大数据思维作业2

金融八班 倪晨豪 181910839

1. 前言

本文选取了中国医药与万科的从2015年1月1日起到2020年12月1日的每日股价为主要数据来源。对中国医药与万科的对数收益率分别进行时间序列建模以及模型检验，选出各自的最优模型并进行比较。

1. 中国医药和万科股票的概述

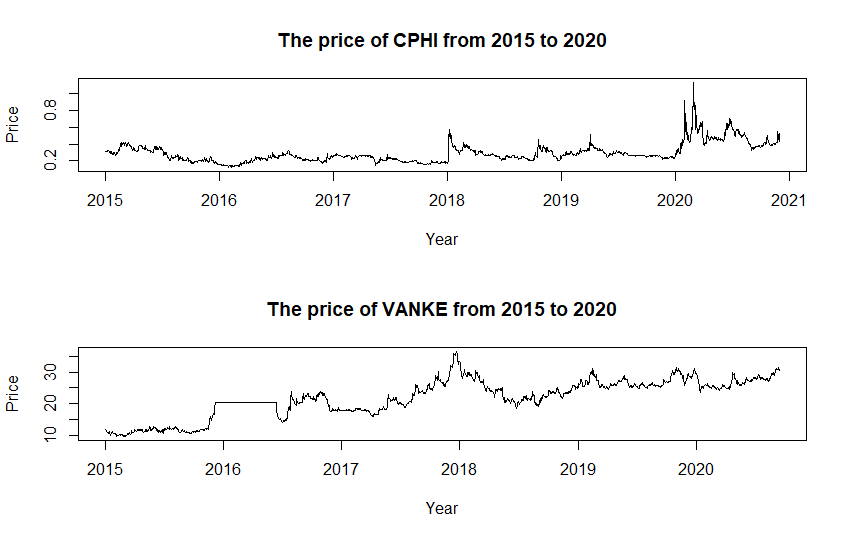


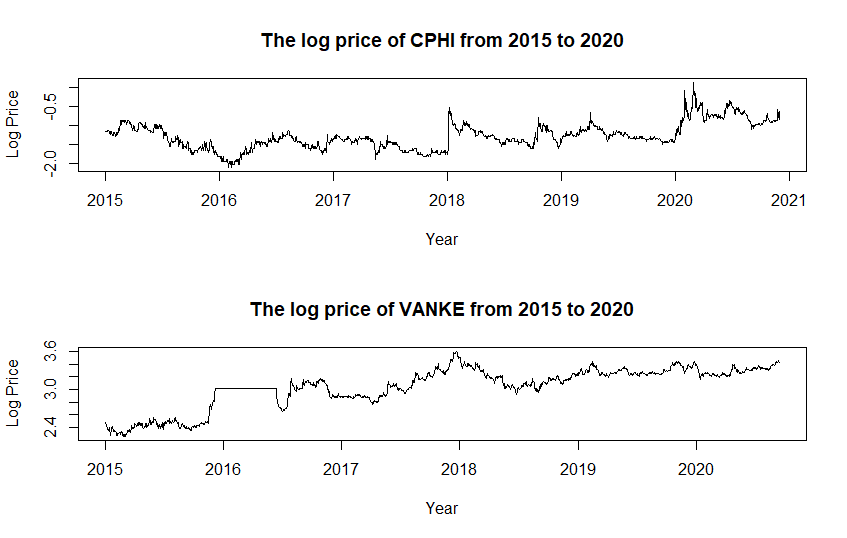
图 1 2015年到2020年中国医药与万科的价格

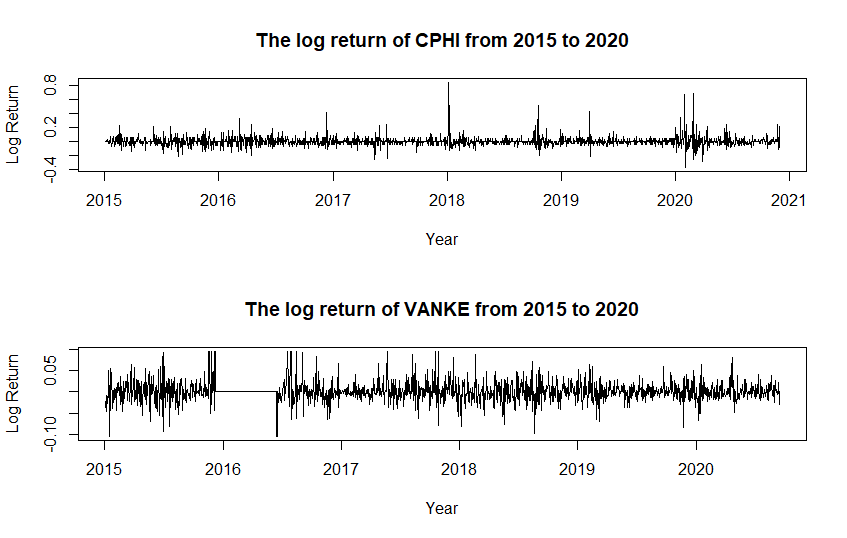
图 2 2015年到2020年中国医药与万科的对数价格

图 3 2015年到2020年中国医药与万科的对数收益率

（一）中国医药股票概述

从中国国药的股价来看可见，在2020年前始终处在较为稳定的位置，实现较低的平均收益率；但是在2020年后股价大幅度上涨，实现了高收益率。可见在疫情期间，医药行业确实实现了短期的崛起，同时也反映出抗疫的重要性。  
可见2020年前的收益率较为稳定，基本在平均值附近浮动，但是在2020年后的较短时间内收益率出现了较大幅度的波动，出现了极端值从而使得收益率的分布产生了肥尾。

（二）万科股票概述

从万科的股价来看，在2018年7月前万科的股价处于大幅度下跌的状态，在2020年1月后经历了两次短期的大幅度下降。从股价上可见疫情对房地产行业带来了一定的冲击，导致股价受挫；此外，前期的万科股价下跌对疫情期间的股价也会产生一定的影响，在对抗风险以及资金充足率上可能准备不足，无法应对系统性风险。

万科收益率的整体波动幅度较大，标准差达到了0.022，但是极端值出现的概率较小，大部分的波动较为稳定。

1. 中国医药的时间序列模型分析

（一）模型选择

首先，对中国医药股票的对数收益率画出ACF以及PACF图并进行定阶，求出AR模型以及MA模型的AIC值；再通过迭代法求出ARMA最优阶数，对三者的AIC值比较并选择中国医药对数收益率的最优时间序列模型。

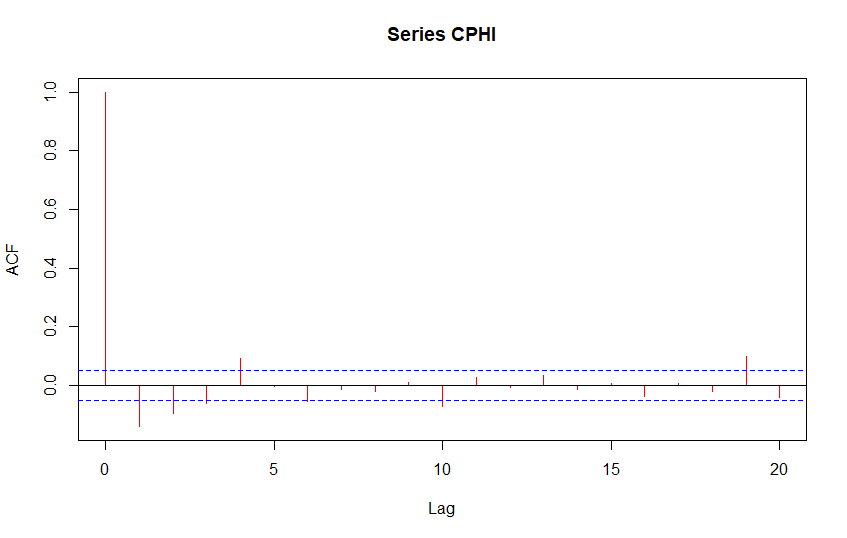


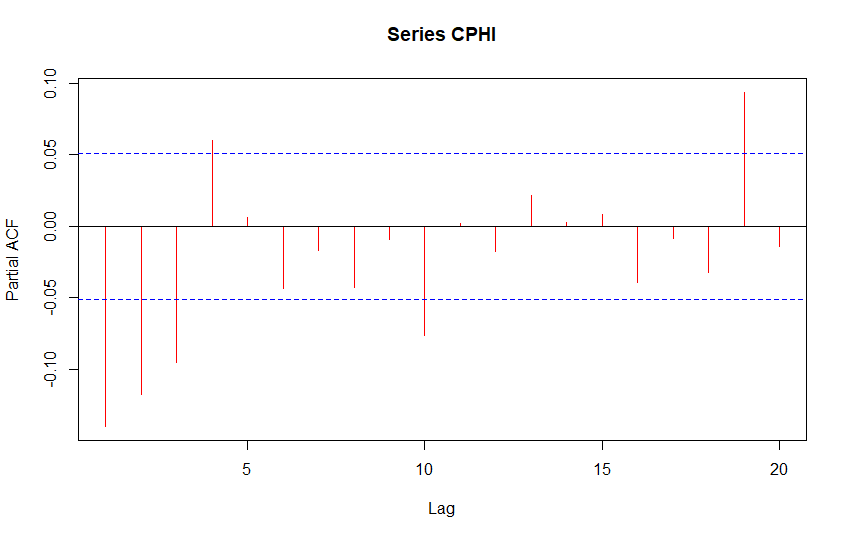
图 4 CPHI的自相关图

图 5 CPHI的偏自相关图

[,1] [,2] [,3] [,4] [,5]  
 [1,] -3615.481 -3615.238 -3619.870 -3622.665 -3630.009  
 [2,] -3616.177 -3617.972 -3625.263 -3623.600 -3629.443  
 [3,] -3624.508 -3626.116 -3624.139 -3622.243 -3627.444  
 [4,] -3624.558 -3624.133 -3631.246 -3629.454 -3627.389  
 [5,] -3622.545 -3622.445 -3629.548 -3627.294 -3633.145

表 1 CPHI的ARMA矩阵

|  |  |  |  |
| --- | --- | --- | --- |
| Model | AR | MA | ARMA |
| AIC | -3628.48396406527 | -3626.03440594644 | -3634.55786794621 |

表 2 CPHI的模型的AIC

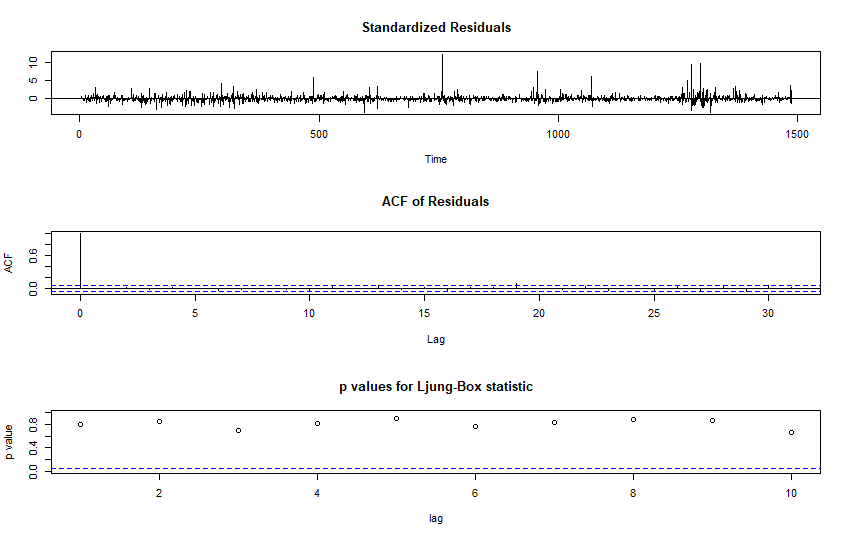
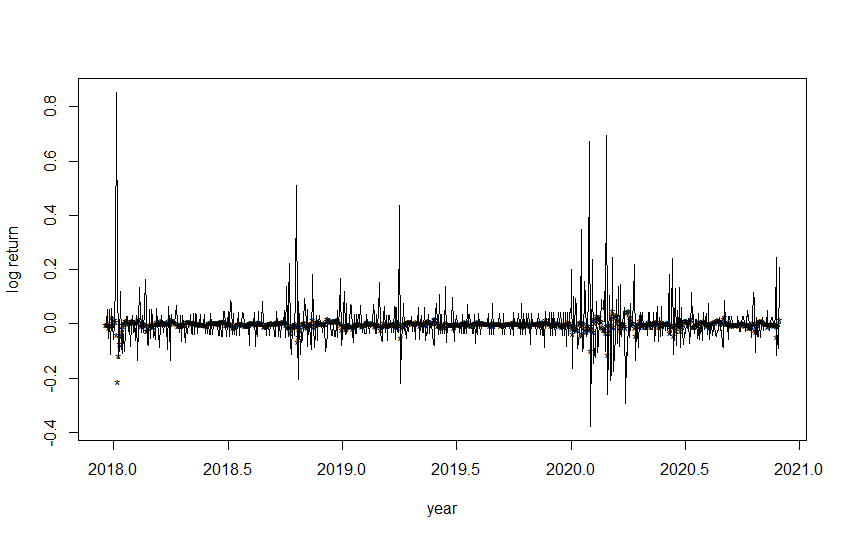


图 6 CPHI的ARMA（5,5）模型的残差检验图

从图4和图5中可得，由ACF以及PACF图像分别对AR以及MA模型定阶为4阶；同时从表1可得ARMA(5,5)为迭代所得ARMA模型中的最优模型；分别比较AR(4)、MA(4)以及ARMA(5,5)模型的AIC值，ARMA(5,5)为中国医药股票对数收益率的最优时间序列模型。如图6所示，ARMA（5,5）的残差不存在自相关性，属于白噪声。因此，选择ARMA（5,5）模型为中国医药的时间序列模型。

（二）中国医药的模型样本外预测

|  |  |  |  |
| --- | --- | --- | --- |
| Model | AIC | RMSFE | MAFE |
| AR(4) | -3628 | 0.08155 | 0.04389 |
| MA(4) | -3626 | 0.08165 | 0.04391 |
| ARMA(1,1) | -3617 | 0.08194 | 0.04387 |
| ARMA(5,5) | -3635 | 0.08187 | 0.04432 |



由上述的图表所示，ARMA(5,5)模型的AIC值最小，AR模型的RMSFE最小，ARMA(1,1)的MAFE最小；由于ARMA(1,1)与AR模型的MAFE非常接近，因此认为AR模型的MAFE也为最小；所以AR模型的表现较好。综上可得，越简单的模型样本外数据表现可能越好。

1. 万科的时间序列模型分析

（一）模型选择

首先，对万科股票的对数收益率画出ACF以及PACF图并进行定阶，求出AR模型以及MA模型的AIC值；再通过迭代法求出ARMA最优阶数，对三者的AIC值比较并选择中国医药对数收益率的最优时间序列模型。

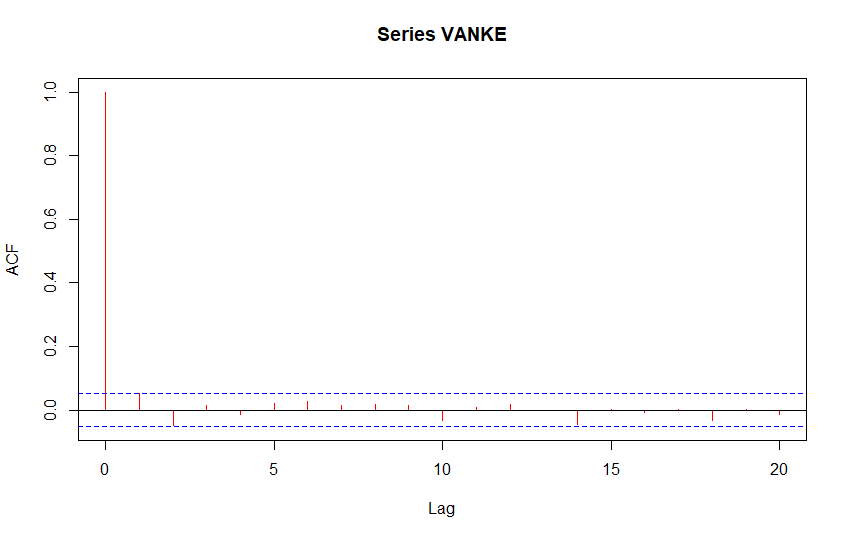


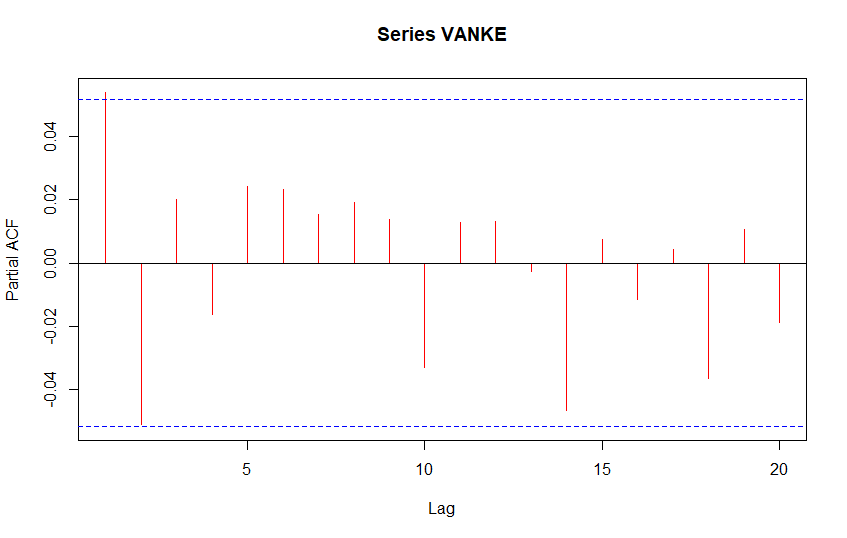
图 7 VANKE的自相关图

图 8 VANKE的偏自相关图

[,1] [,2] [,3] [,4] [,5]

[1,] -6608.959 -6607.091 -6604.824 -6603.157 -6603.028

[2,] -6607.092 -6604.934 -6609.688 -6604.504 -6602.600

[3,] -6605.067 -6603.098 -6601.133 -6603.618 -6598.784

[4,] -6603.167 -6604.468 -6609.545 -6608.282 -6597.819

[5,] -6603.028 -6605.121 -6597.949 -6597.768 -6599.185

表 3 VANKE的ARMA模型矩阵

|  |  |  |  |
| --- | --- | --- | --- |
| Model | AR | MA | ARMA |
| AIC | -6608.90563808812 | -6609.34549968693 | -6608.53392361615 |

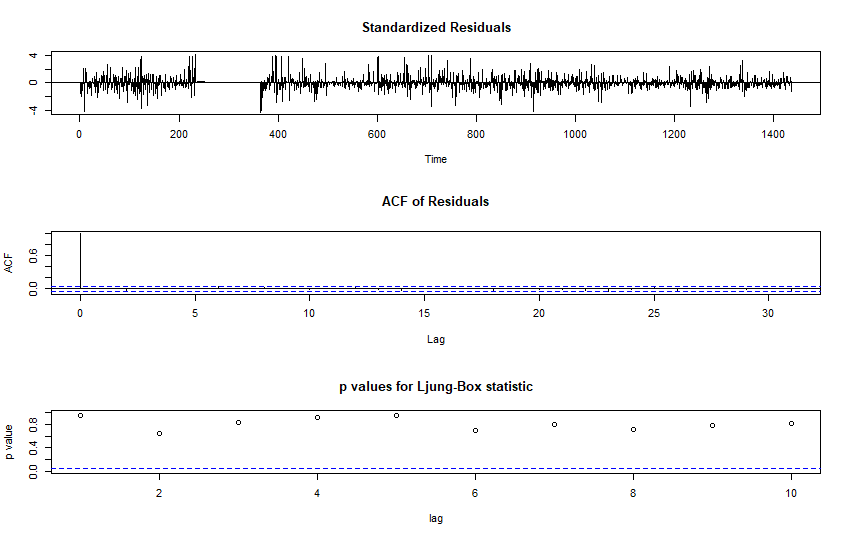
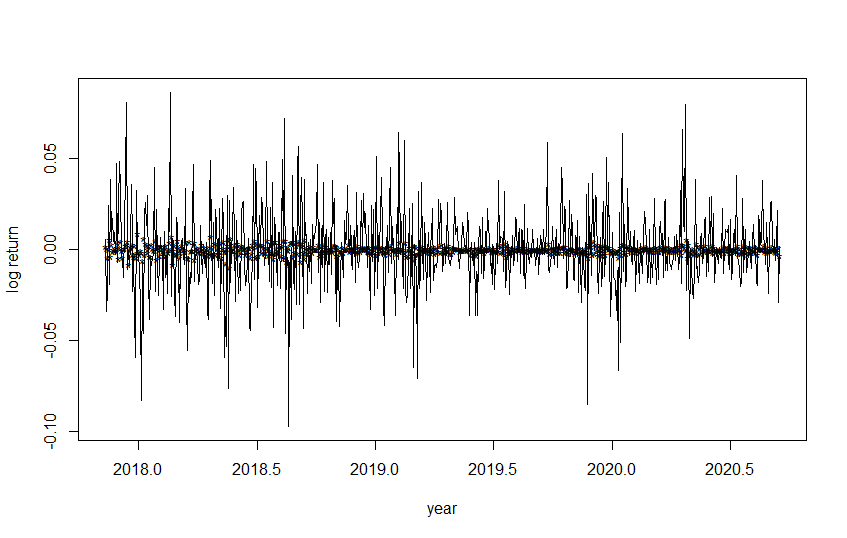
表 4 VANKE的模型的AIC

图 9 VANKE的ARMA（2,3）模型的残差检验图

如图7和图8所示，由ACF以及PACF图像分别对AR以及MA模型定阶为2阶，同时从表3中可得ARMA(2,3)为迭代所得ARMA模型中的最优模型；分别比较AR(2)、MA(2)以及ARMA(2,3)模型的AIC值，ARMA(2,3)为万科股票对数收益率的最优时间序列模型。如图9所示，ARMA（2,3）的残差不存在自相关性，属于白噪声。因此，选择ARMA（2,3）模型为万科最优的时间序列模型。

（二）万科的模型样本外预测

|  |  |  |  |
| --- | --- | --- | --- |
| Model2 | AIC | RMSFE | MAFE |
| AR(2) | -6609 | 0.02216 | 0.01623 |
| MA(2) | -6609 | 0.02213 | 0.01621 |
| ARMA(1,1) | -6610 | 0.02207 | 0.01619 |
| ARMA(2,3) | -6609 | 0.02212 | 0.01618 |



如上述的图表所示，ARMA(1,1)模型的AIC值最小，ARMA(1,1)模型的RMSFE最小，ARMA(2,3)的MAFE最小；ARMA(1,1)的AIC与RMSFE都为最小，同时各模型之间平均绝对误差相差较小；因此，ARMA(1,1)模型表现较好。综上所述，越简单的模型样本外数据表现可能越好。

1. 结论

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Stock | Model | AIC | RMSFE | MAFE |
| CPHI | AR(4) | -3628 | 0.08155 | 0.04389 |
| . | MA(4) | -3626 | 0.08165 | 0.04391 |
| . | ARMA(1,1) | -3617 | 0.08194 | 0.04387 |
| . | ARMA(5,5) | -3635 | 0.08187 | 0.04432 |
| VANKE | AR(2) | -6609 | 0.02216 | 0.01623 |
| . | MA(2) | -6609 | 0.02213 | 0.01621 |
| . | ARMA(1,1) | -6610 | 0.02207 | 0.01619 |
| . | ARMA(2,3) | -6609 | 0.02212 | 0.01618 |

通过AIC准则、RMSFE以及MAFE比较上述两个股票的八个模型的表现可得，万科的模型表现更好，同时越简单的模型样本外表现越好。