

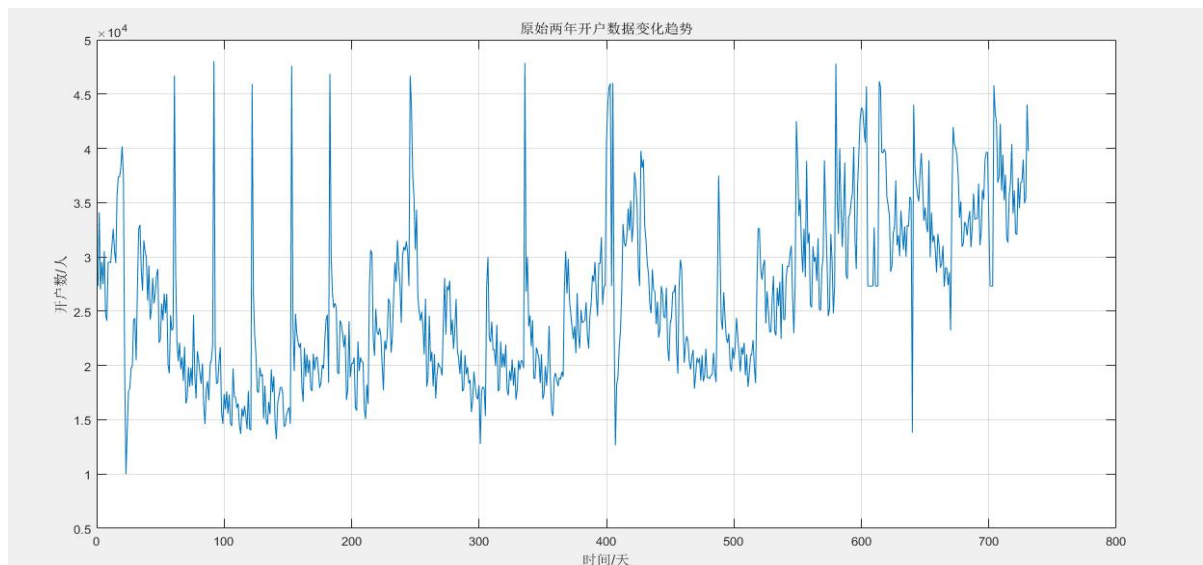
时间序列分析 1

继去趋势后的自/偏相关值估计

2016100104028 李科

1. 数据观察

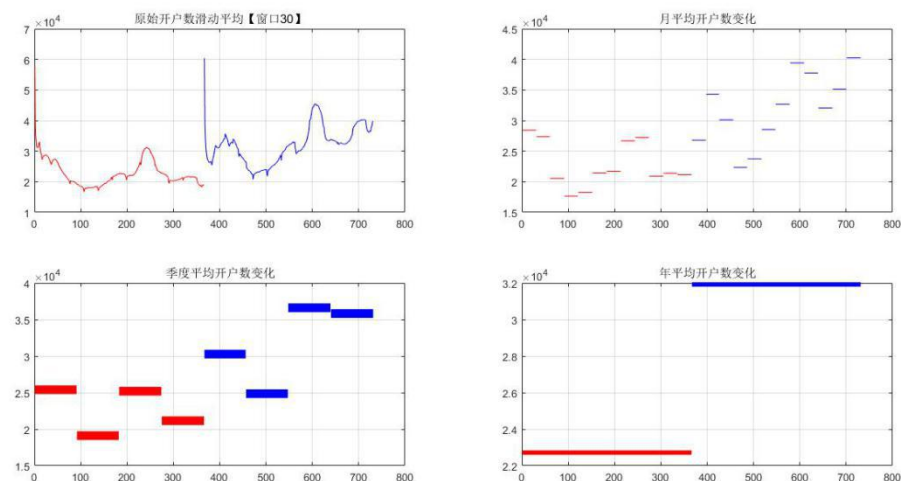
绘制原开户数据的变化趋势如下：



图一.原开户数据趋势变化的观察

可以发现, 开户数在局部区间的极值变化具有一定的周期性。但是数据太多且有点杂乱。

为了更好的观察数据的规律, 拟将数据进行滑动平均处理, 在一定程度上更能看出规律性。并且, 依据两年的月数天数变化, 设置标记每月天数变化的向量, 计算每月, 每个季度, 每年的平均开户数进行观察。计算并绘制得到下图:



图二.开户数据趋势变化周期的观察

上图红色标记为 2012 年，蓝色标记为 2013 年。总的看来均有上升的趋势。图二左上角绘制的是 2012 和 2013 的开户数滑动平均观察（使用 matlab smooth 函数，滑动窗口设置为 30），容易发现在年这个层面上趋势具有较大的相关性，一年显然可以是一个大周期的刻画。但是这个周期太大了，如果选取这个周期做去季节性，显然会损失大量数据。因此，我们想找到一个更小的周期来刻画，其实我们还可以发现图一中有明显的小起伏，经验证起伏平均间隔为 30 即一个月，那么一个月是否为一个周期呢？右上图是按月进行平均的统计值绘制，起伏貌似没有什么规律。但是还是可以作为参考。左下角图绘制的是季度的平均开户数变化，可以发现两个季度为一个周期，因为起伏在两个周期中进行。右下角图绘制的是年平均变化图，如我们所料，随着移动互联网的普及，2013 年开户平均数比 2012 年高。

综上观察，我们可以得出以下结论：

1. 一个月应该是一个最小周期，两个季度是一个中周期，一年为一个确定的大周期。
2. 开户人数有明显的上升趋势。
3. 由观测值我们可以认为存在异常值。

2. 数据的预处理

2.1 缺失值处理

数据中不存在缺失值

2.2 异常值处理

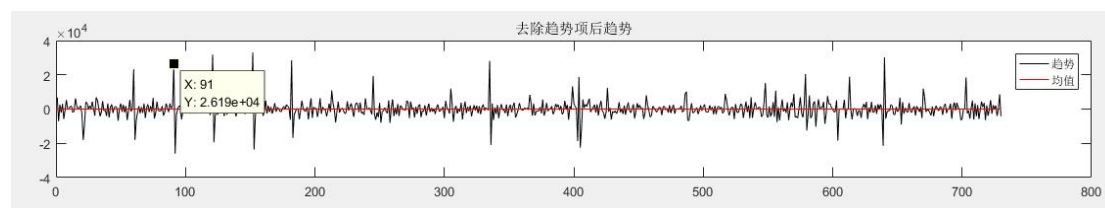
拟采用三倍标准差即拉以达法则筛选异常值，考虑到时间序列的短期影响性，用其周围值平均代替。

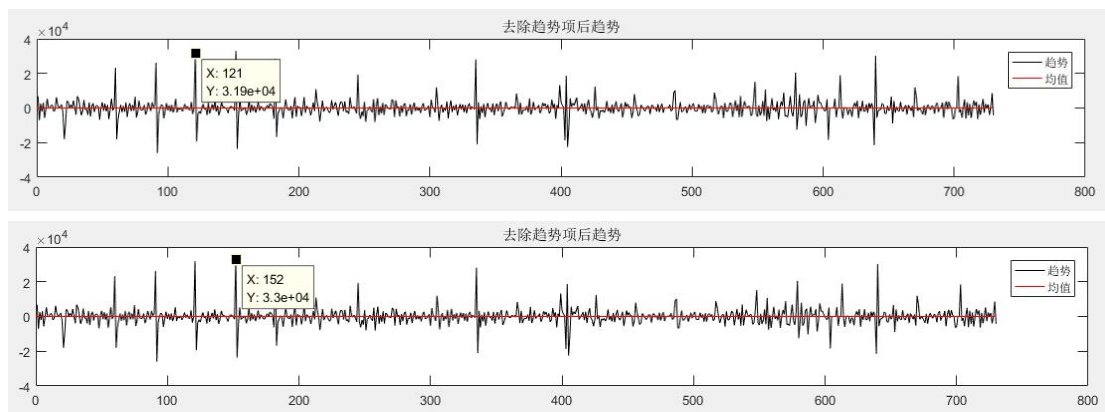
3. 平稳化--去趋势与去周期

进行了简单的观察和数据预处理后，我们开始正式进行时间序列平稳化的操作。考虑到有两年，我们从单独考虑 2012，2013，然后集中考虑 2012 和 2013 来进行处理和检验。

3.1 去趋势

由于数据变化有一定的集中性，且有滑动平均看来近似可以用一次函数拟合。所以采用一次差分的方法去除数据的趋势项。这里采用 diff 函数实现一次差分。去趋势后，我们进行了如下的观察：





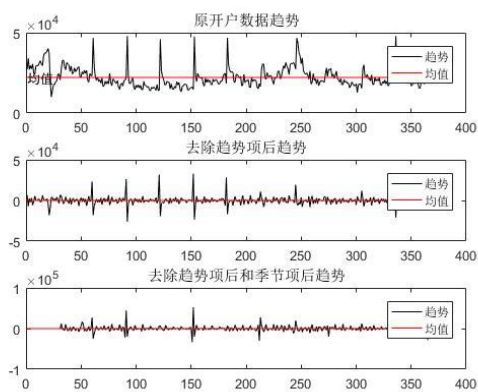
图三.周期为一个月的检验图

我们发现相邻极值点间的时间差距十分接近于一个月，这让我们欣喜万分。

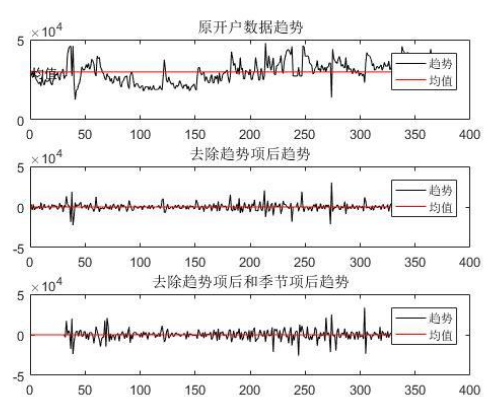
3.2 去周期

由以上的观察数据和去趋势后数据综合考虑，采用一个月作为周期是个不错的选择。这里自定义 d 步差分函数对去趋势后的数据进行去周期的处理。

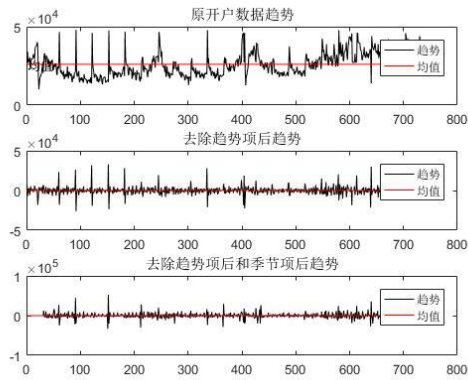
经过 3.1 和 3.2,我们得到如下结果：



图四.2012 年去趋势和去周期比较图



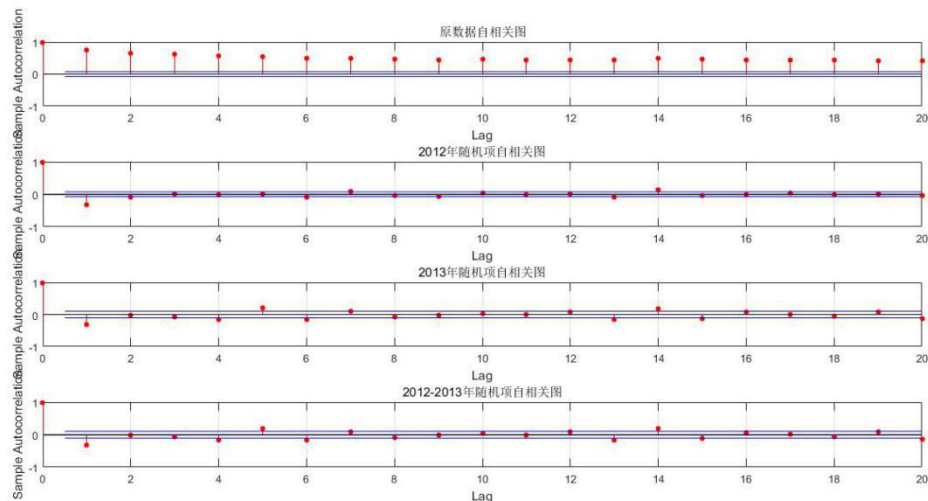
图五.2013 年去趋势和去周期比较图



图六.2012-2013 年去趋势和去周期比较图

肉眼看来，效果还算不错，从均值线看来均值均十分接近于 0。为了更客观的看待平稳的效果，采用自相关图还进行平稳性的验证。

自相关性平稳性检验结果如下：



图七.自相关图检验平稳性

从自相关图看来，经过去趋势和去周期后的数据的确是平稳的。经过以上步骤，我们就成功的将非平稳的时间序列转换为了平稳的时间序列，得到了随机误差项的表示。

* 实验环境：matlab 2016b

附录

1.主函数 detrend_main.m

```
clearvars
%% 读取数据并预处理
open_nums = xlsread('移动通知户开户数.xlsx',1,'B2:B732');
%% 计算月平均，季度平均，年平均并绘图找周期规律,并处理异常值
```

```

figure(1), [months_mean, seasons_mean, years_mean, month_starts, month_ends,
season_starts, season_ends] = Calu_mean(open_nums);
[m,n] = find(abs(open_nums-mean(open_nums))>2*std(open_nums));
open_nums(m) = mean(open_nums);
%% 尝试进行去趋势和去周期,考虑到存在多重周期, 将年份分开按照月为周期去除周期 (其中
每月为多少天根据两年分别不同)
data_2012 = open_nums(1:366);
data_2013 = open_nums(367:end);
figure(2), [detrend_data2012, detrend_deT_data2012] =
Detrend_plot(data_2012,30);% 2012 开户数变化趋势
figure(3), [detrend_data2013, detrend_deT_data2013] =
Detrend_plot(data_2013,31);% 2013 开户数变化趋势
figure(4), [detrend_data, detrend_deT_data] =
Detrend_plot(open_nums,30);% 2012-2013 开户数变化趋势
figure(5), subplot(411), autocorr(open_nums), title('原数据自相关图')% 自相
关图分析
subplot(412), autocorr(detrend_data), title('2012 年随机项自相关图')
subplot(413), autocorr(detrend_deT_data2013), title('2013 年随机项自相关图
')
subplot(414), autocorr(detrend_deT_data2013), title('2012-2013 年随机项自
相关图')
clearvars month_starts month_ends season_starts season_ends data_2012
data2013

```

2, 自定义计算均值和绘图函数 calu_mean.m

```

%% 计算月平均, 季度平均, 年平均意图找寻周期规律
function
[months_mean, seasons_mean, years_mean, month_starts, month_ends, season_s
tarts, season_ends] = Calu_mean(open_nums)
t = 1:length(open_nums);
nums = open_nums;
month_days =
[0,31,29,31,30,31,30,31,31,30,31,30,31,31,28,31,30,31,30,31,31,30,31,
30,31];
season_days =
[0,sum(month_days(2:4)),sum(month_days(5:7)),sum(month_days(8:10)),su
m(month_days(11:13))...
, sum(month_days(14:16)),sum(month_days(17:19)),sum(month_days(20:
22)),sum(month_days(23:25))];
% calu the month_mean nums
months_mean = zeros(1,24);
month_starts = zeros(1,24);
month_ends = zeros(1,24);
for ii = 1:length(month_days)-1

```

```

start_day = sum(month_days(1:ii))+1;
end_day = sum(month_days(1:ii+1));
month_starts(ii) = start_day;
month_ends(ii) = end_day;
months_mean(ii) = mean(nums(start_day:end_day));
end

```

3.去趋势和周期并绘图函数 Detrend_plot.m

```

function [detrend_data,detrend_deT_data] = Detrend_plot(vector,T)
    subplot(3,1,1)
    plot(1:length(vector),vector,'k')
    hold on

    plot(0:length(vector),mean(vector)*ones(1,length(0:length(vector))), '
    r')
    text(1,mean(vector),'均值')
    legend('趋势','均值')
    title('原开户数据趋势')
    % 去趋势项
    detrend_data = diff(vector,1);
    subplot(3,1,2)
    plot(1:length(detrend_data),detrend_data,'k')
    hold on

    plot(0:length(detrend_data),mean(detrend_data)*ones(1,length(0:length
    (detrend_data))), 'r')
    legend('趋势','均值')
    title('去除趋势项后趋势')
    % 去季节项
    detrend_deT_data = zeros(length(detrend_data),1);
    for i=1:length(detrend_data)
        if(i<=T)
            detrend_deT_data(i) = [];
        else
            detrend_deT_data(i) = detrend_data(i)-detrend_data(i-T);
        end
    end
    subplot(3,1,3)
    plot(1:length(detrend_deT_data),detrend_deT_data,'k')
    hold on

    plot(0:length(detrend_deT_data),mean(detrend_deT_data)*ones(1,length(
    0:length(detrend_deT_data))), 'r')
    legend('趋势','均值')

```

```
title('去除趋势项后和季节项后趋势')  
end
```