

Weekly Report

14091203 崔诗颖

November 18, 2017

1 数据说明

1. 数据来源

2007 年至 2016 年 A 股市场每日股票收盘价；其中 2007 年有 792 只股票；2008 年有 816 只股票；2009 年有 821 只股票；2010 年有 829 只股票；2011 年有 856 只股票；2012 年有 901 只股票；2013 年有 930 只股票；2014 年有 935 只股票；2015 年有 978 只股票；2016 年有 1068 只股票。其中交易天数基本都在 242 天左右，有些年份是 245 天。

2. 数据预处理

将每一年的收盘价求对数收益率，即相当于收盘价的一阶对数差分，理论上来说应该是平稳的。

在 *stata* 中用不带截距项的 *D-Fuller* 方法检验每一只股票在当年的平稳性，且设定 95% 的置信区间来寻找非平稳序列

3. 平稳性结果说明

在 2007 年编号为 600137.SH 和 600369.SH 的两只股票不平稳，其缺失数据太多，股票图像如下：

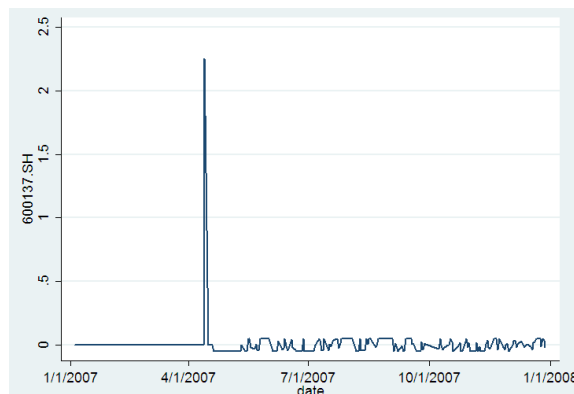


Figure 1: 600137 股票

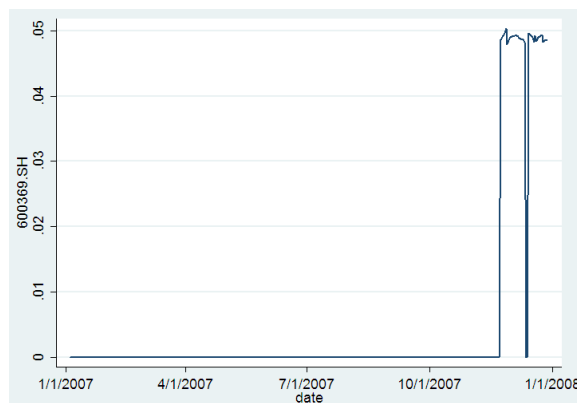


Figure 2: 600369 股票

2008 年的股票全部平稳，而 2009 年有 9 只股票不平稳，它们分别是 600031.SH, 600275.SH, 600276.SH, 600519.SH, 600600.SH, 600694.SH, 600729.SH, 600808.SH, 600868.SH

仔细检查后发现 2009 年 12 月 31 日的所有股票收盘价都十分奇怪，由于代码是学姐给的，这个还需要进一步 check!

2010, 2011, 2012, 2013 的所有股票均平稳，2014 年编号为 600381.SH 的股票序列非平稳，其结果如下：

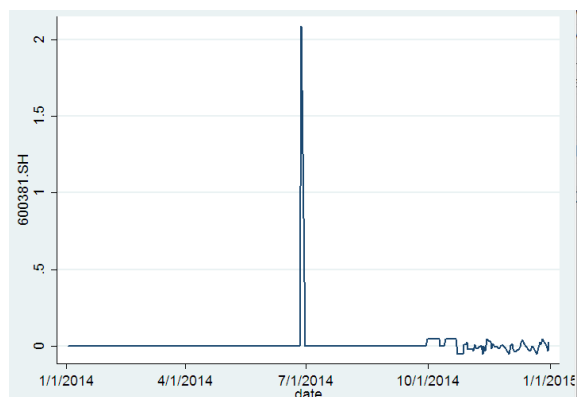


Figure 3: 600381 股票

2015 的所有股票均平稳，2016 年编号为 600145.SH 的股票序列非平稳，其结果如下：

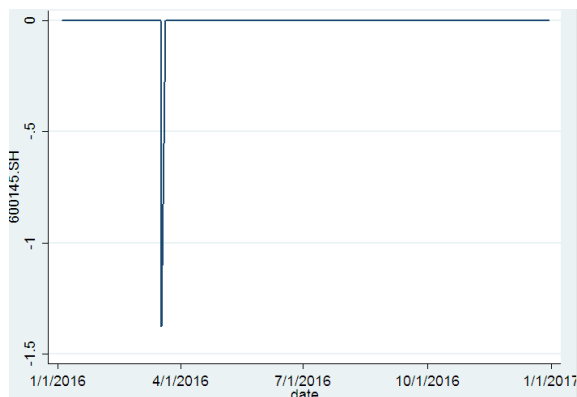


Figure 4: 600145 股票

2 实验过程

该实验的目的是，针对从 2007 年到 2016 年总共 10 年的每只股票，运用 *mgarch dcc* 模型算出两两之间的时变相关系数

1. 检验 *arch* 效应 在进行 *mgarch dcc* 前，需要先检验两只股票间是否存在 *arch* 效应；实验中测试了一阶到六阶 *arch* 效应，置信区间仍设为 95%，从实验结果来看，绝大部分股票收益率序列间都不存在 *arch* 效应，这里存疑，需要和老师讨论。
2. 实验效率 在我最开始直接将 2007 年所有股票间求相关系数时，程序运行了一整个白天加一整晚加一整个白天，还没有运行完，由于我为了节省效率，用了命令 *capture*，隐藏了打印到主界面，我看不到它运行到哪里了，因此决定做一个程序运行时间的测算
2007 年共有 792 只股票，相关系数计算次数为： $792 \times 791 / 2 = 313236$ 次，为了节省效率，将迭代次数手动设为 100，下面测算 *mgarch dcc* 和预测出时变相关系数的单次时间。用 *stata* 程序计时命令 *set rmsg on.* 代码如下：

```
mgarch dcc (B C=), arch(1) garch(1) iterate(100)
```

```
predict h1* if e(sample), correlation
```

单次运行迭代 100 次所需时间为：18 秒，结果如下图

```
r; t=17.80 14:26:03

. predict h1* if e(sample), correlation
r; t=0.03 14:26:03
```

Figure 5: 测算时间图

因此对于 2007 年来说，其总运行秒数（假设没有运行错误跳过的时间）大致应该在：

$$\begin{aligned} time &= 313236 * 18s \\ &= 5638248 / 60min \\ &= 93971 / 60h \\ &\approx 1566hour \end{aligned}$$

3. 处理方式 面对这种情况，我决定每一年都只挑出 *B* 到 *BD* 的 55 个变量序列来进行相关系数测算，并将其相关系数存在 *test.xls* 中。此时，只计算 *B* 到 *BD* 的 *mgarch dcc* 模型，每一年都需要大概 6 小时左右，每天都会测算一年的 *mgarch dcc* 模型。

3 结果说明

源代码和 *xls* 都在 *Weekly report 1* 的文件夹中，从表格中可看出有的序列时变相关系数没有 242 行，只有十几行，经过仔细查阅资料，判断可能原因是，有的序列有连续几个交易天数的缺失，虽然补 0 了但对于算方差仍然有影响，而时变相关系数是通过时变方差算得的，而在计算过程中会出现有的数算不出来的情况，这说明该序列的数据不适合做 *mgarch dcc* 模型。

可以针对目前所有的时变相关系数画出图，但我的疑惑是：即使之后有服务器可以快速把所有的时变相关系数算出来，但是每一年都有三十多万张图，是需要手动看？还是有指标可以用程序判断的吗？

如图，展示为 2016 年的 55 只股票间的时变相关系数表：

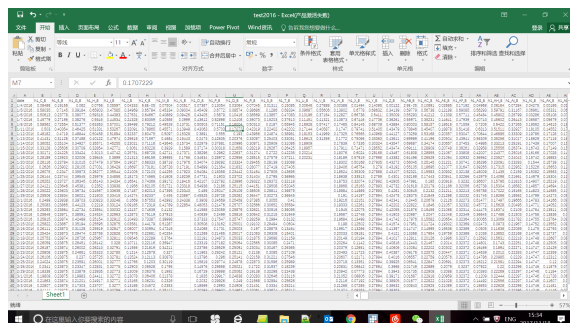


Figure 6: 2016 年表格部分结果展示

时变相关系数图，举个例子，测算 2016 年 X 与 B 列的时变相关系数如下：

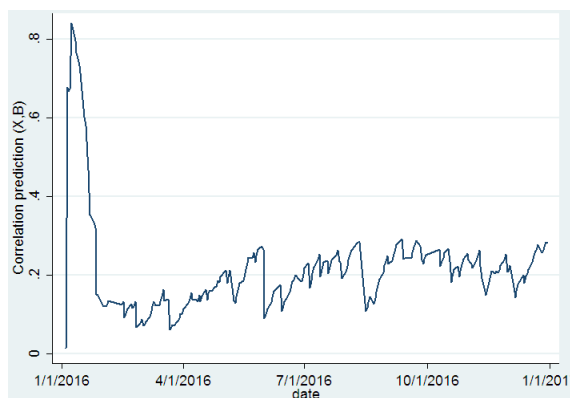


Figure 7: X 与 B 时变相关系数图

但是也有很多序列即使有 242 行，但是在后面的时变系数是一样的，这种情况会出现的原因是，为了节省时间，我强制将迭代次数设置为 100 次，但可能其需要 1000 次才能迭代完成，因此会出现这样的情况，如图 2016 年 E 和 B 列就是这样：

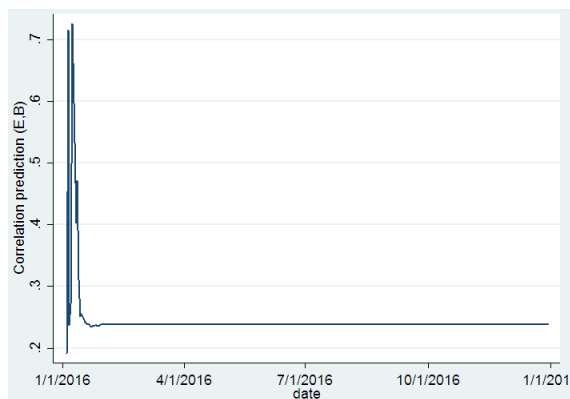


Figure 8: E 与 B 时变相关系数图