickey-Fuller).
- Precise form and critical values depend on deterministic components *and how they appear inside and outside the cointegration relation*.

The Johansen test

Johansen test: Corresponds to multivariate ADF. Begin without deterministic terms and VAR(1). Test rank of Π in

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

If $rank(\Pi) < k$, then there is cointegration. Tests employ eigenvalues (of an estimate of) the matrix Π.

Basic ingredients:

- Likelihood ratio test.
- Multivariate reduced rank regression (RRR).
- Canonical correlation analysis (CCA).

Log-Likelihood: With $\{e_t\} \sim WN(0, \Sigma_e)$ n.i.d.,

$$\begin{aligned}\ell(\alpha, \beta, \Sigma_e) \;&=\; -\frac{Tk}{2}\log(2\pi) + \frac{T}{2}\log\det(\Sigma_e^{-1}) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}(\Delta y_t - \alpha\beta^\top y_{t-1})^\top \Sigma_e^{-1}(\Delta y_t - \alpha\beta^\top y_{t-1})\end{aligned}$$

The Johansen test

Outline of computations in Johansen test:

Step 1: Set up the concentrated likelihood for the cointegration relations.

For known rank r and known β, have least squares solution

$$\hat{\alpha} = \hat{\alpha}(\beta) = S_{01}\beta(\beta^\top S_{11}\beta)^{-1}$$

where

$$S_{ij} = \frac{1}{T}\sum_{t=1}^{T} R_{it}R_{jt}^\top, \quad i,j \in \{0,1\},$$

with $R_{0t} = \Delta y_t$ and $R_{1t} = y_t$. Note that $S_{10} = S_{01}^\top$.

Remark: For a VAR(p), R_{0t}, R_{1t} are residuals from regressions on lagged Δy_t.

The Johansen test

Next, using algebra for partitioned matrices, the covariance matrix is

$$\hat{\Sigma}_e \;=\; \hat{\Sigma}_e(\beta) \;=\; S_{00} - S_{01}\beta(\beta^\top S_{11}\beta)^{-1}\beta^\top S_{10}$$

From this, the *concentrated likelihood* can be shown to be

$$\ell(\beta) \;=\; -\frac{Tk}{2}\log(2\pi) - \frac{Tk}{2} - \frac{T}{2}\log\det(S_{00} - S_{01}\beta(\beta^\top S_{11}\beta)^{-1}\beta^\top S_{10})$$

All this is reduced rank regression (RRR), a multivariate linear regression technique.

The Johansen test

Step 2: Maximization of the log-likelihood.

Step 1 implies that $\ell(\beta)$ is maximized if

$$\det(\hat{\Sigma}_e(\beta)) \;=\; \det(S_{00} - S_{01}\beta(\beta^\top S_{11}\beta)^{-1}\beta^\top S_{10})$$

is minimized. Finding the minimum requires solving the *generalized eigenvalue problem*

$$\det(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}) \;=\; 0.$$

Solutions are $1 \geq \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_k \geq 0$.

The eigenvectors $\hat{\beta}_1, \ldots, \hat{\beta}_r$ corresponding to the r largest eigenvalues define the cointegration relations. These are usually normalized via $\hat{\beta}_i^\top S_{11}\hat{\beta}_i$, $i = 1, \ldots, k$. The minimizer is thus

$$\operatorname*{argmin}_{\beta} \det(\hat{\Sigma}_e(\beta)) \;=\; \det(S_{00}) \prod_{i=1}^{k}(1 - \hat{\lambda}_i)$$

The Johansen test

Remark: All this is essentially *canonical correlation analysis*, a technique from multivariate statistics that aims to find linear combinations for two separate groups of variables that are ‘similar’. The eigenvalues are the squared canonical correlations.

The Johansen test

Step 3: Under the null hypothesis $H(r)$, the concentrated likelihood is

$$\ell(\hat{\beta}) = -\frac{Tk}{2}\log(2\pi) - \frac{Tk}{2} - \frac{T}{2}\log\det(S_{00}) - \frac{T}{2}\sum_{i=1}^{r}\log(1-\hat{\lambda}_i)$$

Likelihood ratio principle yields trace statistic

$$-T\sum_{i=r+1}^{k}\log(1-\hat{\lambda}_i)$$

and maximal eigenvalue or 'lambda max' statistic

$$-T\log(1-\hat{\lambda}_{r+1})$$

In practice, test sequentially for $r = 0, 1, \ldots, k$.

The Johansen test

Limiting distributions are given by trace or maximal eigenvalue of a random matrix involving integrals of Brownian motion.
For $k = 2, r = 1$, it reduces to the square of Dickey-Fuller.

Cointegration

Deterministic Components in Cointegrated VARs

Deterministic components

For simplicity, consider only VAR(1). Distinguish between deterministic terms within and outside the cointegration relation. Most general form is

$$y_t \ = \ \delta_0 + \delta_1 t + \alpha(\beta^\top y_{t-1} - \gamma_0 - \gamma_1 t) + e_t,$$

There are 5 (!) main cases that are discussed in the literature (see Johansen 1995, Pesaran, Shin, Smith 2000, MacKinnon, Haug, Michelis 1999). There is also some confusion . . .

Franses (2001) suggests mainly following two are relevant in empirical work:

- None of the series has trending pattern:

$$\Delta y_t \ = \ \alpha(\beta^\top y_{t-1} - \gamma_0) + e_t$$

- Some or all of the series have trending pattern:

$$\Delta y_t \ = \ \delta_0 + \alpha(\beta^\top y_{t-1} - \gamma_0 - \gamma_1 t) + e_t$$

Cointegration

Example

Example

```
R> library("urca")
R> summary(ca.jo(log(PepperPrice), ecdet = "const", K = 2))
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) ,
  without linear trend and constant in cointegration

Eigenvalues (lambda):
[1] 4.932e-02 1.351e-02 2.082e-17

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  3.66  7.52  9.24 12.97
r = 0  | 13.61 13.75 15.67 20.20
```

Example

```
Eigenvectors, normalised to first column:
(These are the cointegration relations)

         black.l2 white.l2 constant
black.l2   1.0000    1.000    1.000
white.l2  -0.8892   -5.099    2.281
constant  -0.5570   33.027  -20.032

Weights W:
(This is the loading matrix)

        black.l2 white.l2   constant
black.d -0.07472 0.002453 -4.958e-18
white.d  0.02016 0.003537  8.850e-18
```

Cointegration

Practical Issues in Cointegration Testing

Practical issues in cointegration testing

- Most reliable critical values by MacKinnon, Haug and Michelis (1999). Several software packages still use older critical values (e.g., those of Osterwald-Lenum 1992)! This includes the R package **urca**.
- Two different classifications of deterministics: Johansen (1995) and Pesaran, Shin and Smith (PSS) (2000).
- Several software packages have (had?) problems with critical values for some deterministic components (Turner 2011). Reason: confusion wrt. Johansen vs. PSS classification.

Cointegration

Further Methods and Issues

Further methods and issues

- Cointegration-defining equation is static and, for regression purposes, likely misspecified (omitted short-run dynamics).

  Alternative regression methods using single-equation approach:

  - Dynamic OLS (DOLS, Stock and Watson 1993).
    Uses dynamic regression with leads and lags. Needs HAC corrections for tests.
  - Fully modified least squares (FM-OLS, Phillips and Hansen 1991).
    Employs HAC correction.

- Hypothesis testing (e.g., Granger causality) much more complex than in the stationary case. Limit theory is available.