A Survey of Cointegration

Søren Johansen

SETA 2006, April 4-6
Xiamen University

What is cointegration ?

1. Log S&P500 index and futures.

Daily data from January 2, 1989, January 31, 2002

Figure 1 and 2

2. The Danish data for $m_t, y_t, i_t^{deposit}, i_t^{bond}$, quarterly 1974:1 to 1987:3.

Figure 3 and 4

Content of Lectures

1. Definitions and Methodology
2. The Granger Representation Theorem
3. Interpretation of the I(1) model
4. The Statistical Analysis
5. Asymptotic Analysis
6. Further Topics
7. Conclusion

1. DEFINITIONS AND METHODOLOGY

The mathematical formulation of cointegration:

$$\begin{aligned} x_{1t} &= a\sum_{i=1}^{t}\varepsilon_{1i} + \varepsilon_{2t} \sim I(1) \\ x_{2t} &= b\sum_{i=1}^{t}\varepsilon_{1i} + \varepsilon_{3t} \sim I(1) \\ bx_{1t} - ax_{2t} &= b\varepsilon_{2t} - a\varepsilon_{3t} \sim I(0) \end{aligned}$$

1. $x(t)$ is nonstationary and Δx_t is stationary: x_t is $I(1)$
2. $\beta' x_t$ is stationary with $\beta = (b, -a)'$
3. The common trend is $\sum_{i=1}^{t}\varepsilon_{1i}$

Picture 5

Three statistical formulations:

1. The regression formulation

$$\begin{aligned} x_{1t} &= \gamma' x_{2t} + u_{1t} \\ \Delta x_{2t} &= u_{2t} \end{aligned}$$

Engle and Granger (1987), Stock (1987), Phillips and Hansen (1990), Phillips (1991).

2. The autoregressive formulation

$$\Delta x_t = \alpha \beta' x_{t-1} + \varepsilon_t,$$

$\beta = (1, -\gamma')'$.

Johansen (1988), Ahn and Reinsel (1990)

3. The common trends (state space) formulation

$$x_t = \xi \eta' \sum_{i=1}^{t} \varepsilon_i + u_t$$

$\xi = \beta_\perp, \eta = \alpha_\perp$.

Nyblom and Harvey (2000).

Long-run behaviour, short-run behaviour, dynamic model, test for rank.

This lecture summarizes the VAR approach.

Literature

1. Johansen S. (1996) 'Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.' Oxford University Press, Oxford.
2. Juselius, K. (2006), 'The Cointegrated VAR Model: Methodology and Applications.' Oxford University Press, Oxford.
3. Dennis, J., Johansen, S. and Juselius, K. (2006), 'CATS for RATS: Manual to Cointegration Analysis of Time Series.' Estima, Illinois.

(Roughly 1000 pages)

WHICH STATISTICAL MODEL DESCRIBES THE DATA?

General methodology is to build a statistical model

$$f(x, \theta), \theta \in \Theta$$

that

a: contains or ‘embeds’ the economic model

b: contains the probability density that describes the stochastic variation of the data.

What is gained:

1. A coherent framework for formulating and testing the economic hypotheses of interest.

2. The Gaussian likelihood is used for deriving test and estimators but their properties are derived under more general assumptions.

In particular we find

3. Asymptotic distributions of maximum likelihood estimators.

4. Consistent estimates of ‘variances’.

5. Asymptotic distribution of likelihood ratio statistics.

The price paid:

One should carefully check the assumptions underlying the model, otherwise inference may be invalid

2. THE GRANGER REPRESENTATION THEOREM

Error correction formulation of the VAR

$$\begin{aligned}\Delta x_t &= \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t \\ \Pi(z) &= (1-z)I_p - \alpha\beta' z - \sum_{i=1}^{k-1} (1-z) z^i \Gamma_i \\ \det \Pi(1) &= \det \alpha\beta' = 0 \Longrightarrow z = 1 \text{ is a root of } \det \Pi(z) = 0\end{aligned}$$

QUESTION: If the VAR has unit roots and the other roots larger than one, what is the moving average representation?

$I(1)$ condition : $\det(\Pi(z)) = 0 \Longrightarrow z = 1$ or $|z| > 1$ and

$$\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i, \quad \text{det}(\alpha_\perp' \Gamma \beta_\perp) \neq 0$$

THEOREM: If the $I(1)$ condition is satisfied then

$$\begin{aligned}x_t &= C \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{\infty} C_i^* \varepsilon_{t-i} + A, \;\; \beta' A = 0 \\ C &= \beta_\perp (\alpha_\perp' \Gamma \beta_\perp)^{-1} \alpha_\perp'\end{aligned}$$

1. x_t is nonstationary, Δx_t is stationary: x_t is $I(1)$
2. $\beta' x_t$ is stationary: x_t is cointegrating (r relations)
3. The common trends are $\alpha_\perp' \sum_{i=1}^{t} \varepsilon_i$ ($p - r$ trends)

Example

$$\begin{aligned}\Delta x_{1t} &= -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} \\ \Delta x_{2t} &= \quad \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t}\end{aligned}$$

$$\begin{aligned}\Delta(x_{1t} - x_{2t}) &= -\frac{1}{2}(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} - \varepsilon_{2t} \\ &\Longrightarrow x_{1t} - x_{2t} \text{ stationary } (= y_t) \\ \Delta(x_{1t} + x_{2t}) &= \varepsilon_{1t} + \varepsilon_{2t} \\ &\Longrightarrow x_{1t} + x_{2t} \text{ nonstationary random walk } (= S_t)\end{aligned}$$

Granger Representation Theorem

$$\begin{aligned}x_{1t} &= \frac{1}{2}(S_t + y_t) \\ x_{2t} &= \frac{1}{2}(S_t - y_t)\end{aligned}$$

1. x_t is non stationary, Δx_t is stationary
2. $\beta' x_t$ is stationary with cointegration vector $\beta = (1, -1)'$
3. $\sum_{i=1}^{t}(\varepsilon_{1i} + \varepsilon_{2i})$ is a common trend

Figure 5

Figure 7

A strange example

$$
\begin{aligned}
\Delta x_{1t} &= \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \frac{9}{4}\Delta x_{2t-1} + \varepsilon_{1t} \\
\Delta x_{2t} &= -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t}
\end{aligned}
$$

is $I(1)$ and cointegrated.

The sign of the adjustment is not intuitive

The processes do not adjust properly, yet are $I(1)$.

Two applications of the Granger Representation Theorem

1. The role of deterministic terms

$$\begin{aligned}
\Delta x_t &= \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \mu + \varepsilon_t \\
x_t &= C\sum_{i=1}^{t}(\varepsilon_i + \mu) + \sum_{i=0}^{\infty} C_i^*(\varepsilon_{t-i} + \mu) + A \\
x_t &= C\sum_{i=1}^{t}\varepsilon_i + C\mu t + \text{stationary process}
\end{aligned}$$

Thus

1. Linear trend in general
2. If $C\mu = 0$: only constant term

Other deterministics. Additive deterministics

2. Asymptotics

$$\begin{aligned}
x_t &= C\sum_{i=1}^{t}\varepsilon_i + C\mu t + \text{stationary process} \\
\frac{x_{Tu}}{\sqrt{T}} &= C\frac{\sum_{i=1}^{Tu}\varepsilon_i}{\sqrt{T}} + C\mu\frac{Tu}{\sqrt{T}} + \frac{\text{stationary process}}{\sqrt{T}} \\
& \quad\;\; \downarrow w \qquad\qquad\qquad\qquad\qquad\quad \downarrow P \\
& \quad\;\; CW(u) \qquad\qquad\qquad\qquad\qquad\; 0
\end{aligned}$$

3. INTERPRETATION OF THE I(1) MODEL

Interpretation of regression coefficients:

$$x_{1t} = \gamma_2 x_{2t} + \gamma_3 x_{3t} + \varepsilon_t$$

are interpreted via a counterfactual experiment:

Change x_2 by 1, keep x_3 fixed and γ_2 is the effect of x_2 on x_1.

Interpretation of cointegration coefficients:

$$\Delta x_t = \alpha(\beta' x_{t-1} - \beta_0) + \Gamma_1 \Delta x_{t-1} + \varepsilon_t$$

The cointegrating relations are long-run economic relations between the variables.

They are long-run in the sense that they have been there all the time.

The attractor set

$$\{x \in R^p | Cx = \alpha(\beta'\alpha)^{-1}\beta_0\} = \{x | \beta' x = \beta_0\}.$$

is the set of steady states or long-run values:

$$\begin{aligned} x_{\infty . t} &= \lim_{h\to\infty} E(x_{t+h}|x_t) = Cx_t + \alpha(\beta'\alpha)^{-1}\beta_0 \\ (x_t + c)_\infty &= x_{\infty . t} + Cc \end{aligned}$$

The model

$$\Delta x_t = \alpha(\beta' x_{t-1} - \beta_0) + \Gamma_1 \Delta x_{t-1} + \varepsilon_t$$

Consider the cointegrating relation $\beta' x = \beta_0$ solved for x_1 :

$$x_1 = \gamma_2 x_2 + \gamma_3 x_3 + \gamma_0 + \text{stat. process.}$$

A long-run counterfactual experiment:

Change x_t by c, so that in the long run x_1 is changed by γ_2, x_2 is changed by 1 and x_3 is kept fixed.

That is, a long-run change of the form $(\gamma_2, 1, 0)$

This is a vector in sp$(\beta_\perp)$:

$$(1, -\gamma_2, -\gamma_3) \begin{pmatrix} \gamma_2 \\ 1 \\ 0 \end{pmatrix} = 0$$

can be implemented by

$$c = (I_p - \Gamma_1) \begin{pmatrix} \gamma_2 \\ 1 \\ 0 \end{pmatrix}$$

Hypotheses of interest

1. Hypotheses on the rank

$$\begin{aligned}\mathcal{H}_r &: \Delta x_t = \alpha\beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \Phi D_t + \varepsilon_t, \quad \alpha_{p\times r}, \beta_{p\times r}\\ \mathcal{H}_0 &\subset \ldots \subset \mathcal{H}_r \subset \ldots \subset \mathcal{H}_p\end{aligned}$$

2. Hypotheses on β

Some examples for the model with

$$x_t = (m_t^r, y_t^r, i_t^{deposit}, i_t^{bond}, \Delta p_t)'$$

Assume first that $r = 1$:

Inverse velocity is stationary

$$\begin{aligned}\beta' &= (1, -1, 0, 0, 0)\\ \beta' x_t &= m_t^r - y_t^r\end{aligned}$$

Inverse velocity is a function of interest rate spread

$$\begin{aligned}\beta' &= (1, -1, \phi, -\phi, 0)\\ \beta' x_t &= m_t^r - y_t^r + \phi(i_t^{dep} - i_t^{bond})\end{aligned}$$

Assume next that $r = 2$:

Bond rate is stationary and velocity is stationary

$$\begin{aligned}\beta' &= \begin{pmatrix} 0 & 0 & 0 & 1 & 0\\ 1 & -1 & 0 & 0 & 0\end{pmatrix}\\ \beta_1' x_t &= i_t^{bond}\\ \beta_2' x_t &= m_t^r - y_t^r\end{aligned}$$

Identification problem for β

General linear restrictions $(r = 2)$

$$\beta = (H_1\phi_1, H_2\phi_2) \text{ or } R_1'\beta_1 = R_2'\beta_2 = 0, \quad R_i = H_{i\perp}.$$

Wald's identification criterion (rank condition)

$$rank(R_i'\beta) \geq 1, i = 1, 2.$$

Another criterion (independent of the unknown parameter) is

$$rank(R_i'H_j) \geq 1, \text{ for } i \neq j = 1, 2.$$

Hypotheses on α

Some examples for the model with

$$\begin{aligned} x_t &= (m_t^r, y_t^r, i_t^{deposit}, i_t^{bond}, \Delta p_t)' \\ \Delta x_t &= \alpha\beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \Phi D_t + \varepsilon_t \end{aligned}$$

The interests rates are weakly exogenous: Last two rows in α equal to zero

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \phi, \quad \alpha_\perp = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Common trends $\alpha_\perp' \sum_{i=1}^t \varepsilon_i = (\sum_{i=1}^t \varepsilon_{3t}, \sum_{i=1}^t \varepsilon_{4t})$

Shocks, changes and impulse responses

The model

$$\begin{aligned}\Delta x_t &= \alpha\beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_t \\ x_t &= (I_p + \alpha\beta')x_{t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_t\end{aligned}$$

shows that a change in ε_t $(\varepsilon_t \mapsto \varepsilon_t + c)$ is equivalent to a change in x_t $(x_t \mapsto x_t + c)$.

We call ε_t a shock and c a change

Granger Representation Theorem:

$$x_{t+h} = C\sum_{i=1}^{t+h}\varepsilon_i + \sum_{i=0}^{\infty} C_i \varepsilon_{t+h-i} + A, \;\; C = \beta_\perp(\alpha'_\perp(I_p - \Gamma_1)\beta)^{-1}\alpha'_\perp$$

$$\frac{\partial x_{t+h}}{\partial \varepsilon_t} c = (C + C_h)c \to Cc, h \to \infty$$

The impulse response function.

The shocks $\alpha'_\perp \varepsilon_t$ cumulate to the common trends and are the permanent shocks.

The shocks $\alpha'\Omega^{-1}\varepsilon_t$ are independent of the permanent shocks and can be called the transitory shocks.

The decomposition of ε_t

$$\varepsilon_t = \underbrace{\alpha(\alpha'\Omega^{-1}\alpha)^{-1}}_{effect}\underbrace{\alpha'\Omega^{-1}\varepsilon_t}_{trans.\ shock} + \underbrace{\Omega\alpha_\perp(\alpha'_\perp\Omega\alpha_\perp)^{-1}}_{effect}\underbrace{\alpha'_\perp\varepsilon_t}_{perm.\ shock}$$

into transitory and permanent shocks.

The long-run effect of a transitory shock is zero: $C\alpha = 0$.

Structural shocks

$$\begin{aligned}
\Delta x_t &= \alpha(\beta' x_{t-1} - \beta_0) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t \\
x_t &= C \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{\infty} C_i^* (\varepsilon_{t-i} - \alpha\beta_0) \\
x_t &= C B^{-1} \sum_{i=1}^{t} B\varepsilon_i + \sum_{i=0}^{\infty} C_i^* B^{-1} (B\varepsilon_{t-i} - B\alpha\beta_0) \\
B\Delta x_t &= B\alpha(\beta' x_{t-1} - \beta_0) + \sum_{i=1}^{k-1} B\Gamma_i \Delta x_{t-i} + B\varepsilon_t
\end{aligned}$$

$$\begin{aligned}
B' &= (v_1, \ldots, v_p),\ v_i'\varepsilon_t \text{ `structural' shock} \\
B^{-1} &= (w_1, \ldots, w_p), w_i \text{ the effect of `structural' shock} \\
\varepsilon_t &= \sum_{i=1}^{p} w_i v_i' \varepsilon_t
\end{aligned}$$

Two identification problems

1. Identify β by restrictions

2. Identify the remaining coefficients by restrictions on the structural model

4. THE STATISTICAL ANALYSIS

Test for misspecification of the VAR

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi D_t + \varepsilon_t$$

The VAR model assumes

1. Linear conditional mean explained by the past observations and deterministic terms

(Check for unmodelled systematic variation, the choice of lag length, choice of information set (data), possible outliers, nonlinearity, non constant parameters)

2. Constant conditional variance

(Check for ARCH effects, but also for regime shifts in the variance)

3. Independent Normal errors, mean zero, variance Ω

(Check for lack of autocorrelation, distributional form)

4. Estimation of the $I(1)$ model

$$\mathcal{H}_r : \Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi D_t + \varepsilon_t,$$

where ε_t i.i.d. $N_p(0, \Omega)$ and α and β are $(p \times r)$.

Maximum likelihood by reduced rank regression (Anderson 1951) of Δx_t on x_{t-1} corrected for $(\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, D_t)$:

$$\begin{aligned} R_{0t} &= (\Delta x_t | \Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, D_t) \\ R_{1t} &= (x_{t-1} | \Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, D_t) \\ S_{ij} &= T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt} \\ 0 &= \det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}) \end{aligned}$$

$$\begin{aligned} \hat\beta &= (v_1, \ldots, v_r) \\ \hat\alpha &= S_{01}\hat\beta(\hat\beta' S_{11}\hat\beta)^{-1}, \\ \hat\Omega &= S_{00} - S_{01}\hat\beta(\hat\beta' S_{11}\hat\beta)^{-1}\hat\beta' S_{10} \\ L_{\max}^{-2/T}(\mathcal{H}_r) &= |\hat\Omega| = |S_{00}| \prod_{i=1}^{r} (1 - \hat\lambda_i) \\ -2\log LR(\mathcal{H}_r | \mathcal{H}_p) &= -T \sum_{i=r+1}^{p} \log(1 - \hat\lambda_i) \end{aligned}$$

Estimation in the I(1) Model with restrictions on β

Under the hypothesis

$$\mathcal{H} : \beta = H\phi$$

$$\Delta x_t = \alpha\phi' H' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi D_t + \varepsilon_t,$$

Estimate by $RRR(\Delta x_t, H' x_{t-1} | \Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, D_t)$.

This does not work for

$$\beta = (H_1\phi_1, H_2\phi_2)$$

We need a switching algorithm.

5. ASYMPTOTIC ANALYSIS

Asymptotic properties of the process x_t and its product moments for model without deterministic terms.

Three basic results

$$\begin{aligned} \frac{x_{[Tu]}}{\sqrt{T}} = C\frac{\sum_{i=1}^{[Tu]}\varepsilon_i}{\sqrt{T}} &\overset{D}{\rightarrow} CW(u) \\ T^{-2}\sum_{t=1}^{T} x_{t-1}x_{t-1}' &\overset{D}{\rightarrow} \int_0^1 W(u)W(u)'du \\ T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t' &\overset{D}{\rightarrow} \int_0^1 W(dW)' \end{aligned}$$

where W is Brownian motion with variance Ω.

Test for rank

$$\begin{aligned} & -2\log LR(\mathcal{H}_r|\mathcal{H}_p) \\ = \; & -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i) \\ & \xrightarrow{D} tr\{\int_0^1 (dB)B'\left(\int_0^1 BB'\right)^{-1}\int_0^1 B(dB)'\} \end{aligned}$$

where B is standard Brownian motion.

The generalized Dickey-Fuller distribution, tabulated by simulation

Limit invariant to distribution of i.i.d. $(0,\Omega)$ errors.

Likelihood ratio test correctly normalized and the Gaussian LR test has no nuisance parameters in the limit distribution.

Modifications by deterministic terms.

Finite sample modifications

A small sample correction:

$$E_\theta(-2\log LR(\mathcal{H}_r|\mathcal{H}_p)) = a(T,p-r)(1+T^{-1}b(p,\theta)+O(T^{-2}))$$

where $a(T,p-r)$ is simulated and $b(p,\theta)$ is a combination of analytic and simulated terms.

The Dickey-Fuller test: To test that $\mathcal{H}:\pi=\mu_1=0$ in

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{2s}\gamma_i\Delta x_{t-i} + \mu_0 + \mu_1 t + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } N(0,\sigma^2)$$

$$\begin{aligned} E_{\gamma,\mu_0,\sigma^2}(-2\log LR(\mathcal{H})) &= (1+0.12T^{-1}+4.05T^{-2}) \\ &\quad \times(1+\frac{1.72}{T}[s+\frac{\sum_{i=1}^{2s} i\gamma_i}{1-\sum_{i=1}^{2s}\gamma_i}]) \end{aligned}$$

An extra unit root when $1-\sum_{i=1}^{2s}\gamma_i = 0$.

Asymptotics for $\hat{\beta}$ in the cointegrated VAR

$$T\beta'_{\perp}(\hat{\beta}-\beta) \xrightarrow{w} (\int_0^1 GG')^{-1}\int_0^1 G(dV)'$$

$$\begin{aligned} G(u) &= \beta'_{\perp}CW(u) \\ V(u) &= (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W(u) \\ Cov(V(u),G(u)) &= (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\Omega C'\beta_{\perp} = 0 \end{aligned}$$

Hence

1. $\hat{\beta}$ is superconsistent
2. The asymptotic distribution is mixed Gaussian

that is, for fixed G the distribution of $\int_0^1 G(dV)'$ is

$$N_{(p-r)\times(p-r)}(0,(\alpha'\Omega^{-1}\alpha)^{-1}\otimes\int_0^1 GG'du)$$

$$(\int_0^1 GG'du)^{-1/2}\beta'_{\perp}(\hat{\beta}-\beta)(\alpha'\Omega^{-1}\alpha)^{1/2} \xrightarrow{w} N_{p\times r}(0,I_r\otimes I_p)$$

The mixed Gaussian distribution**.**

A simple simulation of the model

$$\begin{aligned}\Delta x_{1t} &= \alpha_1(x_{1t-1} - \gamma x_{2t-1}) + \varepsilon_{1t} \\ \Delta x_t &= \varepsilon_{2t}\end{aligned}$$

where ε_t are i.i.d. $N_2(0, \sigma^2 I_2)$. DGP$(\alpha_1 = -1, \gamma = 1, \sigma^2 = 1)$:

$$\begin{aligned}x_{1t} &= x_{2t-1} + \varepsilon_{1t} \\ \Delta x_t &= \varepsilon_{2t}\end{aligned}$$

MLE by regression:

$$\begin{aligned}\Delta x_{1t} &= \xi_1 x_{1t-1} + \xi_2 x_{2t-1} + \varepsilon_{1t} \\ \hat{\alpha}_1 &= \hat{\xi}_1, \hat{\gamma} = -\hat{\xi}_2/\hat{\xi}_1\end{aligned}$$

Asymptotic results for testing using the Wald test

$$\frac{1}{\hat{\sigma}}\left(\sum_{t=1}^{T}(x_{1t-1} - \hat{\gamma} x_{2t-1})^2\right)^{1/2} (\hat{\alpha}_1 - \alpha_1) \xrightarrow{w} N(0,1)$$

$$\frac{1}{\hat{\sigma}}\left(\sum_{t=1}^{T}\hat{\alpha}^2 x_{2t-1}^2\right)^{1/2} (\hat{\gamma} - \gamma) \xrightarrow{w} N(0,1)$$

$$\frac{1}{\hat{\sigma}}\left(\sum_{t=1}^{T}(x_{1t-1}-\hat{\gamma}x_{2t-1})^2\right)^{1/2}(\hat{\alpha}_1-\alpha_1)\xrightarrow{D}N(0,1)$$

$$\frac{1}{\hat{\sigma}}\left(\sum_{t=1}^{T}\hat{\alpha}_1^2x_{2t-1}^2\right)^{1/2}(\hat{\gamma}-\gamma)\xrightarrow{D}N(0,1)$$

$$T^{-1}\sum_{t=1}^{T}(x_{1t-1}-\hat{\gamma}x_{2t-1})^2\xrightarrow{P}E(x_{1t-1}-\gamma x_{2t-1})^2=\sigma^2(1+\gamma^2)$$

implies we can instead use

$$\frac{\sqrt{T}(\hat{\alpha}_1-\alpha_1)}{\sqrt{1+\hat{\gamma}^2}}\xrightarrow{D}N(0,1),$$

that is, use the (asymptotic) distribution of $\hat{\alpha}$ to construct the Wald test.

$$T^{-2}\sum_{t=1}^{T}\hat{\alpha}_1^2x_{2t-1}^2\xrightarrow{D}\int_0^1W(u)^2du$$

implies that the asymptotic distribution of the test statistic is NOT a function of the asymptotic distribution of the estimator. Thus we need the BOTH $\hat{\beta}$ AND the observed information about $\hat{\beta}$, to construct the Wald test (see See Picture 6)

$$LR=-2\log\frac{L(\theta_0)}{L(\hat{\theta})}\approx(\hat{\theta}-\theta_0)'(-D^2\log L(\hat{\theta}))(\hat{\theta}-\theta_0)=W.$$

6. FURTHER TOPICS

1. Rational expectations
2. Seasonal cointegration
3. Models for explosive roots
4. The I(2) model
5. Non-linear cointegration
6. Panel data cointegration
7. Fractional cointegration

7 CONCLUSION

The development of cointegration started with Granger (1983) and the first results on the relation between cointegration and error correction are in Engle and Granger (1987).

The next 20 years brought many results on cointegration and I have focussed on those that relate to the vector autoregressive model.

I analyse model based inference because one has to make explicit assumptions about the model used and it is therefore a natural part of the methodology to check assumptions.

The theory is now part of many text books and software is available.

Problems to be solved

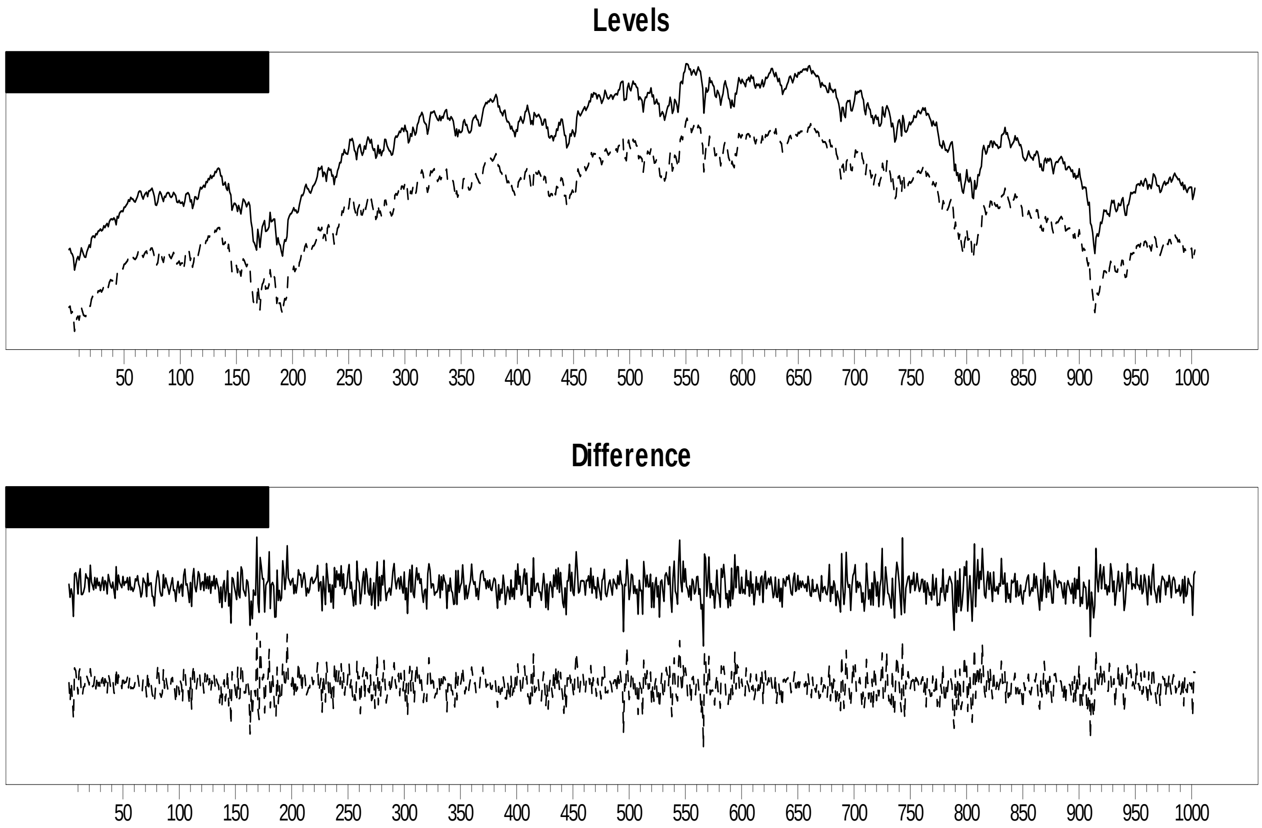
Small sample corrections of the asymptotic results

Detecting and modeling of non constant parameters

Non-linear time models for cointegration

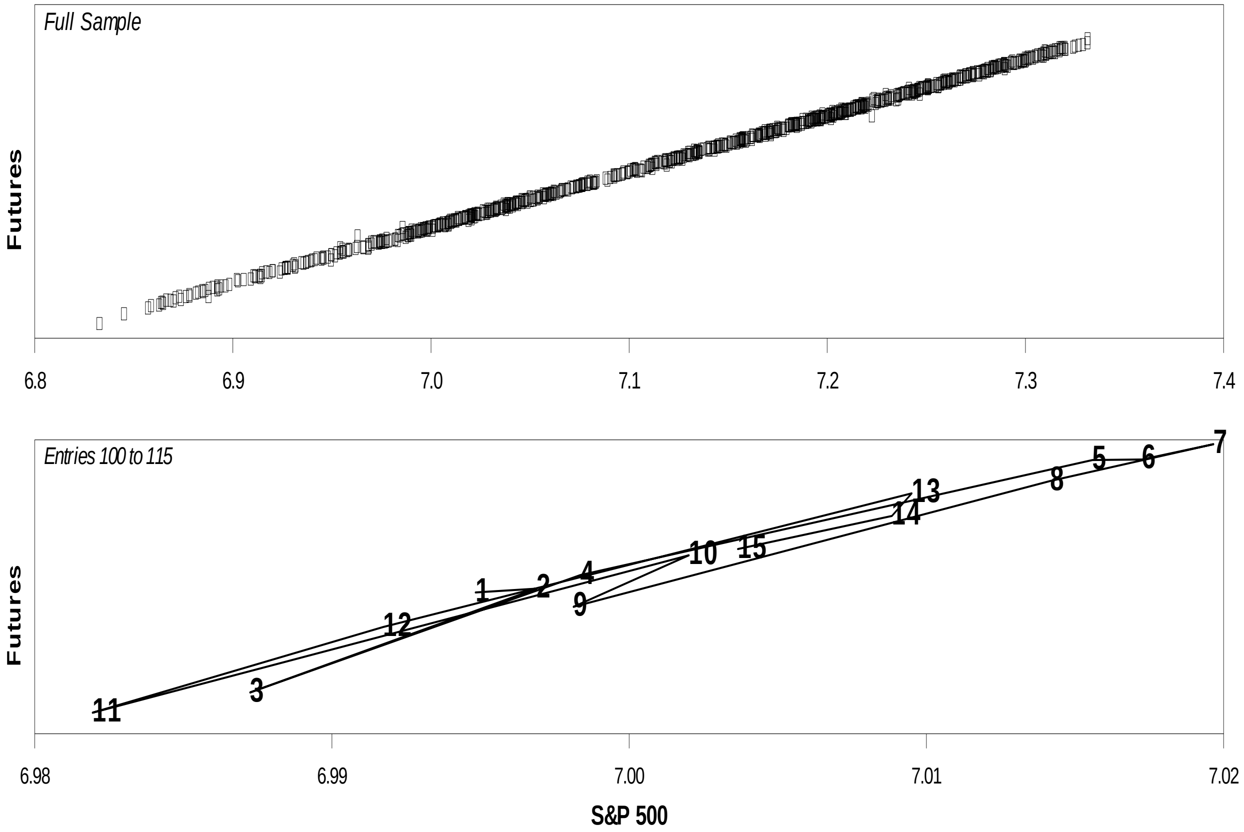
Panel data cointegration

MOST importantly is a development of economic theory, which takes into account the findings of the empirical analyses of non-stationary economic data.

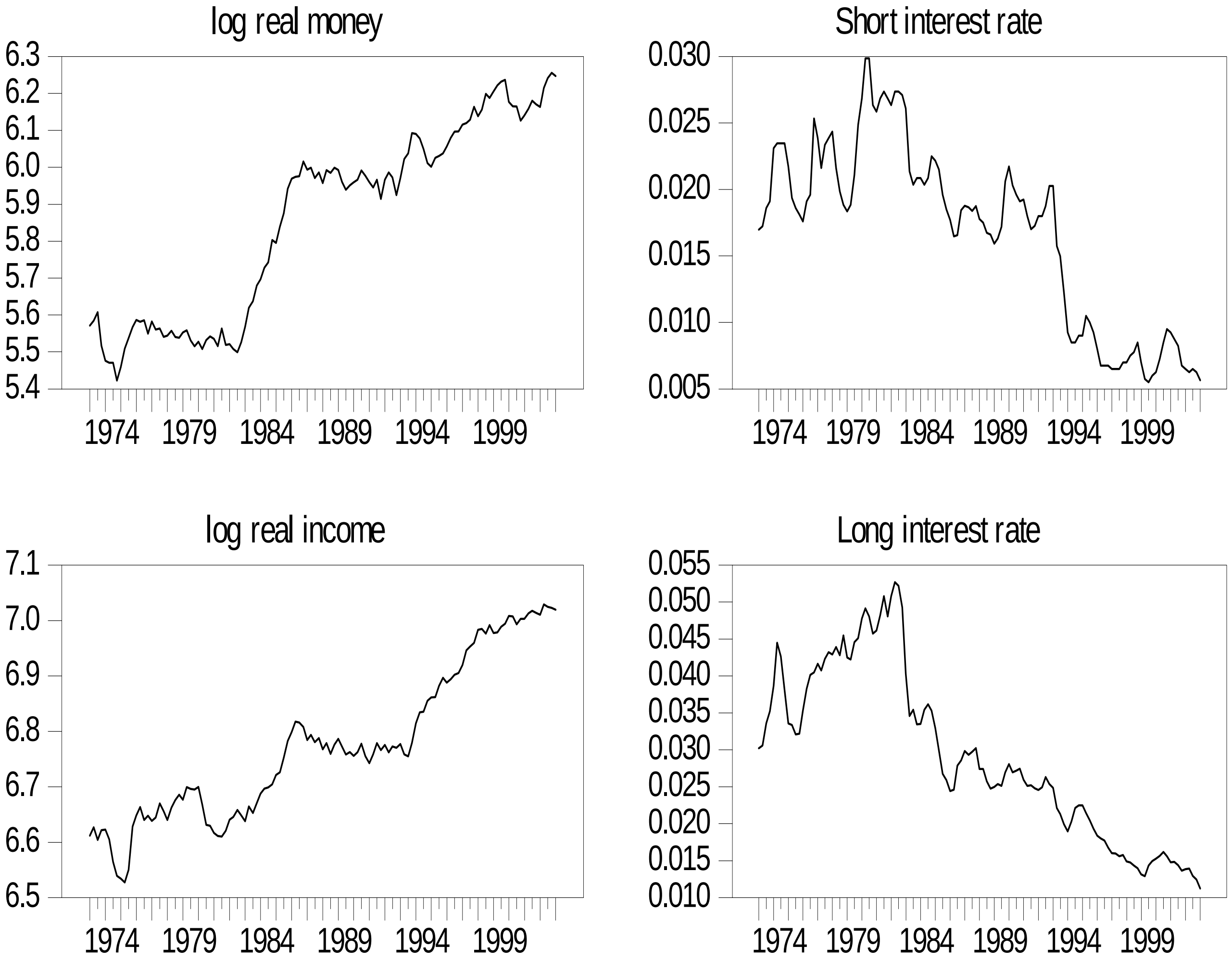

Levels

Difference

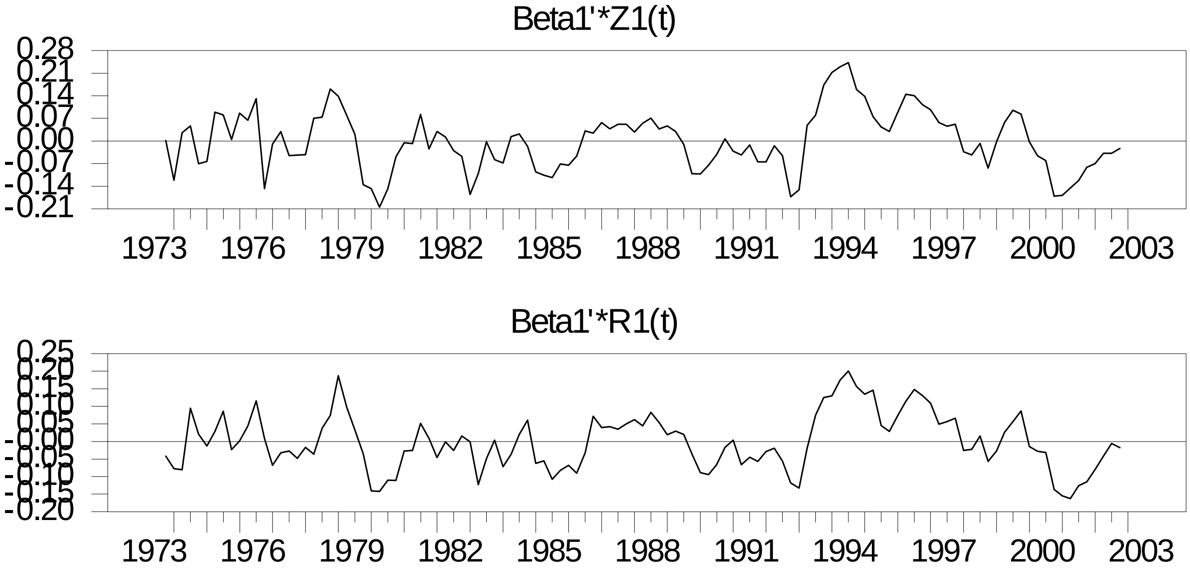
1.The plot of S & P 500 and futures in levels and differences. Daily data from January 2 1989 to January 31 2002



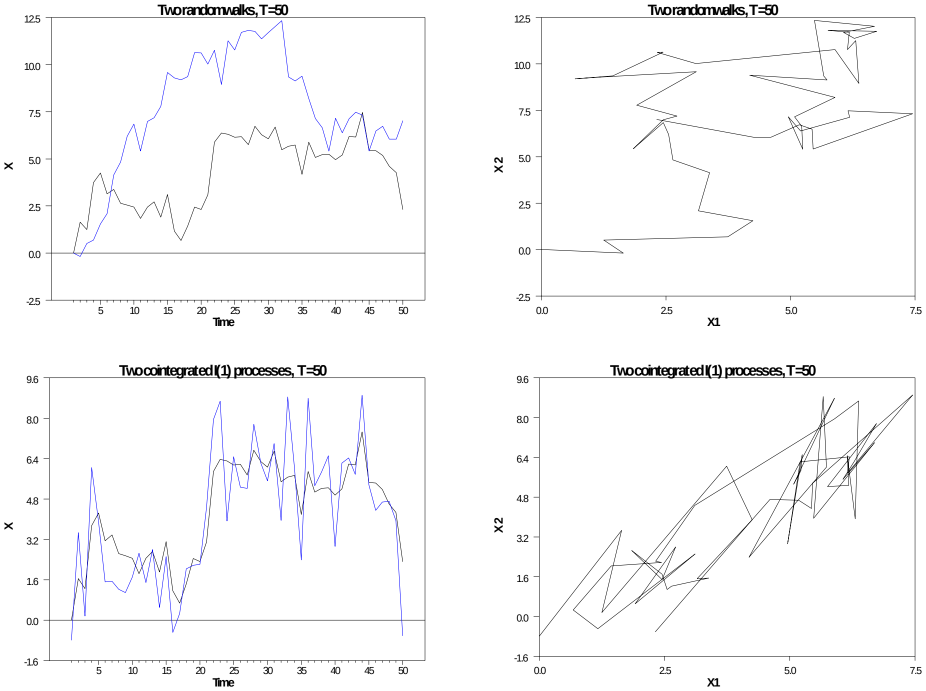
2.The cross plot of futures against S & P 500 for the full data set and for observations 100 to 115



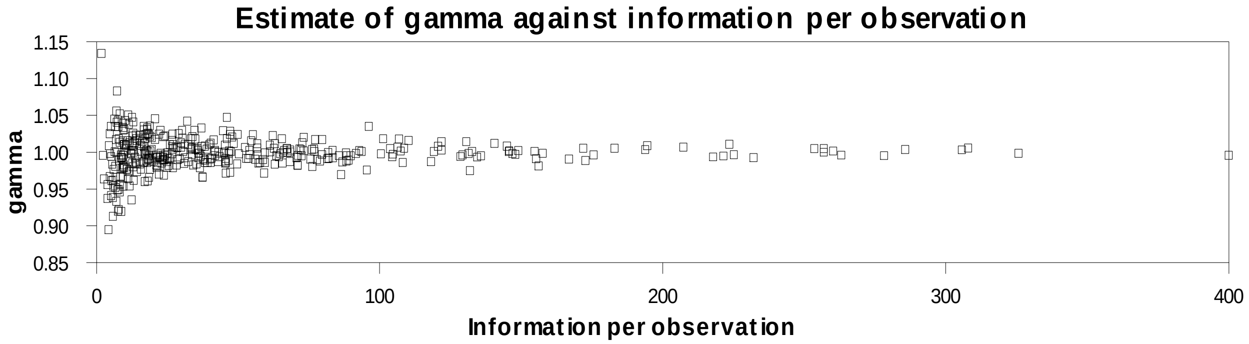
3.Plots of Danish data. Quarterly 1973 2003.

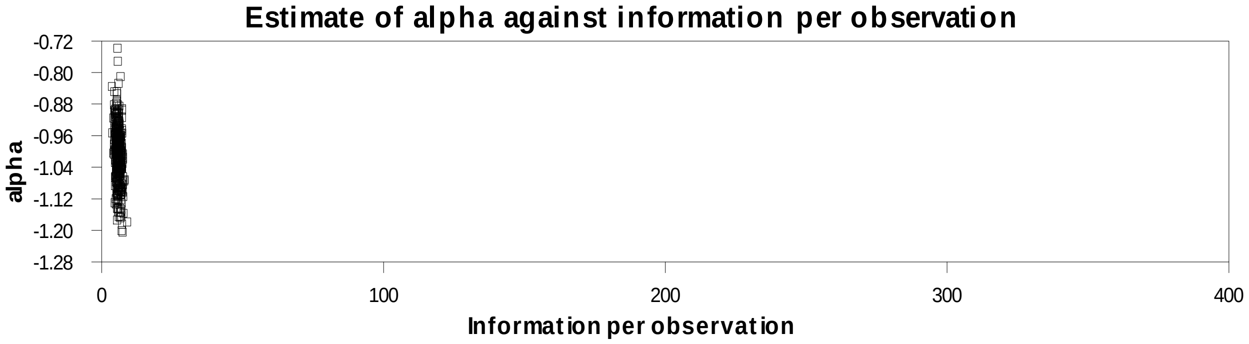


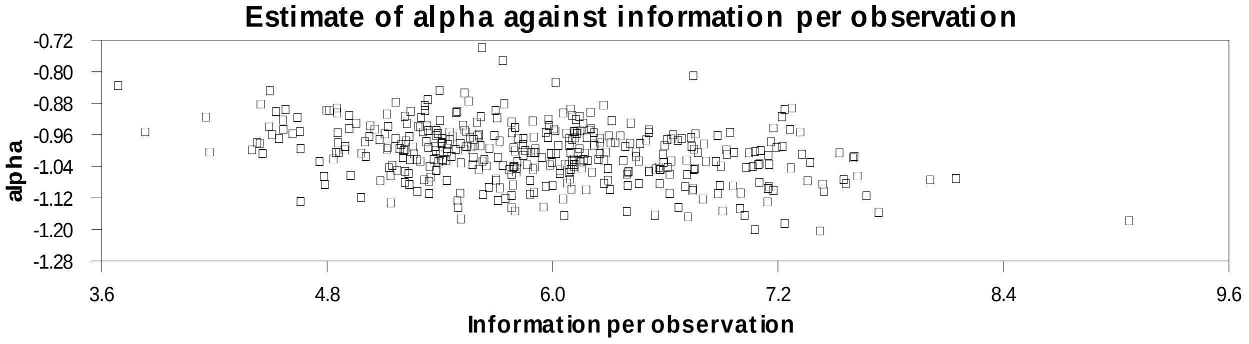
4.The cointegrating relation $m^r - 1.6y^r - 40.9R_m + 24.1R_b - 0.001Tren$



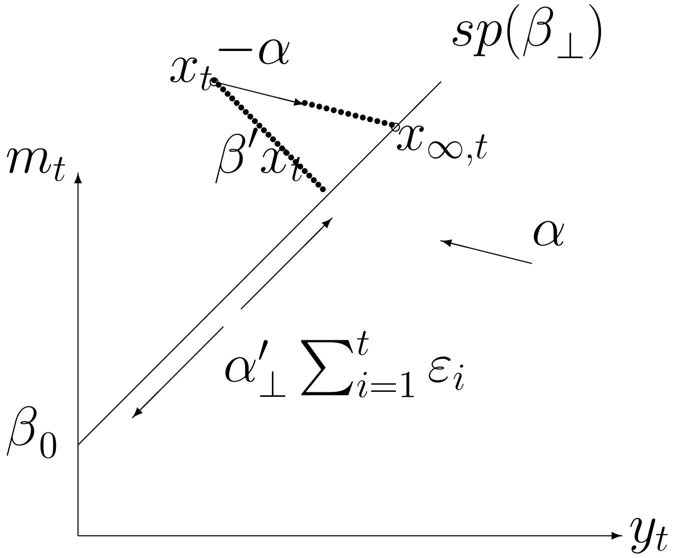
5.Plots of random walks







6.From the model $y_t = \gamma x_{t-1} + \varepsilon_{1t},\ \Delta x_t = \varepsilon_{2t}$, with $T = 100$ we have done 400 simulations



7.The process $x'_t = [m^r_t, y^r_t]$ is pushed along the attractor set by the common trends and pulled towards the attractor set by the adjustment coefficients