FIN3080 Assignment 2

1. 根据CSMAR数据，我做出了如下万科股票的β的时间序列图表：

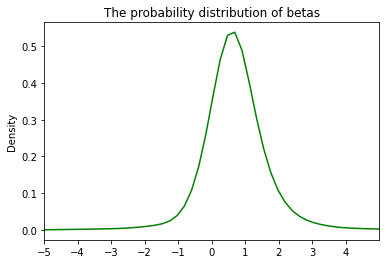
说明：

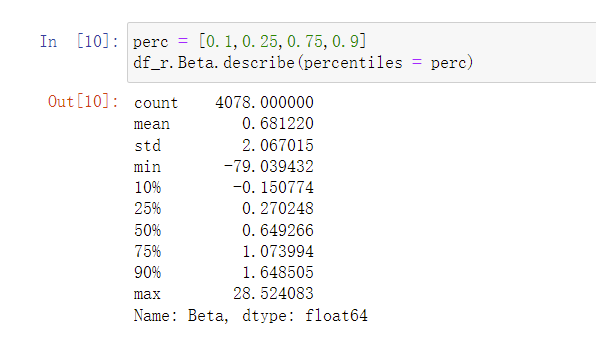
1）本图所用数据均来自CSMAR数据库，其中万科回报率来自股票市场交易-月个股回报率文件，中证800指数回报率来自市场指数-国内指数月行情文件，shibor rate来自银行体系-shibor统计表（日）。

2）原数据使用字段为date, Idxrtn, Mretwd, Shibor，经过适当处理使用statsmodels的rollingOLS模块进行回归，得到2010-01至2022-03的beta数据，后使用matplotlib进行图表建立。具体代码文件名为FIN3080 Homework 2 Q1.ipynb。

3）对于2016-01至2016-07缺失值，我采用减少回归项方法处理。若该时期往前推24个月数据只有18个月数据存在，则采用18个月的数据进行回归。数据缺失的原因有可能是万科股权之争导致的该股停牌，导致该段数据缺失。

发现：2017年至2018年底万科股票β值较其他时期较高，最高值超过4。万科作为一个超大体量的房地产公司，其股票对大盘影响较大，故β值应该大于0，但由于万科在2015年开始深陷股权之争直到2017年结束，其间由于持股股东抛售或买入大量股票而导致股价波动较大，故其与大盘的协方差较大，造成β值在这一时期较大且不稳定。在其他时间阶段与大盘走势相近，由于公司自身原因以及股票市场正常运作造成的股价的小波动体现了β值的上下正常浮动。

2. 根据CSMAR数据，A股市场个股的β的概率密度分布图以及β的百分位值如下：



说明：

1）个股回报率均来自股票市场交易-月个股回报率文件，大盘数据来自第一题中已经处理完成的数据。

2）对数据进行恰当处理后，使用sklearn包中的LinearRegression模块进行分组回归，得到个股2021年6月的预测β值，后使用matplotlib做出图表，以及pandas自带的describe函数得到百分位数。具体代码文件名为FIN3080 Homework 2 Q2.ipynb。

3）最小值与最大值较极端的可能原因是数据量较少，也有公司经营不善导致股价浮动过大的可能。其他分位值均在正常范围之内。

4）所有数据缺失值均执行drop操作，对分组回归无影响。