大数据与金融风险管理

22210980066 秦思淼

Question1

为了找到最优投资组合，首先需要获得求解最优化问题的相关参数，即 50 支股票的平均日收益率以及其协方差矩阵，无风险年利率转换成无风险日利率，期望年利率转换成期望日利率。我所使用的日利率计算方式，以及年利率，日利率转换方法如下：

$$r_t = \log(p_t/p_{t-1})$$

$$r_{day} = r_{annual}/252$$

为了获得这些参数，我使用了滑动窗口来确定五年历史数据的索引以此来找到五年历史数据，然后利用这些数据计算（使用 numpy.mean, numpy.cov）得到股票日利率均值与协方差，然后使用 cvxpy 模块解决如下最优化问题：

$$min_w \; w[1:]^T \, \Sigma \, w[1:]$$

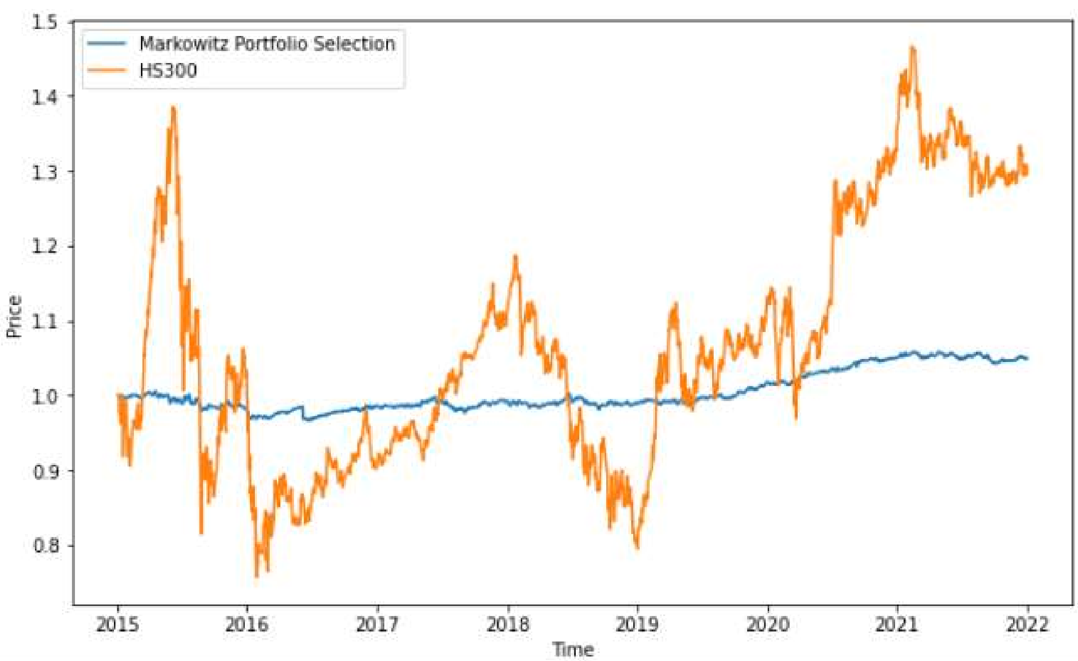
s.t.

$$[1,1,\dots,1]w = 1$$

$$w^T r = expected\_r$$

值得注意的是，上述 w 是包含无风险投资所占比例的 51 维（50+1）向量，相对应地，r 的第 0 维是无风险日利率，r[1:]才是 50 支股票的平均日收益率，而 Σ 是 50 支股票的协方差矩阵，维度是 50 * 50，这也是优化目标里 w 要去掉代表无风险收益所占比例的第 0 维的原因。

如上所述，计算出最优投资组合 w 后，需计算使用该投资组合的 4 个月内每天的收益率，只要使用 w 对该天收益率加权平均即可得到，之后每四个月调整一次投资组合，并一直记录每天的收益率，最后得到从 2015 年 1 月 5 号开始投资后每一天的收益率列表，假设初始资产为单位 1，那么将资产的变化与 HS300 在同一张表内作图，如下图所示：



可以看到，蓝线（即我所解出的最优投资组合曲线）相当平坦，因此我打印了无风险投资所占的比例列表，发现几乎所有的钱都投进了无风险投资，甚至无风险投资的比重超过了 1。

```
In [47]: np.array(w_f).T

Out[47]: array([0.99964017, 0.98130825, 0.99008535, 1.00043741, 1.00723362,
                1.0036278 , 1.00311854, 1.00109496, 0.9935397 , 0.9929121 ,
                0.99229222, 0.99545335, 0.99605659, 0.99301768, 1.00251889,
                1.00894444, 1.03090775, 1.02658006, 1.03195238, 1.03133658,
                1.05177308])
```

因此，我尝试不考虑无风险投资，即只考虑股票的投资组合，最后画出来的曲线相当漂亮。

可以看到，在图的前半部分，Markowitz 解出来的投资组合与大盘趋势基本一致，但在 2020 年后计算出来的投资组合的收益持续走低，与大盘不符，猜测是因为 2020 年后市场发生了较大变化，用过去 5 年的历史数据计算出来的均值与协方差有较大误差，根据过去的历史数据解出的最优投资组合并不是最优投资组合，因此持续亏钱；另一方面，通过查询股票代码，这 50 支所选股票是一些上市已久的传统行业，在近几年的 HS300 的权重占比并不大，因此这 50 支股票的投资组合并不能很好地代表市场。另外，我所采用的 log-rate 计算日收益率可能也有一定影响，log-rate 在日收益率较小时近似等于日收益率，但是市场波动比较大，有时能够达到较大的日收益率，这时候采用 log-rate 就会有一定误差。

Question 2

将 CAPM 视作单因子模型，有以下式子：

$$R_i - R_f = \alpha_i + \beta_i(R_M - R_f) + \varepsilon_i$$

因此，可以沿用第一题的数据，随机选取 6 支股票后，从 stock_return 里选取相应的列得到 6 支股票的 return，并用 HS300_return 代替上式中市场收益率，从而可以用 statsmodels 模块里的类来解线性回归模型参数，并对参数α进行假设检验（t 检验），考虑要不要拒绝 $\alpha = 0$ 的假设。最终结果如图所示：

	code	alpha	beta	p-value for alpha=0	reject
0	601006.SH	-0.000189	0.898469	0.459831	False
1	601107.SH	-0.000302	0.966669	0.365968	False
2	601168.SH	-0.000241	1.072155	0.518450	False
3	601601.SH	0.000106	1.032240	0.767262	False
4	601618.SH	-0.000823	0.958837	0.235810	False
5	601628.SH	-0.000135	1.024329	0.652241	False

从上图可知，对于这 6 支股票我们都没有足够证据推翻原假设。

Question 3

3.1 欧式期权

首先，根据 Cox-Ross-Rubinstein 1979 提供的结果计算上升因子与下降因子：

$$u = \exp\left(\sigma\sqrt{\Delta t}\right)\ ,\ d = 1/u$$

再根据 u,d, 无风险利率 r 计算 p 值：

$$p = (e^{r\Delta t} - d)/(u - d)$$

之后，对于二叉树的每一颗子树，都可以根据其子节点的期望收益决定其父节点的期望收益：

$$f = e^{-rT} \cdot (p \cdot f_u + (1-p) \cdot f_d)$$

如此以来，可以从二叉树的叶子层出发逐步倒推，直到得到根节点的期权定价。另一方面，使用 Black-Scholes 公式计算期权价格的公式如下：

$$p = Ke^{-rT}N(-d_2) - S_0N(-d_1)$$

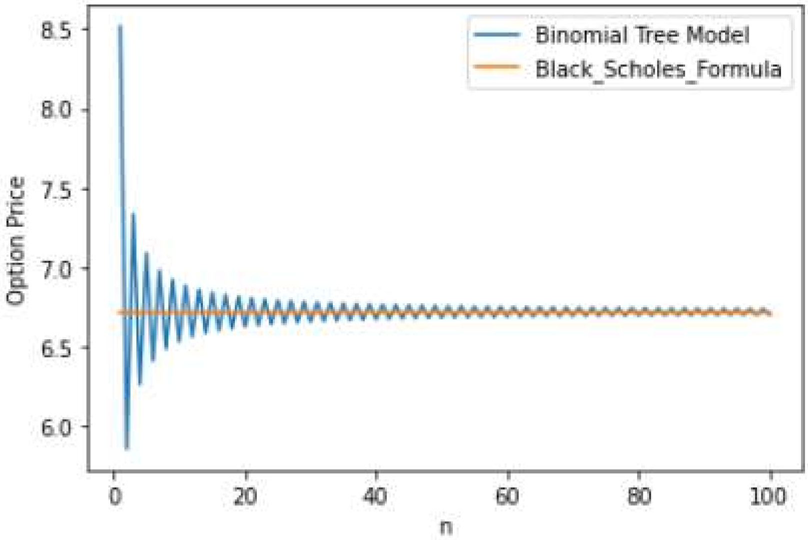
$$d_1 = \frac{\log\left(\frac{S_0}{K}\right) + \left(r - \frac{\sigma^2}{2}\right) \cdot T}{\sigma \cdot \sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

代入题中各值，计算结果如下：

	m=10	m=50	m=100
Binomial Tree Model	6. 534	6. 683	6. 702
Black-Scholes Formula	6. 721		

进一步，让 m 从 1 到 100 遍历，绘制出的期权定价曲线如图所示：



可以看到，二叉树的模型计算出的期权价格近似收敛到 BS 公式计算得到的定价结果。

3.2 美式期权

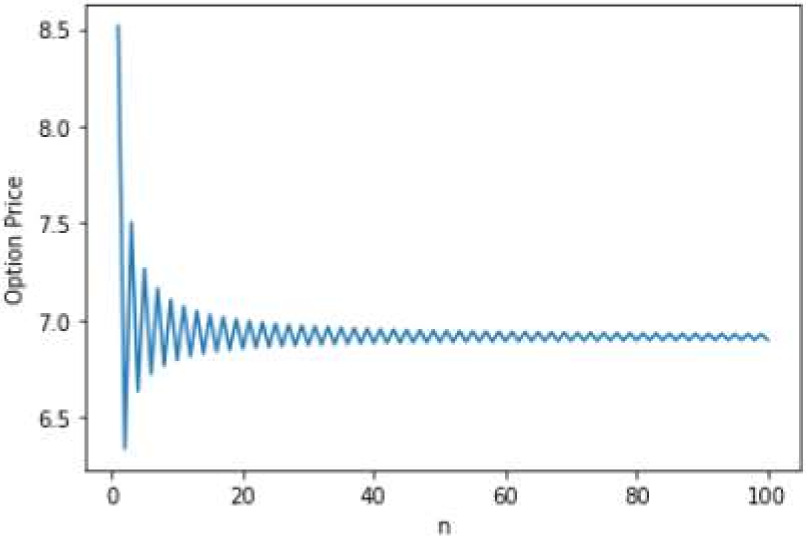
基本类似上述欧式期权的求法，不同之处在于，美式期权要决定是否要提前行权，因此对于一个父节点，需要比较由其子节点加权后按无风险利率折现计算得到的价格是否低于现在马上行权得到的收益，若是则立刻行权，因此父节点期权价格计算公式变成：

$$f_1 = e^{-rT} \cdot (p \cdot f_u + (1-p) \cdot f_d$$

$$f_2 = max\,(0, k-p)$$

$$f = max(f_1, f_2)$$

最后得到的结果如图所示：



$m = 100$ 时美式期权按照二叉树模型的定价是 6.897，略大于 m=100 时欧式期权的定价 6.702，这是美式期权能够提前行权带来的额外收益。