QF 604
ECONOMETRICS OF FINANCIAL MARKETS

LECTURE 8

Professor Lim Kian Guan

SMU Master of Science in
Quantitative Finance 2023-2024 T2

LECTURE OUTLINE

01 Non-Stationary Process

- Consider $Y_t = \theta + \lambda Y_{t-1} + \varepsilon_t , \theta \neq 0$

  where ε_t is a covariance-stationary process with $E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2$, $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0 \;\; \forall k \neq 0$, and $Cov(Y_{t-1}, \varepsilon_t) = 0$.

- Y_t is covariance stationary provided $|\lambda| < 1$.

  If $\lambda = 1$, then, $\quad Y_t = \theta + Y_{t-1} + \varepsilon_t \quad (10.1)$

- Or $(1 - B)Y_t = \theta + \varepsilon_t$, so $(1 - B) = 0$ yields a unit root solution.

  Thus, Y_t contains a unit root and $\{Y_t\}$ is called **a unit root process**.

- By repeated substitution

$$
\begin{aligned}
Y_t &= \theta + (\theta + Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= 2\theta + (\theta + Y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \varepsilon_{t-1} \\
&\;\;\vdots \\
&= t\theta + Y_0 + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots + \varepsilon_2 + \varepsilon_1
\end{aligned}
$$

01 Non-Stationary Process

- A unit root process in Y_t leads to

  Y_t having a **time trend** $t\theta$ and a **stochastic trend** $\sum_{j=0}^{t-1} \varepsilon_{t-j}$.

- If $E(Y_0) = \mu_0$, then $E(Y_t) = \mu_0 + t\theta \neq \mu_0$, provided $\theta \neq 0$.
- The mean of Y_t increases (decreases) with time according to drift $\theta > (<)\ 0$.
- If $Var(Y_0) = \sigma_0^2 < \infty$,

$$Var(Y_t) = \sigma_0^2 + Var\left[\sum_{j=0}^{t-1} \varepsilon_{t-j}\right] = \sigma_0^2 + t\sigma_\varepsilon^2 \neq \sigma_0^2$$

Variance of Y_t increases due to stochastic trend

Hence $\{Y_t\}$ is not covariance-stationary or **non-stationary**

01 Trend Stationary Process

- Suppose random variable Y_t is trend stationary, i.e., stationary about a deterministic time trend $\delta + t\theta$:

$$Y_t = t\theta + \delta + \eta_t \qquad (10.2)$$

η_t: a stationary i.i.d. random variable with zero mean

Then, $Var(Y_t) = Var(\eta_t) = \sigma_\eta^2$

- Then, $Y_t = \theta + Y_{t-1} + \Delta\eta_t \qquad (10.3)$

where $\mathrm{Var}(\Delta\eta_t) = Var(\eta_t - \eta_{t-1}) = 2\sigma_\eta^2$ since η_t is i.i.d.

- Eq. (10.3) looks like the unit root process in Eq. (10.1), but it is NOT

01 Trend Stationary Process

- Eq. (10.3) is not like the unit root process in Eq. (10.1) because the stationary noise term $\Delta\eta_t$ carries a special structure.
- Iterate the process Eq. (10.3) through time:

$$\begin{aligned} Y_t &= \theta + (\theta + Y_{t-2} + \Delta\eta_{t-1}) + \Delta\eta_t \\ &= 2\theta + (\theta + Y_{t-3} + \Delta\eta_{t-2}) + \Delta\eta_t + \Delta\eta_{t-1} \\ &\quad \vdots \\ &= t\theta + Y_0 + \Delta\eta_t + \Delta\eta_{t-1} + \Delta\eta_{t-2} + \cdots + \Delta\eta_2 + \Delta\eta_1 \\ &= t\theta + Y_0 + \eta_t - \eta_0 \end{aligned}$$

where $Var(\eta_t - \eta_0) = 2\sigma_\eta^2$

- If $Var(Y_0) = \sigma_0^2 < \infty$, $Var(Y_t) = \sigma_0^2 + 2\sigma_\eta^2$

  For trend stationary process, Var(Y_t) stays the same even as *t* increases
- Big difference between Eq. (10.3) and Eq. (10.1): $\Delta\eta_t$ does not add up variance, unlike stochastic trend in a unit root process

01 Unit Root vs. Trend Stationary Process

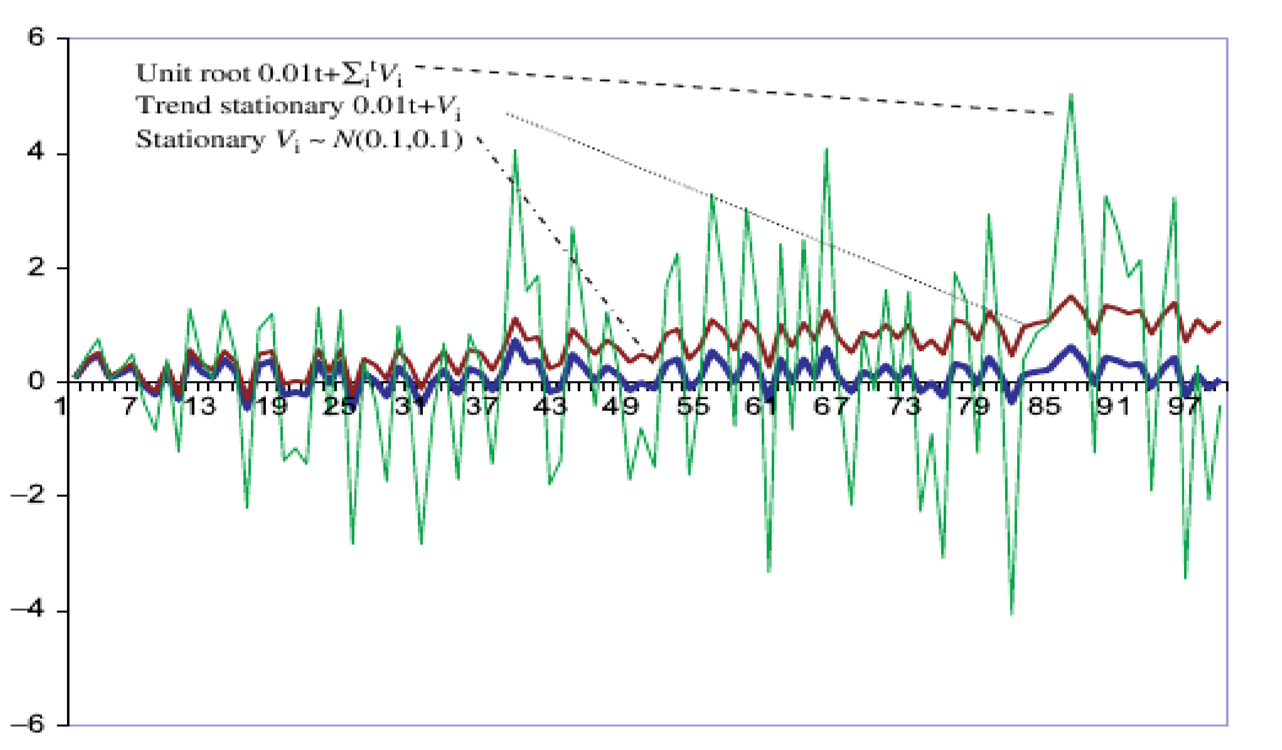


Figure 10.1: Time Series Graphs of Stochastic Processes

- A unit root process contains a deterministic time trend + a stochastic trend $\sum_{j=1}^{t-1} \varepsilon_j$
- A trend stationary process contains a deterministic time trend + stationary noise $\eta_t - \eta_0$
- Both processes display time trend but unit root process displays increasing volatility over time

02 Spurious Regression

- Suppose

$$Y_t = \theta + Y_{t-1} + e_t, \qquad e_t \sim \text{stationary with mean } 0$$

$$Z_t = \mu + Z_{t-1} + u_t, \qquad u_t \sim \text{stationary with mean } 0$$

e_t and u_t are independent of each other, also not correlated with Y_{t-1} and Z_{t-1}

$\{Y_t\}$ and $\{Z_t\}$ are unit root processes with drifts θ and μ, respectively

- Then,

$$Y_t = t\theta + Y_0 + (e_t + e_{t-1} + \cdots + e_1)$$

$$Z_t = t\mu + Z_0 + (u_t + u_{t-1} + \cdots + u_1)$$

- Let Y_0 and Z_0 be independent,

$$Cov(Y_t, Z_t) = Cov(Y_0, Z_0) + Cov\left(\sum_{j=1}^{t} e_j, \sum_{k=1}^{t} u_k\right) = 0$$

02 Spurious Regression

- Y_t , Z_t are independent

  Consider a linear regression of Y_t on Z_t :

$$Y_t = a + bZ_t + \eta_t \tag{10.4}$$

  η_t is independent of Z_t. η_t is a unit root process, $\eta_t = \eta_{t-1} + \epsilon_t$, with stationary ϵ_t, $\mathrm{E}(\epsilon_t) = 0$, $Var(\epsilon_t) < \infty$.

- The slope $b = \frac{Cov(Y_t, Z_t)}{Var(Z_t)} = 0$ since $Cov(Y_t, Z_t) = 0$

- Expand the regression into its time trend and additive stochastic component:

$$t\theta + Y_0 + \sum_{j=0}^{t-1} e_{t-j} = a + b\left(t\mu + Z_0 + \sum_{j=0}^{t-1} u_{t-j}\right) + \left(\eta_0 + \sum_{j=0}^{t-1} \epsilon_{t-j}\right)$$

02 Spurious Regression

- Divide through by t

$$\theta + \frac{Y_0}{t} + \frac{1}{t}\sum e_{t-j} = \frac{a}{t} + b\mu + \frac{bZ_0}{t} + \frac{b}{t}\sum u_{t-j} + \frac{\eta_0}{t} + \frac{1}{t}\sum \epsilon_{t-j}$$

- $\mathrm{Var}\left(\frac{Y_0}{t}\right)$, $\mathrm{Var}\left(\frac{Z_0}{t}\right)$ and $\mathrm{Var}\left(\frac{\eta_0}{t}\right)$ all $\downarrow 0$ as $t \uparrow \infty$

  As t increases, the time-averages of the noise terms in e_t, u_t, ϵ_t converge to zeros.

  Then, in regression Eq. (10.4), $\theta \approx b\mu$, so $b \approx \frac{\theta}{\mu} \neq 0$.

- The regression in Eq. (10.4) between two independent processes produces an estimated slope coefficient $\hat{b}$ which is $\widehat{Cov}(Y_t, Z_t)/\widehat{Var}(Z_t)$.

  Numerator and denominator yield sampling averages giving $\approx \frac{\theta}{\mu} \neq 0$.

- This is termed a **spurious** (seemingly true yet false) regression result: $\hat{b} \neq 0$ is obtained from linear regression when theoretically $b = 0$.

02 Spurious Regression

- The spurious regression above applies also to Y_t and Z_t if they are trend stationary instead of being unit root processes. Consider

$$Y_t = t\theta + \delta + \eta_t$$

$$Z_t = t\mu + \gamma + \xi_t$$

where η_t and ξ_t are mean zero i.i.d. random variables that have zero correlation.

- Even though Y_t and Z_t are not correlated,

$$Y_t = \delta + \theta\left[\frac{Z_t - \gamma - \xi_t}{\mu}\right] + \eta_t = \left(\delta - \frac{\theta\gamma}{\mu}\right) + \frac{\theta}{\mu} Z_t + \left(\eta_t - \frac{\theta}{\mu}\xi_t\right)$$

- So, OLS regression of Y_t on Z_t will give a spurious estimate of $\theta/\mu \neq 0$.
- Spurious non-zero correlation between Y_t , Z_t, even when they are independent processes, comes from their deterministic trend, not stochastic trend

02 Linear Combinations of I(1) Processes

- Suppose

$$Z_t = \mu + Z_{t-1} + u_t, \qquad u_t \sim \text{stationary with mean } 0$$

$$w_t = \gamma + w_{t-1} + \xi_t, \qquad \xi_t \sim \text{stationary with mean } 0$$

are independent unit root processes.

- In general, **a linear combination of the unit root processes** Z_t and w_t, $\boldsymbol{Y_t}$ is also a unit root process:

$$\begin{aligned} Y_t &= c + dZ_t + w_t \\ &= (c+\gamma) + d\mu + dZ_{t-1} + du_t + w_{t-1} + \xi_t \\ &= (\gamma + d\mu) + Y_{t-1} + (du_t + \xi_t) \end{aligned}$$

for constants c, d. Y_t is correlated with Z_t where $d \neq 0$.

02 Spurious Regression of related Unit Root Processes

- Perform OLS of Y_t on Z_t:

$$Y_t = c + dZ_t + w_t$$

where $d \neq 0$ and w_t is independent unit root process.

- OLS estimate of d will involve

$$Cov(Y_t, Z_t) = Cov(c + dZ_t + w_t, Z_t) = d\,Var(Z_t) + Cov(c + w_t, Z_t)$$

The latter is a covariance of two independent unit root processes each with a deterministic trend (and a stochastic trend as well), that produces spurious sampling estimate (seen earlier) that is not zero.

- The sampling estimate of $Cov(Y_t, Z_t)$ under OLS will also be spurious regardless of the value of d, due to $Cov(c + w_t, Z_t) \neq 0$.

02 Spurious Regression of related Unit Root Processes

- When OLS on two **related** unit root processes such as Y_t and Z_t can or cannot be feasible, it has to do with the covariance of the explanatory variable and the residual variable, $\mathrm{cov}(w_t, Z_t)$.

  If both are unit root processes, then there is spuriousness.

- If w_t is a stationary process, and not a unit root process, independent of Z_t.

  Then, the sample estimate of $\mathrm{cov}(w_t, Z_t) = 0$.

  In this case, the OLS estimate of d converges correctly.

- In general, whether unit root processes Y_t and Z_t **are truly related or not**:

$$Y_t = c + dZ_t + w_t$$

  If disturbance w_t has a unit root and is not correlated with Z_t

  Then Not appropriate to perform OLS of Y_t on Z_t since w_t is not stationary

  OLS result will be spurious

03 Unit Root Test

- Suppose $$Y_t = Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \text{stationary} \qquad (10.5)$$

  First difference $\Delta Y_t \equiv Y_t - Y_{t-1} = \varepsilon_t$ stationary

  Y_t: integrated order 1 or $I(1)$ process

  ΔY_t: integrated order 0 or $I(0)$ process, stationary

  In general, $I(k)$ is integrated order k process if after the kth differencing, process first becomes stationary

- Suppose $$Y_t = \delta + Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \text{stationary} \qquad (10.6)$$

  First difference $\Delta Y_t \equiv Y_t - Y_{t-1} = \delta + \varepsilon_t$ is stationary

- Suppose $$Y_t = \delta + \theta t + Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \text{stationary} \qquad (10.7)$$

  First difference $\Delta Y_t = \delta + \theta t + \varepsilon_t \sim$ stationary after detrending

- Eqs. (10.5), (10.6), (10.7) represent unit root processes.

03 Unit Root Test

- The alternative stationary autoregressive hypotheses to Eqs. (10.5), (10.6), (10.7) are respectively:

$$Y_t = \lambda Y_{t-1} + \varepsilon_t \quad (|\lambda| < 1) \tag{10.8}$$

$$Y_t = \delta + \lambda Y_{t-1} + \varepsilon_t \quad (|\lambda| < 1) \tag{10.9}$$

$$Y_t = \delta + \theta t + \lambda Y_{t-1} + \varepsilon_t \quad (|\lambda| < 1) \tag{10.10}$$

- Correspondingly

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t \tag{10.11}$$

$$\Delta Y_t = \delta + \gamma Y_{t-1} + \varepsilon_t \tag{10.12}$$

$$\Delta Y_t = \delta + \theta t + \gamma Y_{t-1} + \varepsilon_t \tag{10.13}$$

where $\gamma = \lambda - 1$.

- For $I(1)$ processes in Eqs. (10.5), (10.6), (10.7), $\gamma \equiv \lambda - 1 = 0$. Then test the null hypothesis of a unit root process, H_0: $\gamma = 0$. Since for stationary alternatives, $|\lambda| < 1$, more negatively estimated $\gamma \Rightarrow$ higher likelihood of stationarity.

03 Unit Root Test

- Specifications (10.11), (10.12), (10.13) are generalized to include lags of ΔY_t to whiten the noise term ε_t (we use the strong case of making residual noise i.i.d.)

$$\Delta Y_t = \gamma Y_{t-1} + \beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} + \cdots + \beta_k \Delta Y_{t-k} + e_t \quad \text{(no constant)} \qquad (10.14)$$

$$\Delta Y_t = \delta + \gamma Y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta Y_{t-j} + e_t \quad \text{(constant)} \qquad (10.15)$$

$$\Delta Y_t = \delta + \theta t + \gamma Y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta Y_{t-j} + e_t \quad \text{(constant and time trend)} \qquad (10.16)$$

where e_t is i.i.d.

- To test if Eqs. (10.14), (10.15), (10.16) are unit root processes, run OLS (for some k)

- If $\hat{\gamma}$ significantly < 0, then reject H_0 of unit root. If not, there is evidence of a unit root process

03 Unit Root Test

Augmented Dickey-Fuller (ADF) tests: Tests using specifications with lagged ΔY_t

- Compute $$x = \frac{\hat{\gamma}_{OLS}}{\text{OLS s.e.}(\hat{\gamma}_{OLS})}$$

  It is the usual formula for *t*-value function, but in this case, is not distributed as Student-t_{T-n} statistic where T is sample size and n is number of parameters

- It is a non-standard nondegenerate distribution.
- For a correctly specified *k* in Eqs. (10.14), (10.15), or (10.16), the probability distribution of the *x*-statistic (independent of *k*) is found by simulations.

  The distribution is reported by Dickey and Fuller.

  See **pages 288 – 290 of book** for an explanation of how this distribution is derived.

03 Unit Root Test

- From the Dickey-Fuller table: if sample size T = 250

(1) No constant: computed x-statistic < -2.58, then reject H_0: $\gamma = 0$ (or $\lambda = 1$) $\rightarrow$ no unit root at 1% significance level

(2) Constant: -3.14 < computed x-statistic < -2.88, then reject H_0: $\gamma = 0$ (or $\lambda = 1$) $\rightarrow$ no unit root at 5% significance level but cannot reject at 2.5% significance level.

(3) Constant and time trend: -3.42 < computed x-statistic < -3.13, then cannot reject H_0: $\gamma = 0$ (or $\lambda = 1$) at 5% significance level

- Another check for unit root: autocorrelation function (ACF) is highly persistent or slow decay for unit root process

Table 10.1: Critical Values for Dickey-Fuller t-Test

	Sample Size	p-Values (probability of a smaller test value)			
Case:	**T**	**0.01**	**0.025**	**0.05**	**0.10**
No constant	25	-2.65	-2.26	-1.95	-1.60
Eq. (10.14)	50	-2.62	-2.25	-1.95	-1.61
	100	-2.60	-2.24	-1.95	-1.61
	250	-2.58	-2.24	-1.95	-1.62
	500	-2.58	-2.23	-1.95	-1.62
	∞	-2.58	-2.23	-1.95	-1.62
Case:	**T**	**0.01**	**0.025**	**0.05**	**0.10**
Constant	25	-3.75	-3.33	-2.99	-2.64
Eq. (10.15)	50	-3.59	-3.23	-2.93	-2.60
	100	-3.50	-3.17	-2.90	-2.59
	250	-3.45	-3.14	-2.88	-2.58
	500	-3.44	-3.13	-2.87	-2.57
	∞	-3.42	-3.12	-2.86	-2.57
Case:	**T**	**0.01**	**0.025**	**0.05**	**0.10**
Constant and	25	-4.38	-3.95	-3.60	-3.24
time trend	50	-4.16	-3.80	-3.50	-3.18
Eq. (10.16)	100	-4.05	-3.73	-3.45	-3.15
	250	-3.98	-3.69	-3.42	-3.13
	500	-3.97	-3.67	-3.42	-3.13
	∞	-3.96	-3.67	-3.41	-3.13

Source: Fuller, W., "Introduction to Statistical Time Series , Second Edition," New York: Wiley, 1996.

04 Purchasing Power Parity

P_t : UK national price index in £
P_t^*: U.S. national price index in USD
e_t : spot exchange rate: number of £ per $

- Absolute purchasing power parity(PPP) version: $P_t = e_t P_t^*$ or $e_t = P_t / P_t^*$

$$\ln P_t = \ln e_t + \ln P_t^*$$

$$d \ln P_t = d \ln e_t + d \ln P_t^*$$

$$\frac{dP_t}{P_t} = \frac{de_t}{e_t} + \frac{dP_t^*}{P_t^*}$$

$$\frac{\Delta P_t}{P_t} = \frac{\Delta e_t}{e_t} + \frac{\Delta P_t^*}{P_t^*} \text{ or } \frac{\Delta e_t}{e_t} = \frac{\Delta P_t}{P_t} - \frac{\Delta P_t^*}{P_t^*}$$

- Relative PPP version: Percent change in exchange rate = Inflation rate difference

04 Purchasing Power Parity

Example:

If U.S. inflation rate is 5%, UK inflation rate is 10%, both over horizon *T* years, then

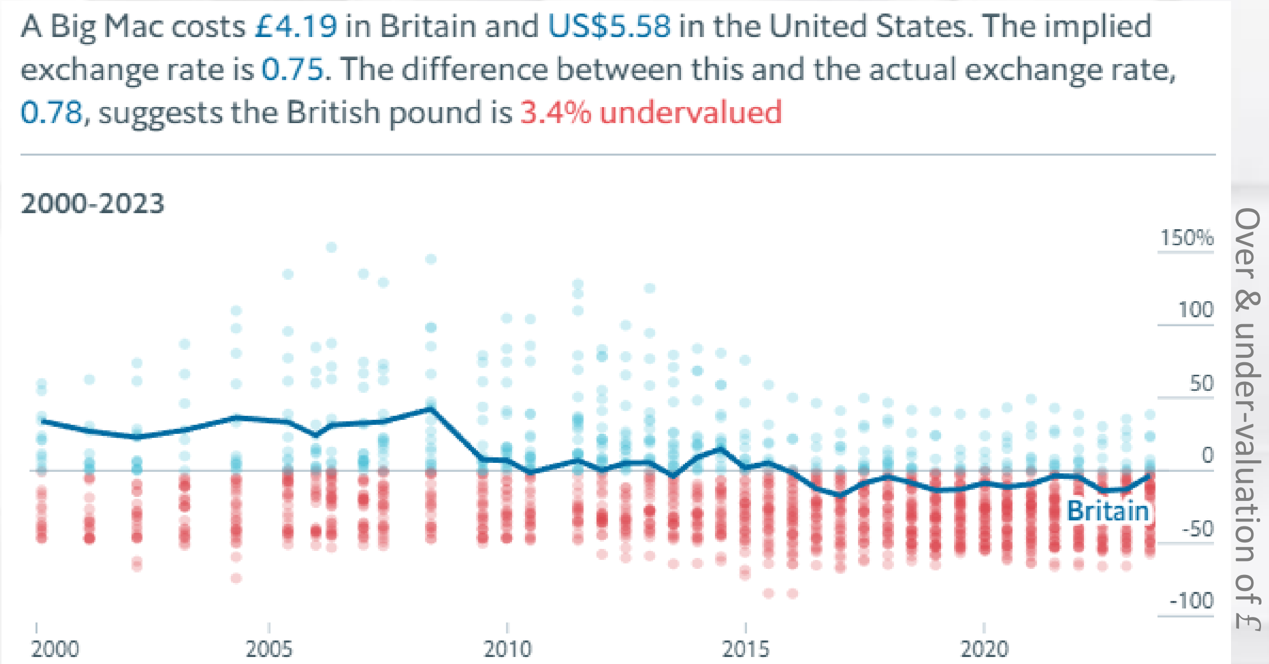
$\Delta e_t/e_t = 10\% - 5\% = 5\%$, \$ is expected to appreciate by 5% against £ over T years.

- The real exchange rate (real £ per \$): $r_t = e_t P_t^* / P_t$

$$r_t = \frac{\text{price of US burger in £}}{\text{price of UK burger in £}}$$

$$\text{or} = \frac{\text{price of US good in £}}{\text{price of UK good in £}}$$

Implication: If US can sell US goods and fetch more Pounds than pay Pounds for similar UK goods, then US has positive or better terms of trade. Real £ per \$ > 1 (£ undervalued)



Source: The Economist, Aug 3, 2023

04 Purchasing Power Parity

- The real exchange rate (log form): $\ln r_t = \ln e_t + \ln P_t^* - \ln P_t$

  Under absolute PPP, $r_t = 1$, $\ln r_t = 0$. But more generally, $\ln r_t$ is a random variable.

- Suppose the linear combination:

$$\ln r_t = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \ln e_t \\ \ln P_t^* \\ \ln P_t \end{pmatrix}$$

  Is stationary and not a unit root process

  Then $\ln e_t$, $\ln P_t^*$ and $\ln P_t$: cointegrated with cointegrating vector $\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}$

- Interpretation of PPP/ Long-run PPP: $\ln r_t$ may deviate from 0 but will over time revert back to its mean at 0.

- If long-run PPP does not hold $\Rightarrow$ $\ln r_t$ may deviate from 0 and not return to it. It can then be described as a unit root process: $\ln r_t = \ln r_{t-1} + \eta_t$ (10.25)

  where η_t is stationary with zero mean

04 Purchasing Power Parity

- Check the validity of the long-run PPP using Eq. (10.25)
- Run OLS on

$$\Delta \ln r_t = \delta + \theta t + \gamma \ln r_{t-1} + \sum_{j=1} \beta_j \Delta \ln r_{t-j} + \xi_t$$

$\Delta\eta_t = \sum_{j=1} \beta_j \Delta \ln r_{t-j} + \xi_t$, ξ_t is i.i.d.

- The null hypothesis of unit root process of $\ln r_t$ is $H_0\colon \gamma = 0$

If unit root is rejected (accepted) $\Rightarrow$ long-run PPP holds (does not hold)

04 Purchasing Power Parity, 1960-2001

Table 10.2: Augmented DF Unit Root Test of $\ln e_t$.
$\Delta \ln e_t = \delta + \theta t + \gamma \ln e_{t-1} + \sum_{j=1} \beta_j \Delta \ln e_{t-j} + \xi_t$. Sample size 42

ADF Test Statistic	−1.776037	1% Critical Value		−4.16
Coefficient	Estimate	Std. Error	t-Statistic	Prob.
γ	−0.280485	0.157928	−1.776037	0.0859
β_1	0.392861	0.192132	2.044747	0.0497
β_2	0.121412	0.185310	0.655185	0.5173
β_3	−0.191780	0.184113	−1.041640	0.3059
β_4	−0.081663	0.187675	−0.435129	0.6666
Constant δ	−0.272398	0.178452	−1.526448	0.1374
Time Trend θ	0.004542	0.003192	1.423116	0.1650
R-squared	0.385694	Mean dependent var		1.017981
Adjusted R-squared	0.262833	S.D. dependent var		0.083637
S.E. of regression	0.071809	Akaike info criterion		−2.260947
Sum squared resid.	0.154697	Schwarz criterion		−1.955179

(1) The ADF test-statistic with constant and trend of -1.7760 > -4.16 at 1% critical level.
(2) It is also greater than -3.18 at 10% critical level.
(3) Cannot reject H_0: $\ln e_t$ during 1960-2001 follows a unit root process.

04 Purchasing Power Parity, 1960-2001

Table 10.6: Augmented DF Unit Root Test of $\ln r_t$.
$\Delta \ln r_t = \delta + \theta t + \gamma \ln r_{t-1} + \sum_{j=1} \beta_j \Delta \ln r_{t-j} + \xi_t$ vide Eq. (10.26). Sample size 42

ADF Test Statistic	-3.096016	1% Critical Value		-4.16
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
γ	-0.663089	0.214175	-3.096016	0.0042
β_1	0.560278	0.192777	2.906358	0.0068
β_2	0.360211	0.185043	1.946631	0.0610
β_3	0.054125	0.188122	0.287711	0.7755
β_4	0.091169	0.188378	0.483969	0.6319
Constant δ	0.006824	0.027781	0.245636	0.8076
Time Trend θ	0.004848	0.002017	2.403897	0.0226
R-squared	0.422015	Mean dependent var		0.002645
Adjusted R-squared	0.306418	S.D. dependent var		0.084680
S.E. of regression	0.070523	Akaike info criterion		-2.297094
Sum squared resid	0.149205	Schwarzcriterion		-1.992326

(1) $\ln r_t$: Cannot reject null of unit root (for constant + time trend) at 10% significance level

(2) ln real exchange rate r_t with a unit root: long-run PPP does not hold and that disequilibrium from PPP or deviations r_t from 0 do not have tendency to revert back toward 0.

04 Purchasing Power Parity, 1960-2001

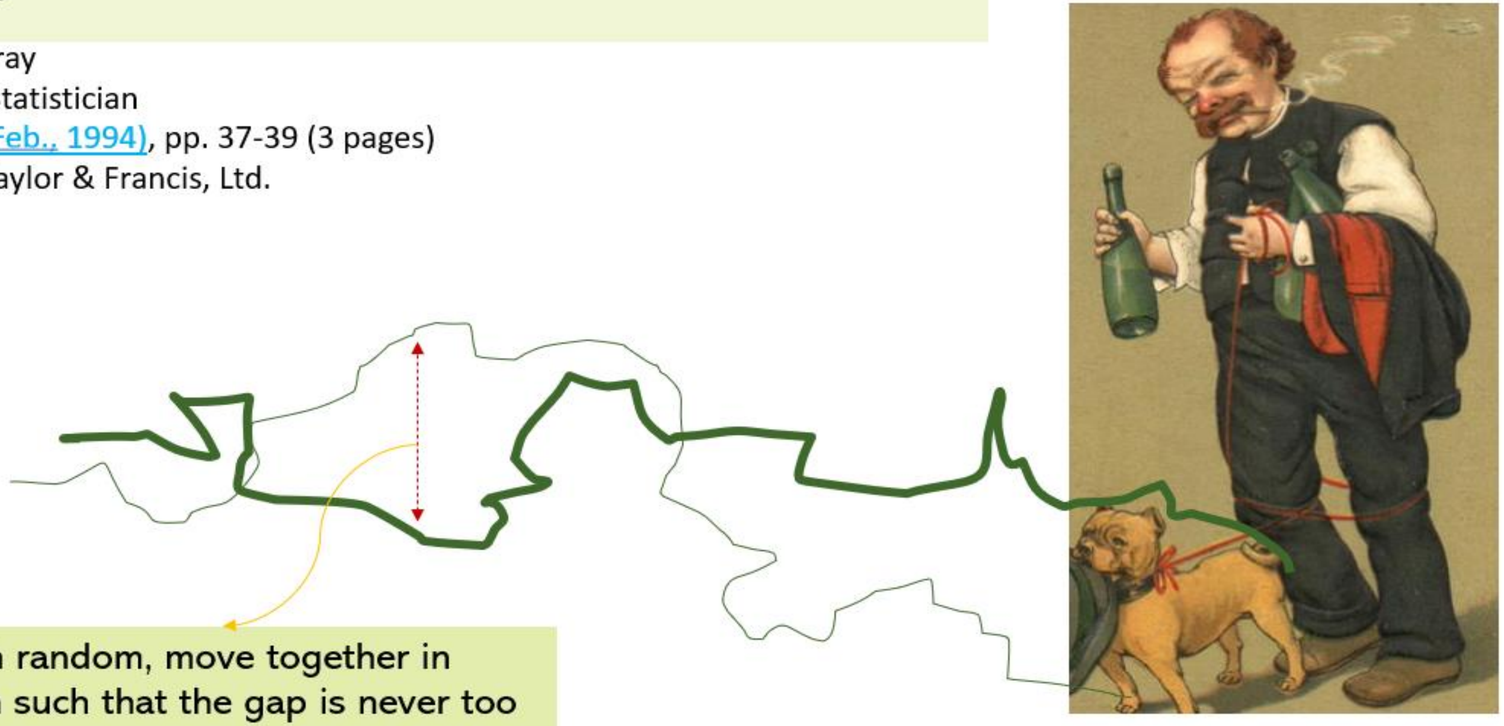
Figure 10.2: Log Real Exchange 1960 till 2001, $\ln r_t = \ln e_t + \ln P_t^* - \ln P_t$

Fig. 10.2 shows log real exchange rate in £ per $ are mostly negative ⇒ better terms of trade and competitiveness favoring UK from 1960 to 2001. £ over-valued.

05 Test of Cointegration

A Drunk and Her Dog: An Illustration of Cointegration and Error Correction

Michael P. Murray
The American Statistician
Vol. 48, No. 1 (Feb., 1994), pp. 37-39 (3 pages)
Published By: Taylor & Francis, Ltd.



From open source web. I could only find picture of a "he"

Paths though random, move together in some fashion such that the gap is never too wide – paths are cointegrated, gap is error correction

05 Test of Cointegration

- If Y_t and Z_t are unit root processes

$$Y_t = c + dZ_t + w_t$$

$d \neq 0$ and w_t is stationary $\Rightarrow Y_t$ and Z_t can be cointegrated with cointegrating vector $(1, -d)$

- In a more general setup, we could have a vector of either unit root $I(1)$ or root-stationary $I(0)$ r.v.'s X_t of order $n \times 1$ whereby there are a number of combinations of the elements that could be co-integrated.

- We can employ the **Johansen maximum eigenvalue test statistic** and/or the **Johansen trace test statistic** to verify the number of cointegrating vectors present in X_t.

05 Test of Cointegration

- Consider two different stochastic processes:

$$X_{1,t} = a_1 + b_1 t + c_1 X_{1,t-1} + d_1 X_{2,t-1} + e_{1,t}$$

$$X_{2,t} = a_2 + b_2 t + c_2 X_{1,t-1} + d_2 X_{2,t-1} + e_{2,t}$$

where $e_{1,t}$, $e_{2,t}$ are independent white noises

$$\Delta X_{1,t} = a_1 + b_1 t + (c_1 - 1) X_{1,t-1} + d_1 X_{2,t-1} + e_{1,t}$$
$$\Delta X_{2,t} = a_2 + b_2 t + c_2\, X_{1,t-1} + (d_2 - 1)\, X_{2,t-1} + e_{2,t}$$

- Univariate stationary $\Delta X_{1,t}$, $\Delta X_{2,t}$ stochastic processes can be stated as Vector Autoregressive (VAR) process

$$\begin{pmatrix} \Delta X_{1,t} \\ \Delta X_{2,t} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} t + \begin{pmatrix} c_1 - 1 & d_1 \\ c_2 & d_2 - 1 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}$$

Or write as $\Delta X_t = \mu_0 + \mu_1 t + \Pi \; X_{t-1} + e_t$

- Suppose $X_{1,t}$ and $X_{2,t}$ are unit root I(1)processes. Hence are $X_{1,t-1}$ and $X_{2,t-1}$.

If $X_{1,t}$ and $X_{2,t}$ are not cointegrated, then $\Pi = 0$, or $c_1 = d_2 = 1$ and $d_1 = c_2 = 0$.

If there is one cointegrating vector, then either one of $(c_1 - 1 \quad d_1)$ or $(c_2 \quad d_2 - 1)$ is zero while the other is not, so Π is of rank 1.

If there are two cointegrating vectors, then rank of Π is 2.

05 Test of Cointegration

- More generally, consider a $n \times 1$ dimension Vector Autoregressive Process VAR(p) with possible constant and time trend:

$$X_t = \mu_0 + \mu_1 t + \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} + a_t \qquad (10.30)$$

where a_t: i.i.d. vector ~ N(0,Ω)

Elements of X_t: either $I(1)$ or $I(0)$

Dimension of Φ_i : $n \times n$

- An (vector) error correction model (ECM) for the VAR(p) process X_t by re-arranging Eq.(10.30):

$$\Delta X_t = \mu_0 + \mu_1 t + \prod X_{t-1} + \Phi_1^* \Delta X_{t-1} + \cdots + \Phi_{p-1}^* \Delta X_{t-p+1} + a_t \qquad (10.31)$$

where $\Phi_j^* = -\sum_{i=j+1}^{p} \Phi_i$ and $\Pi_{n\times n} = \sum_{i=1}^{p} \Phi_i - I_{n\times n}$

- $\sum_{k=1}^{p-1} \Phi_k^* \Delta X_{t-k}$ is the autoregressive distributed lags, captures short-run impact on LHS ΔX_t
- $\prod_{n\times n} X_{t-1} = (\sum_{k=1}^{p-1} \Phi_k - I)\, X_{t-1}$ is the error correction, captures the deviation and adjustment toward the long-run equilibrium since it is stationary and will revert to zeros at some points

05 Test of Cointegration

- **If all elements of X_t are $I(0)$**, then in Eq.(10.31), add X_{t-1} to both sides.

  Then LHS X_t is stationary

  so RHS (except for time-trend) has stationary terms in ΔX_{t-i}'s + (I + ∏) X_{t-1}

  ⇒ Any n linear combinations of X_{t-1} is stationary, hence ∏ **is of full rank n**

- If some elements (at least one) of X_t are $I(1)$:

  LHS ΔX_t is stationary

  If there is no linear combination of X_{t-1} that is stationary, i.e., elements of X_{t-1} are **not cointegrated**, then for RHS to be stationary, **rank (∏) = 0** or ∏ = 0

05 Test of Cointegration

- The intermediate case: if some elements of X_t are $I(1)$, and there **exists $m < n$ number of cointegrating $n \times 1$ vectors**, then can find $m < n$ numbers of independent rows of $\prod$ such that $\prod X_{t-1}$ is stationary.
  - **Rank($\prod$) = m**

- Write Eq.(10.31) for detrended X_ts:

$$\Delta X_t = \mu_0 + \alpha\beta^T X_{t-1} + \Phi_1^* \Delta X_{t-1} + \cdots + \Phi_{p-1}^* \Delta X_{t-p+1} + a_t \qquad (10.32)$$

where $\prod_{n\times n} = \alpha_{n\times m}\beta_{n\times m}^T$ (m ≤ n), so rank of $\prod$ is at most m (m can also be equal to n here). Rank of $\prod$ can still = 0, or < m if the elements in α or β have linear dependencies.

- The log likelihood function of sample observations on detrended X_t is:

$$\ln L = -\frac{Tp}{2} ln(2\pi) - \frac{T}{2}\ln|\Omega| - \frac{1}{2}\sum_{t=1}^{T} a_t^T\, \Omega^{-1} a_t$$

05 Test of Cointegration

- Perform two auxiliary ML (or OLS) regression :

$$\Delta X_t = \mu_c + C_1^* \Delta X_{t-1} + \cdots + C_{p-1}^* \Delta X_{t-p+1} + u_t \qquad (10.34)$$

$$X_{t-1} = \mu_d + D_1^* \Delta X_{t-1} + \cdots + D_{p-1}^* \Delta X_{t-p+1} + v_t \qquad (10.35)$$

From Eqs. (10.34), (10.35), obtain

$$u_t = \alpha\beta^T v_t + a_t$$

- OLS fitted residuals $\hat{u}_t$ and $\hat{v}_t$ from Eqs. (10.34), (10.35), given $\prod = \alpha\beta^T$, are used to construct the maximum likelihood estimator of Ω, $\widehat{\Omega}$

$$\widehat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - \prod \hat{v}_t)\,(\hat{u}_t - \prod \hat{v}_t)^T = \hat{R}_{uu} - \prod \hat{R}_{vu} - \hat{R}_{uv}\prod^T + \prod \hat{R}_{vv}\prod^T$$

where $\hat{R}_{uu} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{u}_t \hat{u}_t^T$, $\hat{R}_{uv} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{u}_t \hat{v}_t^T$, and $\hat{R}_{vv} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{v}_t \hat{v}_t^T$

- Ignoring constants, asymptotically, concentrating out α (see **details** in pages 299 – 304 of book):

$$\ln L = -\frac{T}{2}\ln\left|\widehat{\Omega}\right| = -\frac{T}{2}\ln\left|\hat{R}_{uu}\right|\left|\beta^T\left(\hat{R}_{vv} - \hat{R}_{vu}\hat{R}_{uu}^{-1}\hat{R}_{uv}\right)\beta\right| / \left|\beta^T\hat{R}_{vv}\beta\right|$$

05 Test of Cointegration

- Let the m ≤ n eigenvectors of $\hat{R}_{vv}^{-1}\hat{R}_{vu}\hat{R}_{uu}^{-1}\hat{R}_{uv}$ be estimate $\hat{\beta}_{n\times m}$ and the corresponding eigenvalues be contained in diagonal $\hat{\Lambda}_{m\times m}$ where

$$\left(\hat{R}_{vv}^{-1}\hat{R}_{vu}\hat{R}_{uu}^{-1}\hat{R}_{uv}\right)\hat{\beta}_{n\times m}=\hat{\beta}_{n\times m}\,\hat{\Lambda}_{m\times m}$$

- Then $\hat{\beta}^T\,\hat{R}_{vu}\hat{R}_{uu}^{-1}\hat{R}_{uv}\hat{\beta}=\hat{\beta}^T\ \hat{R}_{vv}\,\hat{\beta}\,\hat{\Lambda}$
- Normalize $\hat{\beta}$ so $\hat{\beta}^T\ \hat{R}_{vv}\,\hat{\beta}\ =I$, then $\hat{\beta}^T\,\hat{R}_{vu}\hat{R}_{uu}^{-1}\hat{R}_{uv}\hat{\beta}=\hat{\Lambda}$

and $$\frac{\left|\beta^T\left(\hat{R}_{vv}-\hat{R}_{vu}\hat{R}_{uu}^{-1}\hat{R}_{uv}\right)\beta\right|}{\left|\beta^T\hat{R}_{vv}\beta\right|}=\frac{\left|I-\hat{\Lambda}\right|}{|I|}=\begin{vmatrix}1-\hat{\lambda}_1 & 0 & 0 & 0\\ 0 & 1-\hat{\lambda}_2 & 0 & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 1-\hat{\lambda}_m\end{vmatrix}$$

- Hence the likelihood becomes

$$\ln L=-\frac{T}{2}\ln\left|I\ -\hat{\Lambda}\right|$$

05 Test of Cointegration

- In In Johansen's **maximum eigenvalue** test:

  H_0: Rank ($\prod$) = m vs H_A: Rank ($\prod$) = m + 1

  Under H_0, the (m+1)th eigenvalue λ_{m+1} = 0, hence ln(1 – λ_{m+1}) = 0.

  Testing H_0 is testing the restriction ln(1 – λ_{m+1}) = 0.

- Restricted log likelihood under H_0 is ln $L^r = -\frac{T}{2}\ln\left|I_{m\times m} - \hat{\Lambda}_{m\times m}\right|$

  Unrestricted log likelihood is ln $L = -\frac{T}{2}\ln\left|I_{m+1\times m+1} - \hat{\Lambda}_{m+1\times m+1}\right|$

- Likelihood ratio test involves the restricted ln L^r and unrestricted ln L

  $$-2\ln\left[\frac{L^r}{L}\right] = T\left(\ln\left|I_{m\times m} - \hat{\Lambda}_{m\times m}\right| - \ln\left|I_{m+1\times m+1} - \hat{\Lambda}_{m+1\times m+1}\right|\right) = -T\ln(1-\hat{\lambda}_{m+1})$$

  or $-(T-p)\ln(1-\hat{\lambda}_{m+1})$ which is positive as eigenvalues are positive and less than 1

- Test statistic $-T\ln(1-\hat{\lambda}_{m+1}) > 0$ has a non-standard (not chi-sq) distribution. If test statistic is too large, reject H_0 , accept H_A

05 Test of Cointegration

- In In Johansen's **trace cointegration** test:

  H_0: Rank ($\prod$) = m vs H_A: Rank ($\prod$) > m

  Under H_0, all (m+1)th to nth eigenvalue = 0.

  Testing H_0 is testing the restriction $\sum_{i=m+1}^{n} \ln(1 - \lambda_i) = 0$.

- Restricted log likelihood under H_0 is $\ln L^r = -\frac{T}{2} \ln \left| I_{m\times m} - \hat{\Lambda}_{m\times m} \right|$

  Unrestricted log likelihood is $\ln L = -\frac{T}{2} \ln \left| I_{n\times n} - \hat{\Lambda}_{n\times n} \right|$

- Likelihood ratio test involves the restricted $\ln L^r$ and unrestricted $\ln L$

  $$-2\ln\left[\frac{L^r}{L}\right] = T\left(\ln\left|I_{m\times m} - \hat{\Lambda}_{m\times m}\right| - \ln\left|I_{n\times n} - \hat{\Lambda}_{n\times n}\right|\right) = -T\sum_{i=m+1}^{n} \ln(1 - \hat{\lambda}_i)$$

  or $-(T-p)\sum_{i=m+1}^{n} \ln(1 - \hat{\lambda}_i)$ which is positive as eigenvalues are positive and less than 1

- Test statistic $-T\sum_{i=m+1}^{n} \ln(1 - \hat{\lambda}_i) > 0$ has a non-standard (not chi-sq) distribution. If test statistic is too large, reject H_0, accept H_A

05 Application

- We illustrate the use of the Johansen method with an application to test the Fisher hypothesis. The Fisher hypothesis posits that in the long run, nominal interest rate and inflation rate move together, so real interest rate (nominal interest rate less inflation rate) are cointegrated.
- In the study by Hjalmassson and Pär Österholm (2007), they used monthly data on US short nominal interest rate i_t and CPI inflation rate π_t from January 1974 to October 2006. They found that the null hypothesis of a unit root cannot be rejected for inflation rate.
- However, they find that the nominal interest rate is "near to unit root". (The power of ADF unit root test is not strong when processes are close to unit root, i.e., the test tends to accept unit root when it may not be.) They then set out to estimate the cointegrating rank of the vector (π_t, i_t). Based on Akaike Information Criterion, they used $p - 1 = 10$ distributed lags.
- As there are only two stochastic processes, the number of cointegration, r, is 0, 1, or 2 between the two variables π_t and i_t. Suppose $r = 0$, then $\Pi = 0$ and there is no cointegration. Suppose $r = 2$, then this is the case where all elements of X_t, i.e. both π_t and i_t here are I(0).

05 Application

The p-values are in brackets.

Table 10.7: Johansen Cointegration Tests

Null Hypothesis	J trace-statistic	J max eigenvalue statistic
$r = 0$	22.045	16.402
	(0.028)	(0.042)
$r = 1$	5.642	5.642
	(0.220)	(0.220)

- Suppose $r = 1$, then $i_t - b\pi_t$ is stationary (ignoring constant and trend) for some $b \neq 0$, i.e. there exists a cointegrating vector $\beta^T = (1\,,\text{-}b)$. This is then consistent with the Fisher hypothesis.
- Using Johansen trace and max eigenvalue tests, the study found one cointegrating vector, i.e. $r = 1$, cannot be rejected, where $\Pi = \alpha\beta^T$ and cointegrating vector β^T has dimension $r \times n$ or 1×2 in this case. Usually both the Johansen trace test and the maximum eigenvalue test proceed from a null of $r = 0$ upward to a null of $r = 1$, and so on, until the null is not rejected. The two test results should be consistent and are complementary.
- Their Johansen tests showed that the null of $H_0 : r = 0$ is rejected for trace test (i.e. alternative is $r > 0$) and rejected for maximum eigenvalue test (i.e. alternative is $r = 1$). Next, their Johansen tests showed that the null of $H_0 : r = 1$ is not rejected for trace test (i.e. alternative $r > 1$ is not acceptable) and not rejected for maximum eigenvalue test (i.e. alternative is $r = 2$ is not acceptable). Hence the test conclusion is that $r = 1$. However, their study indicated that when nominal interest rate is nearly I(1) but could be I(0), the cointegrating test may have low power.

Practice Exercise (not graded)

+PPP.ipynb
POUND_USD.csv
US_CPI.csv
UK_CPI.csv