

基于时间序列的股票投资方案

[摘要] 想要从股票中获取最大的利益，则要对于股票进行分析及预测并制定一定的投资方案，依据股票的历史价格建立适当的模型是一种可行的方法。本文主要阐述基于 HP 滤波以及 ARMA 的模型方法。选取了来自 DC 竞赛网站中关于量化投资竞赛中 roundA 的 111 支不同股票的 488 天收盘价的数据。使用 python 语言，对数据进行 HP 滤波，分离出两个时间序列，分别进行建模，并对模型进行检验及预测，最后筛选出收益较好的股票。

[关键词] 时间序列 ARMA 模型 HP 滤波 python 语言

Investment plan based on time series

[Abstract] It is a feasible way to analyze and forecast the stock and formulate a certain investment plan so that I can get the most benefit from the stock. There is a way to establish an appropriate model based on the historical price of the stock. This paper mainly describes the model method based on HP filter and ARMA. Data from the DC contest website for the 488-day closing price of 111 different stocks in roundA in the quantitative investment competition were selected. By using python, the data are divided into two time-series by applying HP filter. Then, both time-series are modeled separately, and both models will be tested and predicted. Finally, the stocks which have better returns will be selected.

[Key words] Time-series ARMA HP-filter python

目录

引言	1
1. 绪论.....	1
1.1 研究背景及意义	1
1.2 文献综述	1
2. 基础知识.....	2
2.1 时间序列及模型	2
2.1.1 时间序列及基本概念.....	3
2.1.2 ARMA 模型	4
2.1.3 收益率和波动率.....	6
2.1.4 检验及评估方法.....	7
2.2 Hodrick-Prescott 滤波	9
3. 股票价格预测的时间序列模型方法.....	11
3.1 模型结构	11
3.2 模型实证	12
3.3 模型应用	23
总结	24
致谢	25
[参考文献]	26

引言

股票(stock)，简单来说就是股份公司发行给各个股东的所有权凭证（有价证券），其目的在于筹集资金谋发展，根据此凭证，股东可以取得股息和红利，当然股东也可以将其进行转让或买卖从而获利。随着国家综合实力的增强，国人的工资水平以及生活水平不断水涨船高，人们可支配财产也在不断地增加，从而使得人们对财务的管理越加注重，而其中有一些追求高回报的人，会选择股票，基金等高风险的理财产品。股市随之愈发繁荣，越来越多的投资者想要参与其中。当然，从中获取最多的利润是人们最原始和直接的目的。

1. 绪论

1.1 研究背景及意义

要让在股市中获利达到最大化，则需要对投资方案进行一个合理的选择，这就要求人们要对于股票的历史数据或其他因素进行充分有效的分析，从而进行股票的预测，当投资者能够相对比较准确的预测出某股票在某一阶段的价格的时候，无疑可以获得丰厚的收益。

同时，在 1970 年尤金·法玛提出的有效市场假说中表明，只要证券市场的价格充分并且及时反映了全部有价值的信息，则市场的价格真实地代表了证券的价值。然而在中国的证券市场中，因为每个投资者获取到的信息并不相同，并且各种环境因素不断地影响着投资者们，从而使得股价普遍偏离公司的实际价值。因此有效市场假说不适用于我国的证券市场，所以根据股票的历史价格建立适当的模型对股票未来的价格进行合理的预测是可行的。^[1]

那么，建立一个精确度较高和广泛适用于多种类型股票的股票价格预测模型，且在此模型基础上给出合理的股票投资方案，这对于股票投资新手有着一定的指导价值，也为股市老手提供了一些参考。

1.2 文献综述

方子良（1999）提出将时序法应用到股市分析中，并表明将时间序列分析

法应用于股市行情进行技术分析是切实可行的，该方法能够客观真实地反映被研究对象。^[2]

付燕和栗锋（2012）应用 ARMA 模型对我国体育股票价格预测分析，对 2002/1/1 到 2010/6/30 的中体产业股票数据进行模型拟合预测，最终得到模型预测值与实际值的差异非常小。^[3]

孟坤和李丽（2016）应用 ARMA 模型对 2015/1/5 至 2015/12/18 的上证综合指数的日收盘价进行拟合预测，最终结果表明，ARMA 模型方法对于预测时间序列数据具有很好的效果。^[4]

王丹和冯长焕（2018）应用 HP 滤波以及 ARIMA-ARCH 模型对我国 GDP 进行分析和预测，并对 HP 滤波分解前后情况进行比较，提出对分解后的序列进行建模可以提高最终的预测精度。^[5]

吴齐等人（2015）基于 HP 滤波和 ARIMA 模型对我国 GDP 进行分析预测中，其建立的 ARIMA 模型能够很好地反映 GDP 的变动情况，最终预测得到的平均相对误差也比较低。^[6]

赵庆（2014）应用 HP 滤波及 ARMA 模型对国际原油期货的价格进行预测，选取了 2003/8/1 至 2013/4/19 的 WTI 现货价格的周数据进行拟合预测，文章结论表明，该模型能精确拟合数据，预测精度比其他方法有所提高，并且该建模方法更简单易用，方便推广。^[7]

综上，提出将 HP 滤波及时间序列模型 ARMA 应用到本文中，建立一个精确度较高、广泛适用于多种股票且方法较为简易的股票收盘价预测模型。

2. 基础知识

本章主要详细介绍后续用到的知识。

2.1 时间序列及模型^[8, 9, 10, 11]

社会经济活动千变万化，然而许多现象的发展都具有随着时间而演变的特征，所以认识社会经济现象的数量规律离不开对于已知数据的分析。从历史数据中反映现象，是社会经济统计分析中不可分割的重要任务。

2.1.1 时间序列及基本概念

对于一些反映社会、经济和自然等情况的现象，在其某一数量指标进行时间上的观察后得到一组数据，被称为时间序列数据。把这些观察到的数据按照时间顺序依次排列，随之形成的序列叫做时间序列。

定义一组股票价格的时间序列 $P_T = \{p_1, p_2, p_3, \dots, p_T\}$ ， T 为时间点的数量。

简单来说，若股票价格数据在一定的水平线上下进行波动，且波动范围基本不变，认为该时间序列是一个“弱平稳序列”。具体来说，若时间序列 P_T 存在有限的二阶矩且满足以下：(1) $E(P_T) = \mu$ 与 t 无关；(2) $Var(P_T) = \gamma_0$ 与 t 无关；(3) $\gamma_k = Cov(p_{t-k}, p_t)$ $k=1, 2, \dots$ 与 t 无关，则称 P_T 为弱平稳序列。

其中， $\gamma_k = Cov(p_{t-k}, p_t)$ $k=0, 1, 2, \dots$ 为时间序列 P_T 的自协方差函数。

若 P_T 为弱平稳序列， $\{\gamma_k\}$ 为自协方差函数序列。则有

$$\rho(p_{t-k}, p_t) = \frac{Cov(p_{t-k}, p_t)}{\sqrt{Var(p_{t-k})Var(p_t)}} = \frac{\gamma_k}{\sqrt{\gamma_0\gamma_0}} = \frac{\gamma_k}{\gamma_0} \quad (2-6)$$

记 $\rho_k = \frac{\gamma_k}{\gamma_0}$ ，这是 p_{t-k}, p_t 的相关系数且与 t 无关，称 $\{\rho_k\}, k=0, 1, \dots, \rho_0=1$ 为时间序列 P_T 的自相关函数 ACF (Autocorrelation function)。

若 P_T 为弱平稳序列且满足 $\rho_k = 0, k=1, 2, \dots$ ，则称 P_T 为白噪声序列。

在现实生活的经济金融活动中，得到的时间序列大都具有非平稳性，因此要将这些非平稳序列通过平稳化操作后，才能将其用于经典时间序列模型，并建立起一系列模型识别、估计和检验的方法。最常见的平滑化方法为差分运算：若 P_T 是非平稳的序列，则要在估计模型前，将 P_T 进行一阶差分，得到一组新的差分时间序列 $\{\Delta p_t\}$ ，其中：

$$\Delta p_t = p_t - p_{t-1}, \quad t=2, \dots, T \quad (2-7)$$

若一阶差分仍不能满足平稳，则要进行二阶或二阶以上差分，即在之前的差分基础上再进行一阶差分，依次类推。

2.1.2 ARMA 模型

若 P_t 为平稳序列，则有

$$p_t = \sum_{i=1}^p \phi_i p_{t-i} + \varepsilon_t \quad (2-8)$$

其中 $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ 。式(2-8)也被称为 p 阶自回归模型，记作 AR(p)，其表示了 t 时刻的序列值 p_t 可以用前 p 个值进行线性表示，最后加上一个随机干扰项，且干扰项与各个 p_t 相互独立。可引入滞后算子 B，将 AR(p) 模型简记为 $\Phi(B)p_t = \varepsilon_t$ ，其中 $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ 。

由于阶数 p 要进行进一步的确定，据此，引入偏自相关函数的概念，设 X_1, X_2, \dots, X_n, Y 伪随机变量，有

$$L(Y | X_1, \dots, X_n) = \arg \min_{\hat{Y} = b_0 + b_1 X_1 + \dots + b_n X_n} E(Y - \hat{Y})^2 \quad (2-9)$$

称作 X_1, X_2, \dots, X_n 对 Y 的最优线性预测。 $Y - L(Y | X_1, \dots, X_n)$ 称为 Y 在扣除 X_1, X_2, \dots, X_n 影响后的偏相关系数。对于平稳线性时间序列， $n = 1, 2, \dots$ ，有

$$L(X_t | X_{t-1}, \dots, X_n) = \phi_{n0} + \phi_{n1} X_{t-1} + \dots + \phi_{nn} X_{t-n} \quad (2-10)$$

其中 $\phi_{nj}, j = 0, 1, \dots, n$ 与 t 无关，称 ϕ_{nn} 为时间序列 $\{X_n\}$ 的偏自相关系数， $\{\phi_{nn}\}$ 为时间序列 $\{X_n\}$ 的偏自相关函数 PACF (partial autocorrelation function)。

ϕ_{nn} 使用样本进行估计，得到 $\hat{\phi}_{nn}$ 称为样本偏自相关函数。AR 模型的 PACF 具有截尾性，因此可以以此得到阶数 p。

若 P_t 为平稳序列，则有

$$p_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2-11)$$

其中 $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ 。式(2-11)也被称为 q 阶滑动平均模型，记作 MA(q)，其认为 t 时刻的序列值 p_t 可以用前 q 个误差项来表示，最后加上一个随机干扰项。可引入滞后算子 B，将 MA(p) 模型简记为 $p_t = \Theta(B)\varepsilon_t$ ，其中

$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ 。MA 模型的 ACF 具有截尾性，因此可以以此得到阶数 q 。

若 P_t 为平稳序列，则有

$$p_t = \sum_{i=1}^p \phi_i p_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2-12)$$

其中 $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ 。式(2-12)也被称为 (p, q) 阶自回归滑动模型，常记作

ARMA (p, q) ，表示为 t 时刻的序列值 p_t 可以用前 p 个序列值和前 q 个误差值的线性组合加上一个随机项得到。可引入滞后算子 B ，将 ARMA (p) 模型简记为

$$\Phi(B)p_t = \Theta(B)\varepsilon_t, \text{ 其中 } \Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q,$$

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p. \text{ AR 和 MA 的定阶方法在 ARMA 模型上仍可使用。}$$

使用最小二乘估计方法对模型的参数进行估计。对于模型(2-12)，未知参数有 $p+q+1$ 个，记 $\tilde{\beta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$ ， $F_t(\tilde{\beta}) = \phi_1 p_{t-1} + \dots + \phi_p p_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ ，则 $\varepsilon_t = p_t - F_t(\tilde{\beta})$ ，残差的平方和为：

$$Q_t(\tilde{\beta}) = \sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (p_t - \phi_1 p_{t-1} + \dots + \phi_p p_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})^2$$

使 $Q_t(\tilde{\beta})$ 达到最小，得到的参数值为 $\tilde{\beta}$ 的最小二乘估计值。由于随机误差不可观测，所以 $Q_t(\tilde{\beta})$ 不是 $\tilde{\beta}$ 的显性函数，只能借助迭代法来求出模型未知参数的最小二乘估计值。最小二乘估计的精度很高，因为它将时间序列包含的信息进行了充分的使用。

还可以使用极大似然准则来对模型未知参数进行估计。在使用该方法之前，必须知道序列总体的分布函数，但时间序列的总体分布是未知的，所以通常假设时间序列服从多元的正态分布。有 $\tilde{p} = (p_1, \dots, p_n)$ 其似然函数为：

$$\begin{aligned} L(\tilde{\beta}; \tilde{p}) &= f(p_1, p_2, \dots, p_n; \tilde{\beta}) \\ &= (2\pi)^{-\frac{n}{2}} |\sigma_\varepsilon^2|^{-\frac{n}{2}} |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{\tilde{p}' \Omega^{-1} \tilde{p}}{2\sigma_\varepsilon^2} \right\} \end{aligned}$$

对数似然函数为：

$$l(\tilde{\beta}; \tilde{p}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln|\Omega| - \frac{1}{2\sigma_\varepsilon^2} [\tilde{p}' \Omega^{-1} \tilde{p}]$$

求未知参数的偏导数，令其等于 0，得到似然方程组：

$$\begin{cases} \frac{\partial}{\partial \sigma_\varepsilon^2} l(\tilde{\beta}; \tilde{p}) = -\frac{n}{2\sigma_\varepsilon^2} + \frac{S(\tilde{\beta})}{2\sigma_\varepsilon^4} = 0 \\ \frac{\partial}{\partial \tilde{\beta}} l(\tilde{\beta}; \tilde{p}) = -\frac{1}{2} \frac{\partial \ln|\Omega|}{\partial \tilde{\beta}} - \frac{1}{\sigma_\varepsilon^2} \frac{\partial S(\tilde{\beta})}{2\partial \tilde{\beta}} = 0 \end{cases} \quad (2-13)$$

其中， $S(\tilde{\beta}) = \tilde{p}' \Omega^{-1} \tilde{p}$ ， $\Omega = \begin{pmatrix} \sum_{i=0}^{\infty} G_i^2 & \cdots & \sum_{i=0}^{\infty} G_i G_{i+n-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^{\infty} G_i G_{i+n-1} & \cdots & \sum_{i=0}^{\infty} G_i^2 \end{pmatrix}$ ， G 为 ARMA 模型的格

林函数。由于 $S(\tilde{\beta})$ 和 $\ln|\Omega|$ 都不是 $\tilde{\beta}$ 的显示表达式，所以未知参数极大似然估计只能通过复杂的迭代算法得到。由于极大似然估计使用到每一个时间序列数据所提供的信息，从而使得它的估计精度较高，还使得估计还具有许多优秀的统计性质，如：一致性、渐进正态性和渐进有效性等，视其为较优秀一种的参数估计方法。

当我们将时间序列的模型估计并检验完毕后，我们便可根据模型进行后续的预测，即

$$\tilde{p}_{T+1}, \cdots, \tilde{p}_{T+m} = f(p_1, p_2, \cdots, p_T) \quad (2-13)$$

其中，只 \tilde{p}_{T+1} 称为单步预测，若预测 \tilde{p}_{T+m} ， $m > 1$ 则称为 m 步预测， f 为估计得到的函数。

2.1.3 收益率和波动率^[9]

有一组股票价格的时间序列 P_t ， p_t 代表该股票在 t 时刻的价格。

有单期简单净收益率（简单收益率）：

$$R_t = \frac{p_t}{p_{t-1}} - 1 = \frac{p_t - p_{t-1}}{p_{t-1}} \quad (2-14)$$

可得 k 期净收益率（复合收益率）：

$$R_t[k] = \frac{p_t}{p_{t-k}} - 1 = \frac{p_t - p_{t-k}}{p_{t-k}} \quad (2-15)$$

假定价格是连续的变化，定义价格对数变动公式：

$$R_t = \ln \frac{p_t}{p_{t-1}} = \ln p_t - \ln p_{t-1} \quad (2-16)$$

得到历史波动率的估计值：

$$\sigma = \sqrt{\frac{\sum (R_t - \bar{R})^2}{T-1}} \quad (2-17)$$

$$\bar{R} = \frac{1}{T} \sum R_t$$

其中 R_t 是使用对数价格变动公式获得的值。若使用每日的数据进行计算，则 σ 代表日收益率，为对数收益的条件标准差。

2.1.4 检验及评估方法

为判断一个序列是否平稳，需要用到平稳性检验：ADF（Augmented Dicky-Fuller Test）检验。对于一个 AR(p) 过程 $p_t = \phi_1 p_{t-1} + \dots + \phi_p p_{t-p} + \varepsilon$ ，考虑一个基础的模型：

$$p_t = \beta p_{t-1} + \sum_{j=1}^{p-1} \Psi_j \Delta p_{t-j} + e_t \quad (2-18)$$

其中， $\beta = \phi_1 + \phi_2 + \dots + \phi_p$ ， $\Psi_j = -\phi_{j+1} - \phi_{j+2} - \dots - \phi_p$ ，当 $\beta = 1$ 时，就是差分序列 $\{\Delta p_t\}$ 的 AR(p-1) 模型，当 $\beta < 1$ 时，就是 $\{p_t\}$ 的 AR(p) 模型，考虑如下原假设和备择假设：

$$H_0: \beta = 1 \leftrightarrow H_a: \beta < 1$$

取检验统计量：

$$ADF = \frac{\beta - 1}{SE(\hat{\beta}_1)} \quad (2-19)$$

统计量 ADF 充分小则拒绝 H_0 ，序列是平稳的。

构造时间序列模型 ARMA 时，(p, q) 的阶数通过观察序列的 ACF 和 PACF 得

到，这一步尤为重要。但观察到的不一定是最佳的模型阶数，所以一般使用最小信息量准则（AIC, an information criterion）来寻找最佳模型。该衡量准则的指导思想，是考虑拟合程度的似然函数值以及模型中未知参数的个数。拟合程度的似然函数值越大，代表模型拟合效果越好，但这会增加未知参数的个数，这时会导致未知的风险和估计的难度加大，因此要在拟合精度和未知参数个数中进行考究衡量。于是有以下加权函数：

$$AIC = -2\ln(\text{模型极大似然函数值}) + 2(\text{模型中未知参数个数}) \quad (2-20)$$

目标最优函数应该使得 AIC 函数尽可能的小。ARMA(p, q) 的对数似然函数为：

$$l(\tilde{\beta}; p_1, p_2, \dots, p_n) = -\left[\frac{T}{2} \ln(\sigma_\varepsilon^2) + \frac{1}{2} \ln|\Omega| + \frac{1}{2\sigma_\varepsilon^2} S(\tilde{\beta}) \right] \quad (2-21)$$

因为 $\frac{1}{2} \ln|\Omega|$ 有界， $\frac{1}{2\sigma_\varepsilon^2} S(\tilde{\beta}) \rightarrow \frac{T}{2}$ ，有对数似然函数与 $-\frac{T}{2} \ln(\sigma_\varepsilon^2)$ 成正比。模型

未知参数个数为 $p+q+1$ 个，所以 ARMA(p, q) 的 AIC 函数为：

$$AIC = T \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q + 1) \quad (2-22)$$

AIC 虽然很方便，但其也有不足之处，它所确定的阶不能收敛到模型真实的阶，而是比真实的阶要高。因此在 AIC 基础上，提出了 BIC 准则：

$$BIC = T \ln(\hat{\sigma}_\varepsilon^2) + k \ln T \quad (1-1)$$

即用 $\ln T$ 代替了 AIC 中的 2，因为 $\ln T > 2$ 总是成立的，所以使用 BIC 准则得到的阶数一般都要比 AIC 准则得到的阶数要低。

对模型进行识别和估计后，要对拟合好的模型进行必要的显著性的检验：LB 检验，其原理在于考虑得到的模型残差序列是否被识别为白噪声序列，即残差序列是否仍包含了原序列的相关信息。对残差序列进行检验，考虑一下原假设和备择假设：

$$\begin{aligned} H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0, \forall m \geq 1 \\ H_1 : \text{至少存在某个 } \rho_k \neq 0, \forall m \geq 1, k \leq m \end{aligned}$$

ρ 代表残差序列的自相关函数，取检验统计量 LB(Ljung-Box)：

$$LB = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{(T-k)} \sim \chi^2(m), \forall m \geq 1 \quad (2-20)$$

得到统计量 LB 的 p 值，若 p 值小于 0.05，应该拒绝 H_0 ，认为模型残差存在自相关，若 p 值大于 0.05，则接受 H_0 ，说明模型是显著有效的，p 值越大，越能够接受原假设。

最后，得到模型后，进行预测。对于预测出的 $\hat{p}_i, i=1, \dots, m$ ，将其与实际值 p_i 进行比较，定义以下评价指标：

$$MSE = \frac{1}{m} \sum_{i=1}^m (p_i - \hat{p}_i)^2 \quad (2-21)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (p_i - \hat{p}_i)^2} \quad (2-22)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |p_i - \hat{p}_i| \quad (2-23)$$

对于 MSE, RMSE, MAE 来说，其指标越小代表预测误差越小。

2.2 Hodrick-Prescott 滤波^[12, 13, 14]

HP 滤波，广泛使用于宏观经济分析，通过它可得到一些与经济相关的时间序列长期趋势。它最早于 1923 年由英国数学家 Edmund Taylor Whittaker 提出，在 20 世纪 90 年代，经济学家 Robert J. Hodrick 和诺贝尔经济学纪念奖得主 Edward C. Prescott 首次在经济学的领域中推广了这种滤波器。

该滤波把实际数据分解为趋势和周期两种成分。其基本原理在于使用对称的数据移动平均的方法，从变化不定的时间序列数据中，将具有一定趋势变化的、较为平滑的序列分离出来。

假设有一组时间序列 $\{y_1, y_2, \dots, y_T\}$ ，是由趋势部分 $\{\tau_1, \tau_2, \dots, \tau_T\}$ 和周期部分 $\{c_1, c_2, \dots, c_T\}$ 组成，有：

$$y_t = \tau_t + c_t, t=1, 2, \dots, T \quad (2-1)$$

定义一个损失函数 M，当损失函数 M 达到最小值时，所得到的 $\{\tau_1, \tau_2, \dots, \tau_T\}$ 就是我们所求的趋势。

$$\min M = \min \left\{ \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=3}^T [(\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2})]^2 \right\} \quad (2-2)$$

其中 λ 是惩罚参数（平滑指数）。M 中第一部分是对于 c_t 方差的惩罚，第二部分是对 τ_t 平滑度的惩罚，也就是说，HP 滤波是通过衡量使趋势部分在规定的平滑度上保持对 y 的跟踪的程度（即好的拟合），进行对于周期部分的好坏的识别。

注意， λ 的取值直接影响 HP 滤波的结果。当 $\lambda=0$ 时，HP 滤波退化成最小二乘法，此时，要让 M 达到最小值，只有 $y_t = \tau_t$ ，即趋势部分为时间序列本身； λ 越大，得到的趋势部分越平稳；当 λ 趋于无穷的时候，趋势部分接近于线性函数。一般来说，当时间序列数据为每一季度数据时，将 λ 设置为 1600，是月度数据则 λ 设置为 [100000, 140000]，年度数据 λ 设置为 [6, 14]。（ $t=1, \dots, T$ ）

为了求取极值，对损失函数 M 分别对 $\tau_1, \tau_2, \dots, \tau_T$ 进行求导，有：

$$\frac{dM}{d\tau_t} = 0, \text{ 则有:}$$

$$-2(y_t - \tau_t) + 2(\tau_{t+2} - 2\tau_{t+1} + \tau_t) = 0, (t=1, T) \quad (2-3)$$

$$-2(y_t - \tau_t) - 4(\tau_{t+1} - 2\tau_t + \tau_{t-1}) = 0, (t=2, T-1)$$

$$-2(y_t - \tau_t) + 2(\tau_{t+2} - 2\tau_{t+1} + \tau_t) - 4(\tau_{t+1} - 2\tau_t + \tau_{t-1}) + 2(\tau_{t+2} - 2\tau_{t+1} + \tau_t) = 0, (t=3, \dots, T-2)$$

将其整理并向量化，有：

$$\mathbf{y}_T = (\lambda \mathbf{F} + \mathbf{I}_T) \boldsymbol{\tau}_T \quad (2-4)$$

其中 \mathbf{y}_T 和 $\boldsymbol{\tau}_T$ 是 $(T \times 1)$ 的向量， \mathbf{I}_T 是单位向量，有

$$\mathbf{F} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & \dots & 0 \\ -2 & 5 & -4 & 1 & 0 & \dots & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & \dots & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & & & \ddots & & \ddots & \ddots \\ 0 & \dots & \dots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & \dots & \dots & \dots & \dots & 0 & 1 & -4 & 5 & -2 \\ 0 & \dots & \dots & \dots & \dots & 0 & 1 & -2 & 1 & \end{bmatrix}$$

因此依次得到趋势部分 $\boldsymbol{\tau}_T$ 和周期部分 \mathbf{c}_T ，有：

$$\begin{aligned}\tau_T &= (\lambda F + I_T)^{-1} y_T \\ c_T &= y_T - \tau_T\end{aligned}\quad (2-5)$$

此时，我们便得到了关于时间序列的两部分。

3. 股票价格预测的时间序列模型方法

本次数据选取了来自 DC 竞赛网站中关于量化投资竞赛中的 roundA 数据，经过处理后得到了 111 支股票 488 天的日收盘价。构建一个时间序列模型对 111 支股票 488 天之后三天的收盘价进行预测，随后根据预测到的收盘价以及各股票历史的波动情况进行分析，针对不同性格的投资者，从 111 支股票中选出 10 支作为合适的参考投资方案。

3.1 模型结构

首先，给出整体方法的大致流程图：

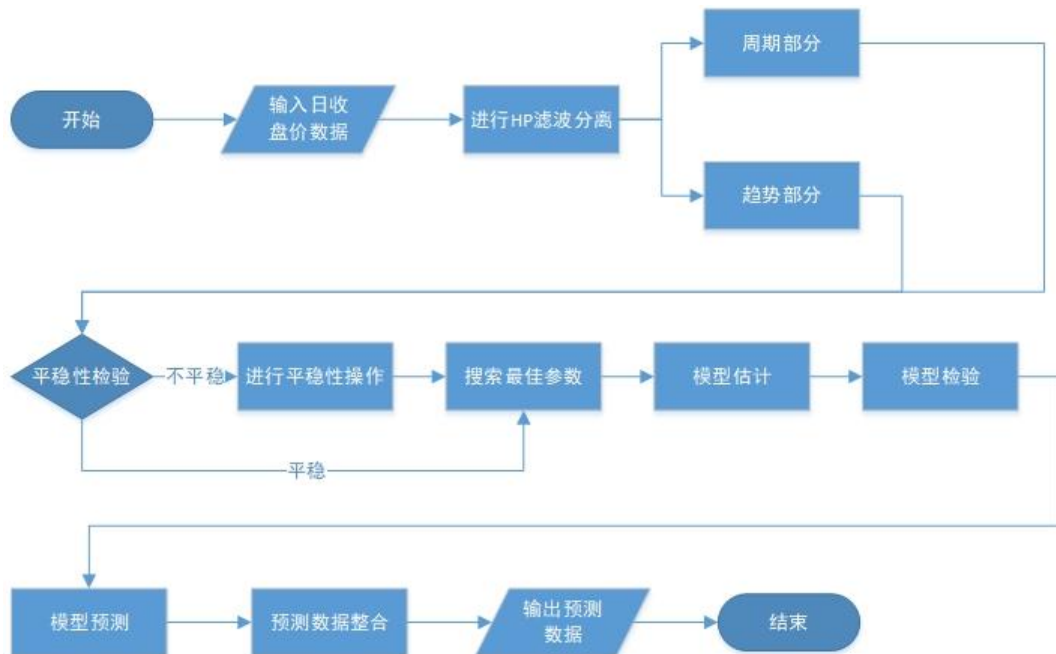


图 3-1 方法流程图

如图所示，对于每一只股票的收盘价时间序列，进行 HP 滤波分离数据，将数据分为周期部分 cycle（记为 CP ）以及趋势部分 trend（记为 TP ），有

$P_T = CP + TP$ ，对于 CP 以及 TP 分别建立 ARMA(p, q) 模型，包括的步骤如下^[15]：

- (1) 平稳性检验：若不平稳，则预处理数据，使其平稳化；
- (2) 定阶：对于模型的 p, q 要确定其大小，使用 AIC 或 BIC 准则；
- (3) 估计：对于模型的参数进行估计；
- (4) 显著性检验：对得到的模型进行 LB 检验；
- (5) 预测：使用检验后的模型进行预测，得到预测数据。

具体建模中，上述步骤需要反复试验调整，一些具体的理论知识已在第二章中给出了总结。

3.2 模型实证

从 111 支股票中，随机选出一支，读入数据，选择前 485 个时间序列数据进行构建模型，其图为：

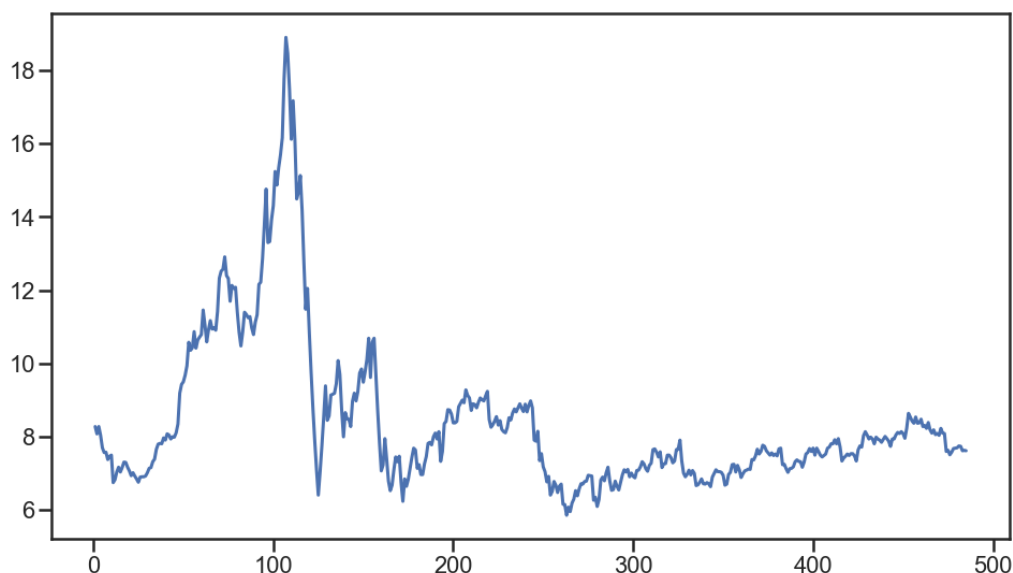


图 3-2 股票代码 stock102602 的收盘价序列图

对其进行 HP 滤波分解，得到两部分数据的时间序列图。蓝色线为 CP 部分，橙色线为 TP 部分。

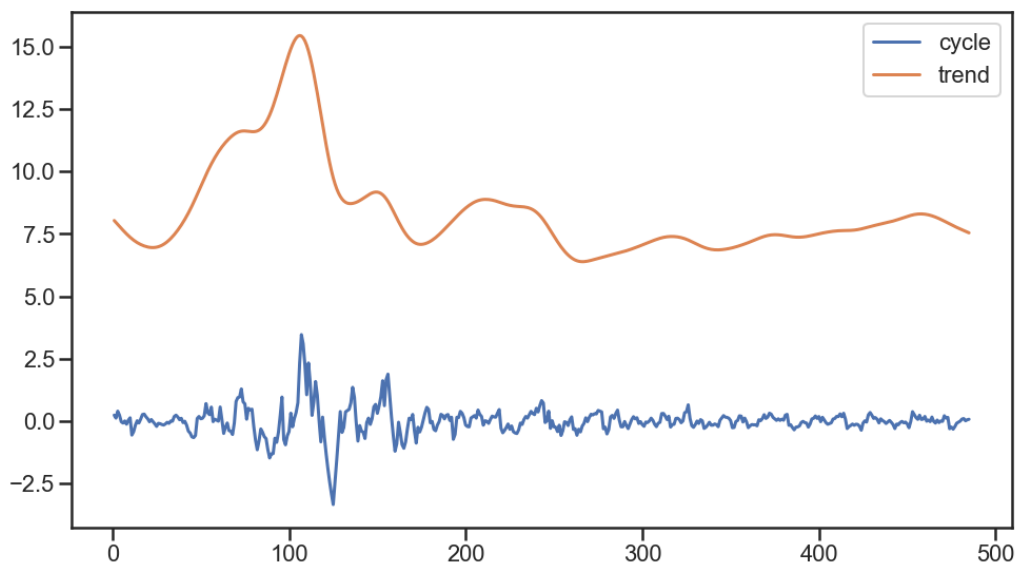


图 3-3 HP 滤波分解后的收盘价序列图

先对 *CP* 部分进行操作，从图上来看，其 *CP* 部分是大致平稳的，对其做 ADF 检验，得到以下检验数据：

表 3-1 *CP* 部分 ADF 检验相关数据

检验统计量	P 值	显著水平	1%	5%	10%
-7.91	4.0×10^{-12}	临界值	-3.44	-2.87	-2.57

p 值小于 0.05，说明检验统计量充分小，则此序列为平稳序列，可对其直接进行模型构建。

对 *CP* 部分进行模型构建，考虑到模型计算量，使用网格搜索的方法对模型的阶数进行确定（BIC 准则），p 与 q 的最大阶数设置为 4，得到以下数据：

表 3-2 *CP* 部分网格搜索定阶（BIC 准则）

p\q	0	1	2	3	4
0	835.66	492.54	—	—	—
1	314.71	304.97	310.84	317.03	320.92
2	306.83	310.87	—	—	—
3	311.11	316.96	276.60	—	—
4	315.95	266.73	307.31	268.50	—

从上表中看出，(4, 1) 为 BIC 准则下的最佳阶数。

使用得到的最佳阶数对模型参数进行估计，为了加快模型估计速度以及得

到较高的准确度和优良的统计性质，将最小二乘估计值作为迭代的初始值，对参数进行极大似然估计的迭代，得到以下数据：

ARMA Model Results						
Dep. Variable:	y	No. Observations:	485			
Model:	ARMA(4, 1)	Log Likelihood	-111.719			
Method:	css-mle	S.D. of innovations	0.303			
Date:	Mon, 08 Apr 2019	AIC	237.439			
Time:	18:29:04	BIC	266.728			
Sample:	0	HQIC	248.946			
	coef	std err	z	P> z	[0.025	0.975]
const	0.0002	0.002	0.083	0.934	-0.004	0.004
ar.L1.y	1.8502	0.045	41.435	0.000	1.763	1.938
ar.L2.y	-1.1028	0.094	-11.780	0.000	-1.286	-0.919
ar.L3.y	0.3763	0.094	4.021	0.000	0.193	0.560
ar.L4.y	-0.1713	0.045	-3.840	0.000	-0.259	-0.084
ma.L1.y	-1.0000	0.005	-194.120	0.000	-1.010	-0.990
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	1.0502	-0.2096j	1.0709	-0.0313		
AR.2	1.0502	+0.2096j	1.0709	0.0313		
AR.3	0.0478	-2.2553j	2.2558	-0.2466		
AR.4	0.0478	+2.2553j	2.2558	0.2466		
MA.1	1.0000	+0.0000j	1.0000	0.0000		

图 3-4 模型拟合后的相关数据

可以看到，常数参数未通过检验，因此将常数参数进行去除再一次进行估计得到以下：

ARMA Model Results						
Dep. Variable:	y	No. Observations:	485			
Model:	ARMA(4, 1)	Log Likelihood	-111.723			
Method:	css-mle	S.D. of innovations	0.303			
Date:	Mon, 08 Apr 2019	AIC	235.446			
Time:	18:30:05	BIC	260.551			
Sample:	0	HQIC	245.310			
	coef	std err	z	P> z	[0.025	0.975]
ar.L1.y	1.8502	0.045	41.434	0.000	1.763	1.938
ar.L2.y	-1.1028	0.094	-11.780	0.000	-1.286	-0.919
ar.L3.y	0.3762	0.094	4.020	0.000	0.193	0.560
ar.L4.y	-0.1713	0.045	-3.840	0.000	-0.259	-0.084
ma.L1.y	-0.9999	0.005	-193.774	0.000	-1.010	-0.990
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	1.0502	-0.2096j	1.0709	-0.0313		
AR.2	1.0502	+0.2096j	1.0709	0.0313		
AR.3	0.0478	-2.2555j	2.2560	-0.2466		
AR.4	0.0478	+2.2555j	2.2560	0.2466		
MA.1	1.0001	+0.0000j	1.0001	0.0000		

图 3-5 模型去除常数拟合后的相关数据

此时模型的拟合已经比较可以了，画出模型的残差图：

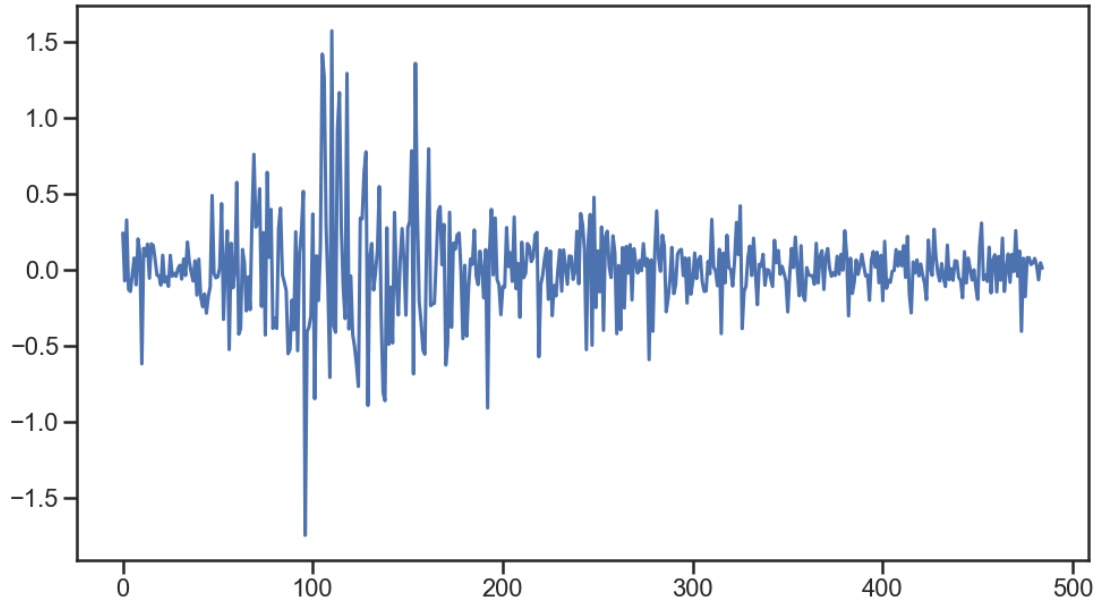


图 3-6 残差图

对残差进行 LB 检验，有：

lag	AC	PAC	Q	P-value
1.0	-0.011549	-0.011549	0.065094	0.798618
2.0	0.004569	0.004569	0.075304	0.963048
3.0	-0.065312	-0.065312	2.165583	0.538759
4.0	-0.011353	-0.011353	2.228876	0.693746
5.0	-0.003596	-0.003596	2.235241	0.815730
6.0	-0.068783	-0.068783	4.568123	0.600270
7.0	0.000384	0.000384	4.568196	0.712490
8.0	0.131497	0.131497	13.130325	0.107444
9.0	-0.038587	-0.038587	13.869171	0.127056
10.0	-0.125291	-0.125291	21.674958	0.016849
11.0	0.021578	0.021578	21.906973	0.025104
12.0	-0.002254	-0.002254	21.909510	0.038547

图 3-7 残差自相关检验

可以看到，Q(1)-Q(9)对应的 p 值都比 0.05 要大，按照 LB 检验方法，可以接受 H_0 ，认为其 $\rho_k = 0$ ，模型的残差序列具有白噪声特征。得到 CP 部分的时间序列模型：

$$p^{(CP)}_t = 1.8502p^{(CP)}_{t-1} - 1.1028p^{(CP)}_{t-2} + 0.3762p^{(CP)}_{t-3} - 0.1713p^{(CP)}_{t-4} + \varepsilon^{(CP)}_t - 0.9999\varepsilon^{(CP)}_{t-1} \quad (3-1)$$

接下来对 *TP* 部分进行操作，从图 3-3 看出，*TP* 部分是不平稳的，由于数据量较大，且 *TP* 部分的数据一般来说也比较大，为了缩小差分对于序列的影响，先对原序列进行对数操作，然后再对其做 ADF 检验，得到以下检验数据：

表 3-3 *TP* 部分取对数后 ADF 检验相关数据

检验统计量	P 值	显著水平	1%	5%	10%
-2.78	0.062	临界值	-3.44	-2.87	-2.57

p 值大于 0.05，说明检验统计量不够小，则此序列为不平稳序列，需要进行平稳化操作后，再对其进行模型构建。

对 *TP* 部分取对数后的序列进行差分操作，进行一阶差分，有：

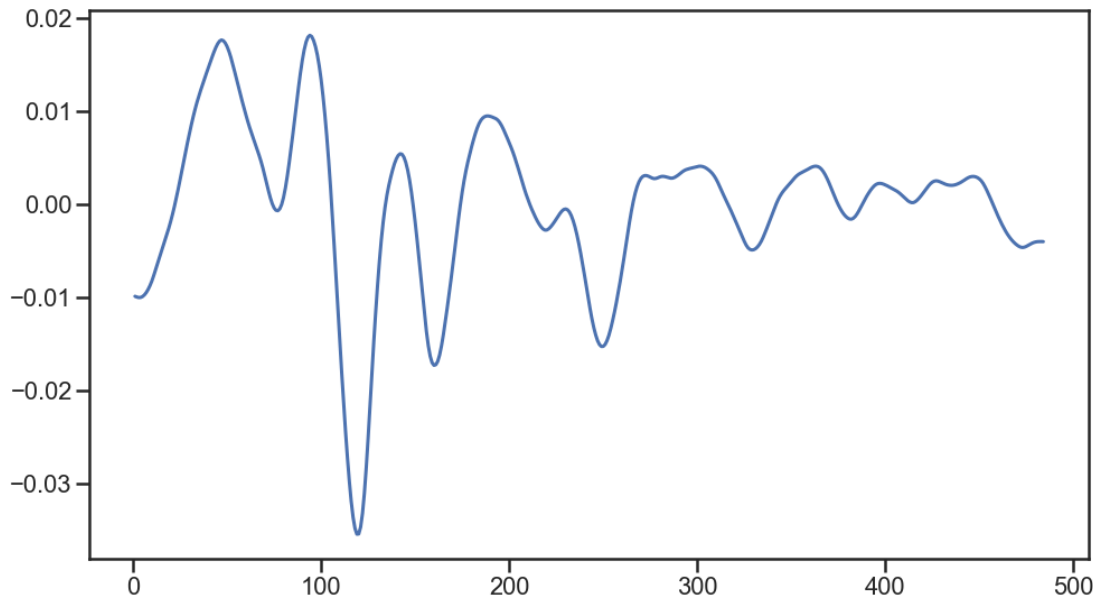


图 3-8 *TP* 部分一阶差分

对其进行平稳性检验，有

表 3-4 trend 部分一阶差分 ADF 检验相关数据

检验统计量	P 值	显著水平	1%	5%	10%
-2.76	0.063	临界值	-3.44	-2.87	-2.57

可以看出，其序列仍不平稳，需要进行再一次的差分操作，有：

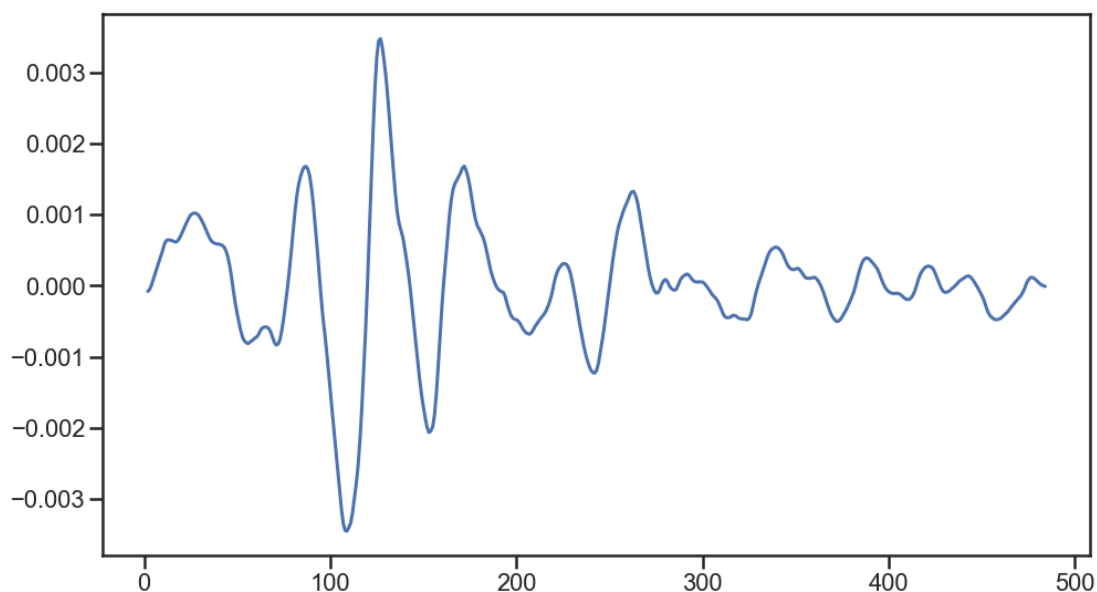


图 3-8 TP 部分二阶差分

对其进行平稳性检验，有

表 3-5 TP 部分二阶差分 ADF 检验相关数据

检验统计量	P 值	显著水平	1%	5%	10%
-6.56	8.33×10^{-9}	临界值	-3.44	-2.87	-2.57

经过二阶差分后的 TP 部分已经平稳了，可进行构建模型操作。

按前述网格搜索 p, q 参数方法，有：

表 3-6 TP 部分网格搜索定阶（BIC 准则）

$p \backslash q$	0	1	2	3	4
0	-5351.15	-6004.58	—	—	—
1	-7165.85	—	—	—	—
2	—	-9022.87	-9068.99	-9072.78	-9072.68
3	—	-9084.43	-9078.38	-9072.26	-9067.87
4	—	-9078.46	—	—	—

从上表中看出，(3, 1)为 BIC 准则下的最佳阶数。

按前述方法使用最佳阶数对模型参数进行估计，得到以下数据：

ARMA Model Results						
Dep. Variable:	diff	No. Observations:	483			
Model:	ARMA(3, 1)	Log Likelihood	4560.754			
Method:	css-mle	S.D. of innovations	0.000			
Date:	Mon, 08 Apr 2019	AIC	-9109.508			
Time:	20:21:35	BIC	-9084.428			
Sample:	0	HQIC	-9099.652			
	coef	std err	z	P> z	[0.025	0.975]
const	1.285e-05	0.000	0.127	0.899	-0.000	0.000
ar.L1.diff	2.4963	0.054	46.140	0.000	2.390	2.602
ar.L2.diff	-2.0564	0.106	-19.335	0.000	-2.265	-1.848
ar.L3.diff	0.5497	0.054	10.229	0.000	0.444	0.655
ma.L1.diff	0.2279	0.063	3.629	0.000	0.105	0.351
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	1.0231	-0.1633j	1.0360	-0.0252		
AR.2	1.0231	+0.1633j	1.0360	0.0252		
AR.3	1.6949	-0.0000j	1.6949	-0.0000		
MA.1	-4.3881	+0.0000j	4.3881	0.5000		

图 3-9 模型拟合后的相关数据

可以看到，常数参数未通过检验，因此将常数参数进行去除再一次进行估计得到以下：

ARMA Model Results						
Dep. Variable:	diff	No. Observations:	483			
Model:	ARMA(3, 1)	Log Likelihood	4560.737			
Method:	css-mle	S.D. of innovations	0.000			
Date:	Mon, 08 Apr 2019	AIC	-9111.475			
Time:	20:22:18	BIC	-9090.575			
Sample:	0	HQIC	-9103.262			
	coef	std err	z	P> z	[0.025	0.975]
ar.L1.diff	2.4948	0.054	46.035	0.000	2.389	2.601
ar.L2.diff	-2.0531	0.107	-19.267	0.000	-2.262	-1.844
ar.L3.diff	0.5478	0.054	10.171	0.000	0.442	0.653
ma.L1.diff	0.2286	0.063	3.636	0.000	0.105	0.352
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	1.0224	-0.1631j	1.0353	-0.0252		
AR.2	1.0224	+0.1631j	1.0353	0.0252		
AR.3	1.7031	-0.0000j	1.7031	-0.0000		
MA.1	-4.3752	+0.0000j	4.3752	0.5000		

图 3-10 模型去除常数拟合后的相关数据

此时模型的拟合已经比较可以了，画出模型的残差图：

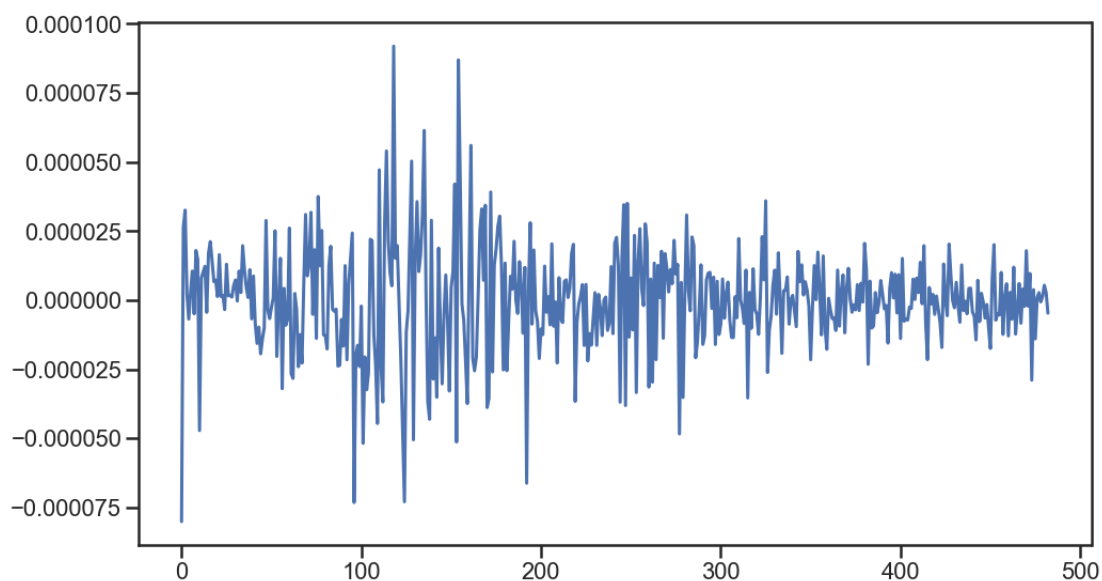


图 3-11 残差图

对残差进行 LB 检验，有：

lag	AC	PAC	Q	P-value
1.0	-0.001760	-0.001760	0.001505	0.969053
2.0	-0.007264	-0.007264	0.027206	0.986489
3.0	0.000937	0.000937	0.027634	0.998788
4.0	0.019736	0.019736	0.218116	0.994468
5.0	-0.046441	-0.046441	1.275080	0.937475
6.0	-0.072142	-0.072142	3.830970	0.699538
7.0	-0.013151	-0.013151	3.916086	0.789390
8.0	0.042374	0.042374	4.801596	0.778556
9.0	-0.032383	-0.032383	5.319838	0.805584
10.0	-0.139141	-0.139141	14.908074	0.135449
11.0	0.031082	0.031082	15.387553	0.165433
12.0	-0.025750	-0.025750	15.717332	0.204528

图 3-12 残差自相关检验

可以看到，Q(1)-Q(12)对应的 p 值都比 0.05 要大，按照 LB 检验方法，可以接受 H_0 ，认为其 $\rho_k = 0$ ，模型的残差序列具有白噪声特征。得到 TP 部分的时间序列模型：

$$p^{(TP)}_t = 2.4948p^{(TP)}_{t-1} - 2.0531p^{(TP)}_{t-2} + 0.5478p^{(TP)}_{t-3} + \varepsilon^{(TP)}_t + 0.2286\varepsilon^{(TP)}_{t-1} \quad (3-2)$$

模型预测，对 CP 模型和 TP 模型进行拟合后与原数据进行比较：

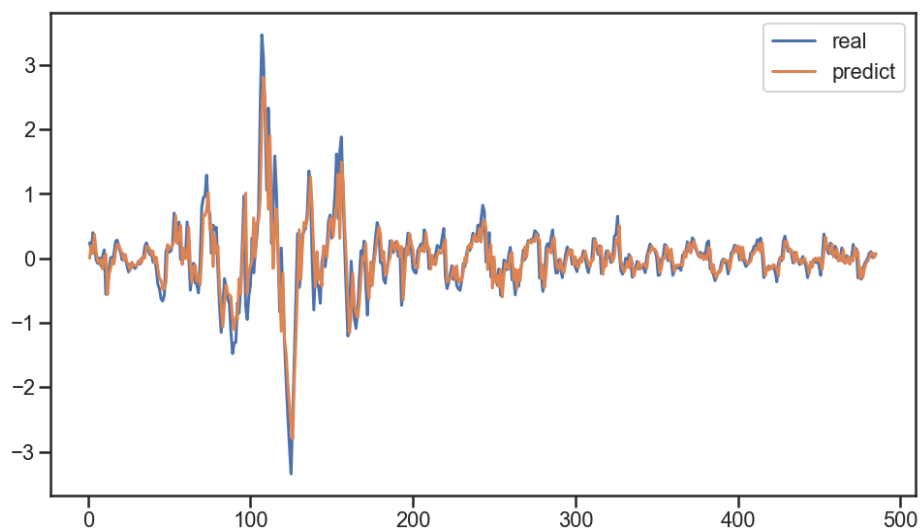


图 3-13 CP 部分模型拟合

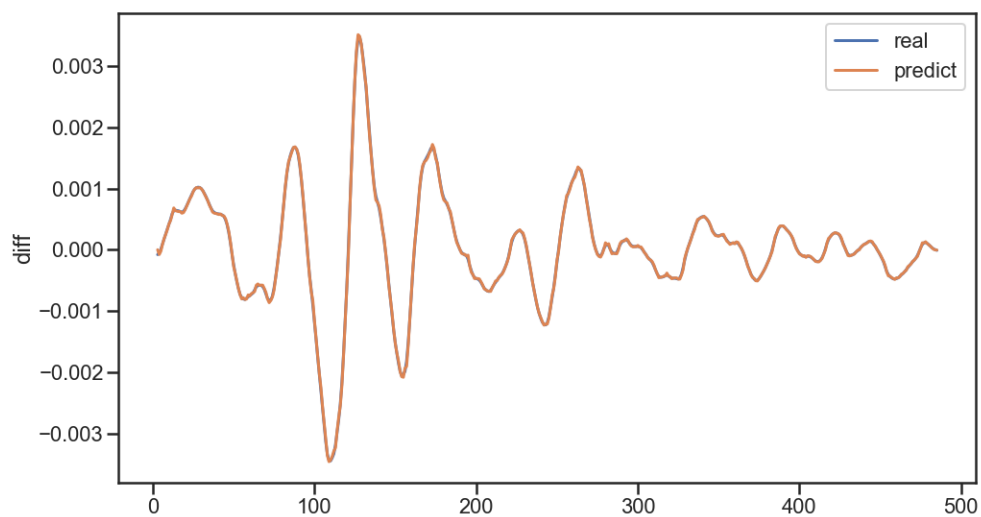


图 3-14 TP 部分模型拟合（数据处理前）

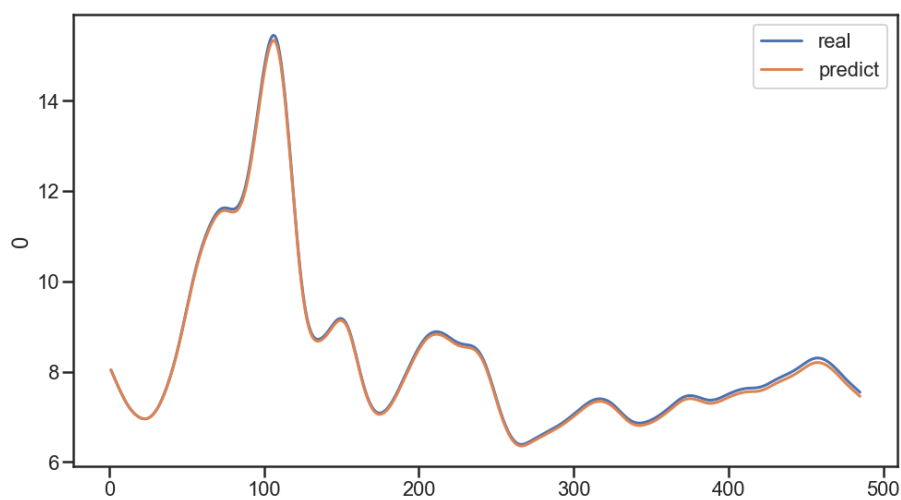


图 3-15 TP 部分模型拟合（数据还原后）

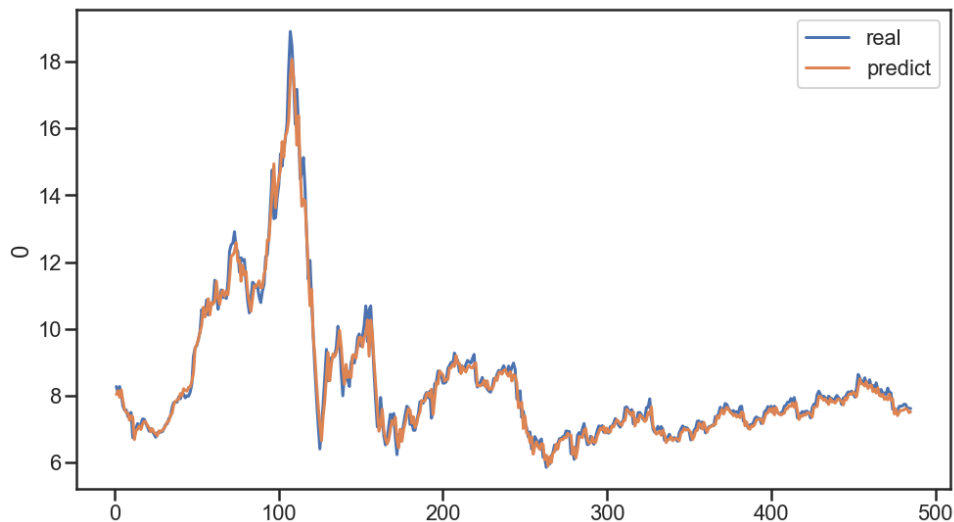


图 3-15 原时间序列拟合

对后三天的数据进行预测， TP 要进行差分还原并与真实值进行比较：

表 3-7 预测值与实际值对比

时间编号	CP	TP	预测值	实际值	差值	相对误差
486	0.0570	7.4257	7.4827	7.68	0.1973	2.57%
487	0.0420	7.3955	7.4375	7.8	0.3625	4.65%
488	0.0335	7.3651	7.3986	7.83	0.4314	5.51%

其中后三日预测数据的 MSE 为 0.1188，RMSE 为 0.3446，MAE 为 0.3303。

使用上述模型方法，随机抽取 10 支股票进行实证：

表 3-8 随机十组股票相关数据

股票编号	MSE	RMSE	MAE
stock100799	0.0243	0.1560	0.1372
stock103117	0.0142	0.1191	0.1132
stock102844	0.0359	0.1894	0.1572
stock100375	0.0004	0.0193	0.0164
stock100042	0.6966	0.8346	0.8224
stock102431	0.0095	0.0975	0.0958
stock100476	0.0593	0.2435	0.2178
stock100055	0.0505	0.2246	0.2215

stock100566	0.0548	0.2341	0.2021
stock101314	0.0330	0.1818	0.1514

使用前述模型方法，对全部 111 支股票进行实证，得到：

表 3-9 111 支股票相关数据

	MSE	RMSE	MAE
总	54.52	45.02	47.67
平均	0.49	0.41	0.43

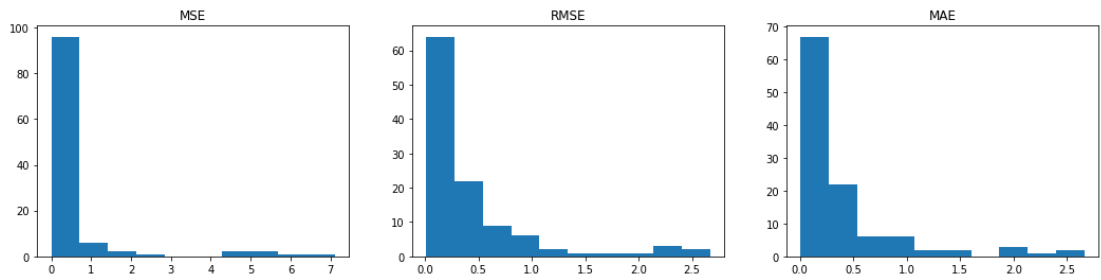


图 3-16 111 支股票预测各误差分布

可以看到，误差分布图 3-16 中有小一部分的预测误差较大，考虑到这些股票可能因为各种因素而使得模型拟合预测效果不佳，实际应用中，考虑从中去除这些股票，不列入方案考虑范围，筛选标准采用去除 MSE 大于 0.5 的股票数据，最终去除了以下 18 支股票：stock103528，stock100042，stock102820，stock101651，stock100591，stock102273，stock101712，stock100683，stock102926，stock100152，stock101758，stock103158，stock102520，stock100497，stock101766，stock103137，stock101003，stock102156。得到去除后股票的相关数据（93 支）：

表 3-10 93 支股票相关数据

	MSE	RMSE	MAE
总	7.43	20.99	19.31
平均	0.08	0.23	0.21

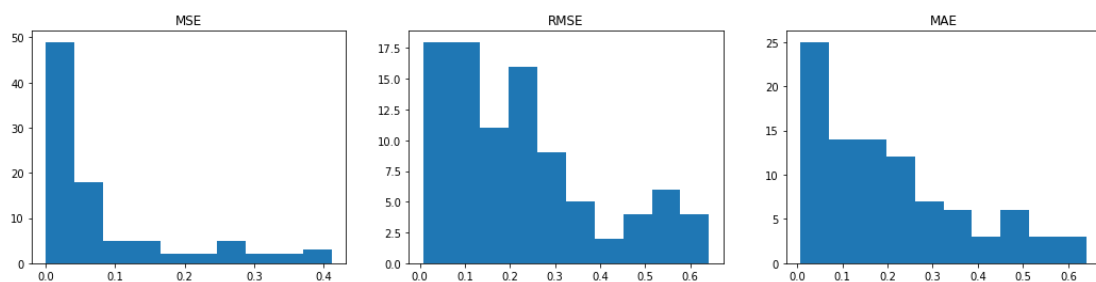


图 3-17 93 支股票预测各误差分布

从表 3-8 和表 3-9 看出，大部分的模型预测的误差较小，但仍存在小部分模型拟合不佳的情况，在表 3-10 和图 3-17 中，将拟合不佳的模型及数据去除，得到的误差情况有了大幅度的改善，且预测误差的分布也比较可观，误差集中的区间范围都比较小。因此认为可以进行下一步的模型实际应用。

3.3 模型应用

对上一小节中筛选出来的 93 支股票，其所有的 488 个时间序列数据进行上一小节的建模方法，并向后三步进行预测，根据预测到的收盘价计算这三天的收益率，以及计算股票的历史波动率。

计算后，进行筛选，筛选出 29 支三天收益率大于 0 的股票，及其波动率如下表所示：

表 3-11 29 支股票收益率和波动率

股票代码	收益率	波动率	股票代码	收益率	波动率
stock102566	0.0132	0.0415	stock102814	0.0035	0.0368
stock103553	0.0027	0.0470	stock102632	0.0500	0.0408
stock102073	0.0011	0.0320	stock100197	0.0006	0.0436
stock103242	0.0467	0.0223	stock100940	0.0025	0.0379
stock101057	0.0147	0.0349	stock100658	0.0180	0.0371
stock103155	0.0073	0.0343	stock100681	0.0221	0.0388
stock102590	0.0885	0.0337	stock102059	0.0008	0.0302
stock103017	0.0539	0.0388	stock101462	0.0016	0.0388
stock100603	0.0215	0.0369	stock101329	0.0050	0.0479
stock101953	0.0038	0.0361	stock102373	0.0154	0.0393

stock102431	0.0007	0.0384	stock101108	0.0108	0.0271
stock103193	0.0156	0.0230	stock100959	0.0078	0.0415
stock102792	0.0281	0.0380	stock101195	0.0836	0.0345
stock103117	0.0097	0.0447	stock100799	0.0061	0.0479
stock101394	0.1091	0.0391	-	-	-

若按照收益率最高进行平均投资，按照收益率进行降序排序后，选择以下十支股票：stock101394、stock102590、stock101195、stock103017、stock102632、stock103242、stock102792、stock100681、stock100603、stock100658。

亦可以对收益率最高的十支股票按照波动率来确定投资比例，例如对于上述十支股票，对各股票波动率进行标准化，得到以下数据：

表 3-11 10 支投资股票及投资比例

stock101394	stock102590	stock101195	stock103017	stock102632
0.08	0.12	0.11	0.08	0.07
stock103242	stock102792	stock100681	stock100603	stock100658
0.20	0.09	0.08	0.09	0.09

可考虑以此作为投资比例。

总结

本文代码编写基于 windows 平台，anaconda 环境下的 Python 语言。提出了基于 HP 滤波和 ARMA 的股票预测模型的主要结构和方法，通过文献以及个人实验表明，该模型方法在预测精度上较普通的 ARMA 模型有所提高，且建模方法简单，易于理解。当然，该模型也存在一些局限性，真实股票数据有着更多不确定的条件因素，而 ARMA 模型是在较为理想的即没有太大波动的情况下对股票进行建模分析预测，并且 ARMA 模型预测步数有限，只能进行短期预测，不能进行长期预测，因为其预测误差会逐步累积，最终使得误差越来越大。

第三章第二小节则先随机选取了一支股票数据进行构建拟合和预测，为了探究该数据构建模型后进行后三步预测的效果，将该股票 488 个数据分为 455

个构建模型数据和 3 个进行误差分析的数据，然后将其推广到所有的 111 支股票数据中，发现大部分的模型预测的误差较小，但仍存在小部分模型拟合不佳的情况。对于拟合不佳的模型，考虑将其模型和数据从中剔除，不列入后续的选择。剔除数据后，剩下 93 支股票，得到的总体平均的预测误差情况有了大幅度的改善，其中平均 MSE 为 0.08，平均 RMSE 为 0.23，平均 MAE 为 0.21。且 93 支股票预测误差的分布也比较可观，误差集中的区间范围比较小，因此认为剩下的 93 支股票对于后三步预测的效果在一定误差范围内是比较准确的。

在上述基础下，第三章第三小节对 93 支股票的 488 个数据进行构建拟合，预测出后三步也就是 489, 490, 491 这三天的预测值，并借此计算这三天的收益率，同时，使用历史数据进行历史波动率的计算，最终，根据股票的收益率，筛选出 29 支三天收益率大于 0 的股票，结合其波动率，提出了一种股票投资方案的思路，选择十支收益率最高的股票，按照其波动率不同而进行投资，一般情况下，波动率越低，投资的比例越高。当然，仍可以结合这两种数据按照个人喜好进行其他方案的选择，例如：选择 5 支波动率最低的股票，确保能够收益，再选择 5 支收益率最高的股票，确保部分高收益；亦可以只选择出收益率和波动率都比较好的一支股票进行投资，等等各种方案，这对于股票市场中的投资者有着一定的参考和指导价值。

致谢

四年时光转瞬即逝，最终还是在不舍中迎来了毕业季，论文即将完成，将为这四年画上一个完美的句号，心中无限感慨。感谢林老师，在论文定题和资料查找的过程中，给予了我很多指导和帮助，让我能够很快的熟悉相关知识背景，在论文撰写、修改以及到最后定稿的过程中，老师也提出了许多宝贵的建议，让我备受启发，不断将文章进行修正以达到最佳。最终，经历了百般困难将论文成功完成。在此，郑重向林老师以及其他在各个方面帮助过我的人表达诚挚和衷心的感谢！

[参考文献]

- [1]王丹,冯长焕.基于HP滤波和ARIMA-ARCH模型的我国GDP分析与预测[J].福建江夏学院学报,2018,8(01):1-7+17.
- [2]孟坤,李丽.基于ARMA模型预测股票价格的实证分析[J].河北北方学院学报(自然科学版),2016,32(05):55-60.
- [3]孙先强.时间序列模型在股票价格预测中的应用[D].云南大学,2016.
- [4]吴齐,杨桂元,戚琦.基于HP滤波和ARIMA模型的我国GDP分析与预测[J].滁州学院学报,2015,17(05):35-38+56.
- [5]赵庆.基于非线性组合模型对石油价格的预测[J].唐山学院学报,2014,27(06):85-88.
- [6]付燕,栗锋.ARMA模型在我国体育股票价格预测中的应用[J].统计与决策,2012(21):101-103.
- [7]方子良.时序法在股市行情技术分析中的应用[J].南京理工大学学报,1999(02):57-61.
- [8]王燕.应用时间序列分析[M].北京:中国人民大学出版社,2005.
- [9]史代敏,谢小燕.应用时间序列分析[M].北京:高等教育出版社,2011.
- [10](美)蔡瑞胸.金融时间序列分析[M].北京:机械工业出版社,2006.
- [11]李东风.金融分析讲义[Z].
- <http://www.math.pku.edu.cn/teachers/lidf/course/fts/ftsnotes/html/ftsnotes/index.html>
- [12]Hyeongwoo Kim. Hodrick-Prescott Filter[Z].2004.
- [13]陈东东,沐年国.基于HP滤波分解的ARMA+BPNN的人民币汇率短期预测[J].经济研究导刊,2018(21):74-77+169.
- [14]Bogdan Corneliu Biolan. A COMPARISON BETWEEN DETRENDING METHODS: HODRICK-PRESCOTT, BAXTER-KING, CHRISTIANO-FITZGERALD FILTERS. A SHORT SURVEY[J]. Review of the Air Force Academy,2018(2).
- [15]孙宇轩.基于小波分析的中国股市时间序列研究[D].北京交通大学,2010.