Modelling A Mean-Reverting Portfolio using the OU Process

Himanshu Babbar
University of North Carolina, Charlotte

1 Abstract

The Ornstein-Uhlenbeck (OU) Process (Vasicek Model in particular) is often used to model the mean-reverting behaviour of the interest rates. Since, a long-short portfolio of a pair of co-integrated risky assets follow a mean-reverting behaviour, we can use an OU Process to model that behaviour. In this paper I model a mean-reverting portfolio of a pair of risky assets using an OU model.

2 Introduction

This paper implements the methodology proposed by Tim Leung et. al. in [1] for modelling a mean-reverting portfolio of a pair of risky assets using the OU process. The portfolio used in this paper is a pair of exchange traded funds (ETFs), viz., iShares Global Energy (IXC) and United States Oil Fund (USO). The parameters for the OU process are estimated using the Maximum Likelihood Estimation (MLE) method.

For the purpose of estimating parameters, historical price data from January-2011 to April-2019 was considered and divided the data into three periods, viz.*Period 1:* January 2011 to December 2014; *Period 2:* January 2015 to April 2019; *Period 3:* January 2011 to April 2019.

The results show that the estimated parameters and the average log-likelihood values for the three periods are very close, and thus, it can be established that the portfolio fits well to the OU model.

3 Methodology

3.1 Theory

The value of a long-short portfolio of a pair of risky assets, at any time $t \geq 0$ is given as:

$$X_t = \alpha S_t^1 - \beta S_t^2$$

where, α is the number of shares of held in risky asset S^1 and β is the number of shares shorted in risky asset S^2 and, S_t^1 and S_t^2 are the prices of asset 1 and asset 2 respectively at time $t \geq 0$. The pair is selected in a way that they form a mean-reverting portfolio. Here, it has been made sure that IXC and USO are co-integrated, and hence, the pair forms a mean-reverting-portfolio.

The level of mean reversion can be improved by adjusting the ratio between α and β, and every pair of $(\alpha\ , \beta)$ results in portfolio values $(x_i^{\alpha,\beta})_{i=0,1,2,\ldots,n}$ realized over an n-day period. An OU Process can be used to fit the observed portfolio values and the parameters can be obtained using the method of Maximum Likelihood Estimation (MLE).

Considering a fixed probability space (Ω, F, P), with historical probability measure P, the SDE for an OU Process is given as:

$$\mathrm{d}X_t = \mu(\theta - X_t)\mathrm{d}t + \sigma \mathrm{d}B_t$$

where,
θ is the long-term mean level.
μ is the speed of reversion.
σ is the instantaneous volatility.
Also, $\mu\ , \sigma > 0\ , \theta \in R$.

Under the OU model, given $X_{t_i-1} = x_{i-1}$, the value for X_{t_i} is given as:

$$X_{t_i} = x_{i-1}e^{-\mu\Delta t} + \theta(1 - e^{-\mu\Delta t}) + \sigma e^{-\mu\Delta t}\int_{t_{i-1}}^{t_i} e^{\mu s}\mathrm{d}B_s$$

where, $\Delta t = t_i - t_{i-1}$. Thus, the conditional probability density of X_{t_i}, at time t_i, given $X_{t_i-1} = x_{i-1}$, is:

$$f(x_i|x_{i-1};\theta,\mu,\sigma) = \frac{1}{\sqrt{2\pi\hat{\sigma}}}exp\left(-\frac{(x_i - x_{i-1}e^{-\mu\Delta t} - \theta(1 - e^{-\mu\Delta t}))^2}{2\hat{\sigma}^2}\right)$$

where,

$$\hat{\sigma}^2 = \sigma^2 \frac{1 - e^{-2\mu\Delta t}}{2\mu}$$

Given the observed portfolio values $(x_i^{\alpha,\beta})_{i=0,1,2,\ldots,n}$, the average log-likelihood function is defined as:

$$\begin{aligned}
&\ell(\theta, \mu, \sigma | x_0^{\alpha,\beta}, x_1^{\alpha,\beta}, \ldots, x_n^{\alpha,\beta}) \\
&= \frac{1}{n} \sum_{i=1}^{n} f(x_i | x_{i-1}; \theta, \mu, \sigma) \\
&= -\frac{1}{2} ln(2\pi) - ln(\hat{\sigma}) - \frac{1}{2n\hat{\sigma}^2} \sum_{i=1}^{n} [x_i - x_{i-1} e^{-\mu\Delta t} - \theta(1 - e^{-\mu\Delta t})]^2
\end{aligned}$$

The maximum of this function can be found where the partial derivatives with respect to every parameter is zero, and this leads to the following,

$$\begin{aligned}
\frac{\partial \ell(\theta, \mu, \sigma)}{\partial \theta} &= \frac{(1 - e^{-\mu\Delta t})}{\hat{\sigma}^2} \sum_{i=1}^{n} [x_i^{\alpha,\beta} - x_{i-1}^{\alpha,\beta} e^{-\mu\Delta t} - \theta(1 - e^{-\mu\Delta t})] = 0 \\
\theta &= \frac{\sum_{i=1}^{n} [x_i^{\alpha,\beta} - x_{i-1}^{\alpha,\beta} e^{-\mu\Delta t}]}{n(1 - e^{-\mu\Delta t})}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell(\theta, \mu, \sigma)}{\partial \mu} &= -\frac{\Delta t e^{-\mu\Delta t}}{\hat{\sigma}^2} \sum_{i=1}^{n} [(x_i^{\alpha,\beta} - \theta)(x_{i-1}^{\alpha,\beta} - \theta) - e^{-\mu\Delta t}(x_{i-1}^{\alpha,\beta} - \theta)^2] = 0 \\
\mu &= -\frac{1}{\Delta t} ln \left(\frac{\sum_{i=1}^{n} [(x_i^{\alpha,\beta} - \theta)(x_{i-1}^{\alpha,\beta} - \theta)]}{\sum_{i=1}^{n} (x_{i-1}^{\alpha,\beta} - \theta)^2} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell(\theta, \mu, \sigma)}{\partial \sigma} &= -\frac{1}{\sigma} + \frac{2\mu}{n(1 - e^{-2\mu\Delta t})} \sum_{i=1}^{n} [x_i^{\alpha,\beta} - x_{i-1}^{\alpha,\beta} e^{-\mu\Delta t} - \theta(1 - e^{-\mu\Delta t})]^2 = 0 \\
\sigma^2 &= \frac{2\mu}{n(1 - e^{-2\mu\Delta t})} \sum_{i=1}^{n} [x_i^{\alpha,\beta} - x_{i-1}^{\alpha,\beta} e^{-\mu\Delta t} - \theta(1 - e^{-\mu\Delta t})]^2
\end{aligned}$$

As observed, the solution for each parameter depends on others. However, θ and μ are independent of σ, and hence, can be solved by substituting one into the other. In order to obtain a solution, let us define the following:

$$
\begin{aligned}
X_x &= \sum_{i=1}^{n} x_{i-1}^{\alpha,\beta}, \\
X_y &= \sum_{i=1}^{n} x_{i}^{\alpha,\beta}, \\
X_{xx} &= \sum_{i=1}^{n} \left(x_{i-1}^{\alpha,\beta}\right)^2, \\
X_{xy} &= \sum_{i=1}^{n} \left(x_{i-1}^{\alpha,\beta}\right)\left(x_{i}^{\alpha,\beta}\right), \\
X_{yy} &= \sum_{i=1}^{n} \left(x_{i}^{\alpha,\beta}\right)^2
\end{aligned}
$$

Using the above definitions, we get:

$$
\theta = \frac{X_y - X_x e^{-\mu\Delta t}}{n(1 - e^{-\mu\Delta t})},
$$

$$
\mu = -\frac{1}{\Delta t} ln \left(\frac{X_{xy} - \theta X_x - \theta X_y + n(\theta)^2}{X_{xx} - 2\theta X_x + n(\theta)^2} \right)
$$

Substituting μ into θ, we get,

$$n\theta = \frac{X_y - \left(\frac{X_{xy}-\theta X_x-\theta X_y+n(\theta)^2}{X_{xx}-2\theta X_x+n(\theta)^2}\right) X_x}{(1 - \left(\frac{X_{xy}-\theta X_x-\theta X_y+n(\theta)^2}{X_{xx}-2\theta X_x+n(\theta)^2}\right))},$$

$$n\theta = \frac{X_y(X_{xx} - 2\theta X_x + n(\theta)^2) - (X_{xy} - \theta X_x - \theta X_y + n(\theta)^2)X_x}{(X_{xx} - 2\theta X_x + n(\theta)^2) - (X_{xy} - \theta X_x - \theta X_y + n(\theta)^2)},$$

$$n\theta = \frac{(X_y X_{xx} - X_x X_{xy}) + \theta(X_x^2 - X_x X_y) + \theta^2 n(X_y - X_x)}{(X_{xx} - X_{xy}) + \theta(X_y - X_x)},$$

$$n\theta(X_{xx} - X_{xy}) - \theta(X_x^2 - X_x X_y) = X_y X_{xx} - X_x X_{xy},$$

$$\theta = \frac{X_y X_{xx} - X_x X_{xy}}{n(X_{xx} - X_{xy}) - (X_x^2 - X_x X_y)}$$

Thus, the optimal parameters under the OU model are given as:

$$\theta^* = \frac{X_y X_{xx} - X_x X_{xy}}{n(X_{xx} - X_{xy}) - (X_x^2 - X_x X_y)}$$

$$\mu^* = -\frac{1}{\Delta t} ln \left(\frac{X_{xy} - \theta^* X_x - \theta^* X_y + n(\theta^*)^2}{X_{xx} - 2\theta^* X_x + n(\theta^*)^2} \right)$$

$$\begin{aligned}(\sigma^*)^2 = \frac{2\mu^*}{n(1 - e^{-2\mu^* \Delta t})} &(X_{yy} - 2e^{-\mu^* \Delta t} X_{xy} + e^{-2\mu^* \Delta t} X_{xx} \\ &- 2\theta^*(1 - e^{-\mu^* \Delta t})(X_y - e^{-\mu^* \Delta t} X_x) + n(\theta^*)^2(1 - e^{-2\mu^* \Delta t})^2)\end{aligned}$$

The maximized average log-likelihood is obtained by using the above optimal parameters in the average log-likelihood function, and is denoted by $\hat{\ell}(\theta^*, \mu^*, \sigma^*)$, and for any α, the optimal value of β, that enhances the level of mean-reversion is denoted by β^* , and is given as:

$$\beta^* = arg \max_{\beta} \hat{\ell}(\theta^*, \mu^*, \sigma^* | x_0^{\alpha,\beta}, x_1^{\alpha,\beta}, \ldots, x_n^{\alpha,\beta}).$$

3.2 Data Used

The theory described in section 3.1 is applied to a pair of Crude Oil ETFs, IXC and USO. IXC is an ETF that tracks the performance of the S&P Global Energy Index, and USO tracks the changes of the price of WTI Crude Oil delivered to Cushing, Oklahoma, as measured by changes in percentage terms of the price of the WTI Crude Oil futures contract on the NYMEX.

Table 1 below shows the values of t-statitic for the Engle-Granger co-integration test obtained for the three periods.

Table 1: t-statistic values using Engle-Granger test for co-integration.

Year	**t-stat**
2011-2014	-7.43
2015-2019	-31.09
2011-2019	-11.91

Thus, it can be confirmed that the pair of these two ETFs form a mean-reverting portfolio.

In order to create a portfolio, IXC is assumed to be S^1 and USO to be S^2. A sum of A dollar(s) is invested in IXC so that we have a long position in IXC with $\alpha = A/S_0^1$ shares. Simultaneously, a short position of $\beta = B/S_0^2$ shares is created in USO.

3.3 Results

The data from January 2011 to April 2019 was divided into three periods.
Period 1: January 2011 to December 2014
Period 2: January 2015 to April 2019
Period 3: January 2011 to April 2019.

For each period a number of different portfolios were created with $A = 1$ and $B/A = 0.001, 0.002, \ldots, 1$ so that the initial portfolio value is non-negative. Then, for each portfolio optimal parameters for the OU model, $\theta^*, \mu^*, \sigma^*$, were calculated using the MLE method as described in section 3.1 and the corresponding value for the maximized average log-likelihood was obtained.

The cash level B and the optimal parameters, $\theta^*, \mu^*, \sigma^*$, of the portfolio with the maximum value of maximized average log-likelihood among all the portfolios is considered to be the one that enhances the mean-reversion level. Figures 1, 2 and 3 below show the average log-likelihood values against the cash amount B for all the three periods mentioned above. The cash level B^* that results in maximum maximized average log-likelihood gives us $\beta^* = B^*/S_0^2$, and the corresponding parameters for the OU model.

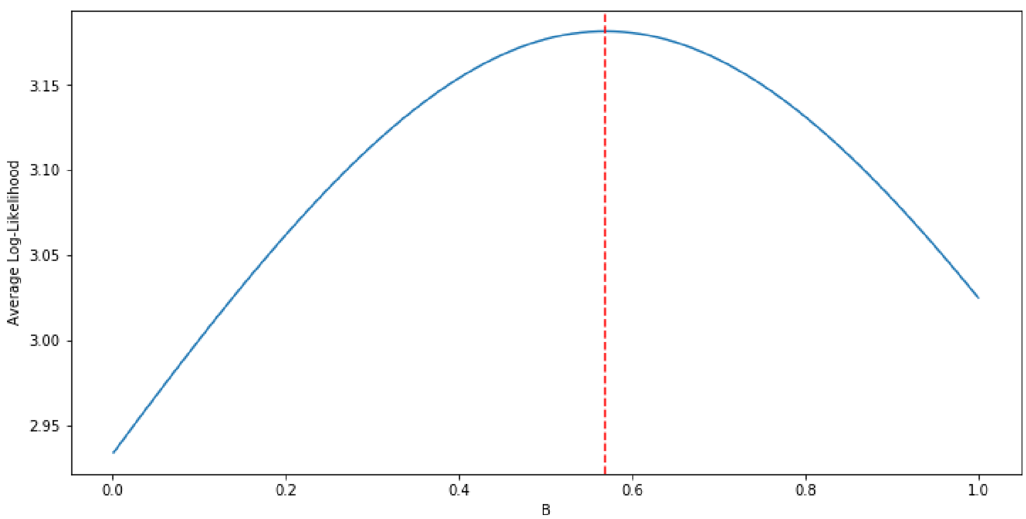

Figure 1: Average log-likelihood against B for Period 1: 2011-2014.

Table 2. below shows the values of the optimal parameters and corresponding value of the maximized average log-likelihood.

Figures 4, 5 and 6 show the portfolio values as obtained based on historical price data for all the three periods. For each portfolio that we invest \$1 in IXC and $-\$B^*$ in USO.

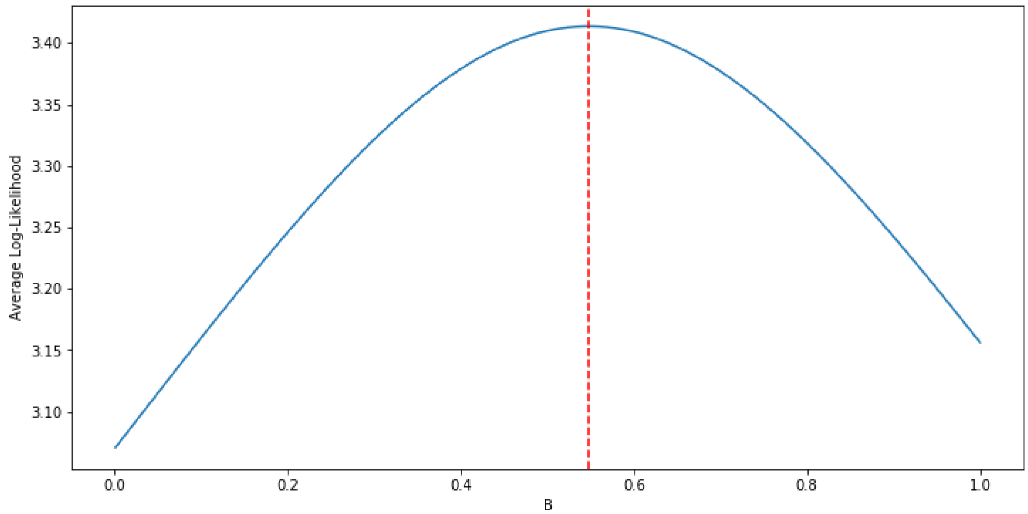

Figure 2: Average log-likelihood against B for Period 2: 2015-2019.

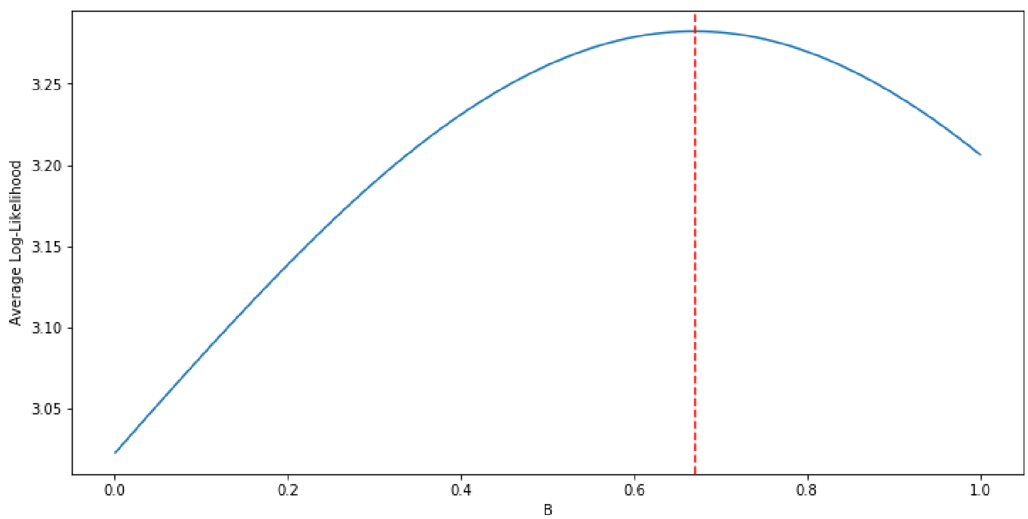

Figure 3: Average log-likelihood against B for Period 3: 2011-2019.

Table 2: MLE Estimates.

	θ^*	μ^*	σ^*	$\hat{\ell}$
2011-2014	0.5426	1.9422	0.1601	3.1812
2015-2019	0.5591	1.9563	0.1269	3.4138
2011-2019	0.5773	0.7904	0.1444	3.2824

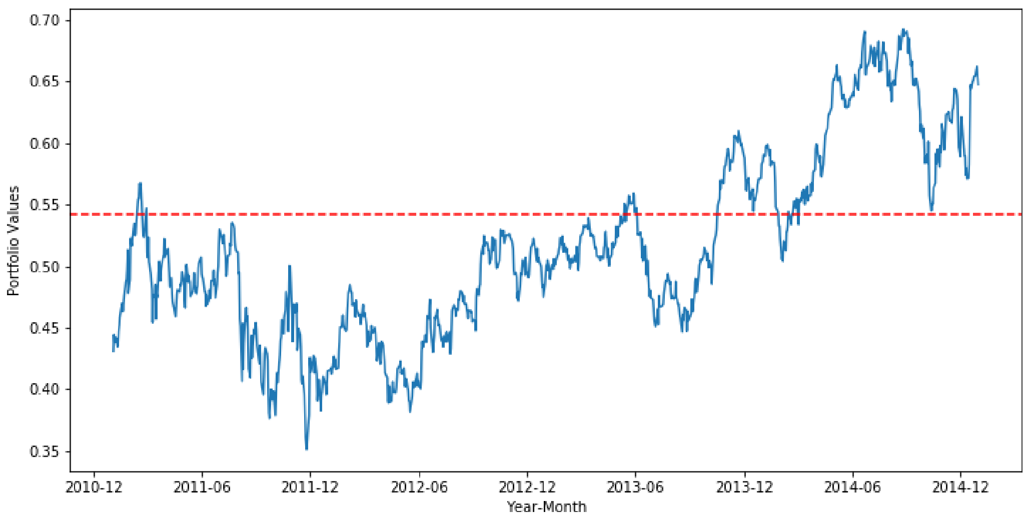

Figure 4: Portfolio values for Period 1: 2011-2014, generated by longing \$1 in IXC and shorting \$0.569 in USO.

4 Conclusion

The results in Table 2. show that the estimated parameters and the average log-likelihood values are very close for the three periods. Thus, it can be concluded that the portfolio fits well to the OU model.

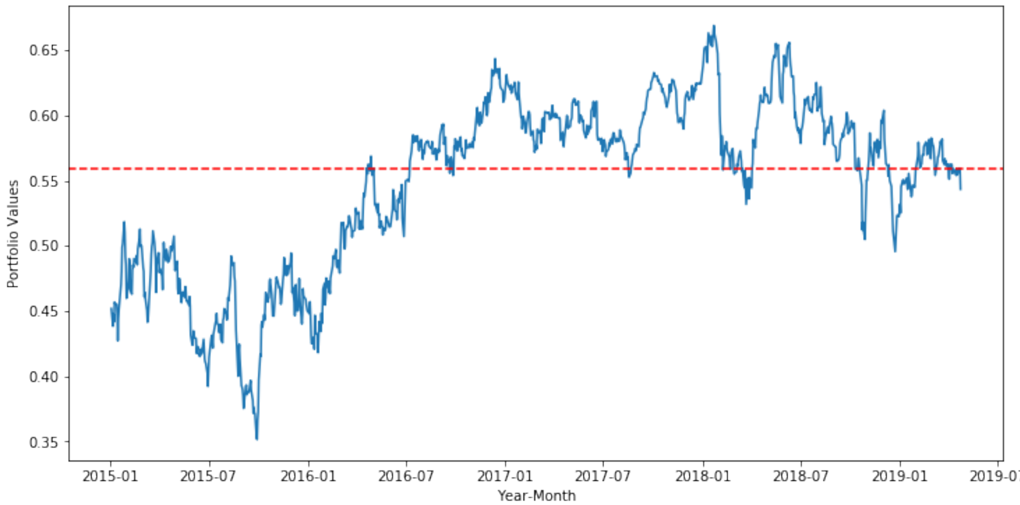

Figure 5: Portfolio values for Period 1: 2011-2014, generated by longing $1 in IXC and shorting $0.548 in USO.

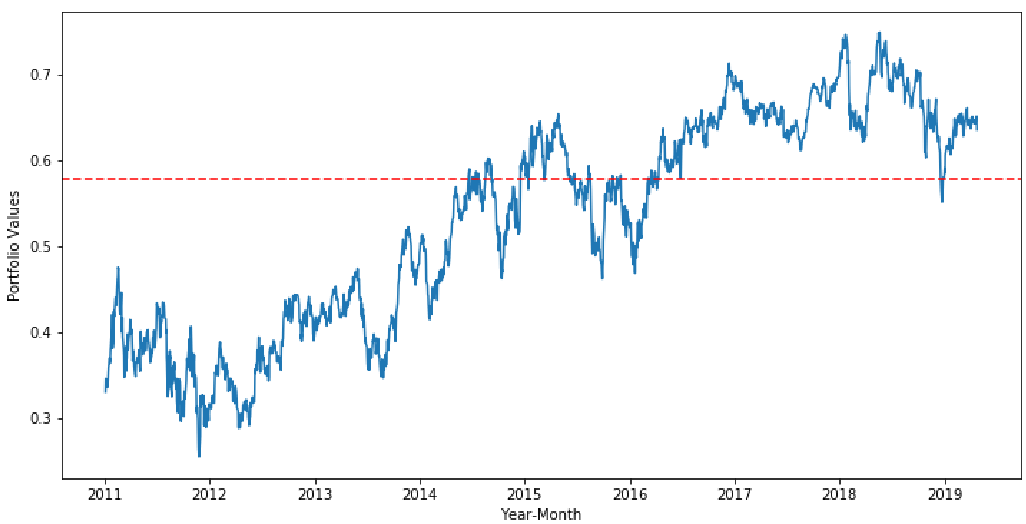

Figure 6: Portfolio values for Period 1: 2011-2019, generated by longing $1 in IXC and shorting $0.67 in USO.

References

[1] T. S. Leung and X. Li, *Optimal Mean Reversion Trading: Mathematical Analysis and Practical Applications*. World Scientific Publishing Company Pte Limited, 2015.

A Python Code

```
# Prepare Data
import pandas as pd
import numpy as np
import os

import warnings
warnings.filterwarnings("ignore")

def get_filtered_data(data, start_yr, end_yr):
    filt_data = data.copy()
    filt_data['Date'] = pd.to_datetime(filt_data['Date'], format='%d/%m/%Y')
    filt_data = filt_data.sort_values(['Date'])
    filt_data = filt_data[(filt_data['Date'].dt.year >= start_yr)
& (filt_data['Date'].dt.year <= end_yr)]
    filt_data = filt_data.set_index(['Date'])
    return filt_data

def get_period_data(etf1, etf2, start_yr, end_yr):
    filtData1 = get_filtered_data(etf1, start_yr, end_yr)
    filtData2 = get_filtered_data(etf2, start_yr, end_yr)

    data = pd.concat([filtData1['Close'], filtData2['Close']], axis=1).dropna()
    data.columns = ['etf1', 'etf2']
    data['etf1_rets'] = np.log(data['etf1']/data['etf1'].shift(1))
    data['etf2_rets'] = np.log(data['etf2']/data['etf2'].shift(1))

    return data

path = '../OilFunds'
```

```
etf1 = pd.read_csv(path+os.sep+'IXC.csv')
etf2 = pd.read_csv(path+os.sep+'USO.csv')

period_1_data = get_period_data(etf1, etf2, 2011, 2014)
period_2_data = get_period_data(etf1, etf2, 2015, 2019)
period_3_data = get_period_data(etf1, etf2, 2011, 2019)

# Check for co-integration
from statsmodels.tsa.stattools import coint

t_stat_1 = coint(period_1_data['etf1_rets'].dropna().values
                 ,period_1_data['etf2_rets'].dropna().values)[0]
t_stat_2 = coint(period_2_data['etf1_rets'].dropna().values
                 ,period_2_data['etf2_rets'].dropna().values)[0]
t_stat_3 = coint(period_3_data['etf1_rets'].dropna().values
                 ,period_3_data['etf2_rets'].dropna().values)[0]
print('Cointegration t-Statistic')
print('2011-2014:', t_stat_1)
print('2015-2019:', t_stat_2)
print('2011-2019:', t_stat_3)

# Find OU Parameters
import matplotlib.pyplot as plt

def get_OU_params(port):
    X_x = np.sum(port.values[:-1])
    X_y = np.sum(port.values[1:])
    X_xx = np.sum((port.values[:-1])**2)
    X_xy = np.sum((port.values[:-1])*(port.values[1:]))
    X_yy = np.sum(port.values[1:]**2)

    n = len(port)
    dt = 1/252

    num = (X_y*X_xx) - (X_x*X_xy)
    den = n*(X_xx-X_xy) - ((X_x**2)-(X_x*X_y))
    theta = (num)/(den)
```

```
    num = (X_xy - (theta*X_x) - (theta*X_y) + (n*(theta**2)))
    den = X_xx - (2*theta*X_x) + (n*(theta**2))
    mu = -(1/dt)*(np.log(num/den))

    alpha = np.exp(-mu*dt)
    t1 = (2*mu)/(n*(1-(alpha**2)))
    t2 = X_yy - (2*alpha*X_xy) + ((alpha**2)*X_xx)
    t3 = 2*theta*(1-alpha)*(X_y - alpha*X_x)
    t4 = n*(theta**2)*(1-alpha)**2
    sigma = np.sqrt(t1*(t2-t3+t4))

    return theta, mu, sigma

def avg_likelihood(params, port):
    n = len(port)
    dt = 1/252
    theta, mu, sigma = params

    sigma_tilde = (sigma**2)*(1-np.exp(-2*mu*dt))/(2*mu)
    sq_term = np.sum((port.values[1:] - (port.values[:-1]*(np.exp(-mu*dt)))\
                     - theta*(1-(np.exp(-mu*dt))))**2)
    l = -(0.5*np.log(2*np.pi)) - np.log(np.sqrt(sigma_tilde))\
                                        - (1/(2*n*sigma_tilde))*sq_term

    return l

def run_test(data):
    A = 1
    B = np.arange(0.001,1.001,0.001)
    l_hat = [0]*len(B)
    th = [0]*len(B)
    max_index = None
    ou_params = [None]*len(B)
    th = [None]*len(B)
    for i in range(len(B)):
        alpha = A/data.etf1.iloc[0]
        beta = B[i]/data.etf2.iloc[0]
        port = alpha*data.etf1-beta*data.etf2
```

```
        ou_params[i] = get_OU_params(port)
        th[i] = ou_params[i][0]
        l_hat[i] = avg_likelihood(ou_params[i], port)
        if max_index is None or l_hat[i] > l_hat[max_index]:
            max_index = i

    plt.rcParams['figure.figsize'] = [12,6]
    plt.plot(B, l_hat);
    plt.axvline(x=B[max_index], color='red', linestyle='--')
    plt.xlabel('B');
    plt.ylabel('Average Log-Likelihood')

    opt_params = ou_params[max_index]
    opt_alpha = A/data.etf1.iloc[0]
    opt_beta = B[max_index]/data.etf2.iloc[0]
    opt_port = opt_alpha*data.etf1-opt_beta*data.etf2
    return (opt_params, avg_likelihood(opt_params, opt_port),
            A, B[max_index], opt_alpha, opt_beta)

# Find OU Parameters for 2011-2014
params_1, max_avg_likelihood_1, A_1, B_1, alpha_1, beta_1\
                                        = run_test(period_1_data)
print('Results for 2011-2014')
print('Optimal Parameters:', params_1)
print('Maximum Average Likelihood:', max_avg_likelihood_1)
print('Optimal Cash Level for IXC, A:', A_1)
print('Optimal Cash Level for USO, B:', B_1)
print('Optimal Alpha:', alpha_1)
print('Optimal Beta:', beta_1)

# Plot portfolio values for 2011-2014
opt_port_1 = alpha_1*period_1_data.etf1-beta_1*period_1_data.etf2
plt.plot(opt_port_1)
plt.axhline(y=params_1[0], color='red', linestyle='--')
plt.xlabel('Year-Month');
plt.ylabel('Portfolio Values');
```

```
# Find OU Parameters for 2015-2019
params_2, max_avg_likelihood_2, A_2, B_2, alpha_2, beta_2\
                                        = run_test(period_2_data)
print('Results for 2015-2019')
print('Optimal Parameters:', params_2)
print('Maximum Average Likelihood:', max_avg_likelihood_2)
print('Optimal Cash Level for IXC, A:', A_2)
print('Optimal Cash Level for USO, B:', B_2)
print('Optimal Alpha:', alpha_2)
print('Optimal Beta:', beta_2)

# Plot portfolio values for 2015-2019
opt_port_2 = alpha_2*period_2_data.etf1-beta_2*period_2_data.etf2
plt.plot(opt_port_2)
plt.axhline(y=params_2[0], color='red', linestyle='--')
plt.xlabel('Year-Month');
plt.ylabel('Portfolio Values');

# Find OU Parameters for 2011-2019
params_3, max_avg_likelihood_3, A_3, B_3, alpha_3, beta_3\
                                        = run_test(period_3_data)
print('Results for 2011-2019')
print('Optimal Parameters:', params_3)
print('Maximum Average Likelihood:', max_avg_likelihood_3)
print('Optimal Cash Level for IXC, A:', A_3)
print('Optimal Cash Level for USO, B:', B_3)
print('Optimal Alpha:', alpha_3)
print('Optimal Beta:', beta_3)

# Plot portfolio values for 2011-2019
opt_port_3 = alpha_3*period_3_data.etf1-beta_3*period_3_data.etf2
plt.plot(opt_port_3)
plt.axhline(y=params_3[0], color='red', linestyle='--')
plt.xlabel('Year-Month');
plt.ylabel('Portfolio Values');
```