Non stationarity and cointegration

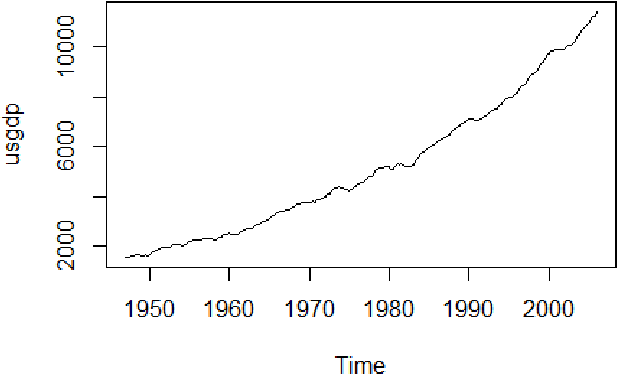
Hugo Harari-Kermadec

EPOG
Econometrics

oct. 18, 2018

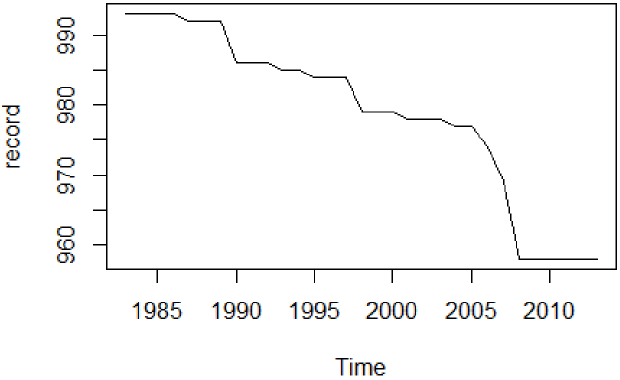
1. Non stationarity
    - GDP
    - Random walk?
2. Dickey-Fuller test strategy
3. Cointegration

US GDP

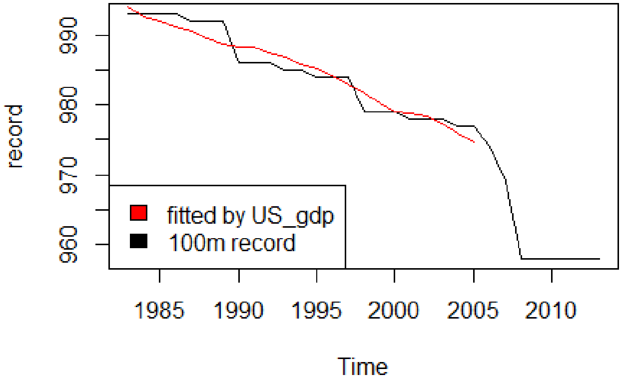


Clearly non stationary.

100m record

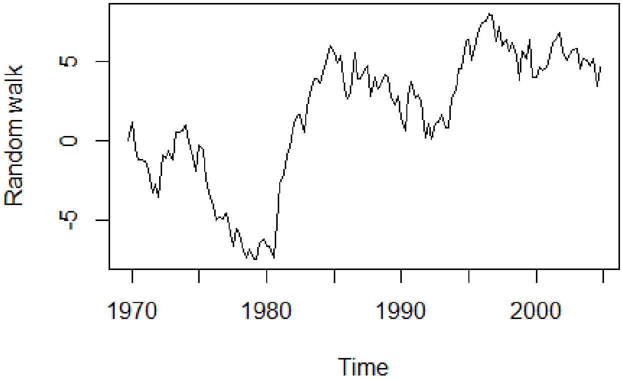


100m record

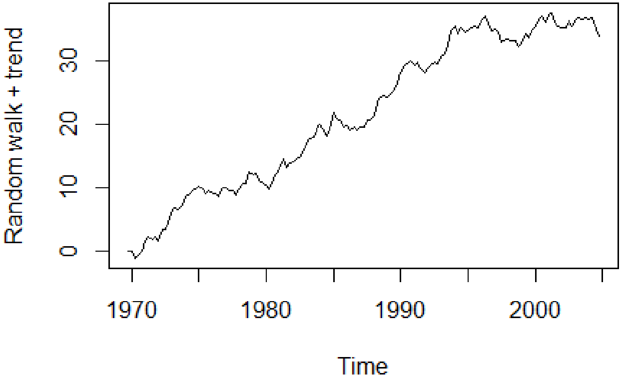

record
990
980
970
960
fitted by US_gdp
100m record
1985
1990
1995
2000
2005
2010
Time

```
Call:
lm(formula = record ~ year_gdp)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.012e+03  1.778e+00  569.40  < 2e-16 ***
year_gdp    -8.513e-04  5.379e-05  -15.83 3.81e-13 ***

Multiple R-squared:  0.9226,Adjusted R-squared:  0.919
```

Random walk



Random walk + trend



Random walk model

AR(1) model

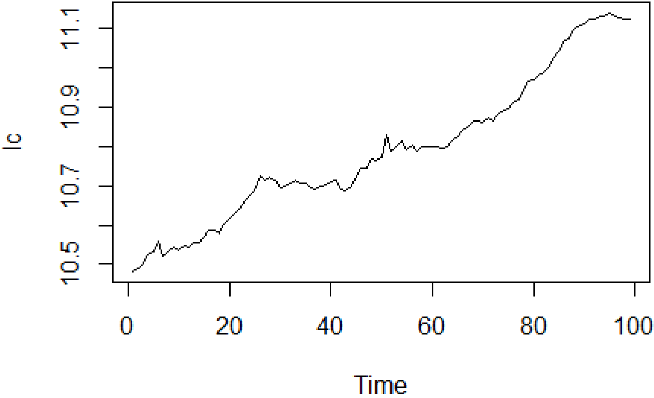
$$X_t = -a_1 \; X_{t-1} + \varepsilon_t$$

$$\Leftrightarrow (1 + a_1 \; L) X_t = \varepsilon_t$$

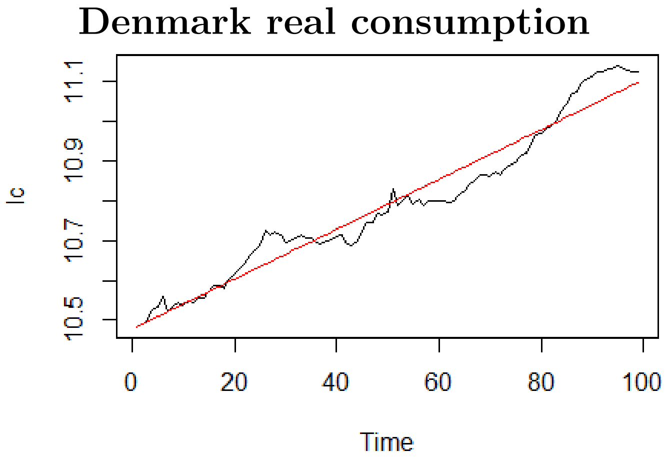
Unit root

$a_1 = -1 \;! \quad \Rightarrow \quad L(X_t) = \varepsilon_t$

Denmark real consumption



Non stationary $\Rightarrow$ Random walk?



Non stationary ⇒ Trend?

Consider the model:

$$\begin{aligned} Z_t &= \alpha + \beta t - a_1 Z_{t-1} + \varepsilon_t \\ (1-L)Z_t &= \alpha + \beta t + (-a_1 - 1) Z_{t-1} + \varepsilon_t \\ (1-L)Z_t &= \alpha + \beta t + \rho Z_{t-1} + \varepsilon_t \end{aligned}$$

Looks like a linear model, but behaves very differently if $\rho = 0$ ("the null").
And again differently if $\alpha \neq 0$ and/or $\beta \neq 0$
So we need a specific test (not a T-test) in each situation.

```
> library(urca)
> df=ur.df(X,type="trend")
> summary(df)

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

               Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.3227245  0.1502083   2.149   0.0327 *
z.lag.1     -0.0329780  0.0166319  -1.983   0.0486 *
tt          -0.0004194  0.0009767  -0.429   0.6680
z.diff.lag  -0.0230547  0.0652767  -0.353   0.7243
```

Last column from linear model → false if unit-root.

$$(1 - L)Z_t = \alpha + \beta t + \rho Z_{t-1} + \varepsilon_t$$

Output continues with adequate stats:

```
Value of test-statistic is: -1.9828 1.8771 2.7371
```

- t-value for $\rho = 0$
- Fisher stat for $(\rho, \alpha, \beta) = (0, 0, 0)$
- Fisher stat for $(\rho, \beta) = (0, 0)$

```
Critical values for test statistics:
      1pct  5pct  10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47
```

$$(1-L)Z_t = \alpha + \beta t + \rho Z_{t-1} + \varepsilon_t$$

```
 Value of test-statistic is: -1.9828 1.8771 2.7371
      1pct  5pct  10pct
tau3 -3.99 -3.43 -3.13
phi3  8.43  6.49  5.47
```

If the test stat for $\rho = 0$ is **larger** than tau3 then accept the unit-root. No absolute values here!
In this example, -1.98>-3.13, we accept $\rho = 0$ at 90%.

- If we accept $\rho = 0$, check that the full model is ok.
  - If the test stat [2.7] for (ρ, β) is larger than phi3 [5.47], then we reject $(\rho, \beta) = (0, 0)$, the full model is ok [No].
  - If not, then $\beta = 0$ and the full model is wrong. Move to model 2 (no trend).
- If we reject $\rho = 0$, it's a classical lineal model, check the trend with the first table.

```
> summary(ur.df(y=lc,type='drift')
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0038899  0.0841706   0.046    0.963
z.lag.1      0.0003199  0.0078044   0.041    0.967
z.diff.lag  -0.1240402  0.1028634  -1.206    0.231

Value of test-statistic is: 0.041 11.1569

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```

$$(1-L)Z_t = \alpha + \rho Z_{t-1} + \varepsilon_t$$

```
Value of test-statistic is: 0.041 11.1569
      1pct  5pct  10pct
tau2 -3.51 -2.89 -2.58
phi1 6.70 4.71 3.86
```

If the test stat for $\rho = 0$ is **larger** than tau2 then accept the unit-root. No absolute values here!
In this example, 0.04>-2.58, we accept $\rho = 0$ at 90%.

- If we accept $\rho = 0$, check that model 2 is ok.
  - If the test stat [11] is larger than phi1 [6.7], then we reject $(\rho, \alpha) = (0, 0)$, model 2 is ok [even at 99%].
  - If not, then $\alpha = 0$ and model 2 is wrong. Move to model 1 (no drift)
- If we reject $\rho = 0$, it's a classical lineal model, check the drift with the first table.

```
> summary(ur.df(y=lc,type='none'))
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

             Estimate Std. Error t value Pr(>|t|)
z.lag.1     0.0006805  0.0001433   4.749 7.24e-06 ***
z.diff.lag -0.1243891  0.1020458  -1.219    0.226

Value of test-statistic is: 4.7485

Critical values for test statistics:
1pct  5pct 10pct
tau1 -2.6 -1.95 -1.61
```

$$(1 - L)Z_t = \rho Z_{t-1} + \varepsilon_t$$

Value of test-statistic is: 4.7485

```
       1pct  5pct  10pct
tau1 -2.6 -1.95 -1.61
```

If the test stat for $\rho = 0$ is **larger** than tau1 then accept the unit-root. No absolute values here!
In this example, 4.74>-1.6, we accept $\rho = 0$ at 90%, model 1 is ok.

All this can be done with more lags in the model (augmented DF model).

You can choose the lags:

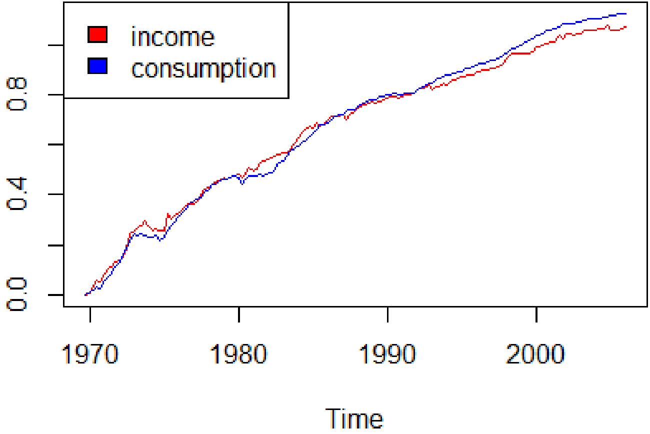
```
summary(ur.df(y=lc,lags=3, type='trend'))
```

or leave it to R:

```
summary(ur.df(y=lc,type='trend',selectlags = "AIC"))
```

All the rest of the DF test strategy remains unchanged.

US Income and Consumption



Clearly twice the same non stationarity.

Cointegration: same random walk in 2 (or) more time series.

$$
\begin{aligned}
X_t &= X_{t-1} + \varepsilon_t \\
Y_t &= aX_t + \nu_t
\end{aligned}
$$

There is a stationary combination: $Y_t - aX_t = \nu_t$ is a white noise.

Cointegration: same random walk in 2 (or) more time series.

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t \\ Y_t &= aX_t + \nu_t \end{aligned}$$

There is a stationary combination: $Y_t - aX_t = \nu_t$ is a white noise.

$$\text{Then } Y_{t-1} = aX_{t-1} + \nu_{t-1} = aX_{t-1} + \left(Y_{t-1} - aX_{t-1} \right)$$

Cointegration: same random walk in 2 (or) more time series.

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t \\ Y_t &= aX_t + \nu_t \end{aligned}$$

There is a stationary combination: $Y_t - aX_t = \nu_t$ is a white noise.

$$\text{Then } Y_{t-1} = aX_{t-1} + \nu_{t-1} = aX_{t-1} + \left(Y_{t-1} - aX_{t-1} \right)$$

$$\text{and } (1-L)\,Y_t = -\left(Y_{t-1} - aX_{t-1} \right) + a(1-L)X_t + \varepsilon_t$$

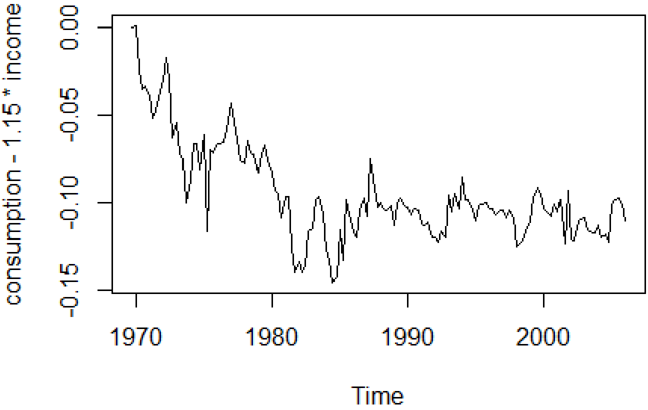
Therefore, the variations of Y_t must compensate the deviations from $Y_{t-1} - aX_{t-1}$ (coef. -1)

$\Rightarrow$ Error Correction Model.

```
>coint=ca.jo(cbind(consumption,income),spec="longrun")
>summary(coint)

Values of teststatistic and critical values of test:
          test 10pct  5pct  1pct
r <= 1 |  8.00  6.50  8.18 11.65
r = 0  | 21.69 12.91 14.90 19.19

Eigenvectors, normalised to first column:
(These are the cointegration relations)
               consumption.l2  income.l2
consumption.l2       1.000000   1.000000
income.l2           -1.153328  -1.023532
```

$r = 0$ is rejected $\rightarrow$ there is at least one relation.
$r \leq 1$ accepted at 95%: there is cointegration.
If it where rejected also, then each cariable is stationary.

US Income and Consumption



The combination is quite stationary