Modelling Long Run Relationships in Time Series

In this lecture. . .

- Financial time series and hedging problems
- Autoregressive process. Testing for stationarity
- How the long-run relationship works: equilibrium correction
- Case Study: cointegration among spot rates (market data)

By the end of this lecture you will be able to . . .

- understand integrated time series
- test for stationarity vs. the unit root PP , ADF
- understand error correction approach to linkage of time series
- estimate cointegration among a pair of time series using the Engle-Granger procedure

Introduction

Cointegration analysis is a powerful tool for investigating equilibrium trends in multivariate time series.

The trends are *common factors*, not statements about direction in asset price.

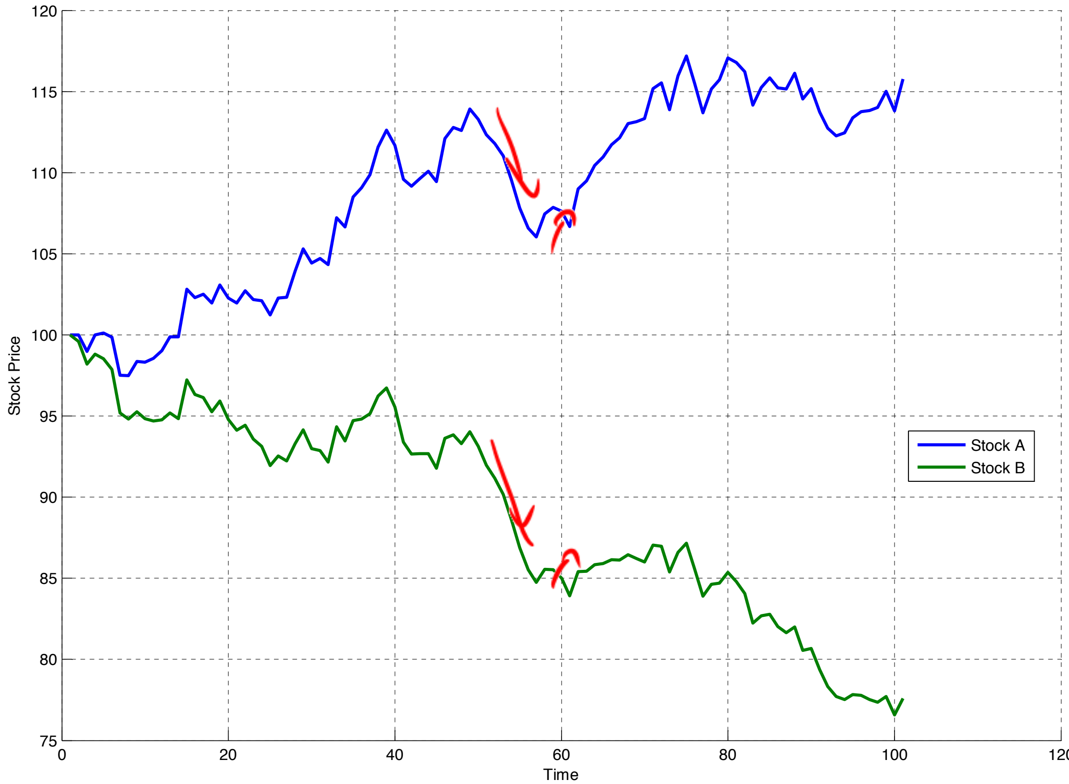
How do we work with empirical time series <u>in levels</u>, such as asset prices, CDS levels, or interest rates?

spurious

- The price levels are non-stationary.
- Unlike with differences or returns, **we can't correlate**.

Correlated Series

These time series are highly correlated but not cointegrated. Their spread possibly has an exponential fit.



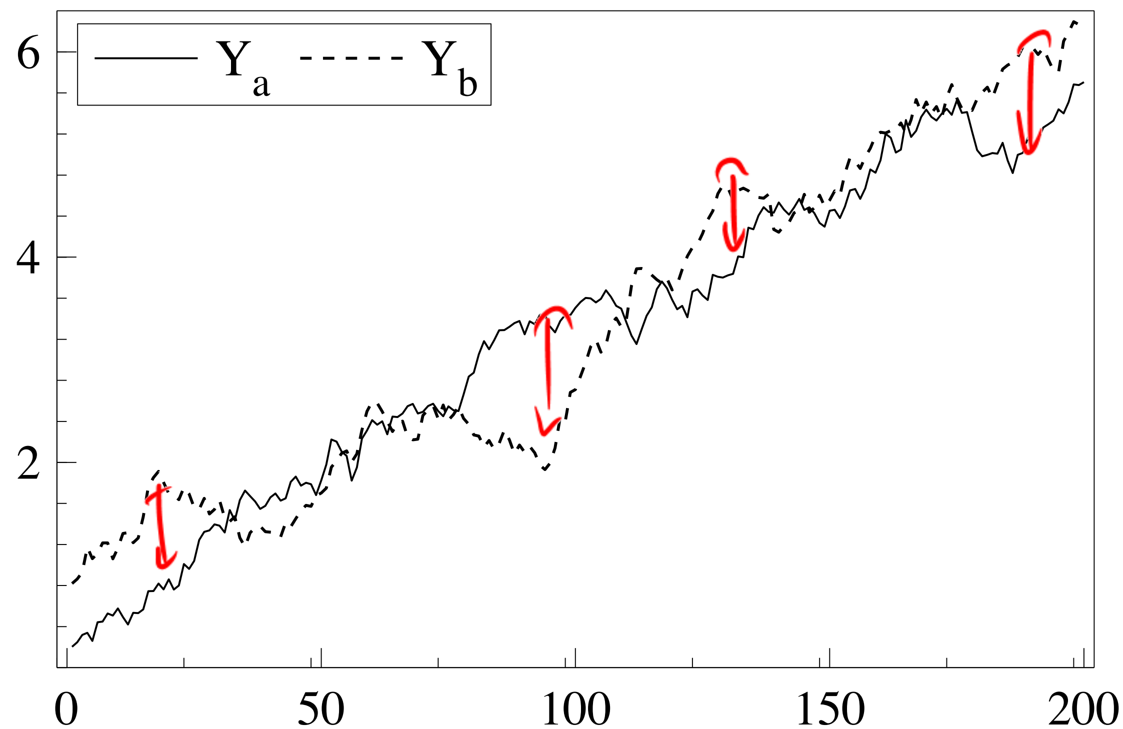
$$\frac{\Delta y_t}{y_{t-1}} = r_y$$

$$\frac{\Delta x_t}{x_{t-1}} = r_x$$

From *Correlation Sensitivity* CQF material.

No linear equilibrium

These series are **not** cointegrated. In fact, their spread contains a unit root. Multicollinearity is potentially irresolvable.



$[y_t - \beta x_t] = e_t$

stationary testing.

From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

Cointegrated Series

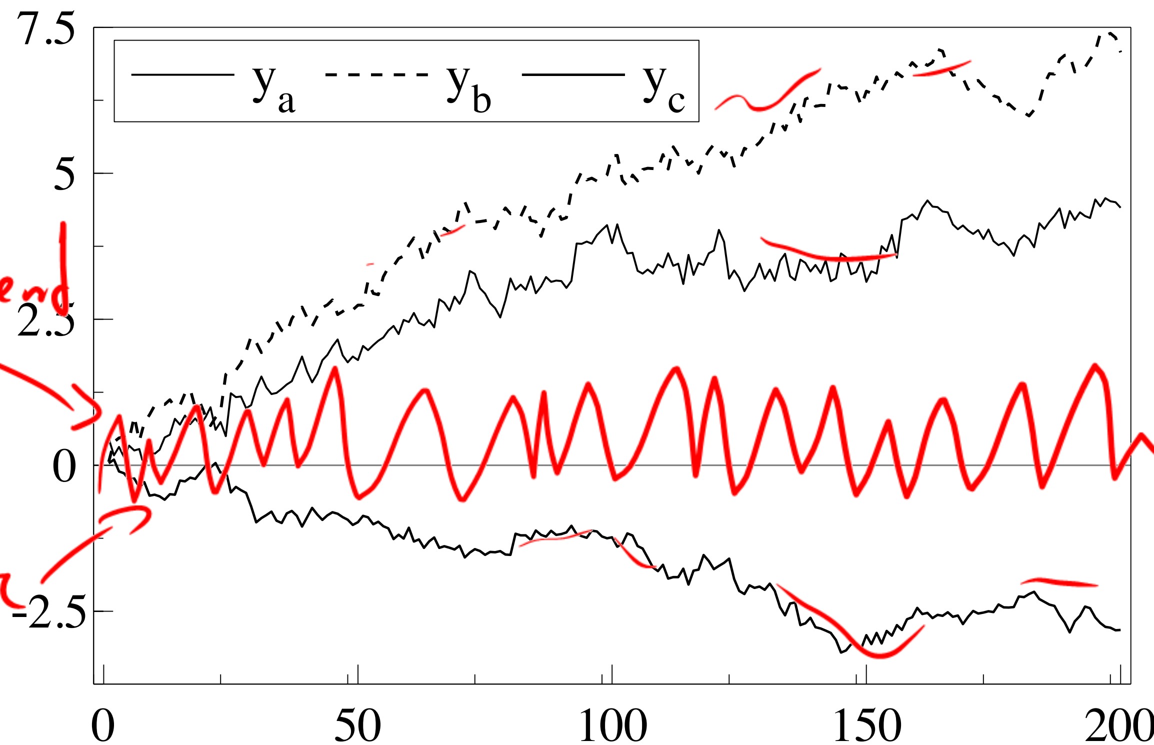
$[1, \beta_1, \beta_2] = \beta_{coint.}$

These series are **cointegrated**. Their linear combination produces a mean-reverting spread (common factor) e_t.

$y_a - \beta_1 y_b - \beta_2 y_c = e_t$

increment

stochastic process



stationary

$W_t - W_{t-1}$

GBM, BM

From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

A Multivariate Linear Combination

"There are fewer feedbacks than variables."

If a linear combination with some *special weights* β'_C produces **a stationary spread**:

$$
\begin{aligned}
e_t &= \beta'_C \boldsymbol{Y}_t \qquad e_t \sim I(0) \qquad (1)\\
&= \pm\beta_1 y_{1,t} \pm \beta_2 y_{2,t} \pm \cdots \pm \beta_n y_{n,t}
\end{aligned}
$$

$y_{i,t} \sim I(1)$

then **cointegration exists**. In a cointegrated system, **the common stochastic trend(s) drive all variables** in the long-run.

Autoregression in $\boldsymbol{Y}_t$ $\Rightarrow$ Error Correction $\Delta\boldsymbol{Y}_t$

Error correction means common term for $\Delta y_{1,t}, \Delta y_{2,t}, \ldots$

{common factor}

Mean-reverting spread

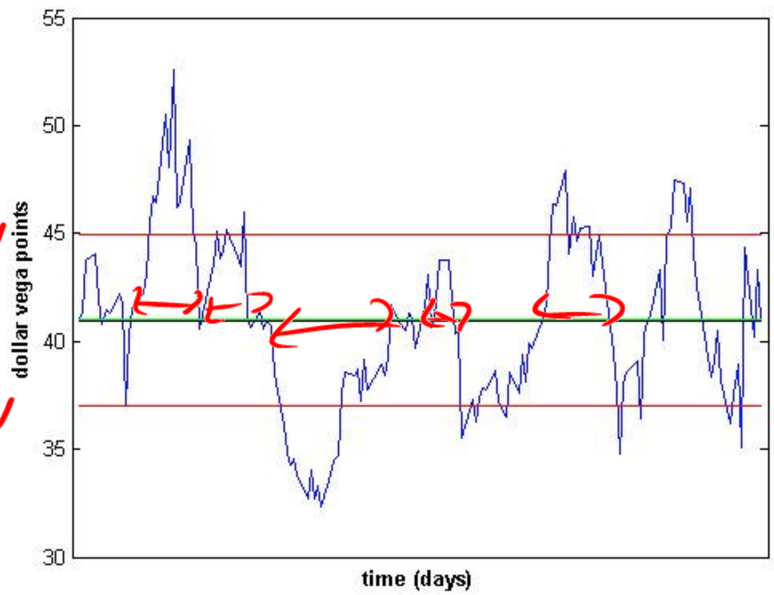
Cointegrating combination produces the special case of **cointegrating residual** $e_t = \beta'_C \boldsymbol{Y}_t$:

- it is stationary $I(0)$ and mean-reverting $\theta \gg 0$.
- Reversion speed $\theta \approx 44$ and bounds are calculated as $\sigma_{OU}/\sqrt{2\theta}$ $= \sigma_{eq}$

OU Process

$1\sigma_{eq}$

$-1\sigma_{eq}$



μ_e

$\theta \to \tau$

$\frac{\ln 2}{\theta}$

From: Diamond (2013). *Learning and Trusting Cointegration*

Cointegrated system

Q: What does the stationarity of e_t implies?

A: It means unit root in each of $y_{i,t}$ gets cancelled by the differencing. The common stochastic process gets removed.

Statistical arbitrage: we can explore mispricing that occurs when asset prices $y_{i,t}$ produce **a disequilibrium** $e_t \neq \mu_e$.

Hedging problem: also resolved if hedging ratios $\boldsymbol{\beta_C}$ give e_t – portfolio $\boldsymbol{\beta}'_{\boldsymbol{C}}\boldsymbol{Y_t}$ is reduced to the stationary spread and common factor(s) made explicit.

Autoregressive model is parsimonious

VAR(1) model provides a single-period forecast.

$Y_{t-2} + \varepsilon_{t-2})$

ARMA(1, 1) $\rightarrow$ MA(∞)

- We start with $\boldsymbol{Y_t} = \boldsymbol{\beta Y_{t-1}} + \boldsymbol{\epsilon_t}$

  AR(1) MA(1)

- $\boldsymbol{Y_t}$ depends on $\boldsymbol{Y_{t-1}}$, $\boldsymbol{Y_{t-1}}$ depends on $\boldsymbol{Y_{t-2}}$, and so on.

For returns... VAR appropriate but forecast is poor
(betas are decreasing exponentially, a moving average process)

For levels... ECM appropriate if cointegration exists
(autocorrelation fading slower than exponential, long memory)

Error Correction

$\Delta \boldsymbol{Y_t}$ is the change in market value between Y_{t-1} and Y_t.

- Naturally, the change is driven by the risk factors we intend to hedge against.

- Cointegration exists where there is a correction of error in $\Delta \boldsymbol{Y_t}$ from the equilibrium, but that is **not forecasting**!

Integrated process (unit root)

Think about the case when autoregressive $\boxed{\boldsymbol{\beta} = \mathbf{1}}$,

- then $\underline{\boldsymbol{Y_t} = \boldsymbol{\epsilon_t} + \boldsymbol{\epsilon_{t-1}} + \boldsymbol{\epsilon_{t-2}} + \ldots + \boldsymbol{Y_0}} = \sum \boldsymbol{\epsilon_i} + \boldsymbol{Y_0}$

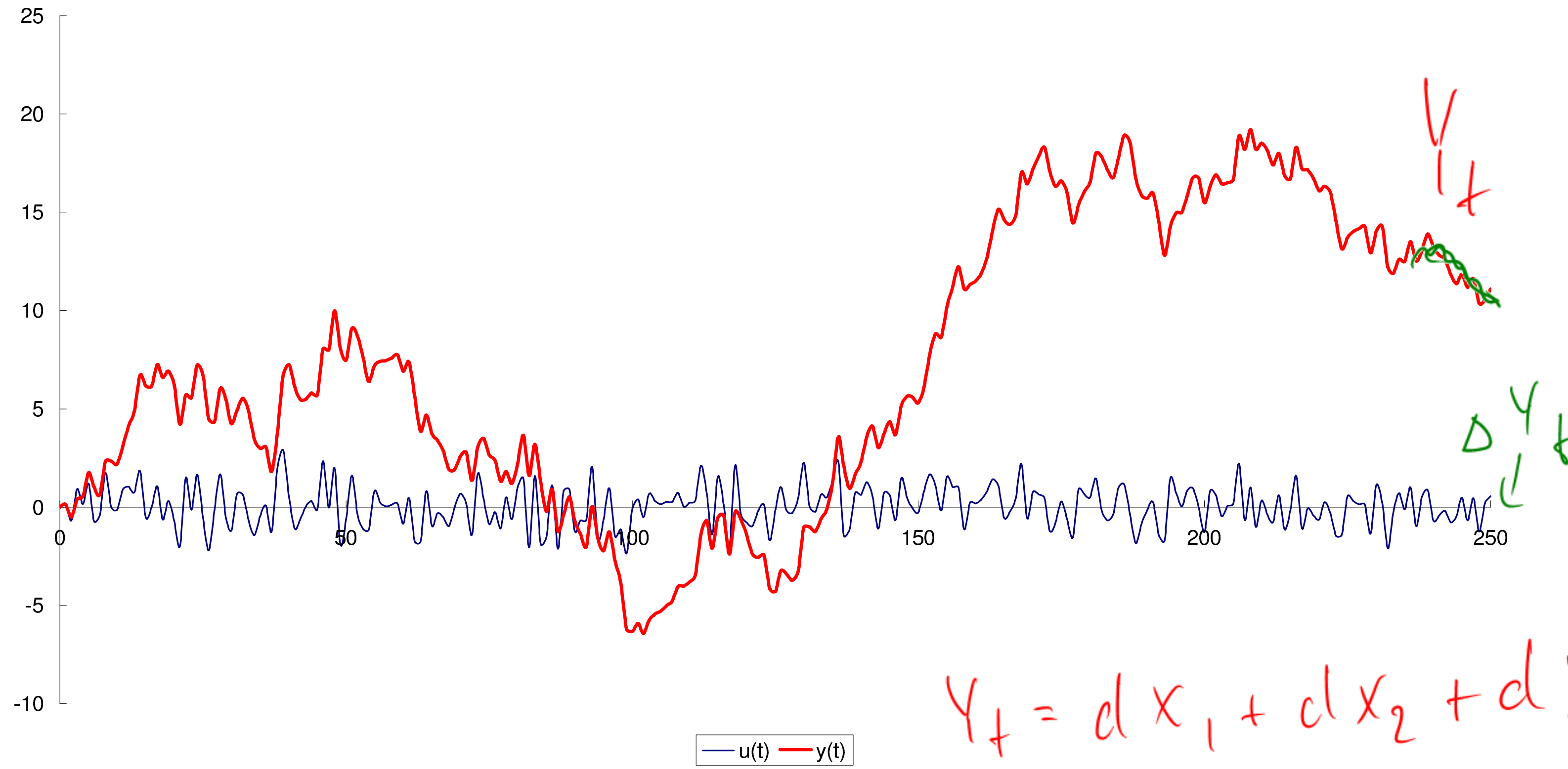
In continuous time, summation becomes an integration, so we say the process is **integrated of order one, I(1)** – or the process has a **unit root**.

Each residual $\epsilon_{t,\tau}$ is an increment of Brownian Motion dW_t

$$\underbrace{\sum \epsilon_t} \quad \Rightarrow \quad \sum dW_t \quad \Rightarrow \quad \epsilon_{t,\tau} \stackrel{D}{=} \underbrace{\int_t^{t+\tau} \boldsymbol{\sigma}\, dW_s}$$

Brownian Motion is **a limiting case** for the integrated series.

Brownian Motion: a random walk



$Y_t = dX_1 + dX_2 + dX_3 \ldots$

This simulated process adds up BM increments $dX_i = \phi_i \sqrt{\tau}$. **It is integrated, stationarity test will confirm the unit root.**

ASIDE Factor in SDE simulation

We use increments of the Brownian Motion, simulated as $\sigma dX_i = \sigma \phi_i \sqrt{\tau}$, to represent *a factor*.

Given that ϕ_i is a random number drawn from the Normal distribution, the factor possesses the same distribution.

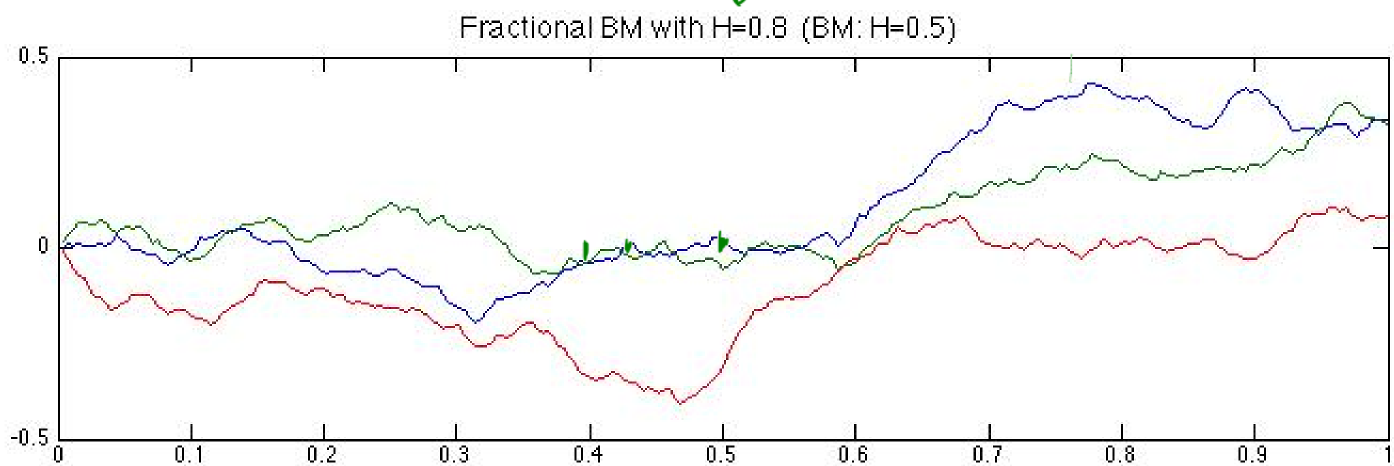
ϕ_i gives the number of standard deviations the variable moves. Large ϕ_i give large shocks.

What is a probability of $\phi_i > 1$, $\phi_i > 3$?

(Answers are 0.1587 and 0.00135)

What if our common factor is **Fractional Brownian Motion**?

interest rates



$\beta^{(t-s)}$

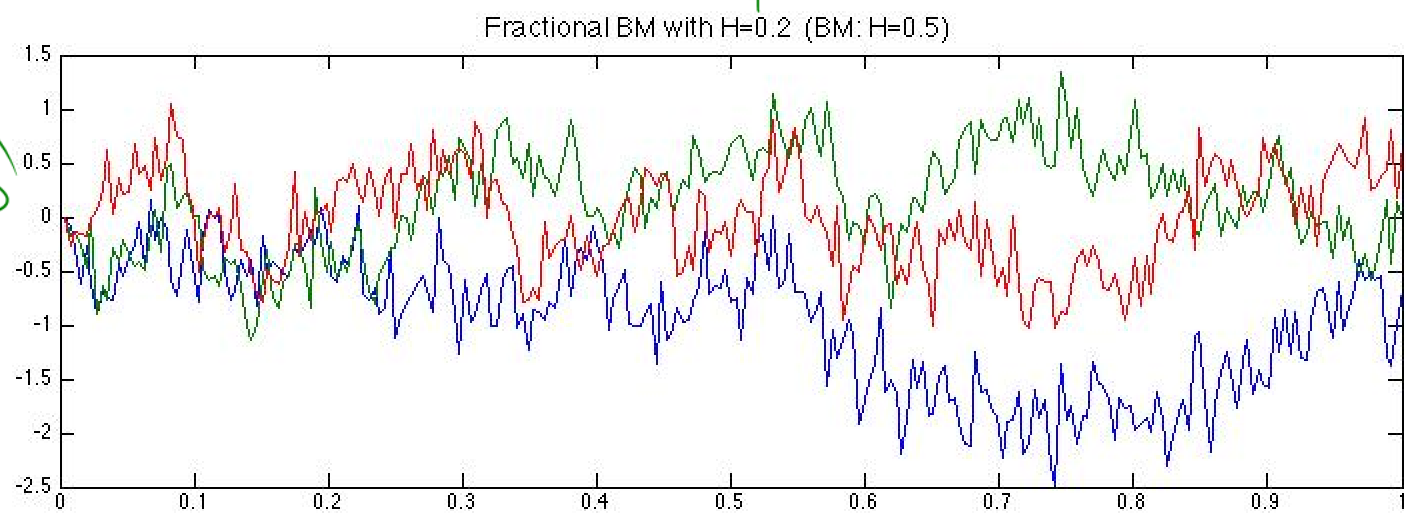
$\beta = 1$

Long Memory

$H = 0.5$

BM

specialty (equity)



From: Algorithm credit to Yingchun Zhou and Stilian Stoev (2005)

Long memory

The dual nature of the Fractional Brownian Motion allows to model integrated series with long memory (e.g., interest rates) by setting the Hurst exponent $H > 0.5$ as well as stationary-like series using the low values of the Hurst exponent $H < 0.2$.

$H = 0.5$ recovers the Brownian Motion, which is an I(1) series.

Long memory: the autocorrelation of the Fractional Brownian Motion decays according to **the power law** τ^{2d-1} which is slower than the exponential decay of the Ornstein-Uhlenbeck $e^{-\theta\tau}$ or $e^{\tau \ln \beta}$ for any stationary process $\beta < 1$.

$$H = d + \frac{1}{2}$$

END OF ASIDE

Testing Random Walk for a unit root

H_0

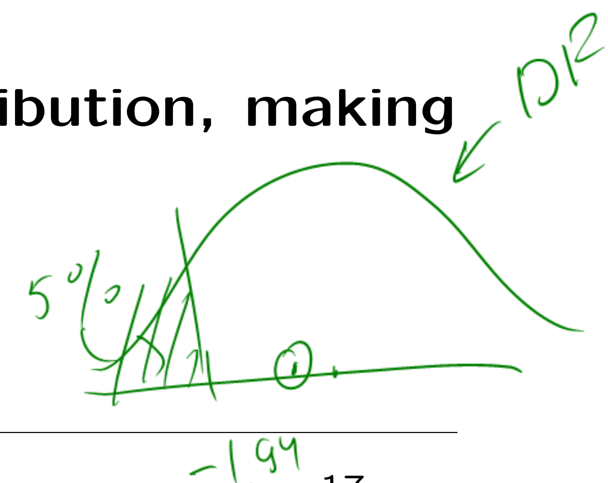
Null Hypothesis: Y_T has a unit root

Exogenous: None

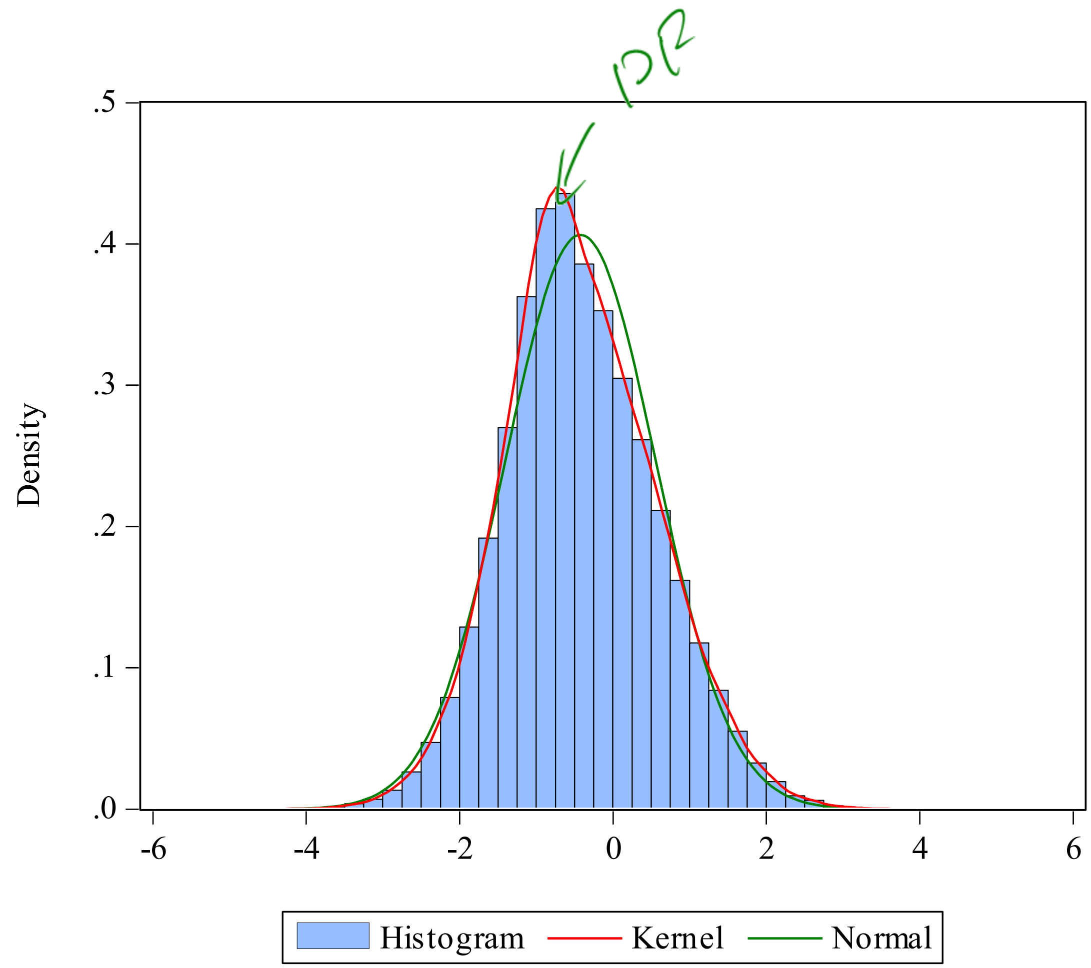
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*	
Augmented Dickey-Fuller test statistic		-0.432663	0.5261	p-value
Test critical values	1% level	-2.574245		
	5% level C.V.	-1.942099		
	10% level DF Distrib.	-1.615852		

DF gives higher critical values than t-distribution, making it difficult to reject a unit root hypothesis.



Bootstrapped Dickey-Fuller distribution



DF distribution is tabulated by Monte Carlo simulation (also called statistical bootstrapping).

Dickey-Fuller Test for unit root

We test for significance of ϕ in the following model:

OLS, usual regression

$$\begin{aligned} \Delta Y_t &= \phi \cancel{Y_{t-1}} + \epsilon_t \\ Y_t - Y_{t-1} &= (\beta - 1)Y_{t-1} + \epsilon_t \end{aligned}$$

If ϕ is not significant the time series has a unit root.

$$\phi = \beta - 1 = 0 \quad \Rightarrow \quad \beta = 1 \quad \Rightarrow \quad \Delta Y_t = \epsilon_t = \sigma W_t$$

Test statistic is calculated as $\frac{\beta}{\text{std error}}$.

But the distribution to compare to is the Dickey-Fuller.

Inference on non-stationary series

1. The **test statistic** for significance of ϕ is calculated as usual.

2. To make ϕ significant, the test statistic has to satisfy the higher critical value.

3. Since the standard error for y_t is under-estimated, we have to use 'the right distribution for the wrong test statistic.' The **critical value** is taken from the Dickey-Fuller distribution.

Conventional critical values (t distribution) lead to over-rejection of H_0 when it is true. The critical value of DF statistic is bootstrapped by generating *iid* residuals ϵ_t for

$$H_0 : \Delta Y_t = (1 - \beta)Y_{t-1} + \epsilon_t$$

$\phi_i \sqrt{2}$

Normal RV

Augmented Dickey-Fuller Test

To improve the Dickey-Fuller procedure, lagged differences Δy_{t-k} 'augment' the test, improving robustness *wrt* serial correlation

$$\Delta y_t = \phi y_{t-1} + \underbrace{\sum_{k=1}^{p} \phi_i \Delta y_{t-k}}_{\Delta y_{t-1} + \Delta y_{t-1}} + \epsilon_t$$

- Insignificant ϕ means unit root for series y_t. DF critical values re-tabulated for each number of lags k setup.

Beware that software-implemented statistical tests offer to add constant '**drift**' or time-dependent '**trend**'. $\Delta y_t = \text{const.}$

$$\overset{(1)}{\Delta y_t} = \overset{(2)}{\phi y_{t-1}} + \sum_{k=1}^{p=1} \overset{(5)}{\phi_i \Delta y_{t-k}} + \cancel{\underbrace{\overset{(3)}{\text{const}} + \overset{(4)}{\beta_t t}}} + \epsilon_t$$

These modifications are your false friends because they create temporary dependency and give **overfitted results**.

Statistical tests implemented in R usually present the underlying regression equation.

So you are able to identify parameters and understand whether excessive specification was used.

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend
                  1         2        3    4       5
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
```

$$\Delta y = y_{t-1} + \text{const} + \beta t + \Delta y_{t-1}$$

Stationarity Tests: overview

There are several formal ways of testing for the presence of a unit root in a non-stationary process.

- We focused on **Dickey Fuller test** that keeps familiar AR(p) model and is popular because of its simplicity and generality.

- A non-parametric Phillips-Perron test transforms t-statistic to further account for autocorrelation and heteroscedasticity effects.

- *Co-integration Regression Durbin Waston Test*, is based on the common Durbin-Watson statistic.

Regression as Equilibrium Model and How Cointegration Works

Static Equilibrium Model

The familiar linear regression is **the** equilibrium model!

$$y = \beta_0 + \beta_1 x$$

In this *static*, stationary y_t and x_t produce **constant** b_g for

$$\Delta y = b_g \Delta x$$

The steady-state of equilibrium transpires through this constant growth rate β_g.

Static Equilibrium Model

CAPM a case of static equilibrium model! linear factor model. It relies on constant beta.

$$\mathbb{E}[r_I] = \beta\left(\mathbb{E}[r_M] - r_{rf}\right) + r_{rf}$$

$$\mathbb{E}[r_I - r_{rf}] = \beta\,\mathbb{E}[r_M - r_{rf}]$$

LFM $\quad \Delta r_I = \beta\,\Delta r_M + \beta_j F_j$

Since regression is involved, CAPM is also a Linear Factor Model.

Asset returns are regressed on Factors $\beta_j F_j$. The factors are linearly independent among themselves.

Equilibrium in STOCHASTIC Models

Assume that y_t and x_t are non-stationary time series **in levels** (e.g., prices/CDS/rates).

The static equilibrium model gives a short run relationship.

$$\Delta y = \beta_g \Delta x$$

Corr $(\Delta y, \Delta x)$

Correlation is estimated among the differences...

Time scale: daily, 10M, etc

What about the relationship in the long run?

If there is a common factor, it must affect *the changes* in y_t.

The same principle as with portfolio factor models (HML, SMB):

we regress returns (differences) on the common factor

$$\underbrace{\Delta y = \beta_g \Delta x}_{\text{Static Eq}} + \underbrace{\text{Factor Term}} + \ldots + \epsilon_t$$

It turns out that the common factor is

$$\hat{e}_t \;=\; y - \hat{b}x - \hat{a}$$

$$\Delta y \approx \Delta x \qquad\qquad \text{and} \qquad \Delta y \approx (y - \hat{b}x)$$

s.t. $\hat{e}_t$ being stationary so that $[1, b]$ is a co-integrating vector.

Equilibrium Correction Model

The model addresses both, the short-run correlation-like $\beta_1 \Delta x_t$, and equilibrium correction working (slowly!) over the long-run

$$\Delta y_t \;=\; \beta_1 \Delta x_t - (1-\alpha)\left(y_{t-1} - b_e x_{t-1} - a_e\right) + \epsilon_t$$

$$\text{where } e_{t-1} = y_{t-1} - b_e x_{t-1} - a_e \quad \text{and} \quad \mathbf{E}[e_{t-1}] = a_e$$

The disequilibrium $e_{t-1} \neq a_e$ is corrected over the long-run.

The speed of correction $-(1-\alpha)$ is inevitably small, but <u>must be significant for cointegration to exist</u>.

Modelling problems

significance ECM

$$\Delta y_t = \beta_1 \Delta x_t - (1-\alpha)e_{t-1} + \epsilon_t$$

$$\Delta x_t = \beta_1 \Delta y_t - (1-\alpha) e_{t-1} + \varepsilon_t$$

- The assumption of x_t **being leading/exogenous/causing** variable.

- Equilibrium-correction mechanism is **linear**: if the ‘error’ e_{t-1} above μ_e the model suggests a small correction downwards (and vice versa).

- Non-unique cointegrating a, b are empirically possible so the speed of correction becomes **a calibrated parameter**

Estimating Cointegration - Pairwise

- **Pairwise Estimation**: select two candidate time series and apply ADF test for stationarity to the joint residual.

  Use the estimated residual to continue with the Engle-Granger procedure.

  Perform the Engle-Granger procedure in both ways,

$$\left[\begin{array}{ccc} \Delta y_t & \text{on} & \Delta x_t \\ \Delta x_t & \text{on} & \Delta y_t \end{array}\right]$$

Cointegration Case B offers R code that re-implements the ECM estimation explicitly. Then, VECM estimation routine is used to analyse further.

Engle-Granger procedure

Step 1. Obtain the fitted residual $\hat{e}_t = y_t - \hat{b}x_t - \hat{a}$ and test for unit root.

- That *assumes* cointegrating vector $\boldsymbol{\beta}'_{Coint} = [1, -\hat{b}]$ and equilibrium level $\mathbb{E}[\hat{e}_t] = \hat{a} = \mu_e$.

ADF lag, no trend

- **If the residual non-stationary** then no long-run relationship exists and regression is spurious.

Step 2. Plug the residual from Step 1 into the ECM equation and estimate parameters β_1, α (with linear regression)

$$\underset{\text{Data}}{\Delta y_t} = \beta_1 \underset{\text{Data}}{\Delta x_t} - (1 - \alpha)\underset{\text{Data}}{\hat{e}_{t-1}}.$$

OLS

- It is required **to confirm the significance for** $(1 - \alpha)$ coefficient.

{ Step 3. Run VECM to get optimal $[1, -\beta_c]$ }

Let's explore the cointegration and specify the appropriate eqn. correction model (ECM) for the spot rates. We will use daily market data from the Bank of England.

Case Study B

Cointegration Estimation Overview (Reference)

1. **Engle-Granger Procedure** to estimate a cointegrating relationship and error correction between a pair of time series.

   - – The choice of leading variable x_t removes uncertainty about cointegrating vector $[1, \beta]$.

     The procedure can work well for a basis relationship.

2. **Johansen MLE Procedure** to estimate a set of cointegrating relationships $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'_C$ in a multivariate setting.

   50 x 50 reduced rank

   - – Multivariate estimation relies on the theorem of the reduced-rank matrix with r linearly independent rows.

     The procedure estimates $\boldsymbol{\beta}'_C$ and then infers calibrated $\boldsymbol{\alpha}$.

Summary

Please take away the following ideas...

- Financial time series are non-stationary and so, can only be analysed with **cointegration** and **error correction**. Δy_t

- Cointegration: if a linear combination of non-stationary time series produces the stationary spread, it means that variables are driven by the stochastic process in common. e_t

- Error Correction: the long-run relationship between two time series transpires as the cointegrated residual in $\Delta \boldsymbol{Y_t}$. $= \dots -\underbrace{(1-\delta)e_{t-1}}_{\text{common factor}}$

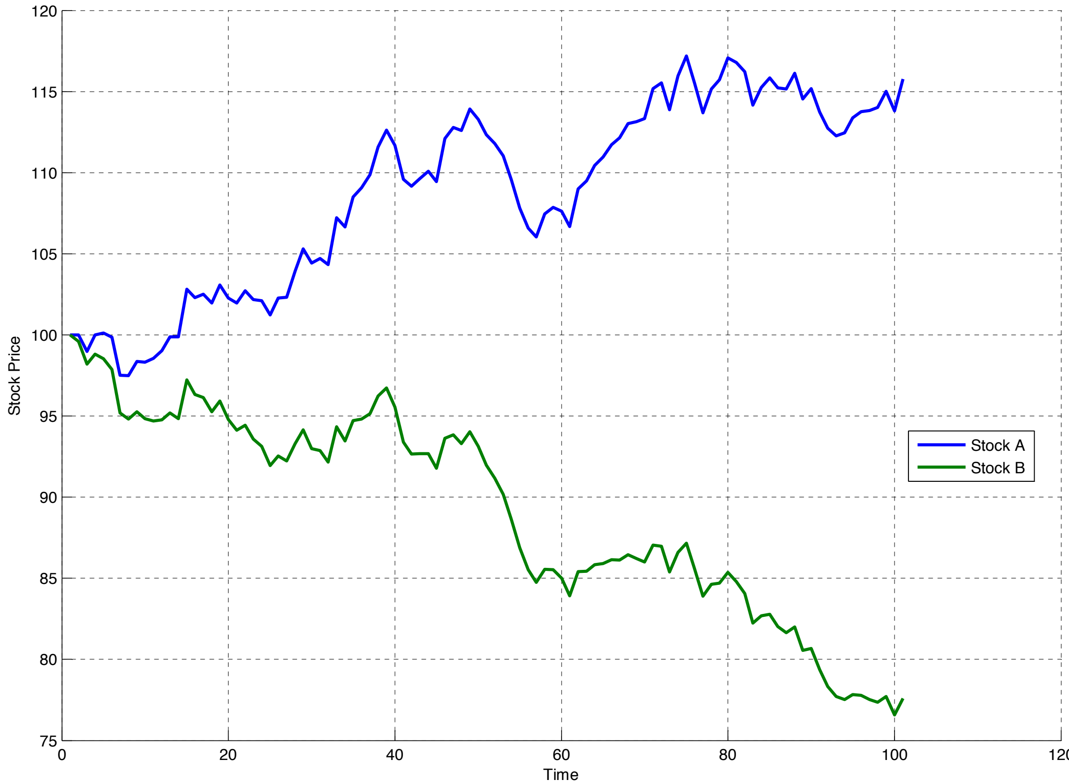
- Use of cointegration for hedging and statistical arbitrage requires checks using multivariate estimation (VECM, Johansen) and base on the idea of a reduced rank.

Cointegrated Series

Extra Slides

Correlated Series

These time series are highly correlated but not cointegrated. Their spread possibly has an exponential fit.

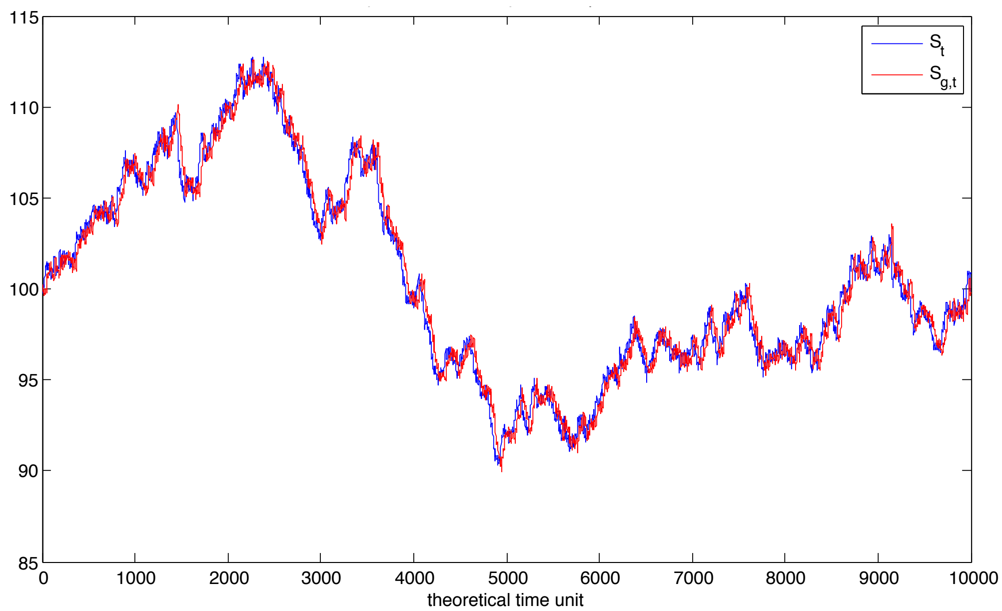


From *Correlation Sensitivity* CQF Lecture.

Cointelation

These series have been generated from the following processes:

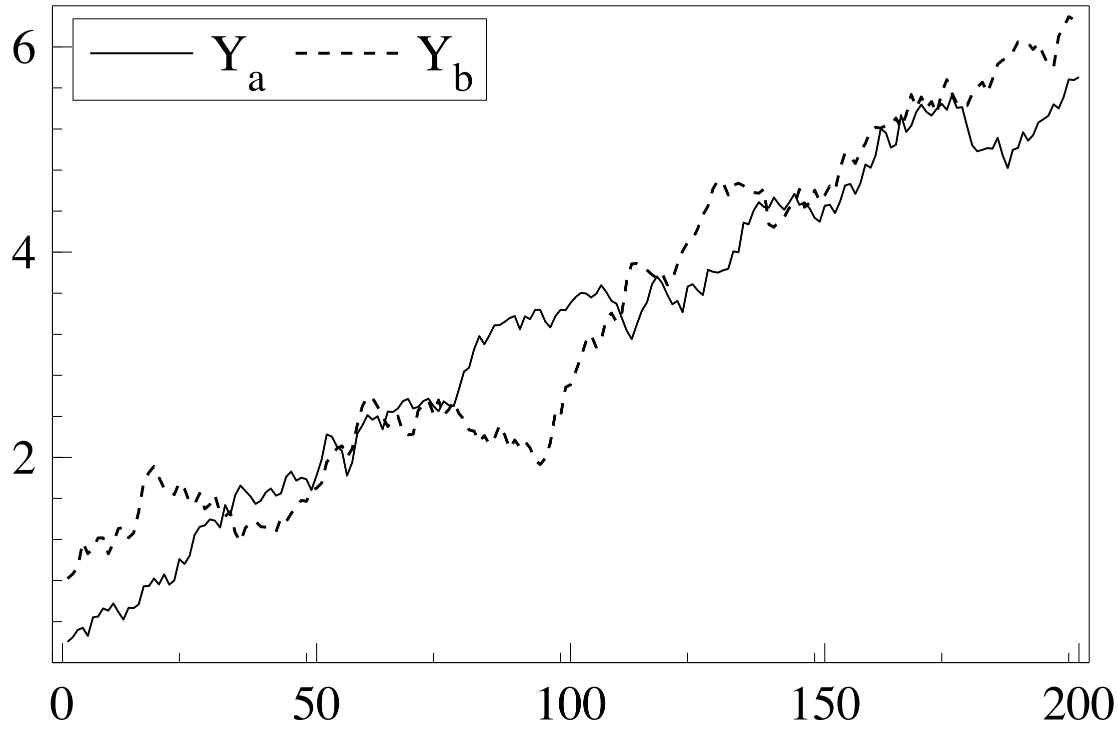
$$\begin{aligned}\frac{dS_t}{S_t} &= rdt + \sigma dW_t \\ dS_{l,t} &= -\theta(S_{l,t} - S_t)dt + \sigma S_{l,t}\left(\rho dW_t + \sqrt{1-\rho^2}dW_t^{\perp}\right)\end{aligned}$$



From Damghani (2014). *Introduction to the Cointelation Model*. CQF Extra

No linear equilibrium

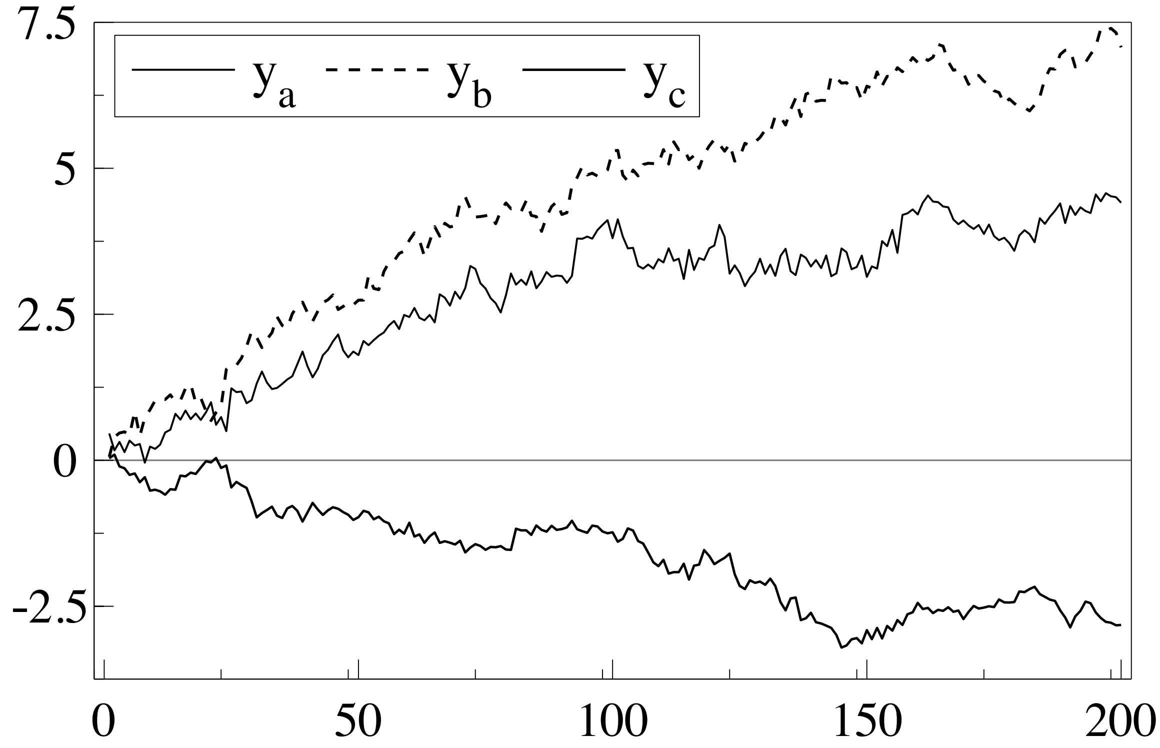
These series are **not** cointegrated. In fact, their spread contains a unit root. Multicollinearity is potentially irresolvable.



From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

Cointegrated Series

These series are **cointegrated**. Their linear combination produces a mean-reverting spread (common factor) e_t.



From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

Cointegrated system

"There are fewer feedbacks than variables."

In a cointegrated system, **the common stochastic trend(s) drive all the related variables in the long-run**.

We are interested to trade **cointegrating residual** e_t which is

- Stationary (has no unit roots) $I(0)$
- Autoregressive $AR(1)$, **not** decomposable as $MA(\infty)$ series
- Mean-reverting $\theta \gg 0$

Common Factor: rates example

The linear combination $\boldsymbol{\beta'_{Coint}Y_t}$ exposes a shared unit root, called *'a stochastic process in common'*.

Think of exposure to the common factor and what it could be.

- Cointegrating vector $[1, -\beta]$ gives hedging ratios for bonds.

- $Z(t;\tau_1) - \beta Z(t;\tau_2) = e_t$ is stationary $I(0)$

- Risk Factor: parallel shift of the yield curve.

A Multivariate Linear Combination

If a linear combination with some *special weights* $\boldsymbol{\beta}'_C$ produces **a stationary spread**:

$$\begin{aligned} e_t &= \boldsymbol{\beta}'_C \mathbf{Y}_t \qquad e_t \sim I(0) \\ &= \pm\beta_1 y_{1,t} \pm \beta_2 y_{2,t} \pm \cdots \pm \beta_n y_{n,t} \end{aligned} \tag{2}$$

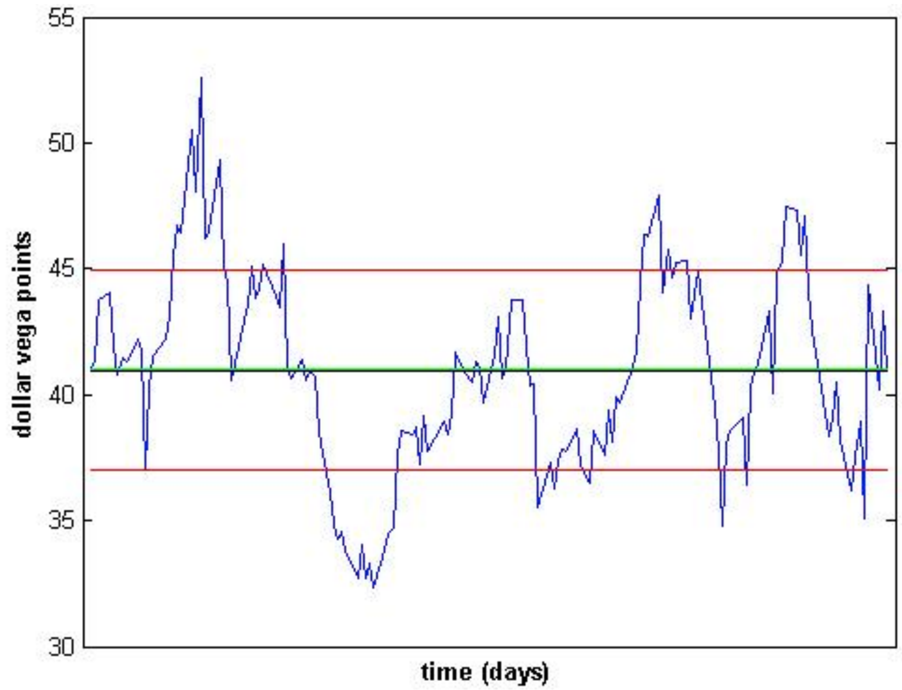
then we can explore mispricing that occurs when asset prices $y_{i,t}$ produce **a disequilibrium** $e_t \neq \mu_e$.

- The cointegration is *alike differencing* among time series.
- Left after the differencing is a **cointegrating residual** e_t. It is stationary $I(0)$ and mean-reverting $\theta > 0$.

Mean-reverting spread

The linear cointegrating combination $\beta'_C Y_t = e_t$ produces a stationary and mean-reverting spread:

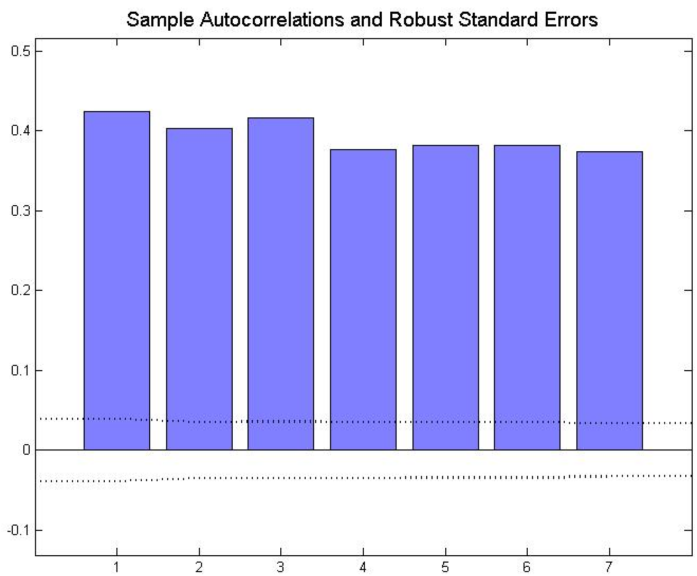
- Reversion speed $\theta \approx 44$ and bounds are calculated as $\sigma_{OU}/\sqrt{2\theta}$

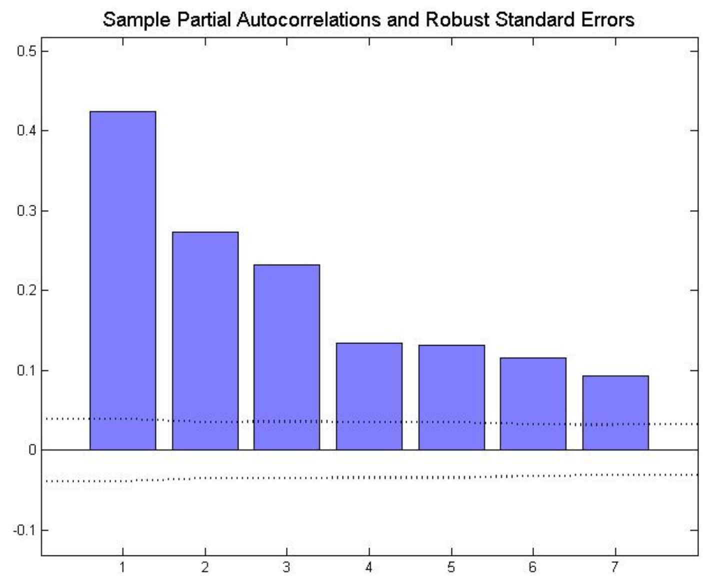


From: Diamond (2013). *Learning and Trusting Cointegration*

ACF and PACF for a high frequency spread e_t

Serial autocorreation $AR()$ is in prominent for e_t.





The process is stationary but ACF has no exponential decay in autocorrelation $\text{Corr}[Y_t, Y_s]$.

Statistical Arbitrage with Cointegration

Statistical arbitrage fundamentals

Makes two claims that **a.** relative mispricing persists and **b.** pricing inefficiencies are identifiable with statistical models.

The product of hedging is a hedging error, and the manageable error behaves like **the cointegrated residual** $e(t)$.

- Otherwise, the hedging error behaves like a random walk (unbounded) due to the unit root in $\boldsymbol{Y_t}$ when $\beta \approx 1$, common to all financial time series in levels.

Quality of mean-reversion

Q: How do we evaluate the quality of mean-reversion and find out entry/exit trade points?

A: We fit the spread to the Ornstein-Uhlenbeck process.

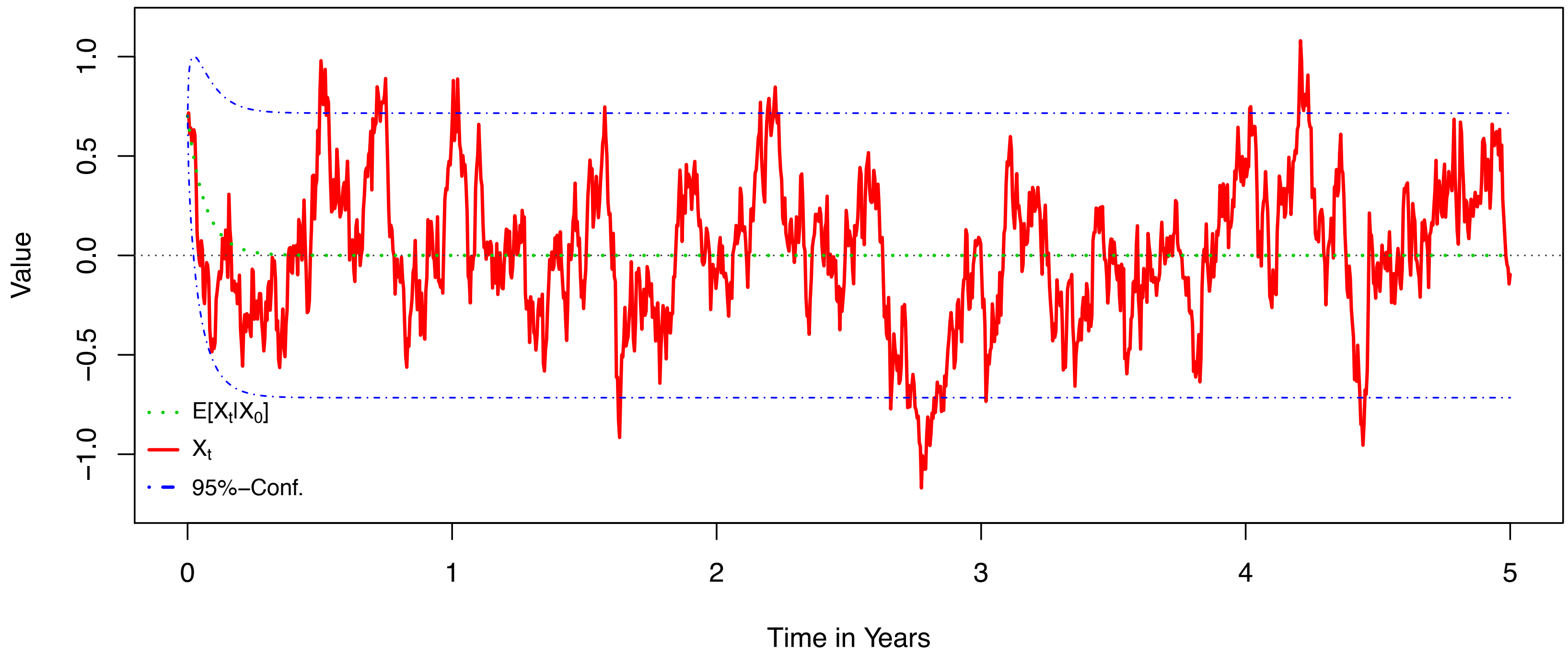
- Quality of mean-reversion is associated with the higher critical values equilibrium-correcting term (empirically).

- Cointegration is a filter on data: mean-reversion is of lower frequency than the data.

E.g., 10 Min data can generate half-life counted in weeks.

Simulated Ornstein-Uhlenbeck process

Here is how the simulated OU process looks like (sample path)



From: Harlacher (2012). *Cointegration Based Statistical Arbitrage*

Mean-reverting but has a different kind of stationarity than an AR(1) process. Why? There is diffusion.

Designing a trade

In order to design an arbitrage trade, one requires the following items of information:

1. **Weights** β'_{Coint} for a set of instruments to obtain the spread.

2. **Speed of mean-reversion** in the spread. For explanation purposes, the speed can be presented as **half-life**, the time between the equilibrium situations, when spread $e_t = \mu_e$.

$$\theta \to \tau$$

3. **Entry and exit levels** defined by σ_{eq}. Optimisation involved.

The inputs allow to backtest P&L and estimate drawdowns.