

时序分析(2) -- 单根检验

如无特殊说明，本系列文章中的数据将使用2012~2017年，分别代表国内股票、香港股票、国内债券和国内货币的四个指数数据。

上一篇文章我们探讨了时序数据的描述性分析和推断性分析。这一节我们主要讨论时序数据的平稳分析和单根检验。

首先我们先介绍平稳性的概念：

时序数据的平稳性属性是其非常重要的特征，上一篇中我们曾经估算指数数据的收益率、标准差、分布等，其实一切的目的都是为了很好的预测时序数据，但想想如果数据的均值和方差是经常有很大的变化的，我们又怎么能够预测呢？

平稳过程定义如下

1. 时序的均值不能是时间的函数。
2. 时序的方差不能是时间的函数。
3. 时序的第 i 项与第 $i+m$ 项的协方差不能是时间的函数。

简单来说，就是时序的均值、方差和自协方差这三个统计特性不能随着时间的变换而变化。

平稳时序相对来说是比较容易预测的，因为我们可以认为未来的统计特性和当前的统计特性是一致的。大部分时序模型都假设至少是协方差稳定的。但实际上，大多数金融时序数据都不是平稳时序数据，所以在实际时序分析时，我们经常需要通过某种方法将其转变为平稳过程。

那么，我们如何得知所要处理的时序数据是否平稳时序呢？这就需要进行统计检验：单根检验。

单根检验(unit root test)就是检查时序数据是否存在单根，现在让我们简要解释一下这个概念：

考虑一个时序过程 $\{y_t, t = 1, 2, 3, \dots, \infty\}$ ，如果可以把它写成一个 p 阶自回归过程：

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \epsilon_t$$

这里， $\{\epsilon_t, t = 0, 1, \dots, \infty\}$ 是串行不相关的，且均值为0，有不变的方差 σ^2 的随机过程。方便起见，不妨设 $y_0 = 0$ ，如果 $m = 1$ 是其特征等式

$$m^p - m^{p-1}a_1 - m^{p-2}a_2 - \dots - a_p = 0$$

，那么该时序过程就被认为存在单位根，或者称为integrated of order 1,记为 $I(1)$ 。

让我们来看一个简单的例子：

一个一阶自回归过程 $y_t = a_1 y_{t-1} + \epsilon_t$ ，当 $a_1 = 1$ 时存在单位根，因为其特征等式 $m - a_1 = 0$ ，是一个非平稳过程。该过程的动量依赖于时间 t

我们可以以初值 y_0 重复带入 $y_t = a_1 y_{t-1} + \epsilon_t$ ，可得到

$$y_t = y_0 + \sum_{j=1}^t \epsilon_j$$

那么， y_t 的方差为

$$\text{Var}(y_t) = \sum_{j=1}^t \epsilon^2 = t\epsilon^2$$

从上式可知，方差依赖于时间t,随着时间的推移，方差必将趋向于无穷大。

如果时序中存在单位根，就说明是非平稳时序；反之就是平稳时序。
下面我们检验四个指数的收益率数据，看是否是平稳时序。

1. 导入必要的python包

In [1]:

```
import warnings
warnings.simplefilter('ignore')
```

In [2]:

```
import pandas as pd
import numpy as np
%matplotlib inline

from finetools.backtest import *
from finetools.datasource import *
from finetools.SimuMultiTest import *

import matplotlib
import matplotlib as mpl
from matplotlib.ticker import FuncFormatter
mpl.style.use('classic')

plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['font.serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False
import seaborn as sns
sns.set_style("whitegrid", {"font.sans-serif": ['simhei', 'Arial']})
sns.set_context("talk")

%load_ext autoreload
%autoreload 2
```

The autoreload extension is already loaded. To reload it, use:

```
%reload_ext autoreload
```

2. 读入数据

In [3]:

```
start = '2012-01-01'
end = '2016-03-01'
```

In [4]:

```
indexs = pd.read_excel('./data/华夏指数.xlsx')
indexs_pv = indexs.pivot_table(index='日期', columns='简称', values='收盘价(元)')
indexs_pv.index = pd.to_datetime(indexs_pv.index, unit='d')
```

In [5]:

```
indexs_pv.columns = ['国内债券', '国内股票', '香港股票', '国内货币']
indexs_pv = indexs_pv[['国内债券', '国内股票', '国内货币', '香港股票']]
indexs_pv.fillna(axis=0, method='bfill', inplace=True)
indexs_sub = indexs_pv.loc[start:end,]
```

国内债券：中债综合财富(总值)指数
国内股票：中证全指
香港股票：恒生指数
国内货币：货币基金

In [6]:

```
indexs_sub.head()
```

Out[6]:

| | 国内债券 | 国内股票 | 国内货币 | 香港股票 |
|------------|----------|----------|-----------|----------|
| 日期 | | | | |
| 2012-01-04 | 141.5160 | 2571.951 | 1166.7726 | 18727.31 |
| 2012-01-05 | 141.5501 | 2513.699 | 1166.9696 | 18813.41 |
| 2012-01-06 | 141.7277 | 2527.247 | 1167.1185 | 18593.06 |
| 2012-01-09 | 141.8669 | 2619.638 | 1167.5058 | 18865.72 |
| 2012-01-10 | 142.0118 | 2713.529 | 1167.6330 | 19004.28 |

In [9]:

```
indexs_logret = indexs_sub.apply(log_return).dropna()
```

In [10]:

```
indexs_logret.head()
```

Out[10]:

| | 国内债券 | 国内股票 | 国内货币 | 香港股票 |
|------------|----------|-----------|----------|-----------|
| 日期 | | | | |
| 2012-01-05 | 0.000241 | -0.022909 | 0.000169 | 0.004587 |
| 2012-01-06 | 0.001254 | 0.005375 | 0.000128 | -0.011782 |
| 2012-01-09 | 0.000982 | 0.035906 | 0.000332 | 0.014558 |
| 2012-01-10 | 0.001021 | 0.035214 | 0.000109 | 0.007318 |
| 2012-01-11 | 0.000188 | -0.002115 | 0.000113 | 0.007740 |

Augmented Dickey-Fuller Testing

- 国内股票

In [11]:

```
from arch.unitroot import ADF
adf = ADF(indexs_logret['国内股票'])
print('国内股票', adf.summary().as_text())
```

国内股票 Augmented Dickey-Fuller Results

```
=====
Test Statistic          -6.483
P-value                  0.000
Lags                     22
=====
```

Trend: Constant

Critical Values: -3.44 (1%), -2.86 (5%), -2.57 (10%)

Null Hypothesis: The process contains a unit root.

Alternative Hypothesis: The process is weakly stationary.

- 香港股票

In [12]:

```
adf = ADF(indexs_logret['香港股票'])
print('香港股票', adf.summary().as_text())
```

香港股票 Augmented Dickey-Fuller Results

```
=====
Test Statistic          -30.856
P-value                  0.000
Lags                     0
=====
```

Trend: Constant

Critical Values: -3.44 (1%), -2.86 (5%), -2.57 (10%)

Null Hypothesis: The process contains a unit root.

Alternative Hypothesis: The process is weakly stationary.

- 国内债券

In [13]:

```
adf = ADF(indexs_logret['国内债券'])
print('国内债券', adf.summary().as_text())
```

国内债券 Augmented Dickey-Fuller Results

```
=====
Test Statistic      -6.013
P-value             0.000
Lags                 14
=====
```

Trend: Constant

Critical Values: -3.44 (1%), -2.86 (5%), -2.57 (10%)

Null Hypothesis: The process contains a unit root.

Alternative Hypothesis: The process is weakly stationary.

- 国内货币

In [14]:

```
adf = ADF(indexs_logret['国内货币'])
print('国内货币', adf.summary().as_text())
```

国内货币 Augmented Dickey-Fuller Results

```
=====
Test Statistic      -4.332
P-value             0.000
Lags                 19
=====
```

Trend: Constant

Critical Values: -3.44 (1%), -2.86 (5%), -2.57 (10%)

Null Hypothesis: The process contains a unit root.

Alternative Hypothesis: The process is weakly stationary.

Augmented Dickey-Fuller Testing 检验法得到: p-value非常小, 这四个时序是弱平稳的。

Dickey-Fuller GLS Testing

- 国内股票

In [15]:

```
from arch.unitroot import DFGLS
dfgls = DFGLS(indexs_logret['国内股票'])
print('国内股票', dfgls.summary().as_text())
```

国内股票 Dickey-Fuller GLS Results

```
=====
Test Statistic            -1.403
P-value                   0.155
Lags                       22
=====
```

Trend: Constant
Critical Values: -2.59 (1%), -1.97 (5%), -1.64 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.

- 香港股票

In [16]:

```
dfgls = DFGLS(indexs_logret['香港股票'])
print('香港股票', dfgls.summary().as_text())
```

香港股票 Dickey-Fuller GLS Results

```
=====
Test Statistic            -3.191
P-value                   0.002
Lags                       19
=====
```

Trend: Constant
Critical Values: -2.59 (1%), -1.97 (5%), -1.64 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.

- 国内债券

In [17]:

```
dfgls = DFGLS(indexs_logret['国内债券'])
print('国内债券', dfgls.summary().as_text())
```

国内债券 Dickey-Fuller GLS Results

```
=====
Test Statistic            -5.993
P-value                   0.000
Lags                       14
=====
```

Trend: Constant
Critical Values: -2.59 (1%), -1.97 (5%), -1.64 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.

- 国内货币

In [18]:

```
dfgls = DFGLS(indexs_logret['国内货币'])
print('国内货币', dfgls.summary().as_text())
```

国内货币 Dickey-Fuller GLS Results

```
=====
Test Statistic            -4.147
P-value                   0.000
Lags                       19
=====
```

Trend: Constant

Critical Values: -2.59 (1%), -1.97 (5%), -1.64 (10%)

Null Hypothesis: The process contains a unit root.

Alternative Hypothesis: The process is weakly stationary.

Dickey-Fuller GLS Testing 结论：除国内股票外，其他都是平稳时序

由于这两种单根检验法在国内股票时序数据上得到的结果不一样，下面我们再采用几种其他单根检验法对国内股票进行重复检验。

Phillips-Perron Testing

- 国内股票

In [19]:

```
from arch.unitroot import PhillipsPerron
pp = PhillipsPerron(indexs_logret['国内股票'])
print(pp.summary().as_text())
```

Phillips-Perron Test (Z-tau)

```
=====
Test Statistic            -29.715
P-value                   0.000
Lags                       22
=====
```

Trend: Constant

Critical Values: -3.44 (1%), -2.86 (5%), -2.57 (10%)

Null Hypothesis: The process contains a unit root.

Alternative Hypothesis: The process is weakly stationary.

KPSS Testing

- 国内股票

In [20]:

```
from arch.unitroot import KPSS
kpss = KPSS(indexs_logret['国内股票'])
print(kpss.summary().as_text())
```

KPSS Stationarity Test Results

```
=====
Test Statistic          0.092
P-value                 0.628
Lags                    22
=====
```

Trend: Constant

Critical Values: 0.74 (1%), 0.46 (5%), 0.35 (10%)

Null Hypothesis: The process is weakly stationary.

Alternative Hypothesis: The process contains a unit root.

Variance Ratio Testing

In [21]:

```
from arch.unitroot import VarianceRatio
vr = VarianceRatio(indexs_logret['国内股票'], 12)
print(vr.summary().as_text())
```

Variance-Ratio Test Results

```
=====
Test Statistic          -10.252
P-value                 0.000
Lags                    12
=====
```

Computed with overlapping blocks (de-biased)

In []:

Phillips-Perron Testing和Variance Ratio Testing认为国内股票是平稳过程，而KPSS Testing认为国内股票不是

总结

本文展示了采用Python语言为四个指数时序数据的收益率做了多种单位根检验，并解释了单位根和平稳过程的概念，也展示了检验的结论。