19) Integration, Cointegration, and Stationarity

Vitor Kamada

June 2018

Reference

Tables, Graphics, and Figures from

https://www.quantopian.com/lectures

Lecture 43 Integration, Cointegration, and Stationarity

Stationary Stochastic Process

$\{x_t : t = 1, 2, ...\}$ is stationary if for every collection of time indices $1 \leq t_1 < t_2 < ... < t_m$, the joint distribution of $(x_{t_1}, x_{t_2}, ..., x_{t_m})$ is the same as the joint distribution of $(x_{t_{1+h}}, x_{t_2+h}, ..., x_{t_{m+h}})$ for all integers $h \geq 1$

- No restrictions on how x_t and x_{t-1} are related to one another
- Any correlation between adjacent terms is the same across all time periods

Covariance Stationary Process

A stochastic process $\{x_t : t = 1, 2, ...\}$ with a finite second moment $[E(x_t^2) < \infty]$ is covariance stationary if

(i) $E(x_t)$ is constant;

(ii) $Var(x_t)$ is constant; and

(iii) for any t, $h \geq 1$, $Cov(x_t, x_{t+h})$ depends only on h and not on t

Weakly Dependent Time Series

A stationary time series process $\{x_t : t = 1, 2, \dots\}$ is **weakly dependent** if x_t and x_{t+h} are "almost independent" as h increases without bound

$$Corr(x_t, x_{t+h}) \rightarrow 0 \text{ as } h \rightarrow \infty$$

LLN and CLT require stationarity and some form of weak dependence

Autoregressive Process of Order One [AR(1)]

$$y_t = \rho_1 y_{t-1} + e_t, \qquad t = 1, 2, ...$$

e_t is an independent and identically distributed (iid) sequence with zero mean and variance σ_e^2

AR(1) **is weakly dependent** if $|\rho_1| < 1$

$$Var(y_t) = \rho_1^2 Var(y_{t-1}) + Var(e_t)$$

$$\sigma_y^2 = \rho_1^2 \sigma_y^2 + \sigma_e^2$$

$$\sigma_y^2 = \frac{\sigma_e^2}{1-\rho_1^2}$$

AR(1) **is weakly dependent** if $|\rho_1| < 1$

$$y_{t+h} = \rho_1 y_{t+h-1} + e_{t+h} = \rho_1(\rho_1 y_{t+h-2} + e_{t+h-1}) + e_{t+h}$$

$$\rho_1^2 y_{t+h-2} + \rho_1 e_{t+h-1} + e_{t+h}$$

$$\rho_1^2 y_t + \rho_1^{h-1} e_{t+1} + \ldots + \rho_1 e_{t+h-1} + e_{t+h}$$

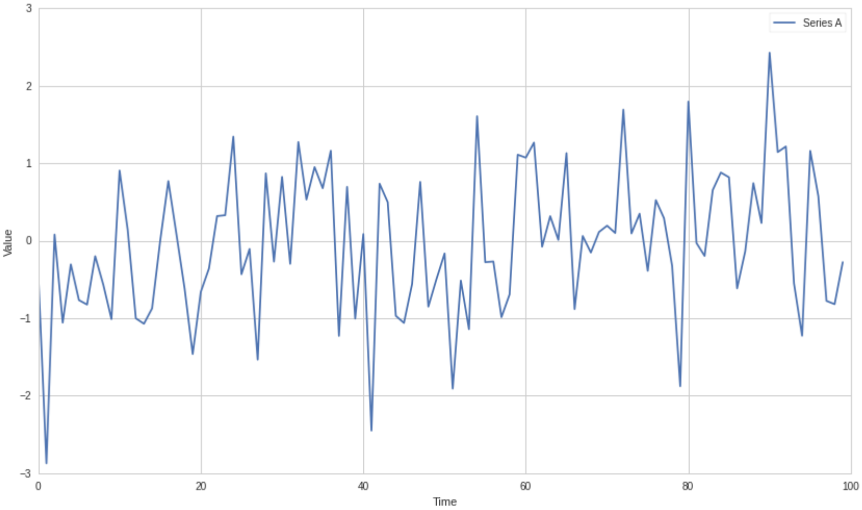
$$Cov(y_t, y_{t+h}) = E(y_t, y_{t+h})$$

$$\rho_1^h E(y_t^2) + \rho_1^{h-1} E(y_t e_{t+1}) + \ldots + E(y_t e_{t+h})$$

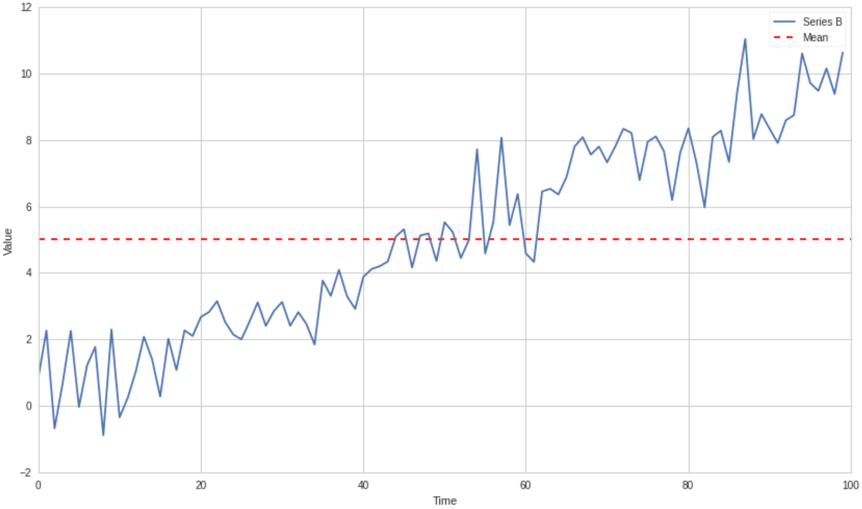
$$\rho_1^h E(y_t^2) = \rho_1^h \sigma_y^2$$

$$Corr(y_t, y_{t+h}) = \frac{Cov(y_t, y_{t+h})}{\sigma_y \sigma_y} = \rho_1^h$$

Stationary Series A $\sim N(0,1)$



Non-Stationary Series B $\sim N(0.1t, 1)$



Dickey and Fuller Test (1979)

$$y_t = \alpha + \rho y_{t-1} + e_t$$

$$H_0 : \rho = 1 \text{ vs } H_a : \rho < 1$$

Augmented Dickey-Fuller Test

Add: $y_{t-2}, y_{t-3} \ldots$

Or/and add Time Trend

from statsmodels.tsa.stattools import adfuller

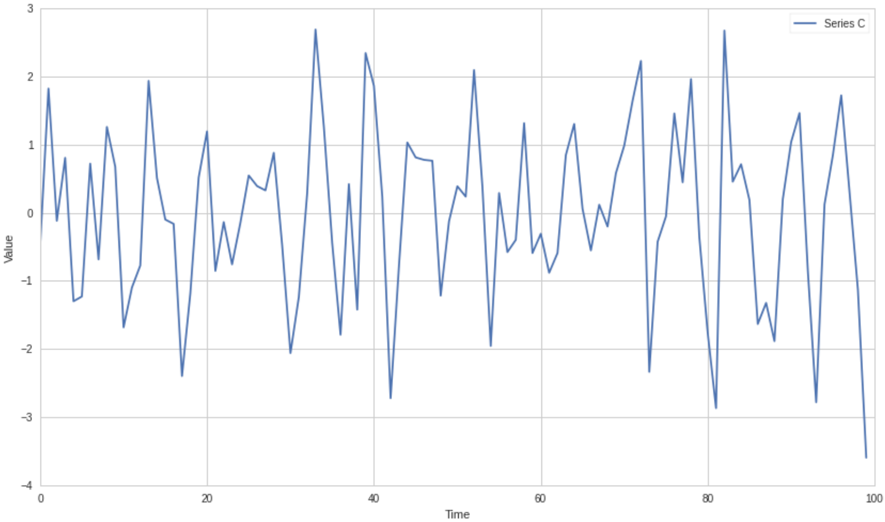
Testing for Stationarity

```
def check_for_stationarity(X, cutoff=0.01):
    # H_0 in adfuller is unit root exists (non-stationary)
    # We must observe significant p-value to convince ourselves that the series
 is stationary
    pvalue = adfuller(X)[1]
    if pvalue < cutoff:
        print 'p-value = ' + str(pvalue) + ' The series ' + X.name +' is likely
 stationary.'
        return True
    else:
        print 'p-value = ' + str(pvalue) + ' The series ' + X.name +' is likely
 non-stationary.'
        return False
```

```
check_for_stationarity(A);
check_for_stationarity(B);
```

```
p-value = 0.000498500723545 The series A is likely stationary.
p-value = 0.948244716942 The series B is likely non-stationary.
```

Series C $\sim N(sin(t), 1)$



p-value $= 0.21959$

Order of Integration: $I(i)$

$$Y_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j} + \eta_t$$

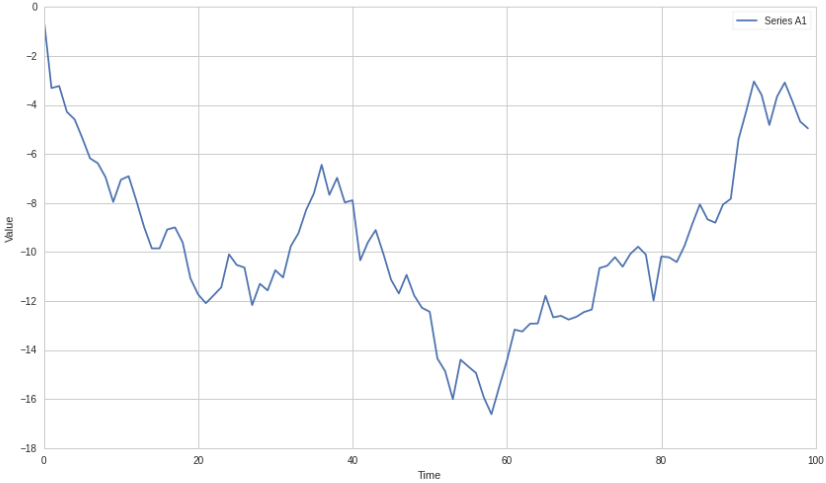
$$I(0) \text{ means } \sum_{k=0}^{\infty} |b_k|^2 < \infty$$

Autocorrelation decays quickly

Stationarity $\rightarrow I(0)$

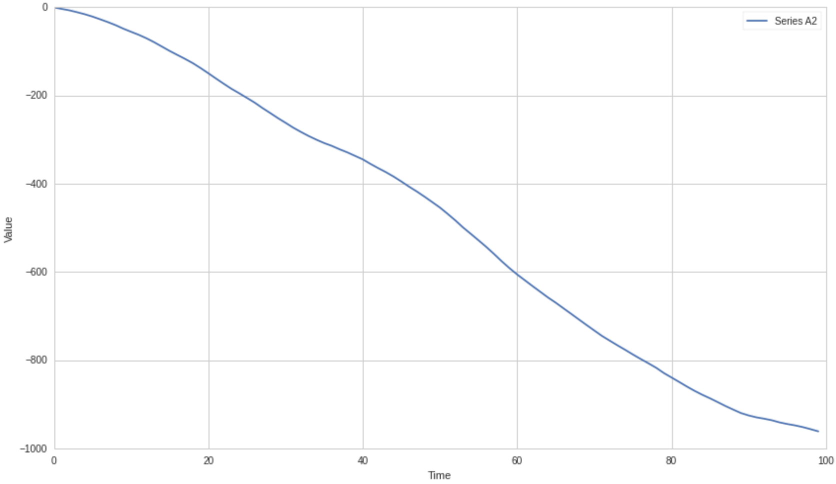
Series A1 is $I(1)$

$$A1 = \text{np.cumsum}(A)$$



Series A2 is $I(2)$

$$A2 = \text{np.cumsum}(A1)$$



Lag Operator (L)

$$LX_t = X_{t-1}$$

$$(1 - L)X_t = X_t - X_{t-1} = \Delta X$$

If X_t is $I(1)$ then

$X_t - X_{t-1}$ is $I(0)$

symbol_list = ['MSFT']

```
prices = get_pricing(symbol_list, fields=['price']
                     , start_date='2014-01-01'
                     , end_date='2015-01-01')['price']
prices.columns = map(lambda x: x.symbol, prices.columns)
X = prices['MSFT']
check_for_stationarity(X);
```

p-value = 0.666326790934

The series MSFT is likely non-stationary.

plt.plot(X.index, X.values)

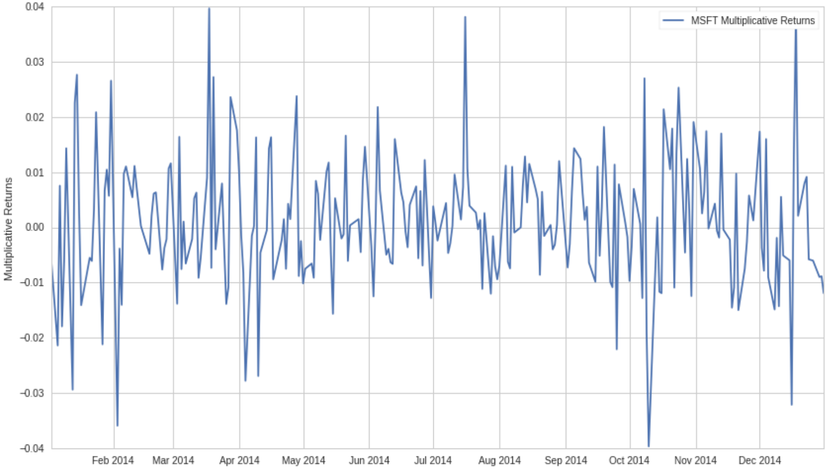
plt.ylabel('Price')

plt.legend([X.name]);

X1 = X.diff()[1:]

$$\text{p-value} = 1.48184901469\text{e-}28$$

X1 = X.pct_change()[1:]



p-value = 8.05657888734e-29

Cointegration

$$(X_1, X_2, ...X_K) \textbf{ are } I(1)$$

A set of time series is Cointegrated, if some linear combination of them is $I(0)$

$$\text{If } 2X_1 + X_2 \text{ is } I(0)\text{, then}$$

they are Cointegrated

Testing for Cointegration

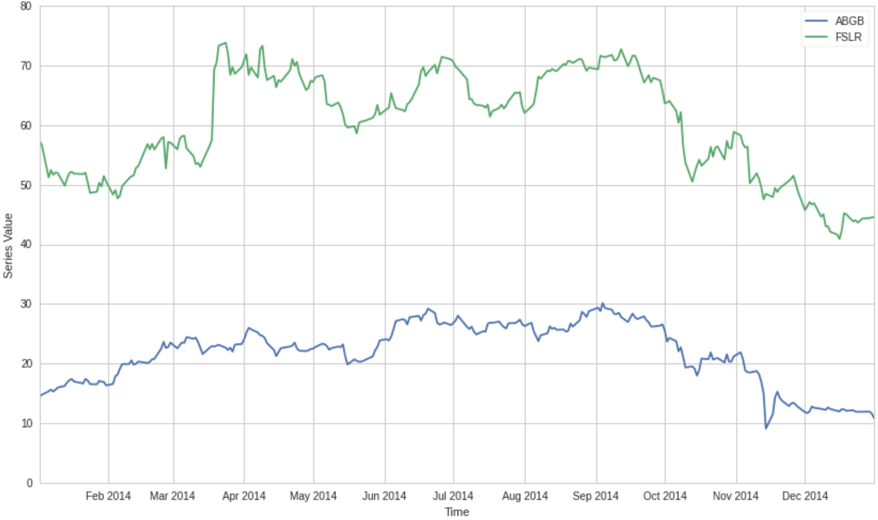
$$X_2 = \alpha + \beta X_1 + \epsilon$$

The combination:

$$X_2 - \beta X_1 = \alpha + \epsilon$$

should be stationary

symbol_list = ['ABGB', 'FSLR']



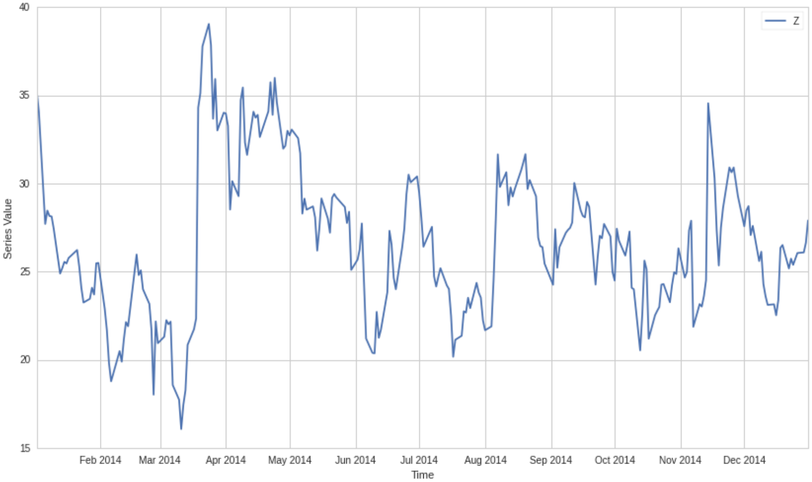
OLS to compute β

```
X1 = sm.add_constant(X1)
results = sm.OLS(X2, X1).fit()
X1 = X1[symbol_list[0]]
results.params
```

const	26.609769
ABGB	1.536686

```
b = results.params[symbol_list[0]]
Z = X2 - b * X1
```

check_for_stationarity(Z)



p-value = 0.00097294855