

# Cointegration in R: Spot Rates Market Data

- Traditionally, cointegration is tested in the very long run
- Case Study A tests for an equilibrium between T-Bill rates and Treasury yields 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.

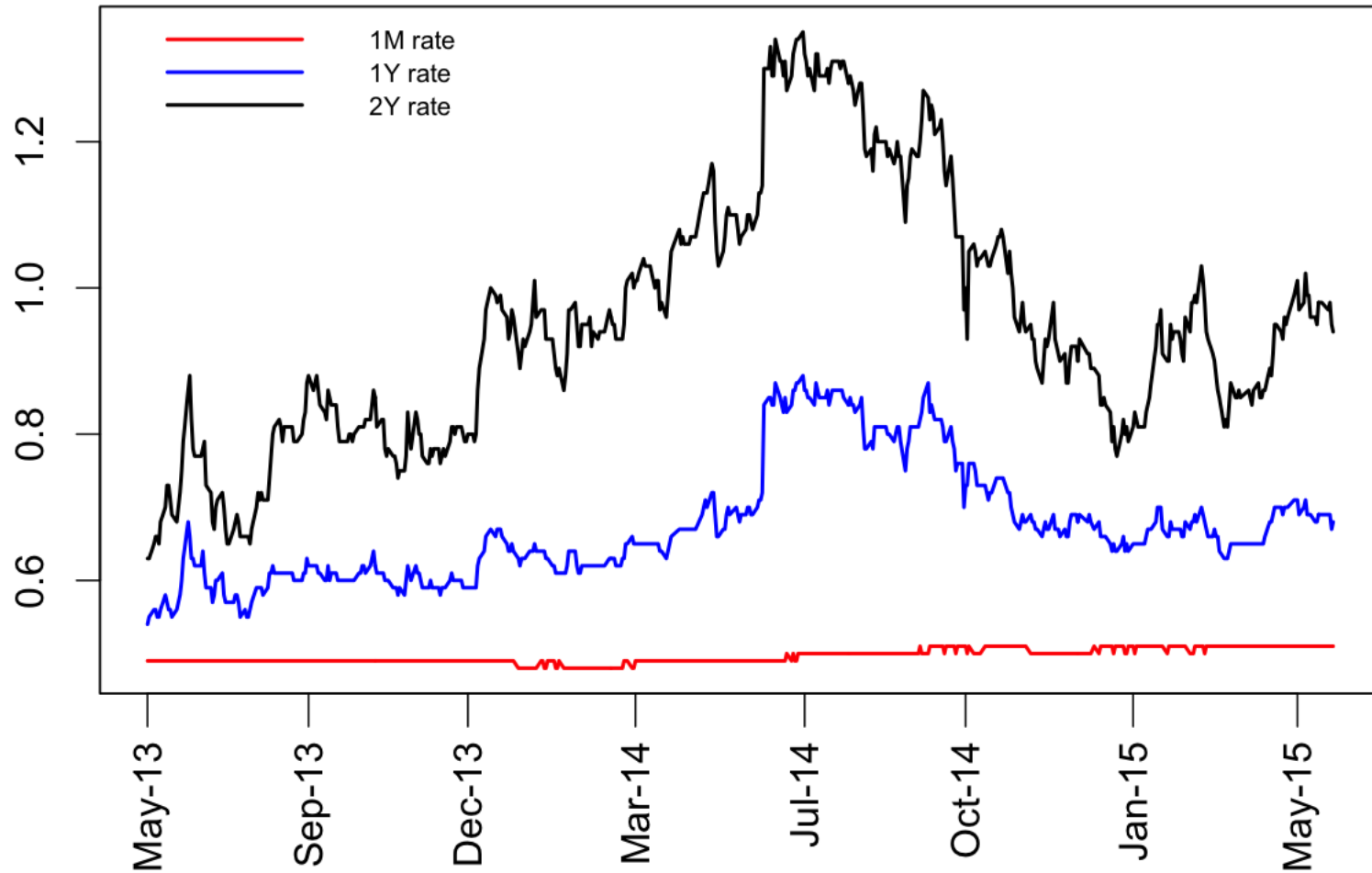
## Spot Curve

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.

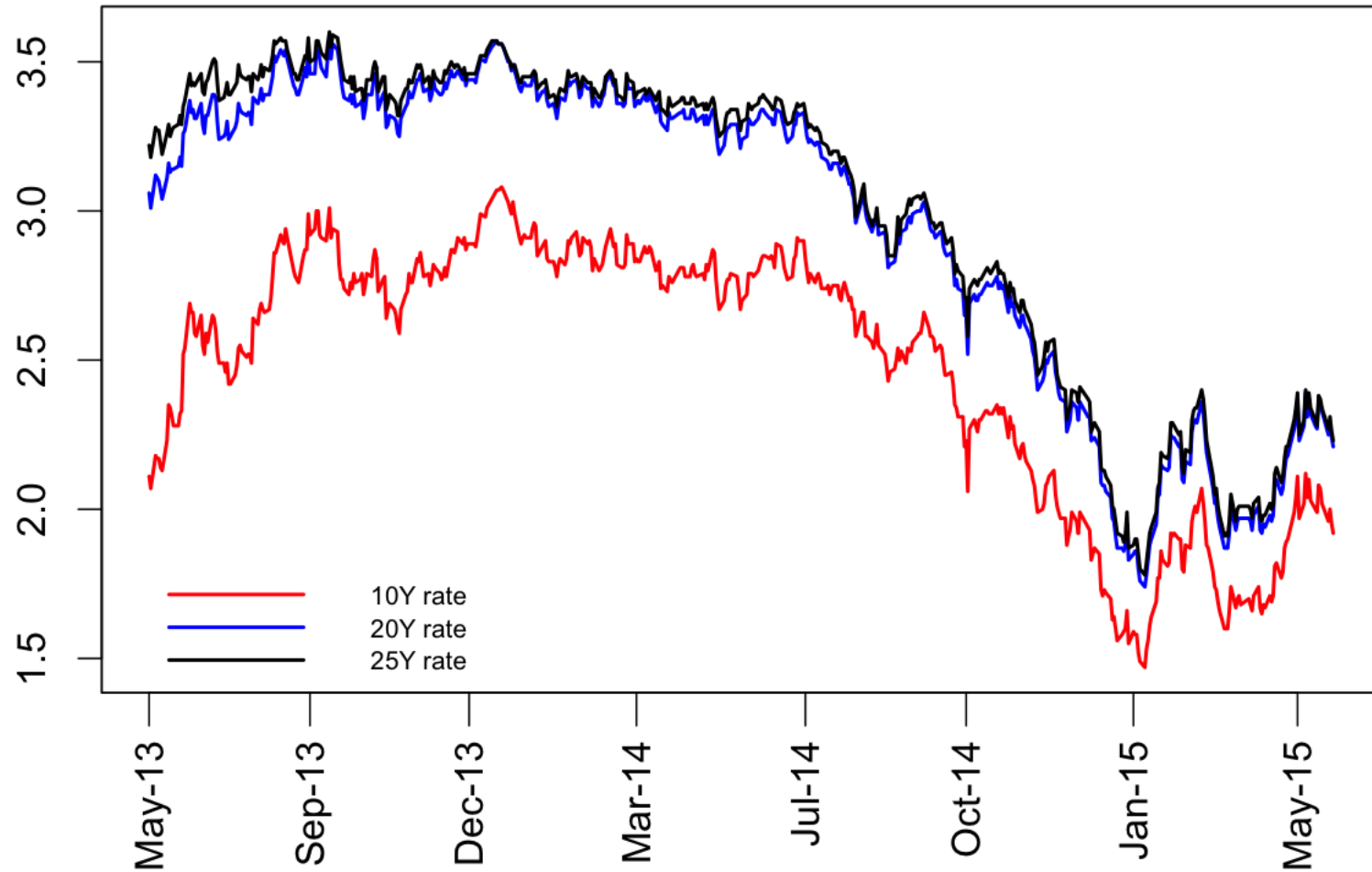
We have to learn the equilibrium-correction mechanics (**ECM**) but it's worthwhile to have a peek from the multivariate test for cointegration.

## Spot Rates at Short End



Certificate in Quantitative Finance

## Spot Rates at Long End



Certificate in Quantitative Finance

## Problems with curve data

1. In that kind of data  $r_t$  at “short end” and  $y_t$  at “long end” **do not** come as cointegrated.

There is simply not enough horizon for a relationship to transpire. Daily movements are noisy.

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

The answer is that cannot decouple movements at the long end that easy.

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

## Engle-Granger preview

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

- First, we test  $r_{10Y}$  and  $r_{25Y}$  for a unit root each.
- Then, we set up a naive cointegrating equation

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

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

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

## Dickey-Fuller Test reminder

Null Hypothesis: time series has a unit root

We assume a linear trend, so  $\Delta Y_t$  will have a constant

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

**If  $\phi$  is not significant 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  $\phi_t$  is the drift.

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

## Unit root in $r_{10Y}$ spot

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

Test regression drift [This means Delta Y=Constant]

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

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.011484	0.012212	0.940	0.347
z.lag.1	-0.004753	0.004848	-0.9805	0.327
z.diff.lag	-0.053306	0.044634	-1.194	0.233

[p-Value of DF test-statistic for -0.9805 is 0.4922]



## Unit root in $r_{25Y}$ spot

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

Test regression drift [This means Delta Y=Constant]

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

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0007984	0.0098612	-0.081	0.936
z.lag.1	-0.0003822	0.0032169	-0.1188	0.905
z.diff.lag	-0.0349222	0.0447136	-0.781	0.435

[p-Value of DF test-statistic for -0.1188 is 0.6514]

## 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  $R^2 \rightarrow 1$ .

Certificate in Quantitative Finance

## Long-run relationship if cointegrated

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

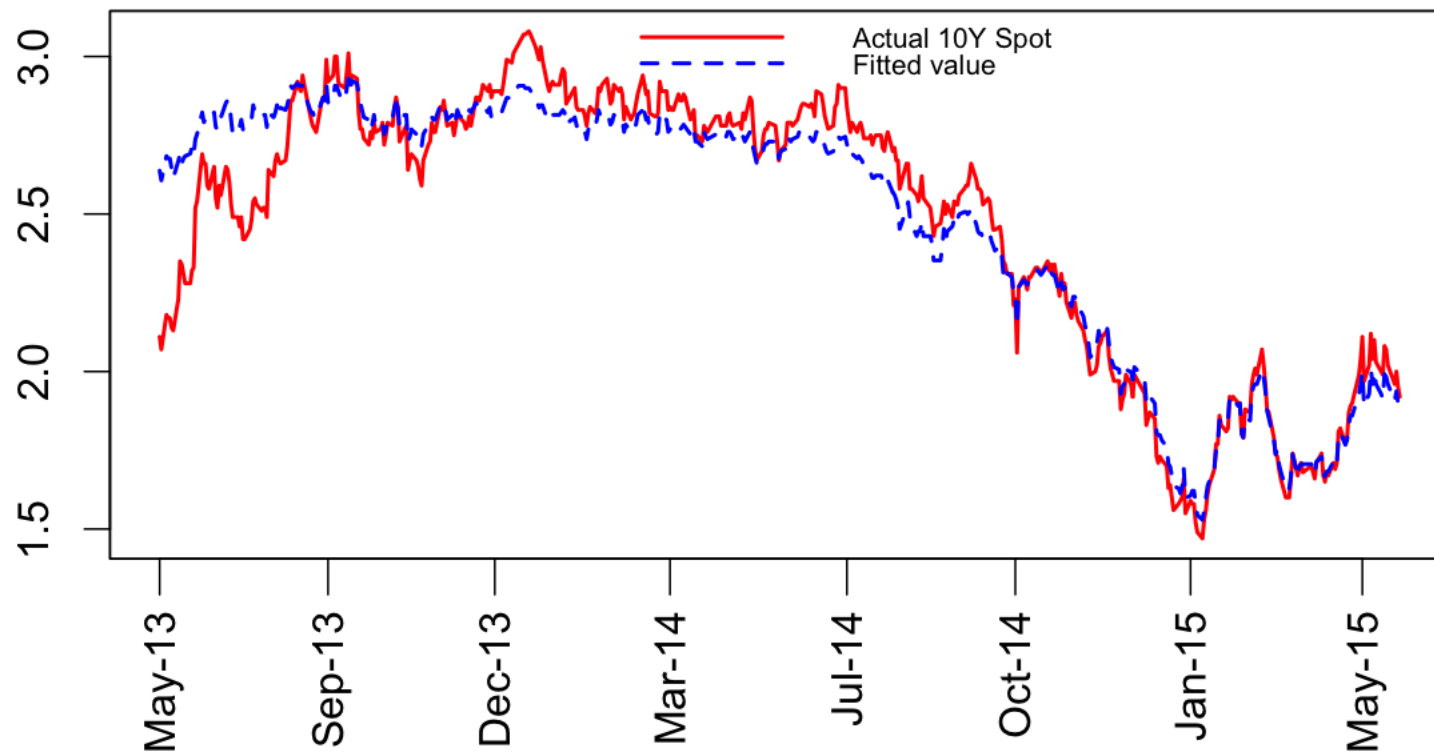
This model is valid only if it produces stationary  $\hat{e}_t$ , so there is co-integration between  $r_{10Y}$  and  $r_{25Y}$

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  $\hat{e}_t$  so we would be differencing actual  $r_{10Y}$  with fitted  $\hat{r}_{10Y}$ .



## Stationarity test for $\hat{e}_t$

```
#####  
# 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]
```

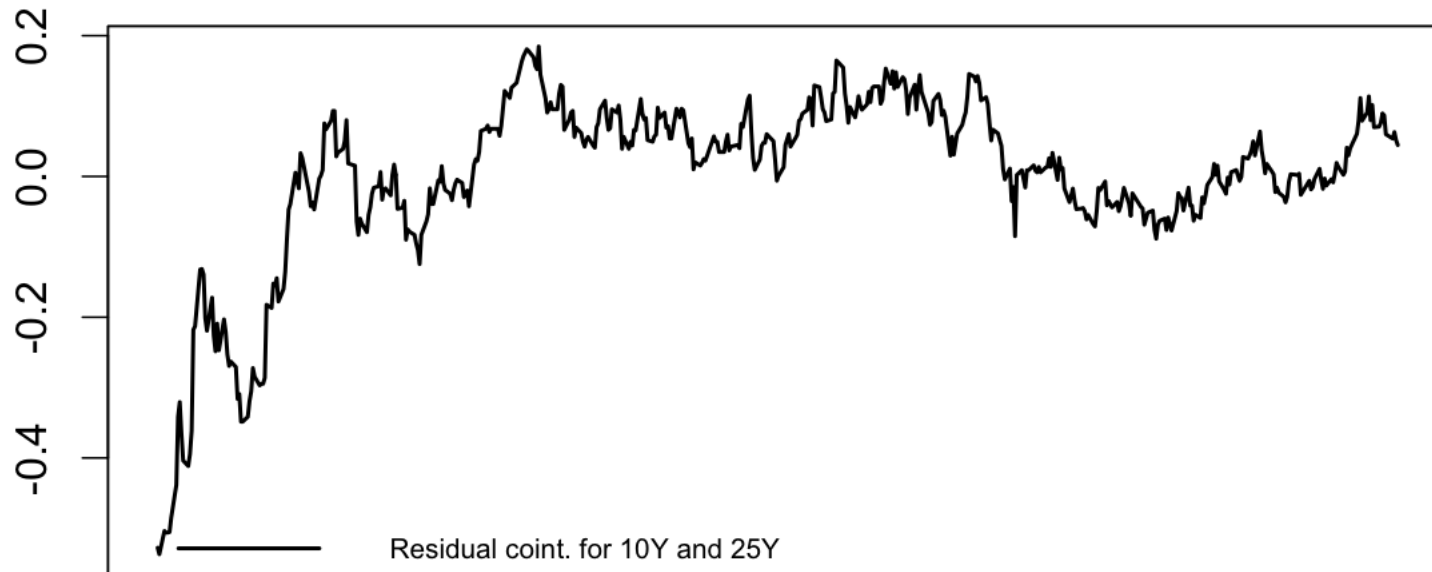
	1pct	5pct	10pct
tau1	-2.58	-1.95	-1.62

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

Certificate in Quantitative Finance

---

## Stationary cointegrating residual $\hat{e}_t$



We have confirmed the stationarity of residual, and therefore, cointegration according to the Engle-Granger procedure.

## Long-run relationship (cointegrated)

ECM estimation [R code provided for your exploration] gives

- **the calibrated 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  $\Delta r_{10Y}$  and  $\Delta r_{25Y}$ . There is co-movement on the short timescale.

For the lower frequency samples, you might find that  $\Delta r_t$  (for the short rate) and  $\Delta y_t$  (for some long-term rate) are cointegrated but correlated weakly negatively.

## Equilibrium Correction Model: two-way

$$\Delta r_{10Y} = 1.086\Delta r_{25Y} - 0.02716e_{t-1} + \epsilon_t$$

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

Residual standard error: 0.01981      Multiple R-squared: 0.8202

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

	Estimate	Std. Error	t value	Pr(> t )
tenorY.diff	0.751627	0.015910	47.243	<2e-16 ***
eq_corr.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.

## Extra Slides

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

## Restricted VECM for $\Delta r_{10Y}$ and $\Delta r_{25Y}$

```
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  $\hat{\epsilon}_t$ .

```
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
```

Certificate in Quantitative Finance

## Comparison to linear regression

A regression between the simple differences  $\Delta r_{25Y}$  and  $\Delta r_{10Y}$  gives us a minimum variance relationship and shows that 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
```

Certificate in Quantitative Finance

## Hedging ratio puzzle

What would you use as a hedging ratio for assets  $r_{10Y}$  and  $r_{10Y}$ ?  
Remember you are not investing in returns.

The choices are:

- 0.7698 from a linear regression of  $r_{10Y}$  on  $r_{25Y}$
- 0.7457 from linear regression on differences  $\Delta r_{25Y}$  on  $\Delta r_{10Y}$
- 0.7516 from ECM of  $r_{25Y}$  on  $r_{10Y}$
- 0.7870 from a cointegrating regression (two-way VECM).