Modelling Long Run Relationships in Time Series

In this lecture. . .

- Financial time series and cointegration
- Integrated series. Testing for stationarity
- How the long-run relationship works: equilibrium correction
- Case Study: cointegration among spot rates (market data)
- Testing for multivariate cointegration *extra*

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

- know why you cannot regress on prices
- understand integrated time series and DF test for unit root
- understand error correction formulation and Engle–Granger procedure
- confirm the long-run relationship between a pair of prices

Introduction

Cointegration analysis is a powerful tool for investigating *common factors* – sources of randomness – among price time series.

$$\text{Price}_A - \beta\,\text{Price}_B$$

is not a random walk! This spread is forecastable.

We have tried to predict individual asset price direction using ML Classifiers.

The proper statistical model 'to predict' one price from another (forecast the spread) is cointegration.

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

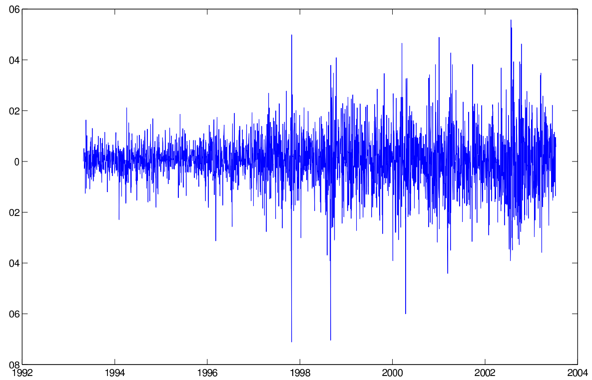
- The price levels are non-stationary.

- Regressing one price series on another is *spurious*.
  R^2 improves with N_{obs}

- Unlike with differences or returns, **we do not correlate**.

What is the underlying model of correlation?

Asset Returns



$$
\begin{aligned}
R_t^{CA=1} &= \beta_{1,0} + \beta_{11} R_{t-1}^{CA} + \beta_{12} R_{t-1}^{FR} + \ldots \beta_{1n} R_{t-1}^{US} + \ldots {}_{t-2} \ldots + \epsilon_{1,t} \\
R_t^{FR=2} &= \beta_{2,0} + \beta_{21} R_{t-1}^{CA} + \beta_{22} R_{t-1}^{FR} + \ldots \beta_{2n} R_{t-1}^{US} + \ldots {}_{t-2} \ldots + \epsilon_{2,t} \\
\ldots & \quad \ldots \\
R_t^{US=n} &= \beta_{n,0} + \beta_{n1} R_{t-1}^{CA} + \beta_{nn} R_{t-1}^{FR} + \ldots \beta_{nn} R_{t-1}^{US} + \ldots {}_{t-2} \ldots + \epsilon_{n,t}
\end{aligned}
$$

For returns... VAR is appropriate but **forecast is poor**.

Forecasting Market Returns

Vector Autoregression **fails** at forecasting daily returns.

For the 2011 data the full model, AIC BIC-tested for optimal lag and stability-checked, forecasting accuracy as follows:

	S&P 500	FTSE 100	HSE	N225
MSE	0.0001	0.0001	0.0001	0.0001
MAPE	1.0175	1.3973	2.5325	1.0111

MAPE results suggest a deviation $O(100\%)$ to $O(200\%)$.

Daily returns for a broad market are a very small quantity.

Static Equilibrium Between Returns

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

$$\mathbb{E}[r_A - r_{rf}] \;=\; \beta\,\mathbb{E}\left[r_M - r_{rf}\right]$$

CAPM assumes the existence of true and constant β.

Aside: betas are estimated by OLS wrt factors (eg, market or market segment returns series) in form of Linear Factor Model:

$$R_t^A \;=\; \alpha + \beta\,R_t^M + \beta_j F^j + \epsilon_t$$

Cointegration in Prices

The static equilibrium in changes in two asset prices – the steady-state means existence of the true and constant β_g, a growth rate.

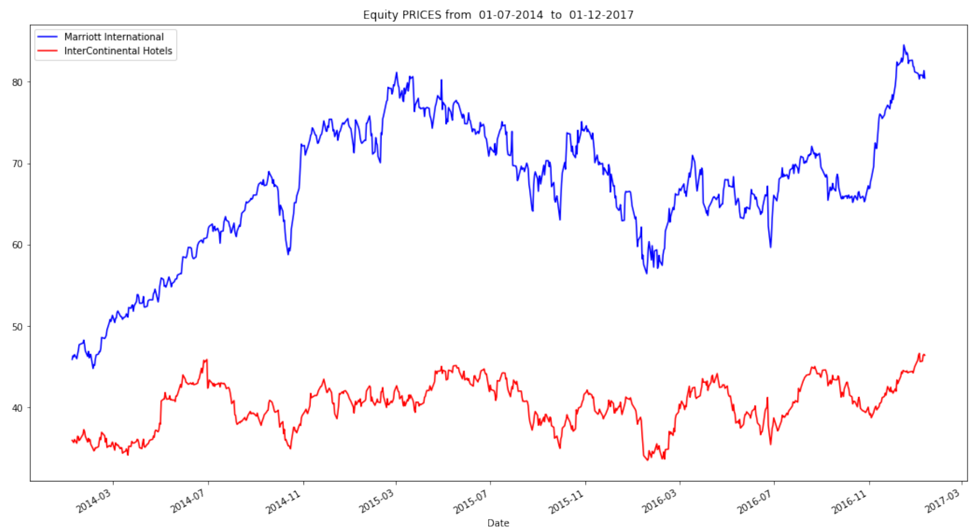
$$\Delta P_t^A = \beta_g \, \Delta P_t^B$$

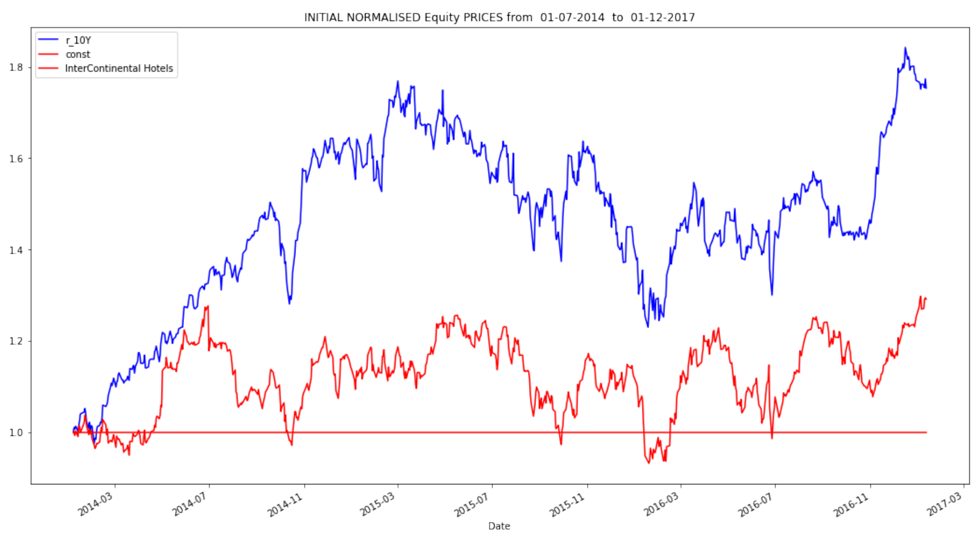
What about **the long run**?

$$\begin{aligned} \Delta P_t^{A=1} &= \beta_{11} \, \Delta P_t^B + \beta_{12} \, \text{Coint\_Factor} \\ \Delta P_t^{B=2} &= \beta_{21} \, \Delta P_t^A + \beta_{22} \, \text{Coint\_Factor} \end{aligned}$$

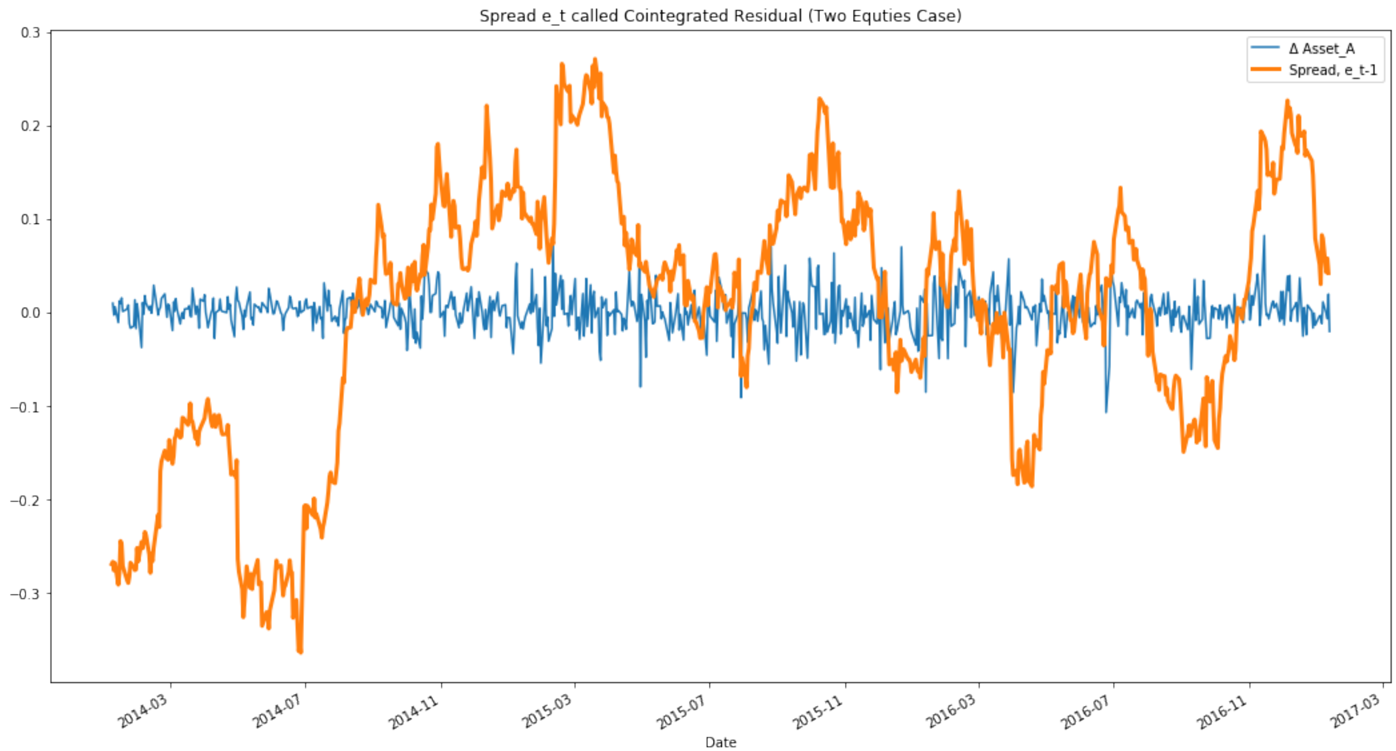
For equities/futures/rate levels... this correction model is appropriate, not naive regression of P_t^A on P_t^B .

Cointegration in Equities





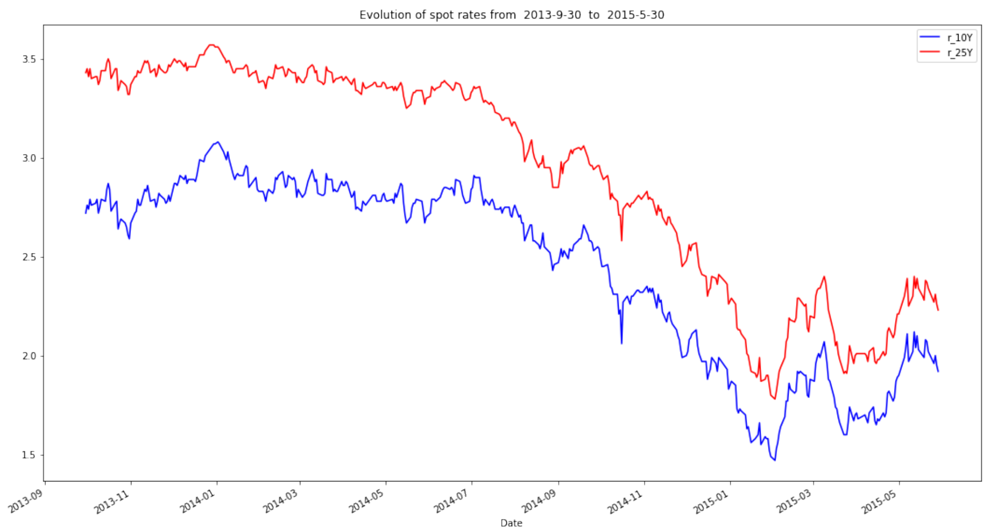
Cointegrated Residual – equities

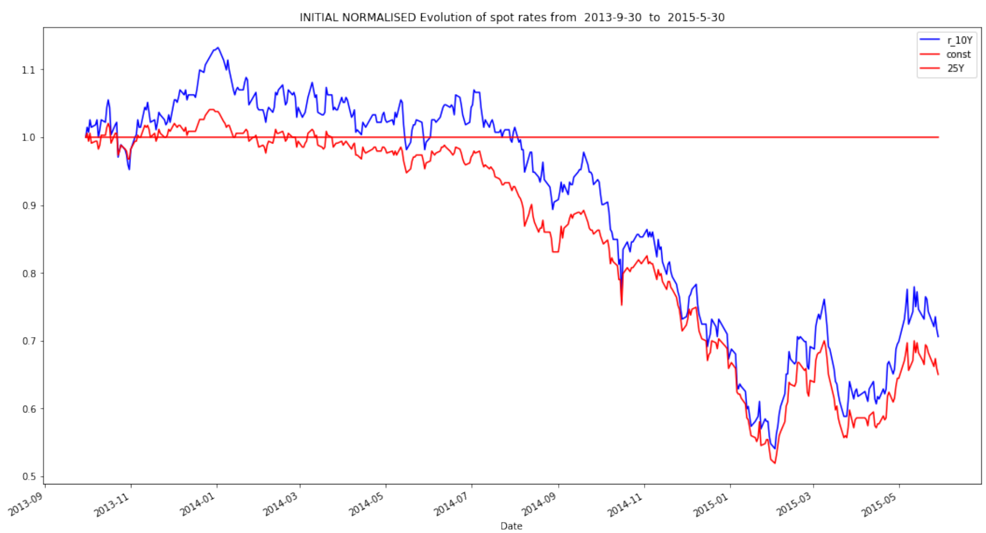


Those two equities proven to be cointegrated over the period – carefully chosen by examination over different time 'windows'.

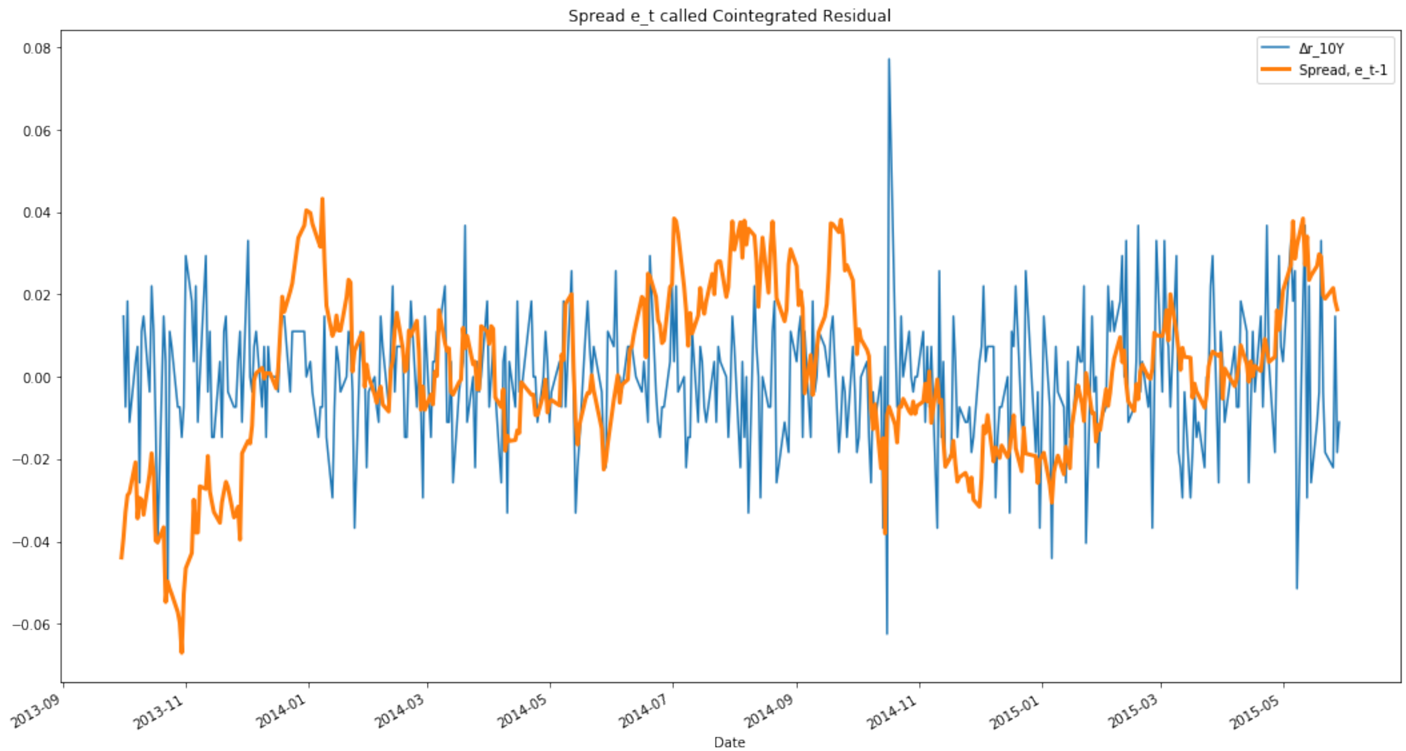
Half-life $< 3-4$ months, it is remarkable but these arb opportunities do exist.

Cointegration in Rates





Cointegrated residual – rates



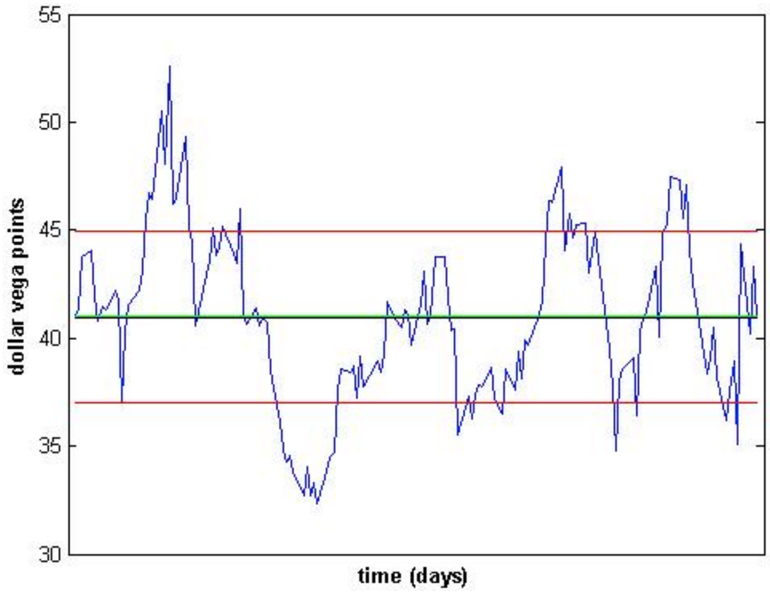
Those two rates r_{10Y} and r_{25Y} proven to be cointegrated over given period.

The residual is $r_{10Y} - \beta_C r_{25Y}$ confirmed to be 1) stationary and 2) common factor in Error Correction.

Dynamic Equilibrium

This mean-reverting spread produced by two co-moving series – is a dynamic equilibrium.

As opposed to static equilibrium which is CAPM β.



Spread example from a hedged basket of VIX Futures. Cointegrated residual $e_t = P_t^{Fut1} - \beta_2 P_t^{Fut2} - \ldots - \beta_n P_t^{Fut\,N}$ is fitted to OU process $(\theta, \mu, \sigma_{OU}, \sigma_{eq})$.

Diamond, R (2013) *Learning and Trusting Cointegration*, WILMOTT.

Uses of cointegration

Cointegration is useful in hedging applications: particularly when hedges are from a different asset class (eg, oil price vs oil-extracting country FX).

- **Equities**: the company and target of its acquisition (particularly when stock swap involved)
- **Commodities**: US heating oil vs. natural gas (competing fuels). Agricultural futures.
- **Rates**: funding rates at short end, long-term rates affecting middle of yield curve.
- **Term Structure**: VIX futures, a very liquid market

For very large samples $> 10Y$, it is likely to find co integration in commodities, global equity indicies.

Granger-Johansen Representation

The stationarity of spread e_t was **a discovery** in statistics.

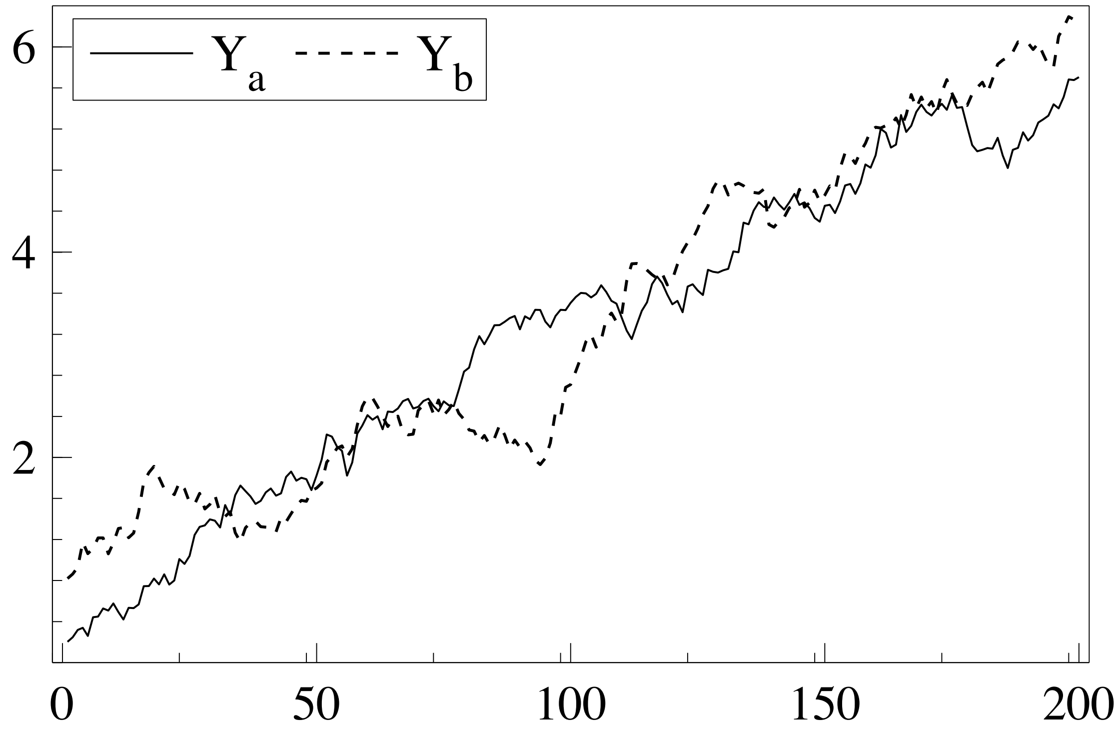
Implies there is a stochastic process (e.g. random walk) that is so similar in two series that it gets removed by linear differencing,

$$
\begin{aligned}
P_t^A &= P_0^A + \underbrace{\sum \epsilon_s^A} && \text{integrated process} \\
P_t^B &= \beta P_{t-1}^A + \epsilon_t^B && \text{regress and assume } \beta \approx 1 \\
P_t^B &= P_0^A - \epsilon_t^A + \epsilon_t^B + \underbrace{\sum \epsilon_s^A}
\end{aligned}
$$

"There are fewer feedbacks than variables."

Explaining Example 1

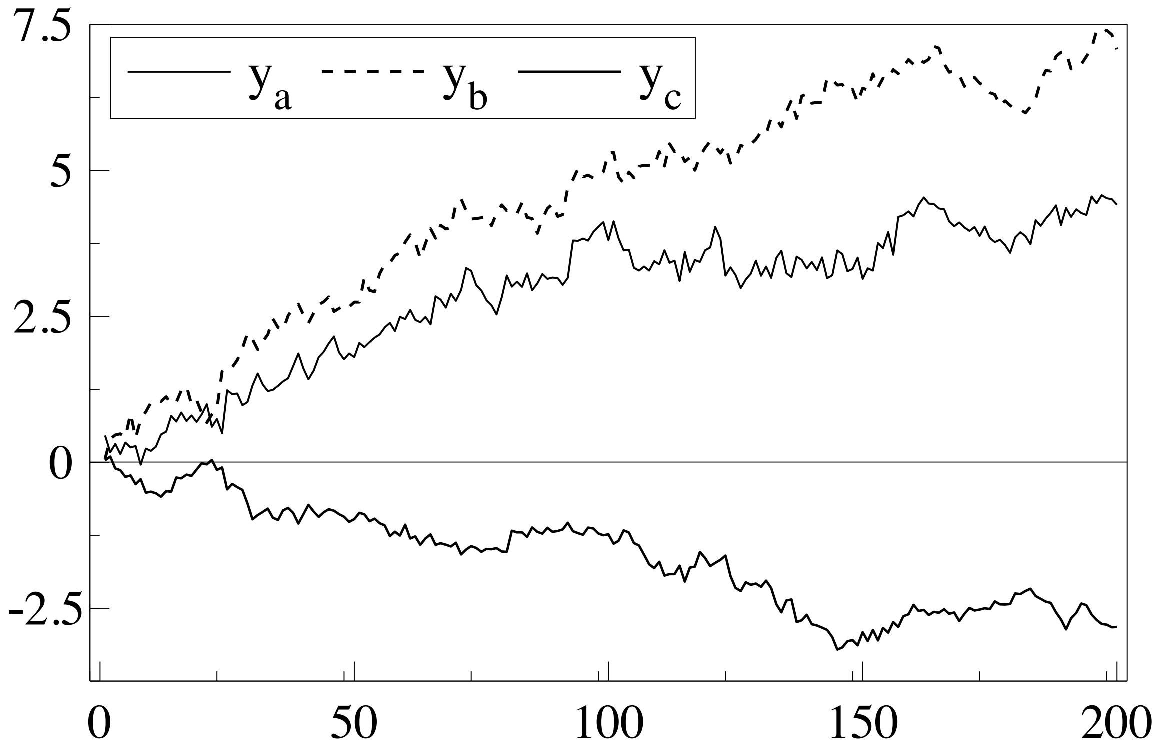
Two series below move together and end up in the similar state, but **not** cointegrated.



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

Explaining Example 2

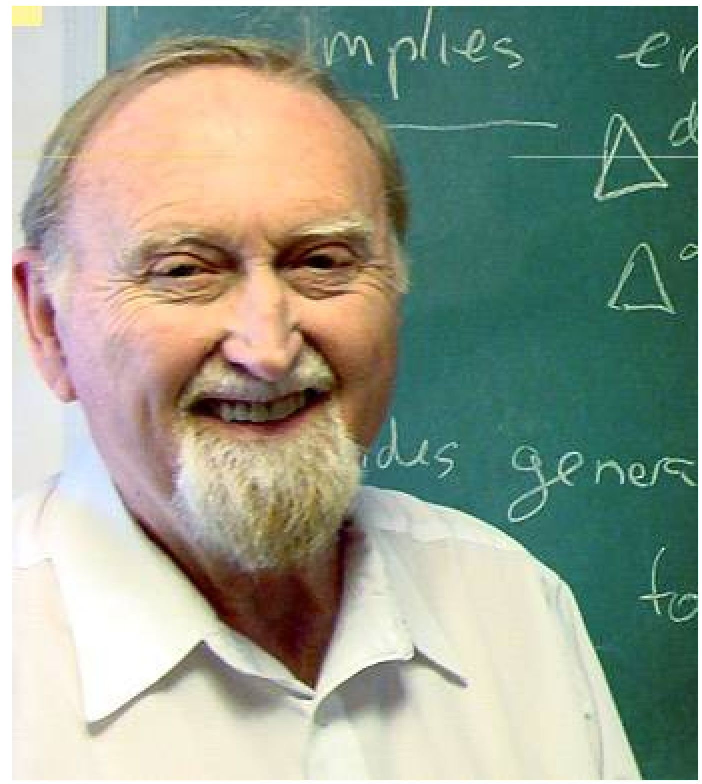
Three series are **cointegrated** and driven by the same common factor $e_t = \boldsymbol{\beta}'_{Coint}\boldsymbol{Y}_t$.



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

Phillips in an interview to *Econometric Theory*: "Clive set out to prove that such linear combinations of integrated variables would in fact remain integrated... In the process, he instead established the conditions under which cointegration could occur:

– when a dynamic system with a reduced-rank feedback matrix must generate integrated data." (Photo: Clive Granger)



Integrated Random Process, Unit Root and Stationarity Testing

Integrated process (unit root)

Take Vector Autoregression model, lag 1 , so VAR(1).

- We start with $\boldsymbol{Y_t} = \boldsymbol{\beta Y_{t-1}} + \boldsymbol{\epsilon_t}$
- $\boldsymbol{Y_{t-1}}$ depends on $\boldsymbol{Y_{t-2}}$, and so, $\boldsymbol{Y_t} = \boldsymbol{\beta}\left(\boldsymbol{\beta Y_{t-2}} + \boldsymbol{\epsilon_{t-1}}\right) + \boldsymbol{\epsilon_t}$.
- Next insertion gives

$$\boldsymbol{Y_t} = \boldsymbol{\beta}\left(\boldsymbol{\beta}\left(\boldsymbol{\beta Y_{t-3}} + \boldsymbol{\epsilon_{t-2}}\right) + \boldsymbol{\epsilon_{t-1}}\right) + \boldsymbol{\epsilon_t}$$

By induction,

$$\boldsymbol{Y_t} = \boldsymbol{\beta^n}\,\boldsymbol{Y_{t-n}} + \sum_{n=1}^{t}\left(\boldsymbol{\beta^{n-1}}\,\boldsymbol{\epsilon_{t-(n-1)}}\right)$$

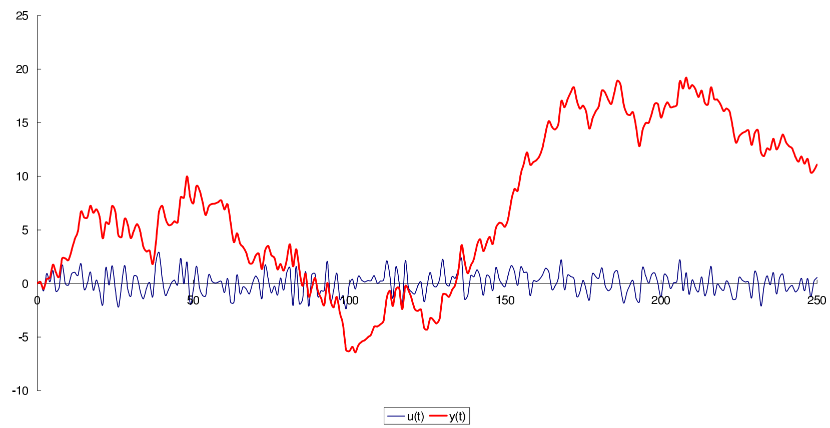
Think about the perfect **unit root** case $\boldsymbol{\beta = 1}$,

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

In continuous time summation becomes an integration

$$\epsilon_{t,\tau} \stackrel{D}{=} \int_t^{t+\tau} \boldsymbol{\sigma} dW_s.$$

We say **the process is integrated** of order one, $Y_t \sim I(1)$.



Simulated random walk (BM) process adds up increments,

$$Y_t = \sum \sigma dW_s = \sum \epsilon_s$$

Testing BROWNIAN MOTION for a unit root

```
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
Test critical values1% level             -2.574245
                    5% level             -1.942099
                   10% level             -1.615852
============================================================
```

DF relies on the higher than t critical values, making it difficult to reject a unit root hypothesis.

Dickey-Fuller Test

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

The alternative hypothesis $H_1 : \phi \neq 0$ (phi significant).

The null hypothesis $H_0 : \phi = 0$ (insignificant, true value is 0).

Custom statistical tests work like this:

1. **Test statistic** for significance of ϕ calculated as usual.

$$\frac{\phi}{\text{std error}}$$

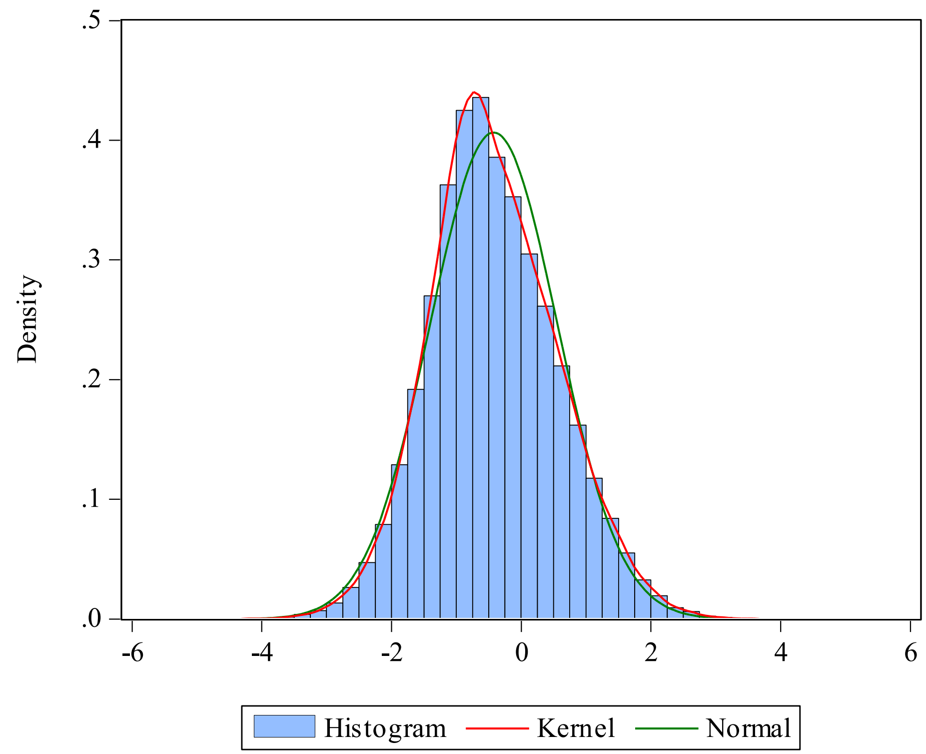
2. Standard error for ϕ is under-estimated, and so we use 'the right distribution for the wrong test statistic.'

   Conventional critical values (t distrib) give to over-rejection of H_0 when it is true.

$$H_0 : \phi = \beta - 1 = 0 \qquad Y_t = \beta^{=1} Y_{t-1} + \epsilon_t \qquad \Delta Y_t = \epsilon_t$$

   Critical value says 't-Statistic' but taken from the Dickey-Fuller as tabulated by MacKinnon (2010 update).

Bootstrapped Dickey-Fuller distribution



DF distribution is bootstrapped (form of Monte Carlo as used in econometrics) by generating *iid* residuals ϵ_t, where $Y_t \sim I(1)$

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

Augmented Dickey-Fuller

Lagged differences Δy_{t-k} improve robustness if there is noticeable serial correlation

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

Insignificant $\phi = \beta - 1 = 0$ means unit root for series y_t. DF critical values re-tabulated for each number of lags k.

Ready statistical tests (R, Matlab, etc) offer to add constant '**drift**' or time-dependent '**trend**'.

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

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 an excessive specification was used.

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

Test regression trend

lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
```

Equilibrium Correction Model (ECM)

How Cointegration Works

Estimating Cointegration - Pairwise

Pairwise Estimation: select two likely candidates to have a stationary spread: gas vs. heating oil futures, two pharmas or hotel chains where one interested in merger.

- **Step 1.** Regress one price P_t^A on another P_t^B, and test the fitted residual by ADF with lag=1. If stationary, proceed.

- **Step 2.** Confirm significance of correction term in the eqns for ΔP_t^A, ΔP_t^B.

- Step 3. Fit the stationary spread to OU SDE solution (by autoregression) to evaluate mean-reversion: μ_e, θ, σ_{eq}.

Engle-Granger Procedure - Fitted Residual

Step 1. Obtain the fitted residual and test it for stationarity

$$P_t^A = \beta_0 + \beta_1 P_t^B + \epsilon_t$$

$$\widehat{e}_t = P_t^A - \widehat{\beta}_C P_t^B - const \qquad \beta_1 = \widehat{\beta}_C$$

- **If the residual non-stationary** then no long-run relationship exists and regression is spurious. Use MacKinnon (2010) for $\tau_c, N = 2$. 'No-constant' assumption in testing eqn is attractive but highly unrealistic.

- Use Augmented Dickey Fuller test with lag=1.

Engle-Granger Procedure - Error Correction

Step 2. Plug the stationary fitted residual $\widehat{e}_{t-1}$ from Step 1 as shifted into the error correction linear regression eqn and confirm statistical significance of its coefficient.

$$\Delta P_t^A = \phi \Delta P_t^B - (1-\alpha)\,\widehat{e}_{t-1}$$

$$\Delta P_t^A = \phi \Delta P_t^B - (1-\alpha)\left(P_{t-1}^A - \beta_C P_{t-1}^B - \mu_e\right)$$

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

- Correction to equilibrium comes in very small moves as $(1-\alpha) \ll 1$.

$$\Delta P_t^A = \phi_{shortrun}(P_t^B - P_{t-1}^B) + \phi_{longrun}(P_{t-1}^A - \beta_C P_{t-1}^B)$$

Planners vs. Hedgers

EC model has Vector Autogregression representation

$$\Delta P_t^{A=1} = \beta_{11}\,(e_{t-1} - \mu_e) + \beta_{12}\,\Delta P_t^B + \beta_{13}\,\Delta P_{t-1}^A$$

$$\Delta P_t^{B=2} = \beta_{21}\,\text{CointFactor} + \beta_{22}\,\text{StaticEq} + \beta_{23}\,\text{Augment}$$

$$\boldsymbol{\Delta P_t} = \boldsymbol{\Pi}\,\boldsymbol{P_{t-1}} + \sum_i \boldsymbol{\Gamma_i}\,\boldsymbol{\Delta P_{t-i}} + \boldsymbol{\mu_0}$$

1. Notation in bold is known as matrix form VECM (constant μ_0, no trend).
2. Coint_Factor has non-unique $[1, -\beta_C]$ cointegrating weights (uncertainty, see Hedging Puzzle choices in Case Study)
3. The challenge to robustness is 'sudden' shift $\boldsymbol{\mu}_e^{Old} \rightarrow \boldsymbol{\mu}_e^{New}$.

This is not forecasting!

Implementation in Python

```
from statsmodels.tsa.stattools import coint
coint(PriceA, PriceB)

import statsmodels.tsa.stattools as ts
ts.adfuller(CointResidual)
```

ts.adfuller() gives a rudimentary output for DF Test for stationarity.

Critical values were fixed to MacKinnon(2010) after been wrong for about 2011-2017!

Requirement (TS Topic): implement Engle-Granger procedure from the first principles. Enclosed R code gives a complete example.

VECM in Python

```
import statsmodels.tsa.vector_ar.vecm as cajo
johansen_test = cajo.coint_johansen(Prices, 0, 2)
```

Python routines output will be very similar to VECM output from R routines (package *urca*), such as *cajorls()*.

- please see Cointegration Case extra slides (about p.58).

VECM seem to be included from statsmodels v0.11.0. However, to install dev version, use *git()* instead of *pip*. Refer to the source code comments to understand inputs and outputs.

`https://www.statsmodels.org/dev/generated/statsmodels.tsa.vector_ar.vecm.VECM.html`

Implementation in R: Multivariate Cointegration

Johansen Procedure tests for the number (how many) of cointegrated relationships.

It is a powerful **screening tool**: to identify dependencies among and within sections of the yield curve, segments of the market, special M&A situations.

The workhorse is *ca.jo()* function from the **R package** *urca*.

```
          test 10pct  5pct  1pct
r <= 6 |   4.67  7.52  9.24 12.97
r <= 5 |   5.87 13.75 15.67 20.20
r <= 4 |   9.78 19.77 22.00 26.81
r <= 3 |  24.98 25.56 28.14 33.24
r <= 2 |  44.91 31.66 34.40 39.79
r <= 1 |  46.88 37.45 40.30 46.82
r = 0  | 101.10 43.25 46.45 51.91
```

cajorls() presents the output as a set of familiar OLS equations with EC term, separate line for each price.

Vector Autoregression to Multivariate Cointegration [EXTRA]

Remember, we learned Vector Autoregression, an endogenous system of equations for **Returns**. We can't use VAR on **Prices** – that would be a spurious model.

Prices can be tied up with special error correction equations:

Changes in prices ΔP_t can have a special correction, that makes prices to move together over the long term.

$$\Delta P_t \;=\; \Pi\, P_{t-1} + \Gamma_1 \Delta P_{t-1} + \epsilon_t$$

$\Longrightarrow$ Π must have **a reduced rank**, otherwise *rhs* will not balance *lhs*. Differences ΔP_t will not equate to non-stationary, random prices P_t on *rhs*.

Now, to make this look alike Engle-Granger, we decompose coefficients $\mathbf{\Pi} = \boldsymbol{\alpha}\, \boldsymbol{\beta}'_C$

$$\begin{aligned}
\boldsymbol{\Delta P_t} &= \mathbf{\Pi}\, \boldsymbol{P_{t-1}} + \mathbf{\Gamma \Delta} \boldsymbol{P_{t-1}} + \boldsymbol{\epsilon_t} \\
\boldsymbol{\Delta P_t} &= \boldsymbol{\alpha}\, \underbrace{(\boldsymbol{\beta}'_C\, \boldsymbol{P_{t-1}} + \boldsymbol{\mu_e})} + \mathbf{\Gamma_1 \Delta} \boldsymbol{P_{t-1}} + \boldsymbol{\epsilon_t}
\end{aligned}$$

μ_e is called 'a restricted constant' or deterministic trend.

$$(n \times n) \;=\; (n \times r) \,\times\, (r \times n)$$

Above are dimensions for $\mathbf{\Pi} = \boldsymbol{\alpha}\, \boldsymbol{\beta}'_C$. r columns of vectorised $\boldsymbol{\beta}'$ are linearly independent – cointegrating vectors.

Eigenvalues of $\mathbf{\Pi}$ are utilised to compute both, Trace Statistic and Max Eigenvalue Statistic.

Sequential Testing for Cointegration Rank

r	lambda	1-lambda	ln(1-lambda)	Trace	CV trace	MaxEig	CV MaxEig
0	0.0167	0.9833	-0.0168	105.7518	103.8473	44.8038	40.9568
1	0.0094	0.9906	-0.0094	60.9479	76.9728	25.1283	34.8059
2	0.0046	0.9954	-0.0046	35.8197	54.0790	12.3440	28.5881
3	0.0038	0.9962	-0.0038	23.4757	35.1928	10.2469	22.2996
4	0.0031	0.9969	-0.0031	13.2287	20.2618	8.3510	15.8921
5	0.0018	0.9982	-0.0018	4.8777	9.1645	4.8777	9.1645

- Trace statistic $H_0 : r = r^*$, and $H_1 : r > r^*$. **Stop at $r^* = 1$**

$$LR_{r^*} = -T \sum_{i=r^*+1}^{n} \ln(1 - \lambda_i)$$

- Maximum eigenvalue statistic $H_0 : r = r^*$, and $H_1 : r = r^* + 1$

$$LR_{r^*} = -T \ln(1 - \lambda_{r^*+1})$$

Cointegrating Vector Estimators β'_{Coint}

	1	2	3	4	5	6	7
Canada	6.78395	-1.96320	-9.07554	7.03629	2.56142	6.25519	-2.08045
France	4.86921	4.86043	-2.08623	-7.28739	2.28808	-1.59825	-1.60875
Germany	-15.76001	-5.94947	0.12170	3.34469	-0.01972	-4.04040	4.24522
Japan	-1.22250	5.52024	-0.70856	1.03285	-0.17938	-0.08242	1.76463
UK	27.19903	-13.06796	-0.55980	-0.36245	-1.03954	-1.76308	0.23821
US	-10.25644	13.17254	7.00734	-0.56186	-5.15207	2.16214	-2.37646
Const	-117.01015	-5.47002	59.45116	-32.77753	5.05186	-8.11528	-7.19582

- $n - 1$ columns are linearly dependent on the 1st column.

- $r = 1$ columns of $\boldsymbol{\beta}$ are cointegrating vectors, take the first column and standardise it (row echelon form).

$$\begin{bmatrix} 1 & 0.7178 & -2.3231 & -0.1802 & 4.0093 & -1.5119 & -17.2481 \end{bmatrix}$$

The allocations $\hat{\beta}'_{\boldsymbol{Coint}}$ provide a mean-reverting spread.

Cointegration Estimation

1. **Engle-Granger Procedure** for a pair of time series.

   - Cause/effect (leading variable) can be established and removes uncertainty about non-unique cointegrating weights $[1, \beta_C]$.

   - Estimation of various basis, eg, tenor basis between rates..

2. **Johansen Procedure** for a set of cointegrating relationships in a multivariate setting.

   - Relies on theorem for a reduced-rank matrix with r linearly independent rows.
     First, $\boldsymbol{\beta}'_C$ estimated, second $\boldsymbol{\alpha}$ are inferred making this a calibration.

Summary

Please take away the following ideas...

- Financial time series are non-stationary and so, naïve linear correlation and regression are spurious models with $R^2 \approx 1$.
- Large samples and cross-validation only amplify the issue.
- **Cointegration**: linear combination of time series produces a stationary spread. Long-term relationship relies on the common stochastic process.
- Engle-Granger procedure is a stepping stone which assumes that process (naïve coint. residual). Multivariate cointegration properly checks the reduced rank.
- To construct a trade: apply the special statistical arb techniques, such as fitting to OU process, and backtest.

Case Study B: Pound Sterling Spot Rates

- Traditionally, cointegration is tested in the very long run
- We had Case Study A testing for an equilibrium between US T-Bills and Treasuries over the horizon of 1960-2010.

  HOWEVER

- As quants we have to look for co-movement in the current, frequent market data.

We will use this opportunity to get introduced to R.

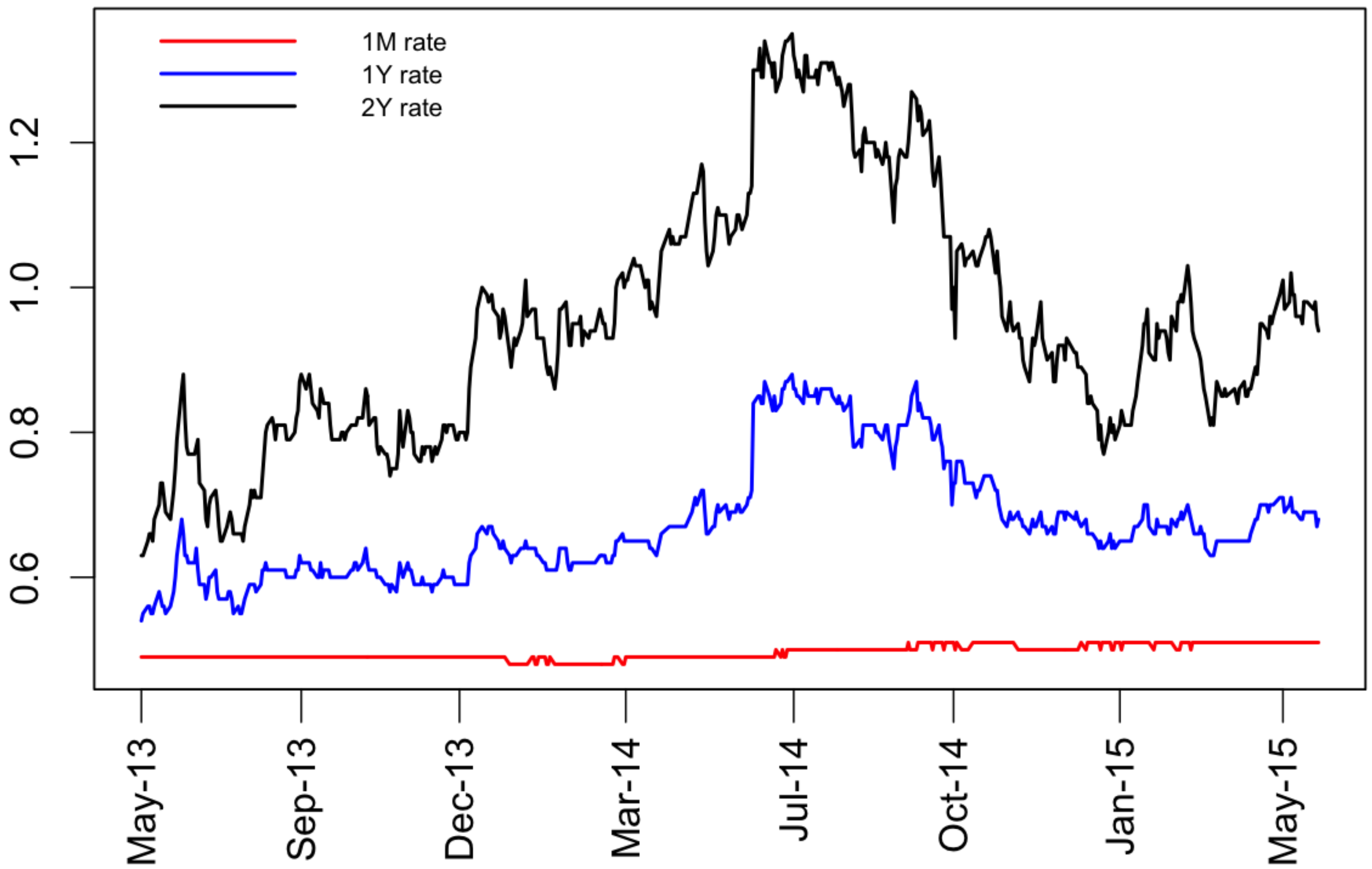
Spot Curve

The Bank of England provides the daily yield curve data. It makes sense to consider smaller windows of the long timeframe:

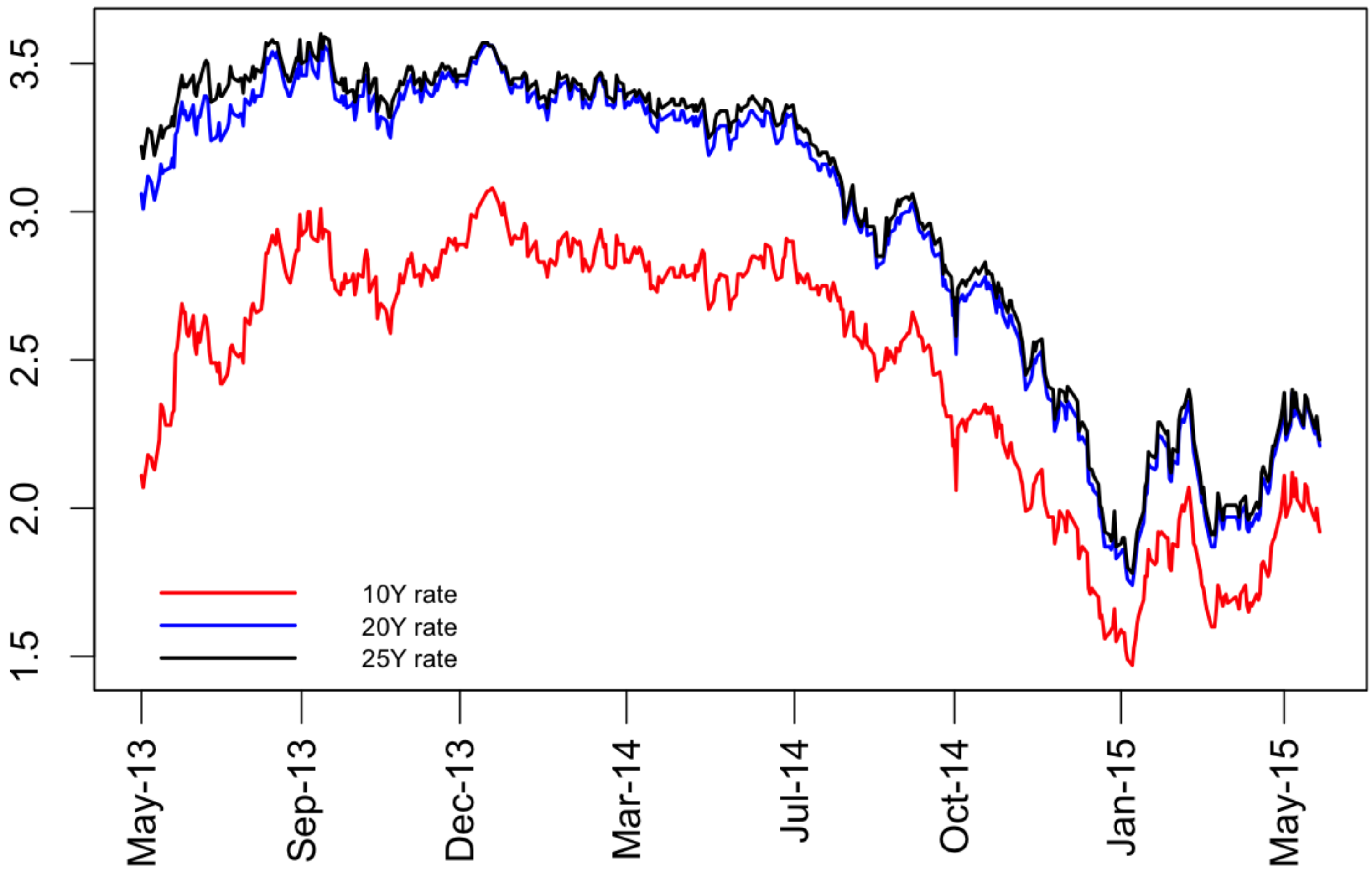
- two-year window May 2013 – May 2015 (charts below) vs.
- all data from from Jan 2005 to May 2015.

We have to learn the equilibrium-correction mechanics (**ECM**) but it's worthwhile to have a peek from the multivariate test for cointegration.

Spot Rates at Short End



Spot Rates at Long End



Problems with curve data

1. r_t at the short end (0.8Y, 1Y, 2Y) and y_t at the long end (7Y, 10Y, 20Y) **do not** come as cointegrated in samples of two-three year period.

   There is simply not enough horizon for a cointegrated relationship to transpire.

2. Let's play a game: **Which long-end rates are co-integrated?** Choose pairs among 10Y, 20Y, 25Y.

   Can't decouple that easily.

   Similar pattern comes up for the short end, if all data included in the testing.

   Parallel to that, short rates have independent co-movement.

Engle-Granger preview

Let's choose a model with **10Y and 25Y tenors** because of their importance as benchmarks.

- We set up a naive cointegrating equation

$$r_{10Y} = \beta\, r_{25Y} + e_t \qquad \Rightarrow \qquad \hat{e}_t = r_{10Y} - \beta\, r_{25Y}$$

- We test this estimated residual $\hat{e}_t$ for stationarity by CADF.

  If the residual is stationary, it means that r_{10Y} and r_{25Y} have a unit root **in common**, removable by differencing

Long-run relationship r_{10Y} on r_{25Y}

```
lm(formula = curve2.this$X10 ~ curve2.this$X25)

                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.15878    0.03132    5.07  5.6e-07 ***
curve2.this$X25  0.76980    0.01023   75.28  < 2e-16 ***

Residual standard error: 0.1231 on 504 degrees of freedom
Multiple R-squared:  0.9183,Adjusted R-squared:  0.9182

Residuals:
     Min       1Q   Median       3Q      Max
-0.53675 -0.03449  0.01926  0.07920  0.18461
```

As usual, regressing one non-stationary series on another gives *extremely* significant coefficients. Large N_{obs} makes $R^2 \to 1$.

Long-run relationship if cointegrated

$$\hat{r}_{10Y} = 0.159 + 0.77\, r_{25Y} + \hat{e}_t$$

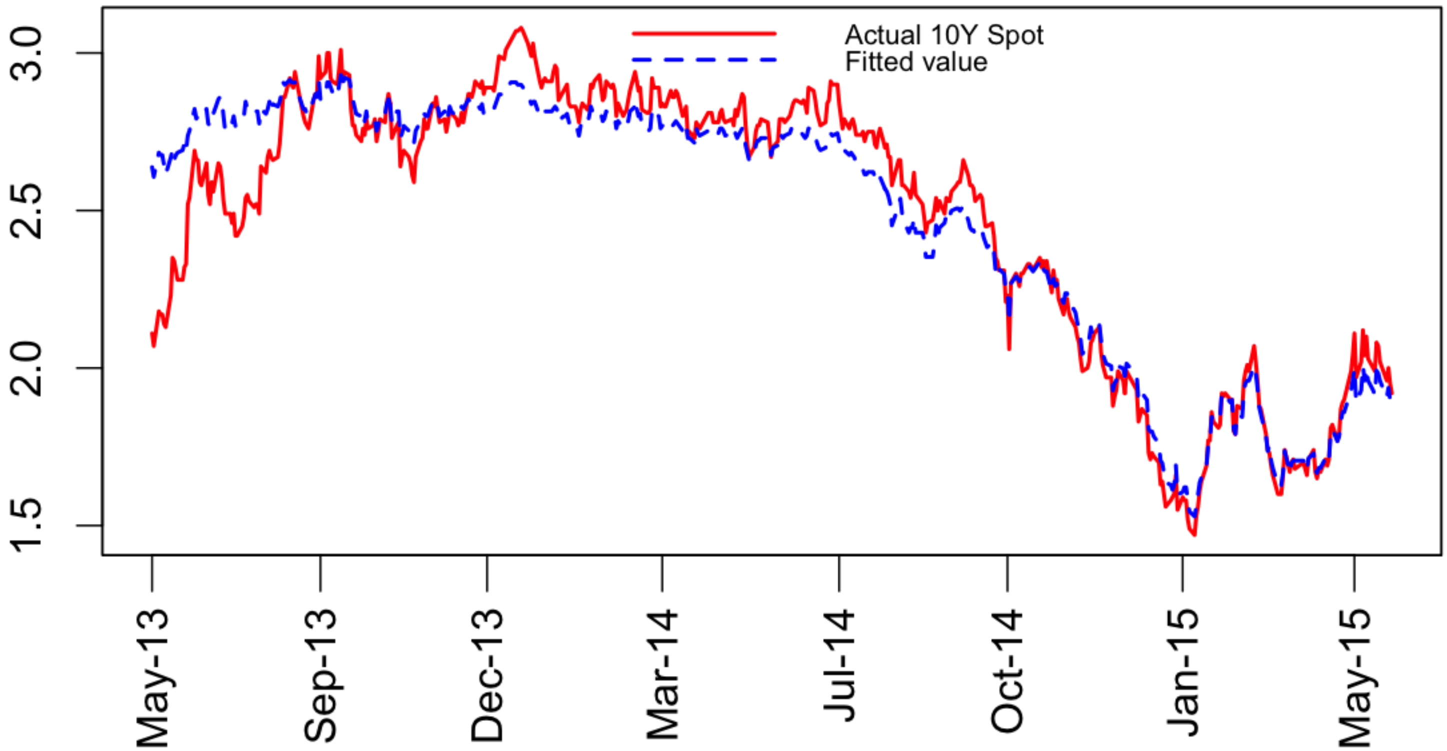
This model is valid only if it produces stationary $\hat{e}_t$, so there is co-integration between r_{10Y} and r_{25Y}

It only works in the context of the equilibrium correction over the long-run, producing stationary and mean-reverting residual:

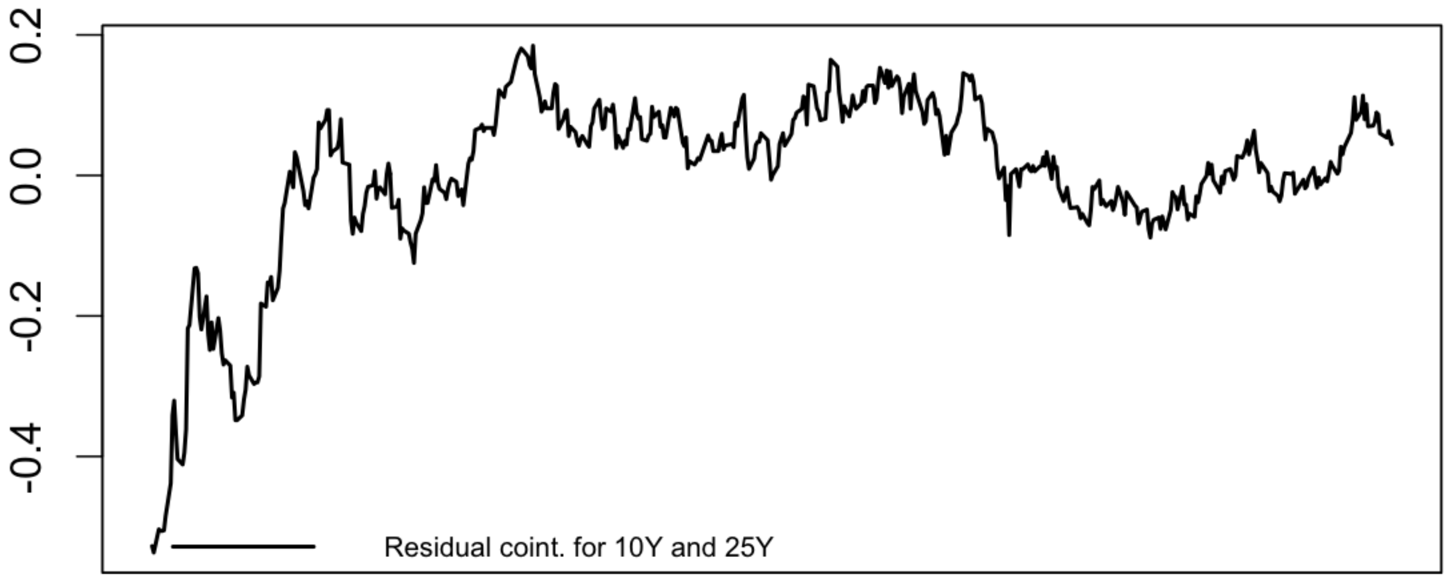
$$\hat{e}_t = r_{10Y} - (0.159 + 0.77\, r_{25Y}).$$

Linear regression FIT for r_{10Y}

Our linear model aims to obtain $\widehat{e}_t$ so we would be differencing actual r_{10Y} with fitted $\widehat{r}_{10Y}$.



Stationary cointegrating residual $\widehat{e}_t$



We will confirm the stationarity of residual, and proceed with forming error-correction equations.

Stationarity test for $\widehat{e}_t$

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

lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.038559   0.008548  -4.511 8.06e-06 ***
z.diff.lag -0.042376   0.043711  -0.969    0.333

[DF test-statistic is -4.5107, for which critical values]
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

Residual standard error: 0.02318 on 502 degrees of freedom
Multiple R-squared:  0.04071,Adjusted R-squared:  0.03689
```

Dickey-Fuller Test reminder

Null Hypothesis: time series has a unit root

We assume a linear trend, so ΔY_t will have a constant

$$\Delta Y_t = \text{Const} + \phi Y_{t-1} + \phi_1 \Delta Y_{t-1}$$

If ϕ is insignificant the time series has a unit root.

We can augment the test equation with more lags in $\phi_k \Delta Y_{t-k}$ or time-dependence $\phi_t t$ where ϕ_t is the drift.

That is likely to increase significance. However, beware you might be innocently introducing time dependence (growth/decrease) where there is none.

Long-run relationship (cointegrated)

ECM estimation [R code provided for your exploration] gives

- **the calbirated parameter** of interest is the speed of correction towards the equilibrium $(1 - \alpha)$

  It is inevitably small but **must be** significant for cointegration to exist.

- We have quite good correlation between differences Δr_{10Y} and Δr_{25Y}. There is co-movement on the short timescale.

  For the lower frequency samples, you might find that Δr_t (for the short rate) and Δy_t (for some long-term rate) are cointegrated but correlated weakly negatively.

Equilibrium Correction Model: two-way, two residuals

$$\Delta r_{10Y} = 1.086 \Delta r_{25Y} - 0.02716 \, e^{10Y}_{t-1} + \epsilon_t$$

```
             Estimate Std. Error t value Pr(>|t|)
tenorX.diff   1.085090   0.022986  47.206  < 2e-16 ***
ec_term.lag -0.027164   0.007202  -3.772 0.000181 ***

Residual standard error: 0.01981   Multiple R-squared: 0.8202
```

$$\Delta r_{25Y} = 0.752 \Delta r_{10Y} - 0.01206 e^{25Y}_{t-1} + \epsilon_t$$

```
             Estimate Std. Error t value Pr(>|t|)
tenorY.diff   0.751627   0.015910  47.243   <2e-16 ***
ec_term1.lag -0.012059   0.004851  -2.486   0.0132 *

Residual standard error: 0.01649   Multiple R-squared: 0.8175
```

Summary

Please take away the following ideas...

- this case of evolution of spot rates at different tenors is a case of a basis relationship,
- so imposing a long-run relationship and using Engle-Granger procedure has more statistical power,
- r_{10Y} and r_{25Y} series each have a unit root,
- it turns out that by differencing these time series, the unit root got cancelled and a stationary residual obtained,
- that means the time series are co-integrated.

Case Extra Slides

- Restricted VECM from Johansen Procedure
- Engle-Granger Procedure for r_{25Y} on r_{10Y} (other way)
- Linear regression on differences Δr_{25Y}, Δr_{10Y}
- Hedging ratio puzzle

Restricted VECM for Δr_{10Y} and Δr_{25Y}

```
cajorls(johansen.test)

lm(formula = substitute(form1), data = data.mat)

          X10.d     X25.d
ect1      -0.05842  -0.02647
X10.dl1   -0.13888  -0.09543
X25.dl1    0.07943   0.06495

[Cointegrating Equation (EC term)]
               ect1
X10.l2     1.0000000
X25.l2    -0.7870489
constant  -0.1435463
```

Long-run relationship r_{25Y} on r_{10Y} (other way)

The linear model $r_{25Y} = \beta\, r_{10Y} + \epsilon_t$ only aims to obtain $\widehat{e}_t$.

```
lm(formula = curve2.this$X25 ~ curve2.this$X10)

                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.05686    0.03989   1.425    0.155
curve2.this$X10  1.19295    0.01585  75.285   <2e-16 ***

Residual standard error: 0.1532 on 504 degrees of freedom
Multiple R-squared:  0.9183,Adjusted R-squared:  0.9182
F-statistic:  5668 on 1 and 504 DF,  p-value: < 2.2e-16

Residuals:
     Min        1Q    Median        3Q       Max
-0.18591 -0.08516 -0.03819  0.02177  0.65373
```

Stationarity test for $\hat{e}_t$ (other way)

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

lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.033920   0.007759  -4.372  1.5e-05 ***
z.diff.lag -0.038024   0.043779  -0.869    0.386

[DF test-statistic is -4.3718, for which critical values]
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

Residual standard error: 0.02619 on 502 degrees of freedom
Multiple R-squared:  0.03792,Adjusted R-squared:  0.03409
```

Comparison to linear regression

OLS on simple differences Δr_{25Y} and Δr_{10Y} gives min variance relationship – cointegration plays a completely separate role.

```
lm(formula = diff(curve2.this$X25) ~ diff(curve2.this$X10) + 0)

                      Estimate Std. Error t value Pr(>|t|)
diff(curve2.this$X10)  0.74570    0.01581   47.16   <2e-16 ***

Residual standard error: 0.01657 on 504 degrees of freedom
Multiple R-squared:  0.8153, Adjusted R-squared:  0.8149

Residuals:
      Min        1Q    Median        3Q       Max
-0.081683 -0.010172 -0.002371  0.007629  0.050172

cor(diff(curve2.this$X25), diff(curve2.this$X10))
[1] 0.903719
```

Hedging ratio puzzle

What would you use as a hedging ratio for assets r_{10Y} and r_{10Y} in presence of cointegration between them?

Multiple Choice:

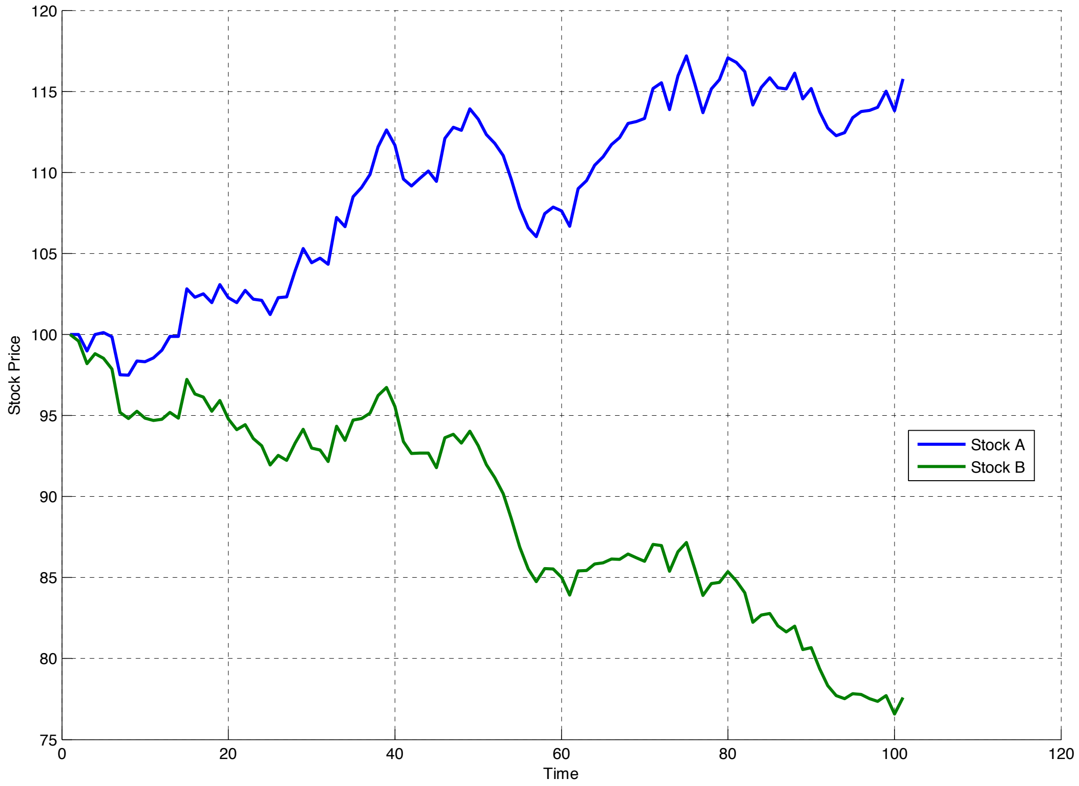
- 0.7698 from linear regression of r_{10Y} on r_{25Y}
- 0.7457 from linear regression on differences Δr_{25Y} on Δr_{10Y}
- 0.7516 from correction equation of Δr_{25Y}
- 0.7870 from a coint regression VECM output.

Time Series Cointegration

Extra Slides

Correlated Series

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



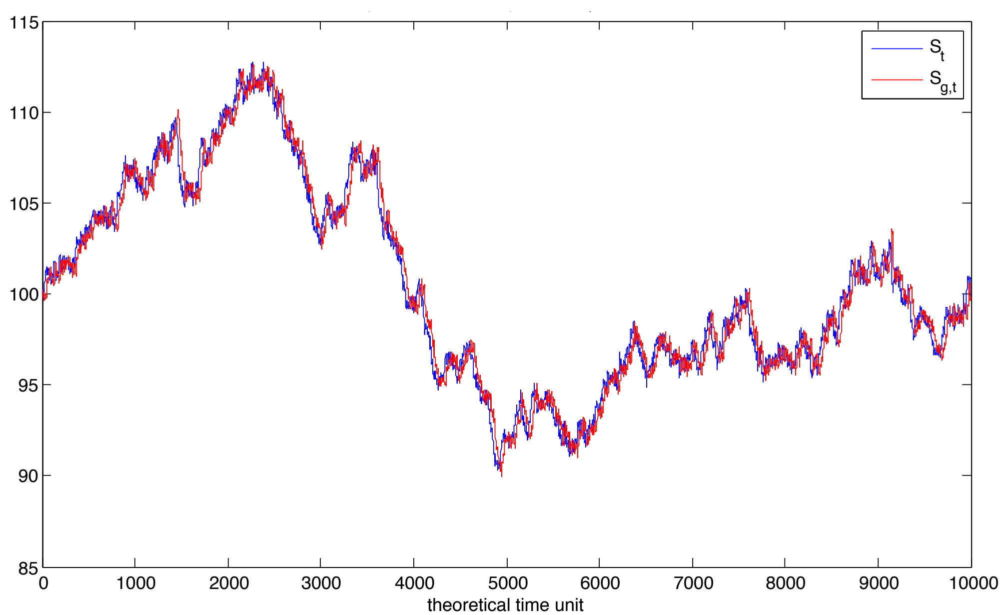
From *Correlation Sensitivity* CQF Lecture.

Cointelation

These series have been generated from the following processes:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t$$

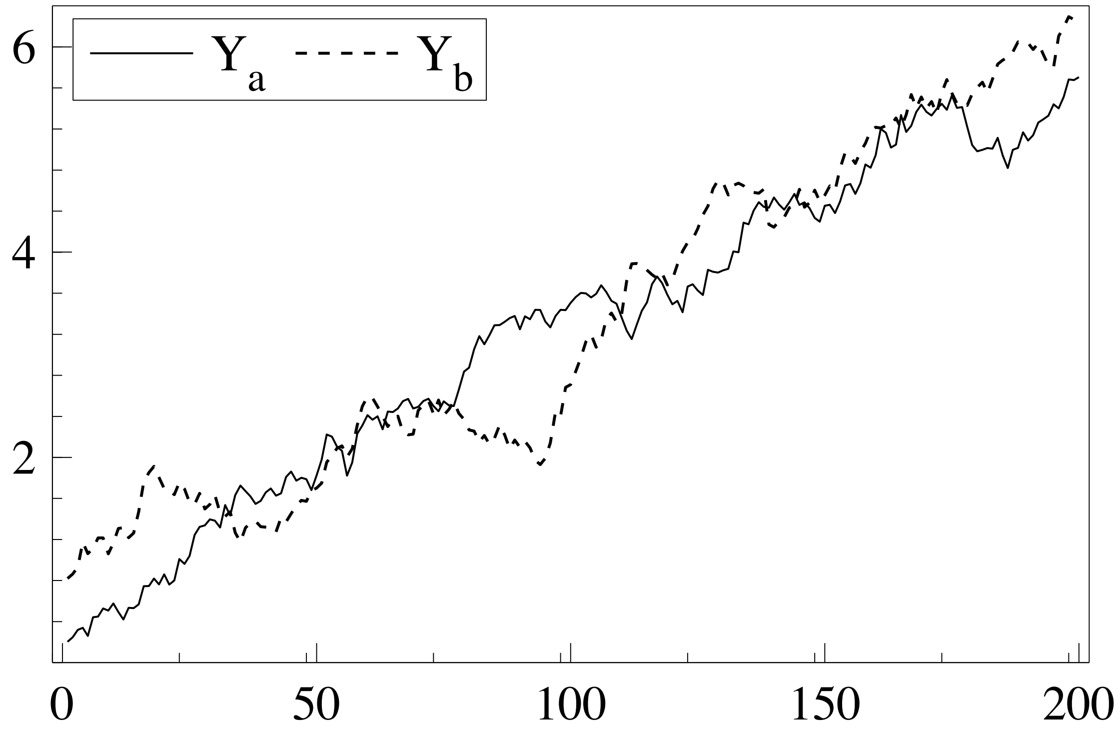
$$dS_{l,t} = -\theta(S_{l,t} - S_t)dt + \sigma S_{l,t}\left(\rho dW_t + \sqrt{1-\rho^2}dW_t^{\perp}\right)$$



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

No linear equilibrium

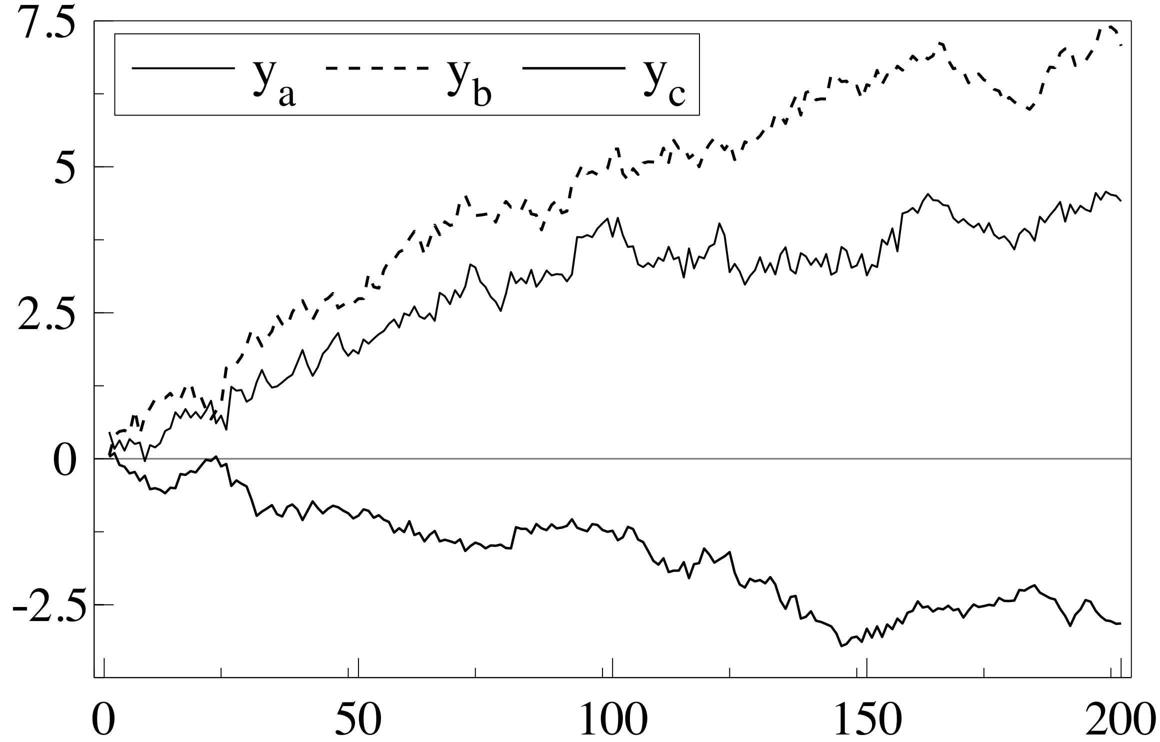
These series are **not** cointegrated. In fact, their spread contains a unit root. Multicollinearity is potentially irresolvable.



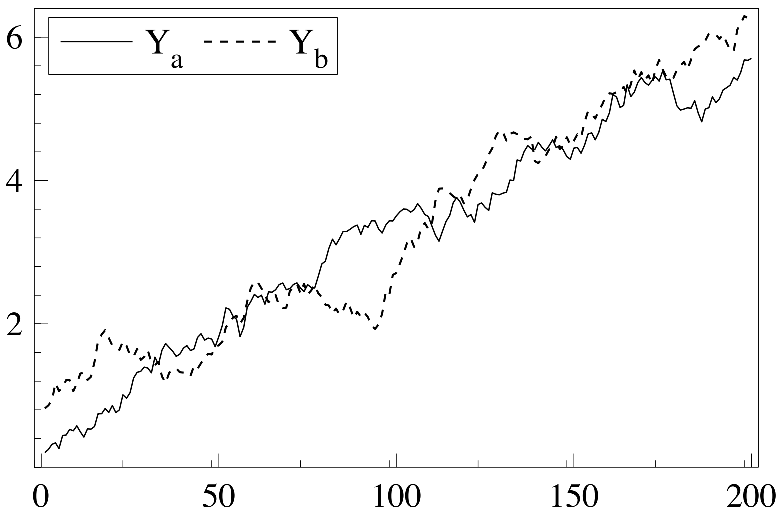
From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

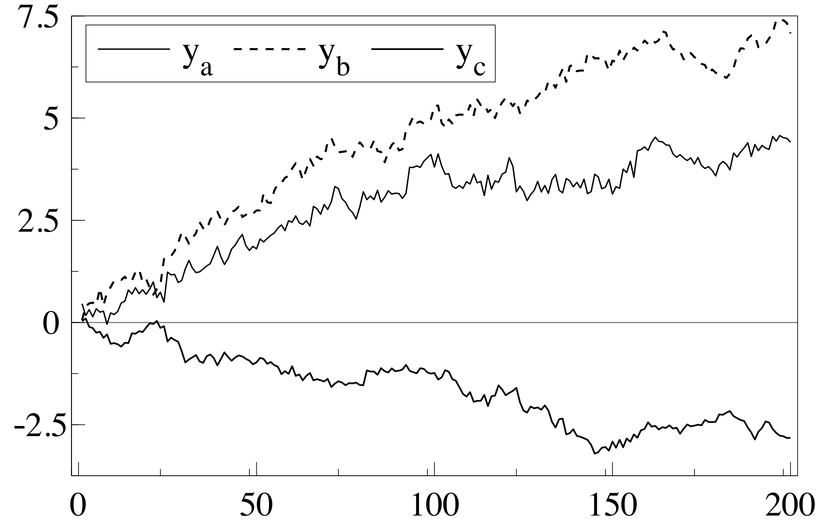
Cointegrated Series

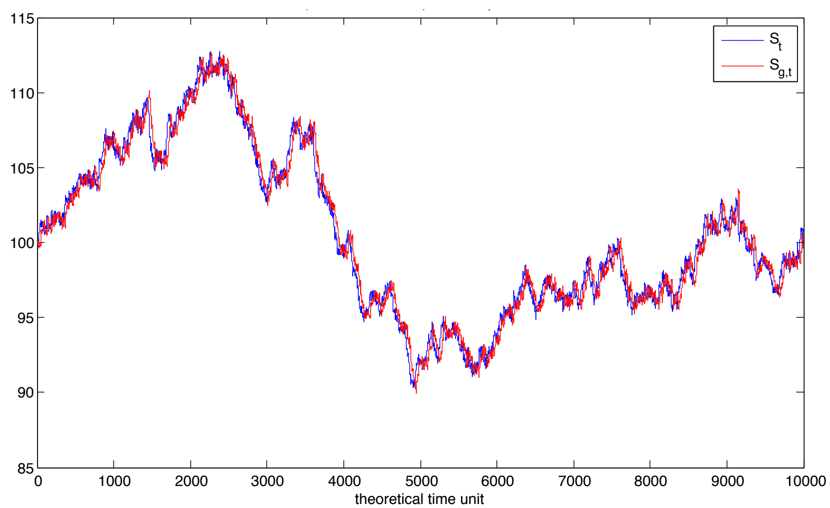
These series are **cointegrated**. Their linear combination produces a mean-reverting spread (common factor) e_t.

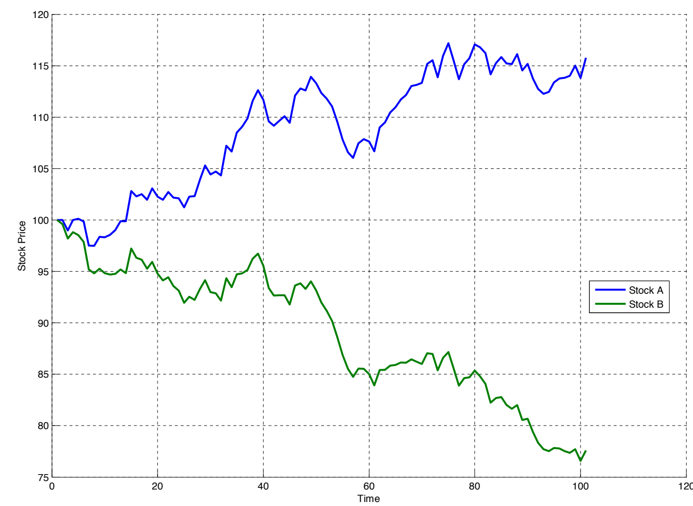


From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*


Y_a
Y_b
0
50
100
150
200
2
4
6


y_a
y_b
y_c
0
50
100
150
200
7.5
5
2.5
0
-2.5


S_t
$S_{g,t}$
theoretical time unit
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
85
90
95
100
105
110
115


Stock A
Stock B
Stock Price
Time
0
20
40
60
80
100
120
75
80
85
90
95
100
105
110
115
120

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 $\boldsymbol{\beta}'_{Coint}\boldsymbol{Y_t}$ 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**:

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

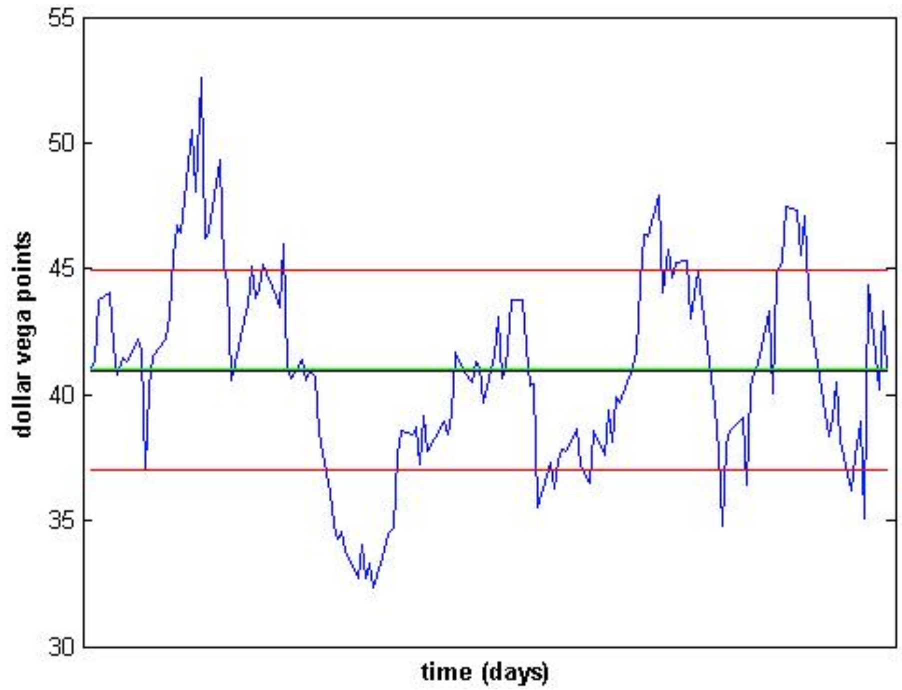
then we can explore mispricing that occurs when asset prices $y_{i,t}$ produce **a disequilibrium** $e_t \neq \mu_e$.

- 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 $\beta'_C \mathbf{Y}_t = e_t$ produces a stationary and mean-reverting spread:

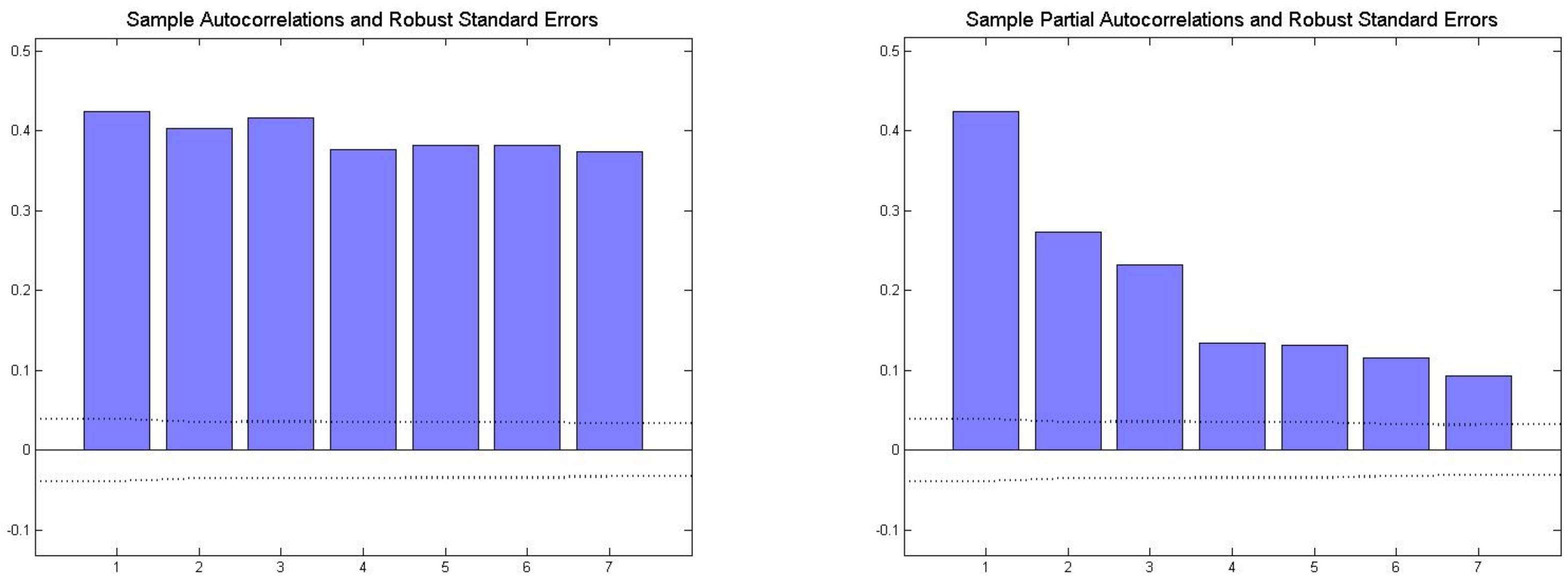
- Reversion speed $\theta \approx 44$ and bounds are calculated as $\sigma_{OU}/\sqrt{2\theta}$



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

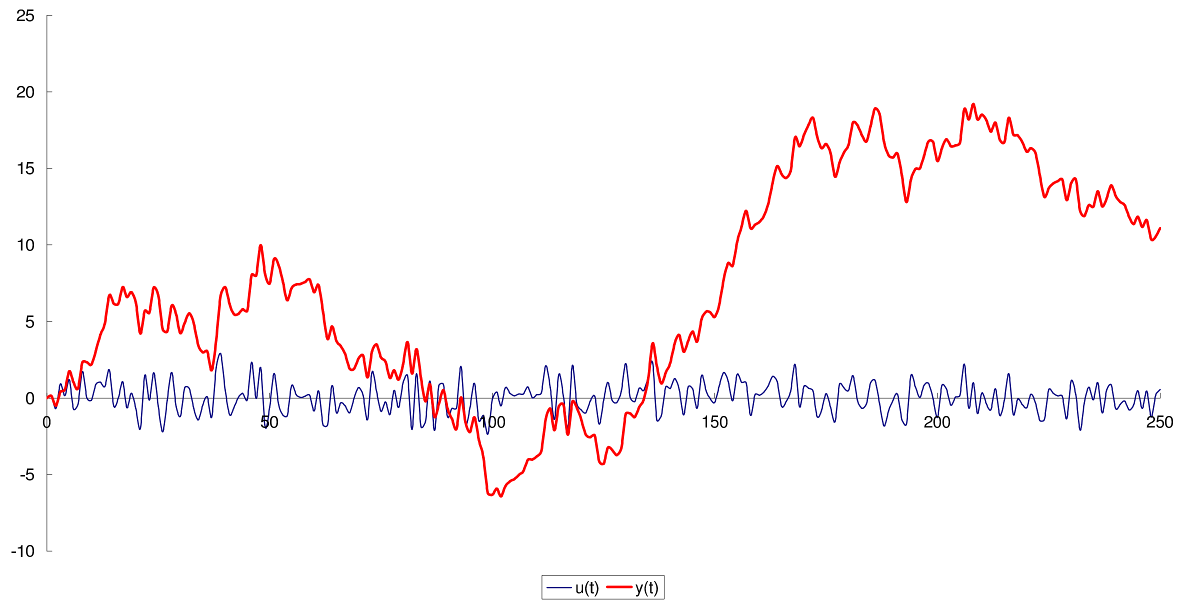
ACF and PACF for a high frequency spread e_t

Here, serial autocorreation $AR()$ noticeable for e_t.



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

ASIDE. Brownian Motion as Integrated Process

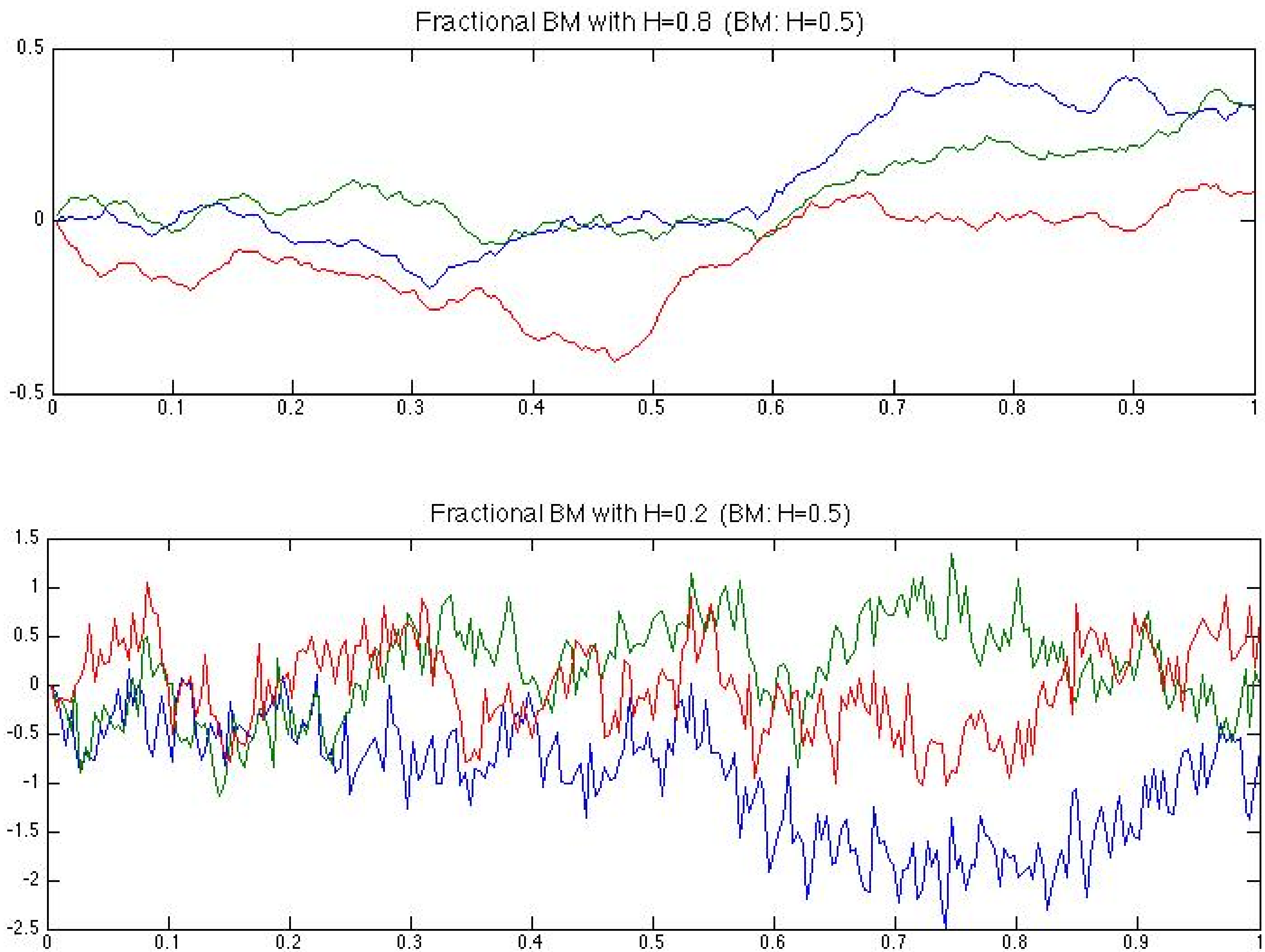


Regression residual $\epsilon_{t,\tau}$ is an increment of Brownian Motion and so, $Y_t = \sum \epsilon_s$ is the Brownian Motion under the unit root condition that, $\boldsymbol{\beta} = \mathbf{1}$.

$$\epsilon_{t,\tau} \stackrel{D}{=} \int_t^{t+\tau} \boldsymbol{\sigma} dW_s$$

It is integrated, stationarity test will confirm the unit root.

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



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

Long memory

Fractional BM decays according to **the power law** τ^{2d-1} which is slower and delivers **long memory** for $H \gg 0.5$.

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

Hurst exponent $H > 0.5$ opens modelling of integrated series with long memory (think evolution of r_t and 1M LIBOR).

$H = 0.5$ recovers the Brownian Motion $\sim I(1)$.

Stationary-like series (simulated) have low values of Hurst exponent $H < 0.2$.

Stationary process $\beta < 1$ has exponentially decaying autocorrelations Corr $\approx e^{\tau \ln \beta}$. OU process has $e^{-\theta\tau}$.

END OF ASIDE

Dynamic Equilibrium in Econometrics

(for low-frequency, macro data)

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.

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

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

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.

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

Correlation is estimated among the differences...

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

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

It turns out that the common factor is

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

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

s.t. $\widehat{e}_t$ 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

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

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

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

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

Modelling problems

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

- 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,

$$\Delta y_t \quad \text{on} \quad \Delta x_t$$

$$\Delta x_t \quad \text{on} \quad \Delta y_t$$

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 $\widehat{e}_t = y_t - \widehat{b}x_t - \widehat{a}$ and test for unit root.

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

- **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)

$$\Delta y_t = \beta_1 \Delta x_t - (1-\alpha)\widehat{e}_{t-1}.$$

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

Equilibrium Correction Mode (review)

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.

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

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

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

Stationarity Tests: reference

There are several ways of testing for unit root (the process being non-stationary).

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

- Non-parametric *Phillips-Perron* test transforms t-statistic to further account for autocorrelation and heteroscedasticity.

- *Cointegration Regression Durbin-Watson* replaces Engle-Granger procedure (presented next). It contains of one simplistic regression (one price on another) and Durbin-Watson test on the residuals.

Statistical Arbitrage

A Few Fundamentals

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.

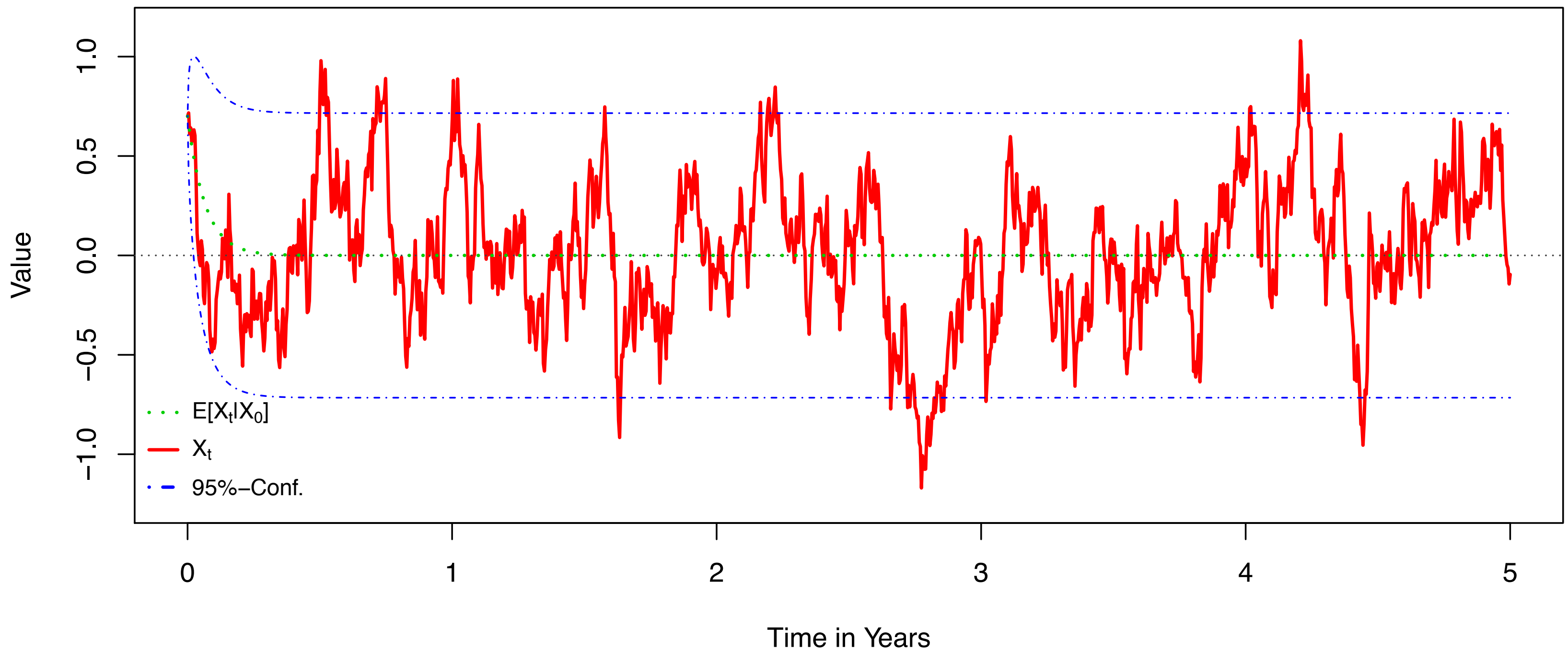
- Quality of mean-reversion is associated with the higher critical values equilibrium-correcting term (empirically).

- 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.