

Currencies_1_2014_to_2015_in_sample

November 12, 2018

1 Pair Trading:

1.0.1 This project is to extract potential opportunities of pair trading for various futures.

1.0.2 Data is based on minute prices from Quantopian. This Python code is for Quantopian.

1.0.3 Steps:

1. Compute sample correlation of in-sample future returns.

2. Compute and plot simple price spreads of any two futures for further observations. For example, for futures X and Y, if series of (X's returns - Y's returns) looks like a stable series, then this implies that it is possible for us to generate stable returns by going long a share of X and going short a share of Y.

3. Compute and plot division price spreads of any two futures for further observations.

4. Find pairs with significant (p-value < 0.05) co-integration.

5. If future X and future Y has a significant in-sample co-integration, apply OLS regression model to get coefficients between X's returns and Y's returns during in-sample periods. For example, in-sample period is 2014-2015. Then, get coefficient and intercept between X's returns and Y's returns in 2014-2015, such as ($Y's\ return = 0.1 + 0.5 * X's\ return$).

6. Compute, plot and check residues in validation period. For example, for future X and future Y, compute residues as ($Y's\ return - 0.1 - 0.5 * X's\ return$) in validation period, 2016-2017.

If the residues look like stable series, this further confirms that we may make use of the pair relationships to form a pair trading strategy.

2 Currency Futures

2.1 In-sample: 2014-2015 validation: 2016-2017

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
```

```

from quantopian.research.experimental import continuous_future, history
import statsmodels.tsa.stattools as ts
from statsmodels.tsa.stattools import adfuller
import seaborn as sns

```

```

In [3]: Currencies={
        'currencies': {
            'jpy': continuous_future('JY'),
            'jpy_emiini': continuous_future('JE'),
            'cad': continuous_future('CD'),
            'mxn': continuous_future('ME'),
            'aud': continuous_future('AD'),
            'nzd': continuous_future('NZ'),
            'gbp': continuous_future('BP'),
            'chf': continuous_future('SF'),
            'eur_emiicro': continuous_future('EU'),
            'euro_fx': continuous_future('EC'),
            'euro_fx_emiini': continuous_future('EE')
        }
    }

```

```

In [12]: # Get data of different futures
def get_data(future_dic,start_date, end_date,fields_type='price',frequency_type='minu
    df = pd.DataFrame()
    future_list = future_dic.values()[0]
    for future_name, future_data in future_list.items():

        data = history(future_data, start = start_date, end = end_date, fields = field
        data = pd.DataFrame(data)
        data.columns=[future_name+'_'+frequency_type]
        df = pd.concat([df,data],axis=1)
    return df

# Calculate simple price spread
def spread_subtraction(df,times=1):
    future_name = df.columns

    delta_df = pd.DataFrame()

    for i in range(len(future_name)):
        for j in range(i+1,len(future_name)):
            y_future = future_name[i]
            x_future = future_name[j]

            delta_df[y_future+' - '+x_future] = df[y_future] - times*df[x_future]

    return delta_df

```

```

# Calculate price division
def spread_division(df):
    future_name = df.columns
    delta_df = pd.DataFrame()

    for i in range(len(future_name)):
        for j in range(i+1, len(future_name)):
            y_future = future_name[i]
            x_future = future_name[j]

            delta_df[y_future+' / '+x_future] = df[y_future] / df[x_future]

    return delta_df

# Plot time series
def plotting(df):
    future_name = df.columns
    num = len(future_name)

    for i in future_name:
        plt.figure(figsize=(10,3))
        plt.plot(df[i])
        plt.title(i)
    plt.show()

# Get adf test results to determine whether time series is stationary
def adf_test(df):
    future_name = df.columns
    adf_df = pd.DataFrame()

    for i in future_name:
        adftest = adfuller(df[i])
        adf_df.ix[i, 'Test Statistic'] = adf_test[0]
        adf_df.ix[i, 'p-value'] = adf_test[1]
        adf_df.ix[i, 'Test Statistic'] = adf_test[0]

        for key, value in adftest[4].items():
            adf_df[i, 'Critical Value (%s)'% key] = value

    return adf_test

# Find paris with significant cointegration
def find_cointegrated_paris(df, pvalue_level=0.05):
    future_name = df.columns
    n = len(future_name)
    pvalue_matrix = np.ones((n,n))

```

```

pairs = []

for i in range(len(future_name)):
    for j in range(i+1, len(future_name)):
        y_future = future_name[i]
        x_future = future_name[j]

        result = ts.coint(df[y_future], df[x_future])

        pvalue = result[1]
        pvalue_matrix[i, j] = pvalue

        if pvalue < pvalue_level:
            pairs.append((y_future, x_future))
return pvalue_matrix, pairs

# Get ols parameters
def ols_in_sample(df, pairs):
    future_name = df.columns
    ols_df = pd.DataFrame()

    for i in range(len(future_name)):
        for j in range(i+1, len(future_name)):
            y_future = future_name[i]
            x_future = future_name[j]

            if (y_future, x_future) in pairs:
                reg = sm.add_constant(df[x_future])
                results = sm.OLS(df[y_future], reg).fit()

                name = str(y_future+' vs '+x_future)

                ols_df.loc[name, 'cons'] = results.params[0]
                ols_df.loc[name, 'coef'] = results.params[1]

    return ols_df

# Get residues of validation data based on in-sample ols parameters
def ols_validation(df, ols_df, pairs):
    future_name = df.columns
    ols_validation_df = pd.DataFrame()
    for i in range(len(future_name)):
        for j in range(i+1, len(future_name)):
            y_future = future_name[i]
            x_future = future_name[j]

            if (y_future, x_future) in pairs:

```

```

        name = str(y_future+' vs '+x_future)

        cons = ols_df.loc[name,'cons']
        coef = ols_df.loc[name,'coef']
        ols_validation_df[y_future+' vs '+x_future+'_residues'] = df[y_future-

    return ols_validation_df

```

2.2 1. In-sample Period: 2014-2015 Validation Period: 2016-2017

```
In [5]: future_list = Currencies.values()[0].keys()
```

```
In [6]: data = get_data(Currencies, '2014-01-01', '2015-12-31')
```

2.2.1 1.1 Correlation: 2014-01-01 to 2015-12-31

```
In [7]: correlation =data.corr()
        correlation
```

```
Out [7]:
```

| | eur_emicro__minute | nzd__minute | jpy__minute | \ |
|-----------------------|--------------------|-------------|-------------|---|
| eur_emicro__minute | 1.000000 | 0.871055 | 0.958945 | |
| nzd__minute | 0.871055 | 1.000000 | 0.875685 | |
| jpy__minute | 0.958945 | 0.875685 | 1.000000 | |
| euro_fx__minute | 0.999996 | 0.871235 | 0.958669 | |
| mxn__minute | 0.915263 | 0.943894 | 0.898045 | |
| aud__minute | 0.927362 | 0.956298 | 0.915378 | |
| chf__minute | 0.804660 | 0.764675 | 0.815733 | |
| euro_fx_emini__minute | 0.999971 | 0.871314 | 0.958730 | |
| gbp__minute | 0.931758 | 0.783519 | 0.917068 | |
| jpy_emini__minute | 0.959213 | 0.875957 | 0.999990 | |
| cad__minute | 0.925965 | 0.932713 | 0.886450 | |

| | euro_fx__minute | mxn__minute | aud__minute | chf__minute | \ |
|-----------------------|-----------------|-------------|-------------|-------------|---|
| eur_emicro__minute | 0.999996 | 0.915263 | 0.927362 | 0.804660 | |
| nzd__minute | 0.871235 | 0.943894 | 0.956298 | 0.764675 | |
| jpy__minute | 0.958669 | 0.898045 | 0.915378 | 0.815733 | |
| euro_fx__minute | 1.000000 | 0.915755 | 0.927524 | 0.804652 | |
| mxn__minute | 0.915755 | 1.000000 | 0.980788 | 0.766600 | |
| aud__minute | 0.927524 | 0.980788 | 1.000000 | 0.774723 | |
| chf__minute | 0.804652 | 0.766600 | 0.774723 | 1.000000 | |
| euro_fx_emini__minute | 0.999971 | 0.915594 | 0.927641 | 0.804832 | |
| gbp__minute | 0.931424 | 0.846321 | 0.881833 | 0.792271 | |
| jpy_emini__minute | 0.958941 | 0.898460 | 0.915519 | 0.815446 | |
| cad__minute | 0.926494 | 0.978414 | 0.973001 | 0.752621 | |

| | euro_fx_emini__minute | gbp__minute | jpy_emini__minute | \ |
|--------------------|-----------------------|-------------|-------------------|---|
| eur_emicro__minute | 0.999971 | 0.931758 | 0.959213 | |
| nzd__minute | 0.871314 | 0.783519 | 0.875957 | |
| jpy__minute | 0.958730 | 0.917068 | 0.999990 | |

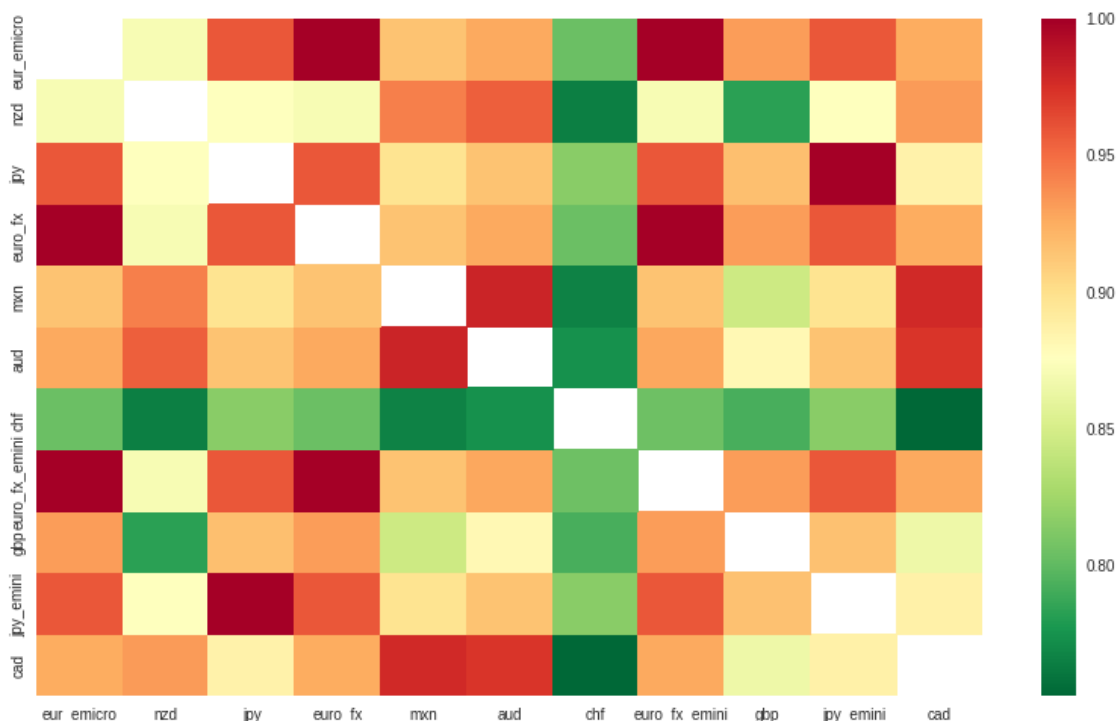
| | | | |
|---------------------|----------|----------|----------|
| euro_fx__minute | 0.999971 | 0.931424 | 0.958941 |
| mxn__minute | 0.915594 | 0.846321 | 0.898460 |
| aud__minute | 0.927641 | 0.881833 | 0.915519 |
| chf__minute | 0.804832 | 0.792271 | 0.815446 |
| euro_fx_emi__minute | 1.000000 | 0.932270 | 0.958997 |
| gbp__minute | 0.932270 | 1.000000 | 0.916705 |
| jpy_emi__minute | 0.958997 | 0.916705 | 1.000000 |
| cad__minute | 0.926645 | 0.865113 | 0.886955 |

| | cad__minute |
|---------------------|-------------|
| eur_emi__minute | 0.925965 |
| nzd__minute | 0.932713 |
| jpy__minute | 0.886450 |
| euro_fx__minute | 0.926494 |
| mxn__minute | 0.978414 |
| aud__minute | 0.973001 |
| chf__minute | 0.752621 |
| euro_fx_emi__minute | 0.926645 |
| gbp__minute | 0.865113 |
| jpy_emi__minute | 0.886955 |
| cad__minute | 1.000000 |

2.2.2 Heatmap of correlation: Red High correlation

In [8]: `sns.heatmap(correlation,xticklabels=future_list, yticklabels=future_list,cmap = 'RdYlGn')`

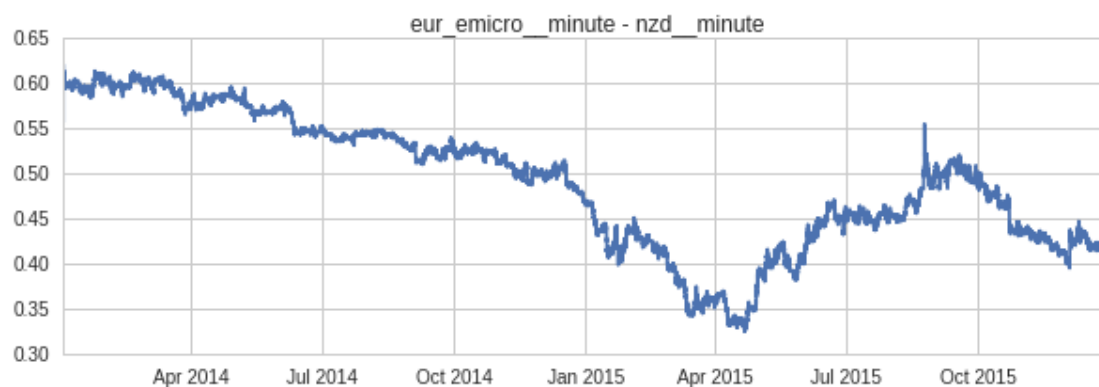
Out[8]: `<matplotlib.axes._subplots.AxesSubplot at 0x7fdc2447fbd0>`

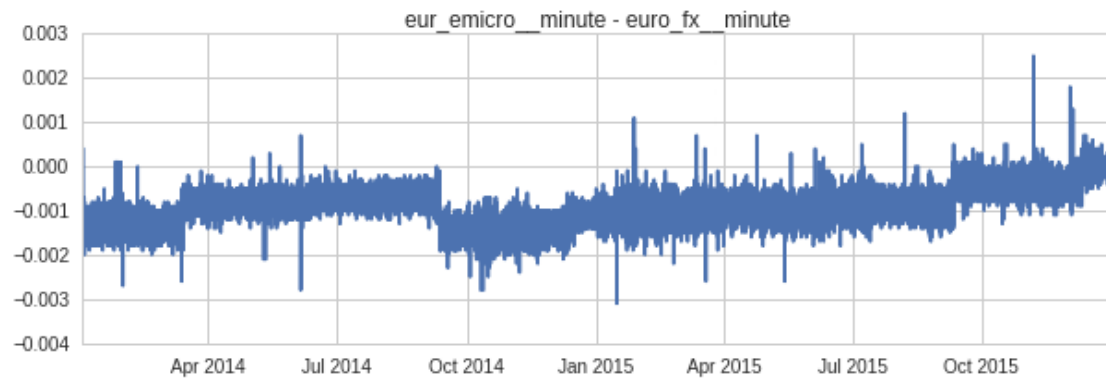


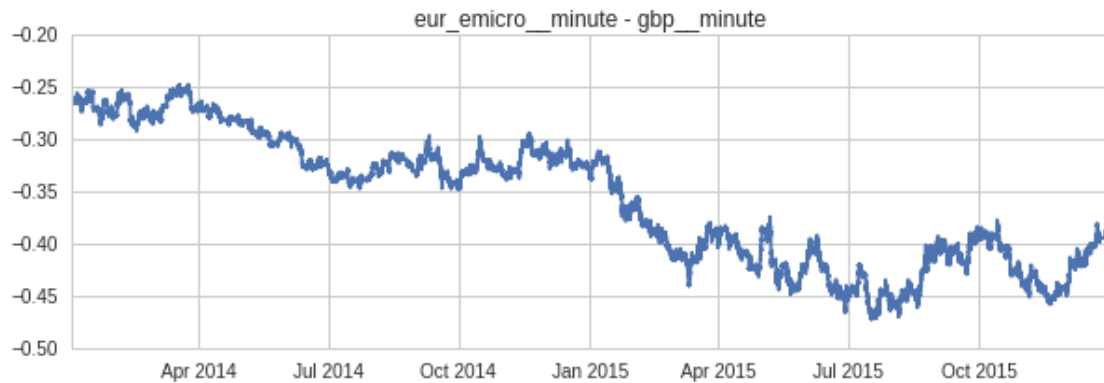
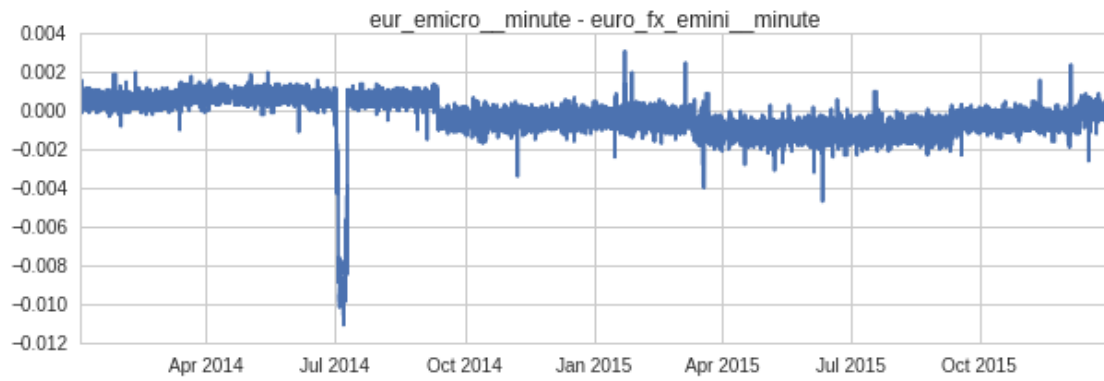
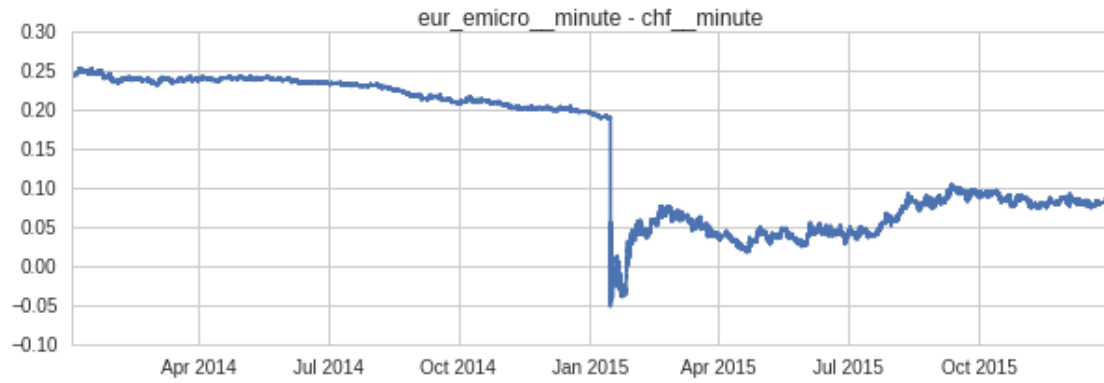
2.2.3 1.2 Price Speard

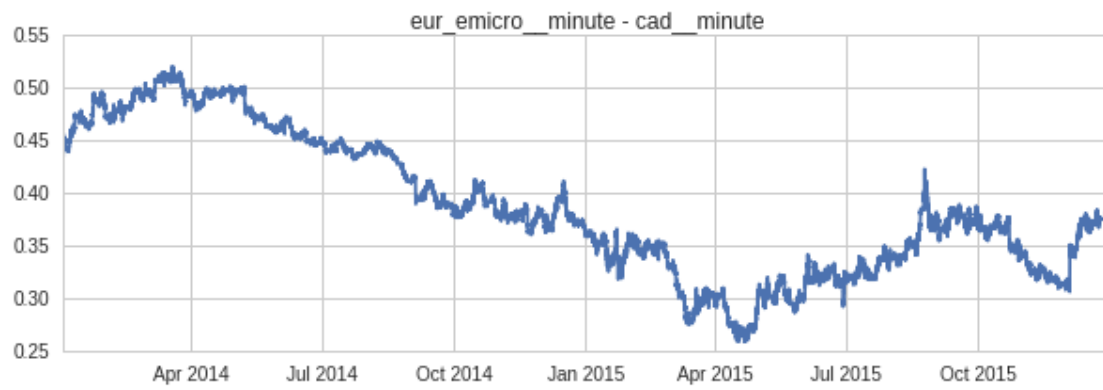
```
In [9]: a = spread_subtraction(data)
        plotting(a)
```

```
/usr/local/lib/python2.7/dist-packages/matplotlib/pyplot.py:516: RuntimeWarning: More than 20
max_open_warning, RuntimeWarning)
```



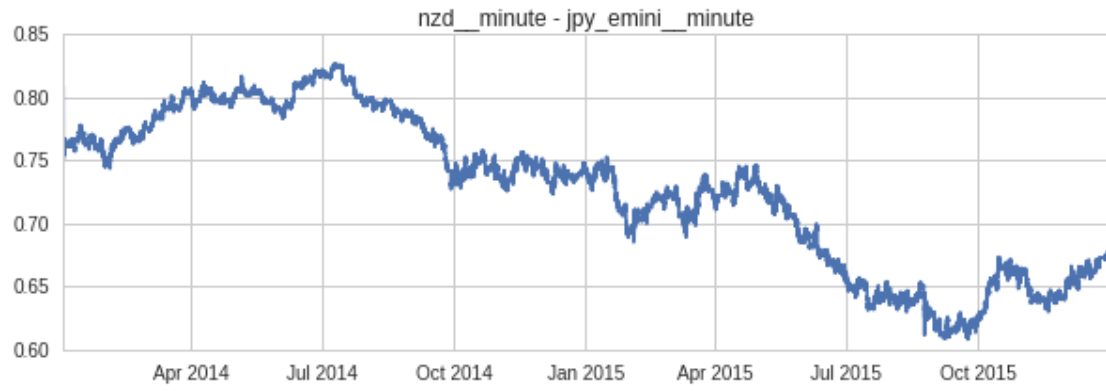




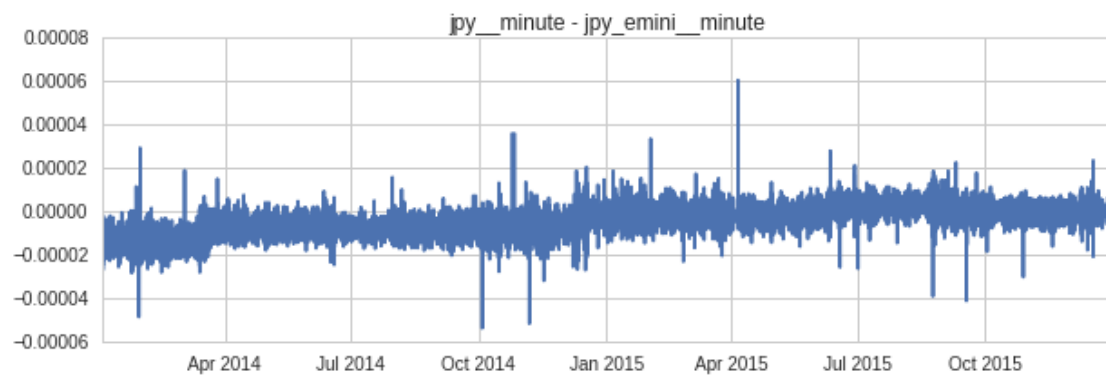




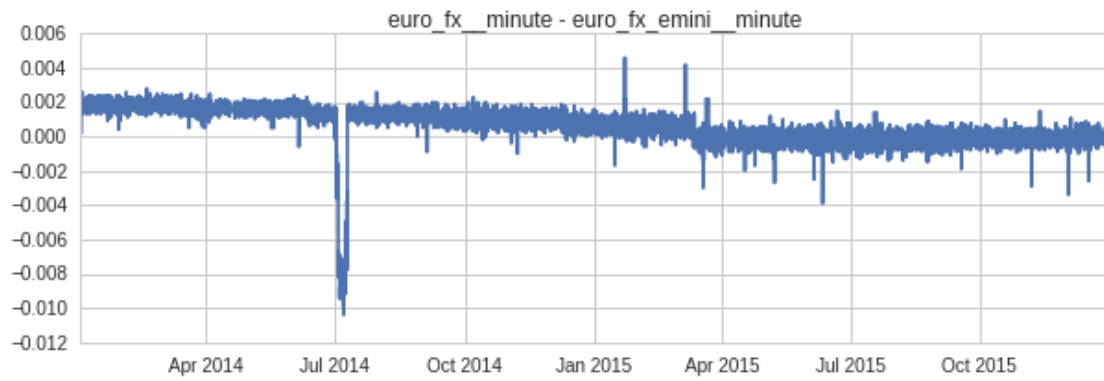






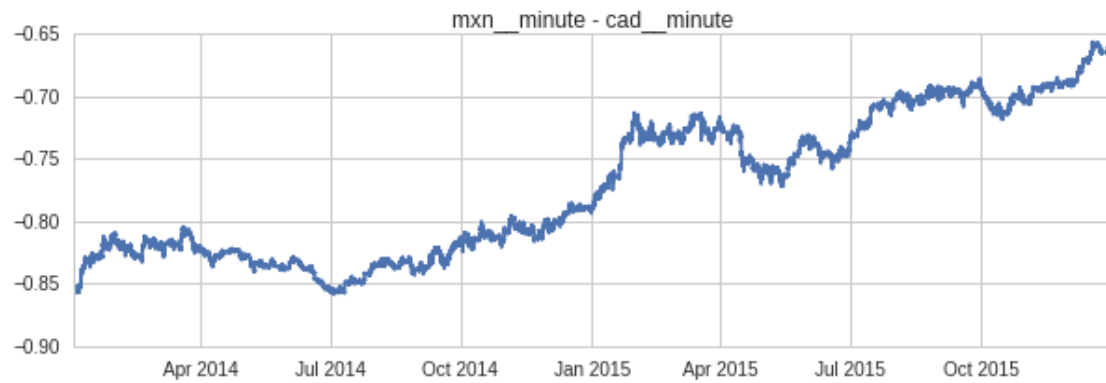


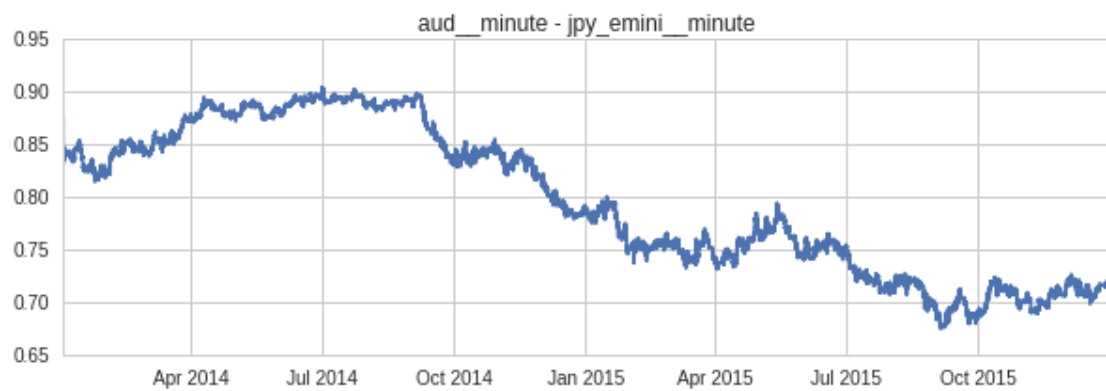


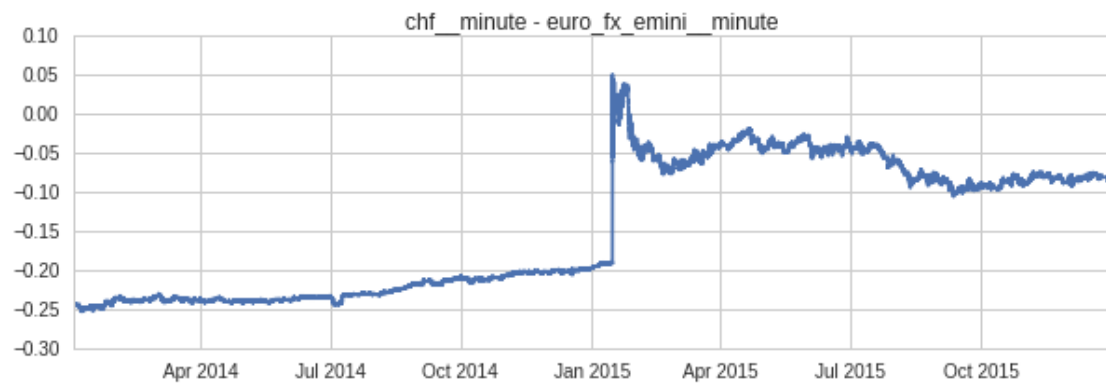
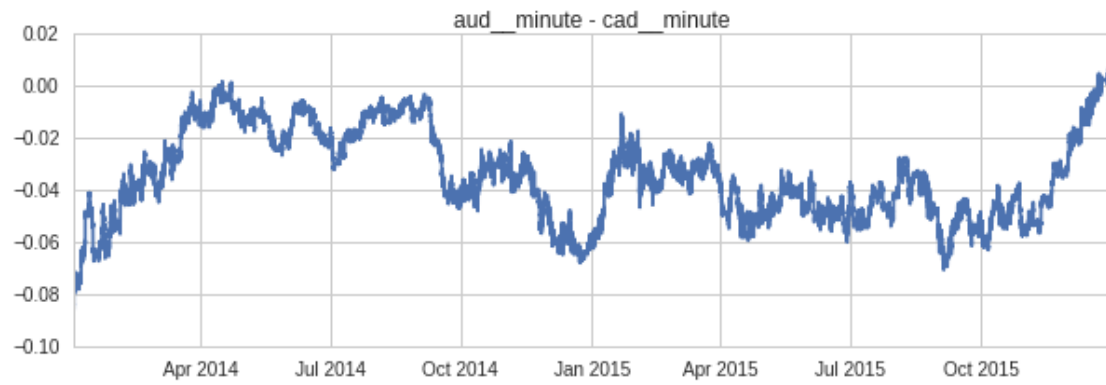




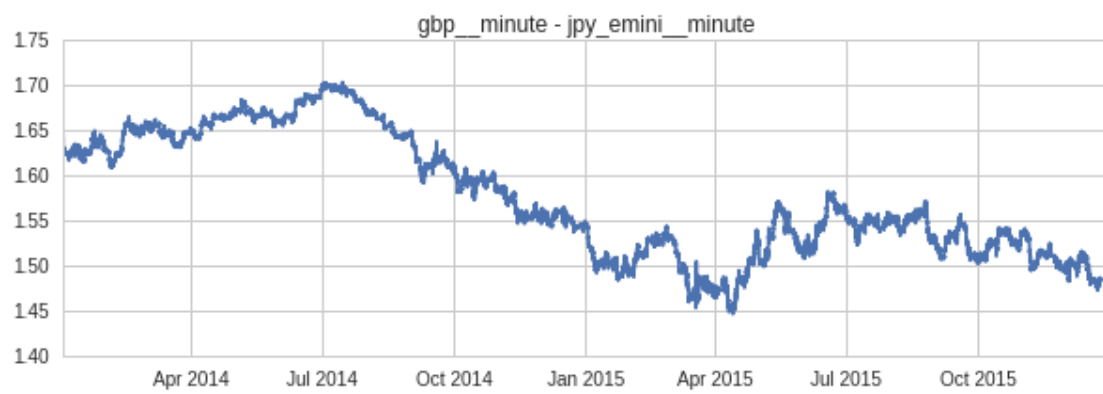
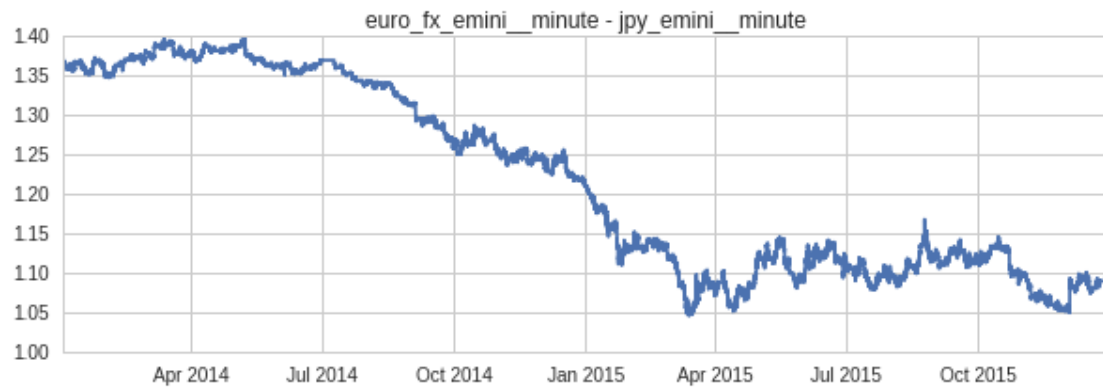


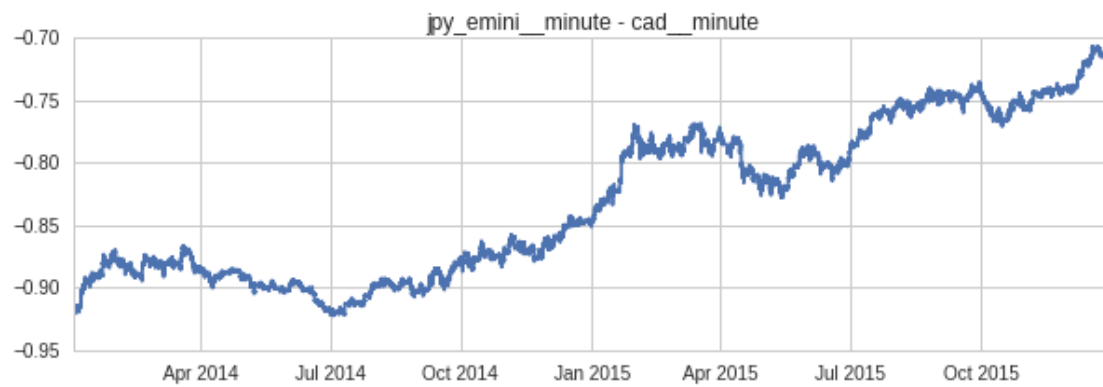










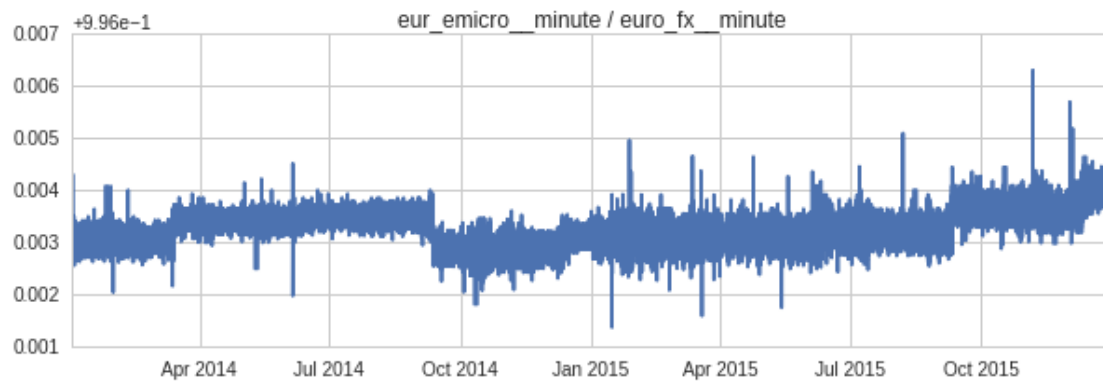


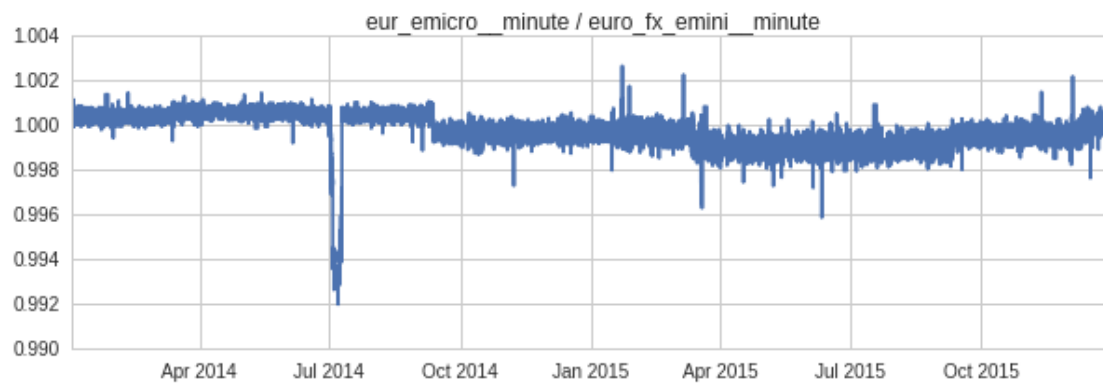
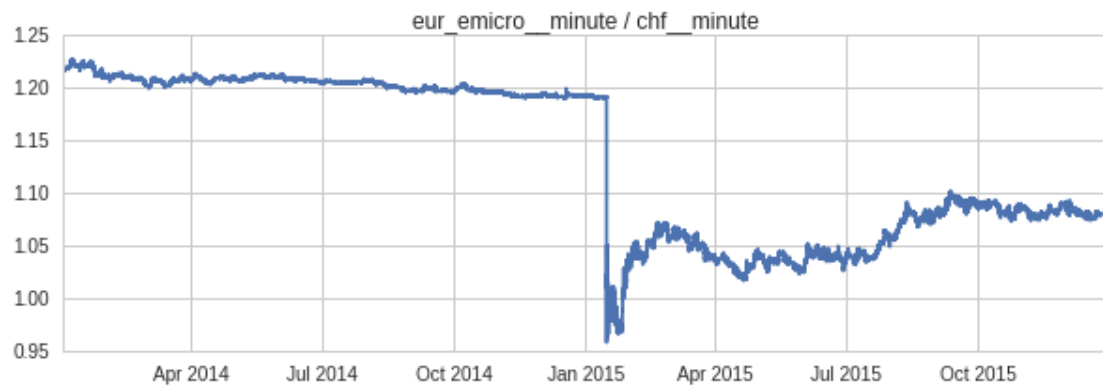
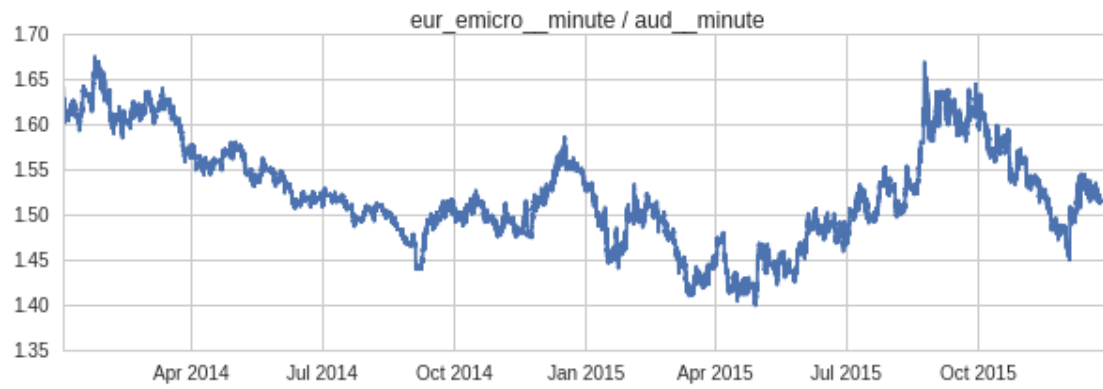
2.2.4 1.3 Price Division

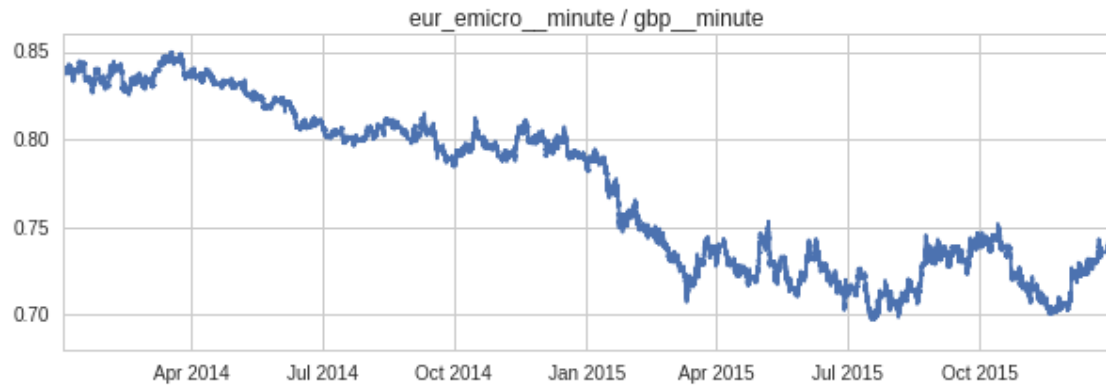
```
In [7]: b = spread_division(data)
        plotting(b)
```

/usr/local/lib/python2.7/dist-packages/matplotlib/pyplot.py:516: RuntimeWarning: More than 20
max_open_warning, RuntimeWarning)





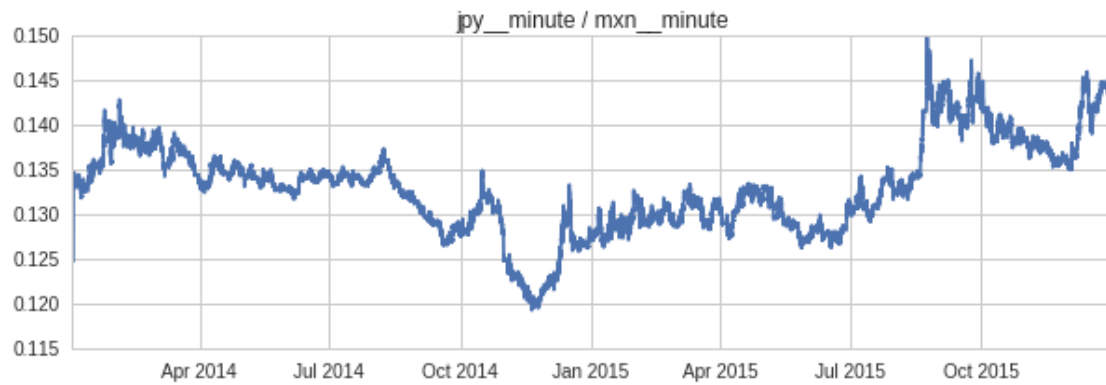




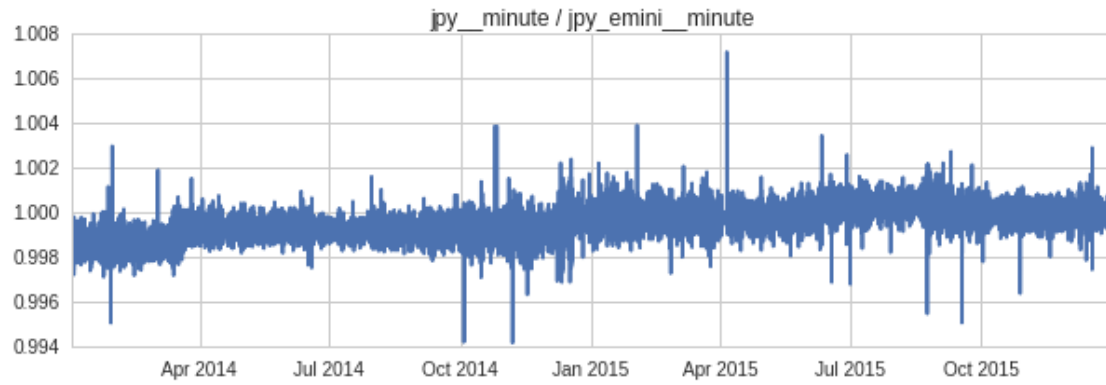


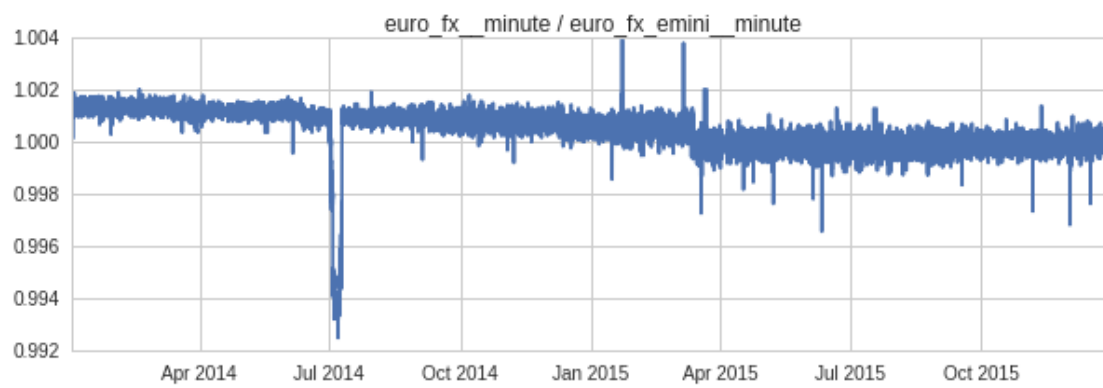










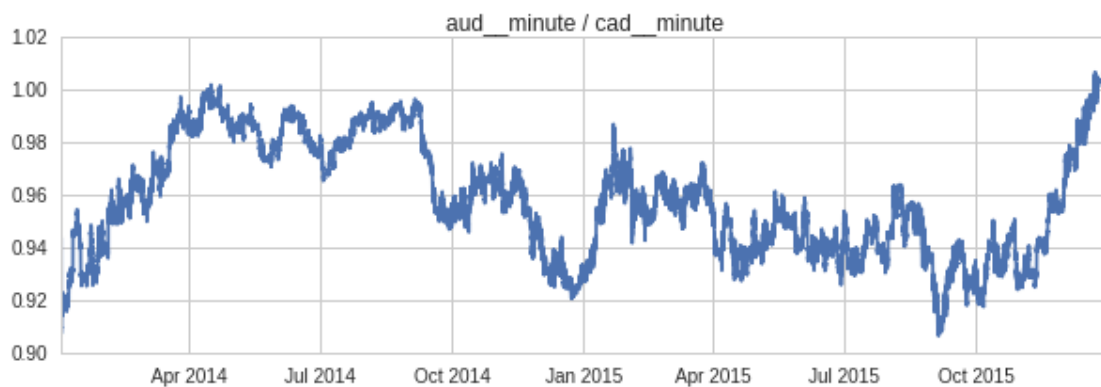






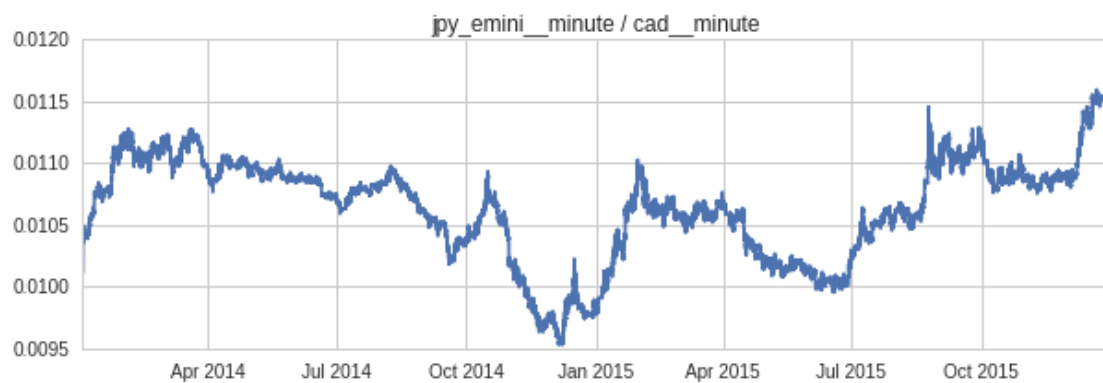












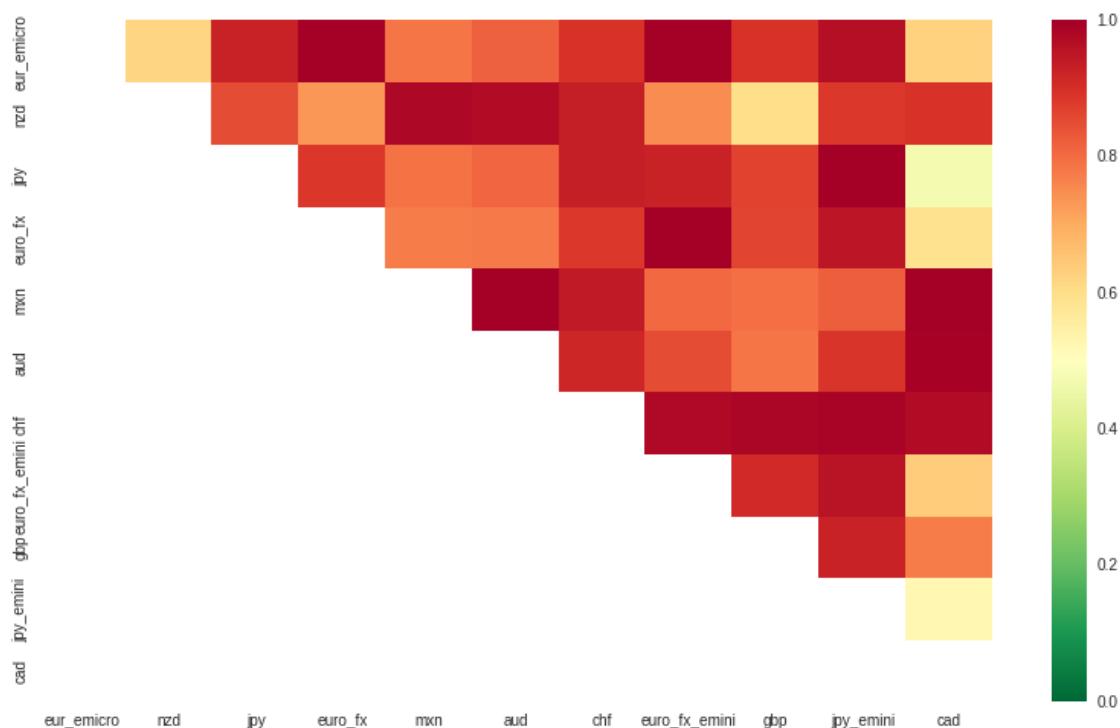
2.2.5 1.4 Cointegration (Green color squares imply lowest p-value and significant cointegration)

In [8]: pvalues, pairs = find_cointegrated_pairs(data)

2.2.6 Heatmap of cointegration: Green Significant cointegration

```
In [9]: sns.heatmap(1-pvalues,xticklabels=future_list, yticklabels=future_list,cmap = 'RdYlGn_r')
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fde095718d0>
```



2.2.7 1.5 Pairs with significant cointegration

```
In [10]: pd.DataFrame(pairs)
```

```
Out[10]:
```

| | 0 | 1 |
|----|--------------------|-----------------------|
| 0 | eur_emicro__minute | euro_fx__minute |
| 1 | eur_emicro__minute | euro_fx_emini__minute |
| 2 | eur_emicro__minute | jpy_emini__minute |
| 3 | nzd__minute | mxn__minute |
| 4 | nzd__minute | aud__minute |
| 5 | jpy__minute | jpy_emini__minute |
| 6 | euro_fx__minute | euro_fx_emini__minute |
| 7 | euro_fx__minute | jpy_emini__minute |
| 8 | mxn__minute | aud__minute |
| 9 | mxn__minute | cad__minute |
| 10 | aud__minute | cad__minute |
| 11 | chf__minute | euro_fx_emini__minute |
| 12 | chf__minute | gbp__minute |
| 13 | chf__minute | jpy_emini__minute |

```

14             chf__minute           cad__minute
15 euro_fx_emi__minute   jpy_emi__minute

```

2.2.8 1.6 OLS In-sample: 2014-2015, Validation:2016-2017

2.2.9 (Only pairs with significant cointegration are included)

```

In [13]: # Get ols parameters using in-sample data
ols_result = ols_in_sample(data,pairs)

# Get residues using validation data based on in-sample ols parameters
validation_data = get_data(Currencies,start_date='2016-01-01', end_date='2017-12-31')
n=ols_validation(validation_data, ols_result,pairs)
plotting(n)

```

