

统计建模与数据分析实训 课程

论文（报告、案例分析）

院 系 信息学院

专 业 统计学

班 级 14 统计 1 班

学生姓名 王晓璠

学 号 1421320022

任课教师 韩嵩 高和鸿

2017 年 6 月 29 日

目录

- 一、绪论3
 - (一) 背景及意义.....3
 - (二) 国内外文献综述4
 - 1、国外研究现状.....4
 - 2、国内研究现状.....4
- 二、模型描述6
 - (一) ARIMA 模型理论.....6
 - (二) 单位根检验.....7
 - (三) 模型识别.....7
 - (四) 参数估计.....7
 - (五) 模型检验.....8
 - (六) 动态预测与静态预测、预测偏差8
- 三、建模过程8
 - (一) 数据选取.....8
 - (二) 序列平稳性检验9
 - 1、序列平稳性的图示检验.....10
 - 2、ADF 检验.....10
 - (三) 模型的识别与选择11
 - (四) 模型的检验与确定12
- 四、模型的拟合与预测14
- 五、结论及缺陷分析.....15
 - (一) 研究结论.....15
 - 1、建模结果.....15
 - 2、预测结果.....15
 - (二)对策与不足16
- 参考文献.....15

上证A股指数走势预测研究

——基于时间序列模型

14 统计 1 班 王晓璠 1421320022

摘要：本文以时间序列模型为理论基础，选择上证 A 股指数为研究对象，对上证 A 股指数在 2015 年 6 月到 2017 年 6 月的月度数据，用 Eviews 软件进行实证分析。在此，本文采用时间序列分析中的一种常见模型，即 ARIMA 模型进行相关的分析和预测，对变量的发展规律进行了研究，通过研究分析计算得到预测值与实际值的相对误差范围均达到要求，因此，可以采用 ARIMA 模型做静态统计预测。并对未来三个月的上证 A 股指数做短期预测。

关键字：上证 A 股指数；时间序列分析；ARIMA 模型；统计预测

一、绪论

（一）背景及意义

中国的股票市场自从上海证券交易所 1990 年 12 月 19 日开始正式营业，至今有 20 多年的发展史，已经形成了与我国经济发展相适应、相结合的特色发展道路。中国股票市场经过不断的发展与完善，已经取得巨大的成绩。股市是国民经济的晴雨表。我国的股票市场作为中国证券市场的重要组成部分，对中国经济的发展和社会的稳定起着重要作用。

中国的股市对国家经济的影响确实很大。另一方面，从投资者的角度看，在当前的经济形势下，银行储蓄实际利率为负。投资方法有很多种，整体上可以分为实物投资和资本投资(即证券投资)。细分主要包括房地产投资、天使投资、创业投资、私募股权投资、买股票、买基金、买保险、债券投资、期货投资、外汇投资、贵金属投资、理财产品投资等。股票投资具有重要的指导性作用。

股票市场是一个复杂的非线性系统，市场受到来自政治，社会，经济，心理等方面的影响，因而对其运动行为很难建模。但是，正如技术分析所假设的“市场是有趋势可循的，市场价格反映了一切，历史往往会重演”，这也就是说明尽管复杂，但市场还是隐含着某些规律性。股价的历史轨迹形态对未来价格趋势特别是短期趋势有着重要的预测价值，这不仅得到市场上许多技术分析者的支持而且一些研究结果也证实了这一点。股票的价格走势直接影响着投资者的经济利益，以及不同行业的景气状况，也影响和反映着国家的宏观经济政策。因此正确的建立相关的模型进行股票价格预测有着重大意义，本文选取的变量是能够反映 A

股的综合波动趋势的上证 A 股指数, 目的在于试图找到一种较为理想的模型可以以一定的精确度来描叙现实股票市场价格波动的现象, 并得到一些有意义的结论。合理的分析股市不管是对投资者还是国民经济的健康稳定发展都是有很重要的意义。

(二) 国内外文献综述

1、国外研究现状

衡量证券分析师优劣的重要依据之一是其对上市公司盈利预测的准确性。美国股市相对成熟, 具有多年的盈利预测数据, 特别是 FirstCall 的数据库收集几十年卖方分析师的预测。在美国早期的研究中, 研究人员主要关注盈利的时间序列属性。Brown and Rozeff(1978)^[1] 分析师的早期研究表明盈利预测比时间序列模型的预测更加准确, 主要是由于证券分析师的预测具有时间的优势和信息的优势。Brown et al.(1987)^[2] 在剔除时间优势因素后, 发现证券分析师预测仍然比时间序列模型准确, 并且与股票收益的关联性较时间序列模型紧密。

Barefield and Comiskey(1975)^[3] 通过分析证券分析师对 100 家上市公司的盈利预测数值以及上市公司实际公布的盈利数据后发现, 证券分析师的预测值比实际值显著要高。证券分析师的预测乐观倾向的存在也被 Fried&Givoly(1982)^[4] 和 O'Brien(1988)^[5] 的研究证实。Calderon(1993)^[6] 利用 1976—1988 年间分析师预测的每股收益值为样本, 来对证券分析师的盈利预测特性进行研究, 研究结果发现证券分析师的盈利预测数据经常超过实际盈余, 这种现象在 20 世纪 80 年代尤其明显。

2、国内研究现状

由于股票市场受到政治因素、利率水平、国家政策、物价水平、国际环境、收益情况、经营状况和投资心理等许多因素的影响, 内部规律及其复杂, 是一个非常复杂的非线性非平稳的动态系统, 很难进行准确的预测, 因此很多研究者对股票的预测也做了很多相关的工作, 提出了很多的预测方法。

其中, 杨枫(2012)^[7] 认为关于证券分析师预测偏差产生的原因有如下两点: 证券分析师的经济动机和认知能力。经济动机是指证券分析师是理性的, 其中包括策略性预测偏差假说(该假说认为由于受到经济利益的驱动, 证券分析师会发布不真实的业绩预测, 进而对投资者产生误导); 还有选择性偏差假说(该假说认为预测结果正确, 但是受到证券分析师主观选择的影响, 在看好股票的时候才会发布盈利预测, 导致结果正偏); 以及效用函数假说(真实、未选择的预测, 预测产生偏差的原因是基于最大化或最小化自身效用函数的考虑)三方面。而认真能力是与前几种假说相对的, 即证券分析师是非理性的, 以及证券分析师的认知能力也存在不足, 导致预测结果出现偏差。

孙赞生(2016)^[8]尝试在 SV 模型中引入宏观经济因子,构建了 SV-MAE 模型,同时设计了用于 SV-MAE 模型的贝叶斯分析的相应 MCMC 算法,深入地研究了中国股市的波动特征,具有重要的现实意义。

崔远远、文忠桥(2015)^[9]通过建立 ARIMA 模型采用一步向前静态预测的方法对股市上证综合指数进行预测。其优势在于对样本取对数后消除了时间序列中的自相关和异方差,同时使得预测值接近实际值且效果良好。

李志杰(2015)^[10]介绍了应用广泛的 ARMA 模型(由自由回归模型 AR 和滑动平均模型 MA 混合而成的模型)的定义以及原理,ARMA 模型能够很好的处理线性问题;其次介绍了处理非线性问题的 BP 神经网络模型,对于非线性时间序列的预测能够得到很好预测结果;最后介绍了两者的组合模型,并且对股票的价格波动进行了更为准确的预测。

卢金长(2015)^[11]主要是通过对上证综合指数预测这一实证分析,对 BP 神经网络和小波神经网络和 SVM 回归进行了分析和比较。分别对上证综合指数预测、预测均方误差 MSE 和运算效率等方面进行了对比。通过对比,可以发现 BP 神经网络是小波神经网络的基础,小波神经网络是在 BP 神经网络的基础上衍生而得到的。小波神经网络在计算效率方面比较低,在符合要求的情况下,小波神经网络的准确性优于 BP 神经网络。而性能最优的方法却是支持向量机,无论是在预测准确程度还是在计算效率方面都领先于神经网络。

纪华东(2014)^[12]提出传统的单一模型对上证指数收益率的预测基本是无效的,面对股票市场的强烈波动,单一模型并不能做出比较合理的预测,在实际中无法应用。因此,融合了 ARMA 模型在线性关系捕捉的优势,还有 NN 神经网络强大的非线性映射能力,集成模型对波动强烈的时间序列能够有更加稳定、准确的预测能力,可以应用到实际问题当中。

郝杰(2014)^[13]建立了基于上证指数数据样本的改进小波神经网络的预测模型,并将它们的预测结果与 BP 神经网络预测模型、小波神经网络预测模型和 BFO-BP 神经网络预测模型三种现有模型比较,其结果表明改进后的 BFO-WNN 与 AIW-PSO 小波神经网络模型的预测效果显著优于现有三种预测模型。

黄喆(2009)^[14]将其近三年的上证指数数据分为四段后,分析了 RBF 神经网络在上证综指预测中不同时间段下的适用性,找到了适合各段时间的训练样本数,并在最后比较了单步预测与滚动预测的预测精度,对上证指数的一段数据进行了预测和拟合。

曾毅成(2006)^[15]使用基于 BP(Back Propagation)算法的神经网络进行预测,证明了我国上证指数长期以来存在的分形特征,指出了使用神经网络对其走势进行预测的可行性。

近几年，随着人工智能的快速发展，神经网络在模式识别、图像处理、数据挖掘等领域中都有着广泛的应用。股票价格数据的内在结构比较复杂，通常存在很多不确定因素。大部分理想的数据都属于线性序列，对于股票价格这种非线性序列来说，神经网络具有很大的优势。但因其可解释性较低没能力来解释自己的推理过程和推理依据。而时间序列是从数量上揭示某一现象的客观规律，分为微观层面和宏观层面，从一定角度揭示变化和时间之前的关系，从本质上发现其中的规律，挖掘其内在的信息。对于时间序列来讲，最主要的就是揭示其隐藏的变化规律，从而客观的认识事物内在本质。所以本文主要使用时间序列模型进行预测。

二、模型描述

（一）ARIMA 模型理论

Box-Jenkins 方法(博克思-詹金斯法)——ARIMA 模型。其基本思想是：某些时间序列是依赖于时间 T 的一组随机变量，构成该时间序列的单个序列值。虽然具有不确定性，但整个序列的变化却有一定的规律可以用相应的数学模型近似描述。该方法不考虑以经济理论为依据的解释变量的作用，而是依据变量本身的变化规律，利用外推机制描述时间序列的变化，能达到最小方差意义下的最优预测，是一种精度较高的时序短期预测方法。

Box-Jenkins 方法用变量 X_t 自身的滞后项，以及随机误差来解释该变量，具体形式可表达成 $ARIMA(p, d, q)$ 。其中 p 表示自回归过程阶数, d 表示差分的阶数, q 表示移动平均过程的阶数。

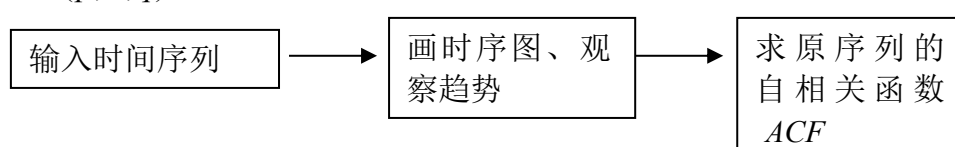
若时间序列是平稳的，则可直接运用 ARIMA 模型：

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

若时间序列是非平稳的，则需要经过 d 阶差分，将非平稳时间序列转换成平稳时间序列。其中，模型的假设条件为：①线性假设；② ε_t 是白噪声序列；③假定： $E(X_t, \varepsilon_s) = 0 (t < s)$ ，并且模型满足适用条件，即适用于一个平稳、零均值时间序列的短期预测。

基于 ARIMA 模型在面对非平稳时序序列预测中的良好效果，本研究将其用于上证 A 股指数平均值的预测中，有效的提高了预测精度。

本文通过建立 ARIMA 模型对上证 A 股指数平均值 X_t 进行拟合，同时利用本文建立的 ARIMA 模型对一段时间内的上证 A 股指数平均值 X_t 进行预测。本文对 $ARIMA(p, d, q)$ 模型的建立过程进行如下图介绍：



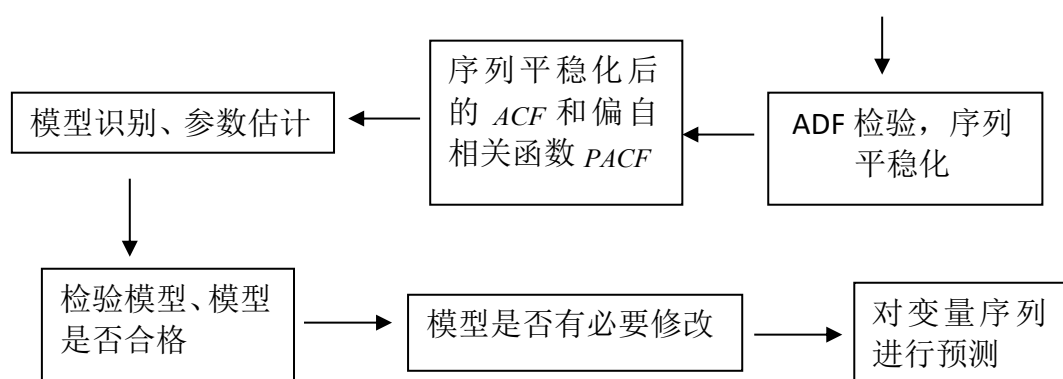


图 ARIMA(p, d, q) 模型的建立过程

(二) 单位根检验

对时间序列进行分析的前提是保证序列的平稳性，而非平稳的时间序列参与回归建模分析，会导致伪回归问题，因此一般情况下进行分析之前，需要对原变量序列进行 ADF 单位根检验，判断序列的平稳性。

ADF 检验的原假设为 H_0 ：时间序列含有单位根。如果序列为平稳序列，则可以继续建模；如果序列为非平稳序列，则需要差分处理或者采用协整等分析方法。

(三) 模型识别

我们引入自相关系数和偏自相关系数这两个统计量来识别 $ARMA$ 模型的系数特点和模型的阶数。若平稳序列的偏相关函数是截尾的，而自相关函数是拖尾的，可断定序列适合 AR 模型；若平稳序列的偏自相关函数和自相关函数均是拖尾的，则序列适合 $ARMA$ 模型。自相关函数成周期规律的序列，可选用季节性乘积模型。自相关函数规律复杂的序列，可能需要作非线性模型拟合。

在平稳时间序列自相关函数和偏自相关函数上初步识别 $ARMA$ 模型阶数 p 和 q ，然后利用 AIC 定则准确定阶。 AIC 准则：最小信息准则，同时给出 $ARMA$ 模型阶数和参数的最佳估计，适用于样本数据较少的问题。目的是判断预测目标的发展过程与哪一随机过程最为接近。因为只有当样本量足够大时，样本的自相关函数才非常接近母体的自相关函数。具体运用时，在规定范围内使模型阶数从低到高，分别计算 AIC 值，最后确定使其值最小的阶数是模型的合适阶数。

(四) 参数估计

确定模型阶数后，应对 $ARMA$ 模型进行参数估计。本文采用最小二乘法 OLS 进行参数估计。

（五）模型检验

完成模型的识别与参数估计后，对估计结果进行诊断和检验。通过检验可以发现模型是否合理。若不合理，需要进行模型修正。其中，检验拟合的模型是否合理包括如下两点：一是检验模型参数的估计值是否具有显著性；二是检验模型的残差序列是否为白噪声。参数估计值的显著性检验是通过 t 检验完成的 Q 检验。其原假设是： $H_0: \rho_1 = \rho_2 = \dots = \rho_k$ 即模型的误差项是一个白噪声过程。判别规则为：

若 $Q \leq \chi^2_{\alpha}(k-p-q)$ ，则接受 H_0 。

若 $Q > \chi^2_{\alpha}(k-p-q)$ ，则拒绝 H_0 。

其中 α 表示检验水平。

（六）动态预测与静态预测、预测偏差

模型预测方法主要有两种，分别是动态预测方法和静态预测方法。动态预测方法是指，除了第一个预测值是用解释变量的实际值预测外，其后各期预测值都采用递归预测的方法，即用滞后的被解释变量的前期预测值带入预测模型来预测下一期的预测值。静态方法是指用解释变量的真实值来进行预测。只要真实值数据可以获得就可以使用静态预测方法。本文将用静态预测的方法对上证 A 股指数平均值进行预测分析。

预测偏差是通过回归方程或其他统计方程得到的。在回归中，给定自变量的观测值，通过回归系数可以计算出相应因变量的预测值。因此，预测都带有误差。

三、建模过程

（一）数据选取

本文原始数据来源于东方财富网 (<http://data.eastmoney.com>)，通过整理可以获得月度上证 A 股指数平均值，其中，选取了 2015 年 6 月到 2017 年 6 月的数据。由于 2015 年 6 月金融危机的影响，股票大盘指数暴跌，给投资者带来了严重的损失，本文通过对金融危机时期 A 股指数走势的分析，对危机后的中国股市做一个大致的了解，并对指数走势进行短期预测。其中，用 X_t 来表示月度 A 股指数序列。原始数据如表 1 所示。

表 1 原始数据

日期	X_t
2015 年 6 月	4798.0202
2015 年 7 月	3848.2492

2015 年 8 月	3594.0173
2015 年 9 月	3127.9923
2015 年 10 月	3342.4785
2015 年 11 月	3561.1920
2015 年 12 月	3546.0441
2016 年 1 月	2983.3718
2016 年 2 月	2803.1900
2016 年 3 月	2908.6127
2016 年 4 月	3005.0718
2016 年 5 月	2859.8628
2016 年 6 月	2899.1867
2016 年 7 月	3016.8832
2016 年 8 月	3050.0210
2016 年 9 月	3036.4658
2016 年 10 月	3085.9883
2016 年 11 月	3195.5148
2016 年 12 月	3155.8410
2017 年 1 月	3136.3257
2017 年 2 月	3211.4586
2017 年 3 月	3240.6875
2017 年 4 月	3213.2640
2017 年 5 月	3094.5787
2017 年 6 月	3136.7955

（二）序列平稳性检验

由于建立 ARMA 模型的序列必须是平稳序列，所以首先要对序列的平稳性进行检验，若为平稳序列，则可以建立 ARMA 模型；若为非平稳序列，只能对其差分后平稳的序列进行 ARMA 模型的建立，即 ARIMA 模型。

通常情况下，可以通过对序列的一些属性图示粗略的判断序列的平稳性，但是这样通过目测结果可能不够准确，所以一般要使用严格的时间序列平稳性检验方法——单位根检验。

1、序列平稳性的图示检验

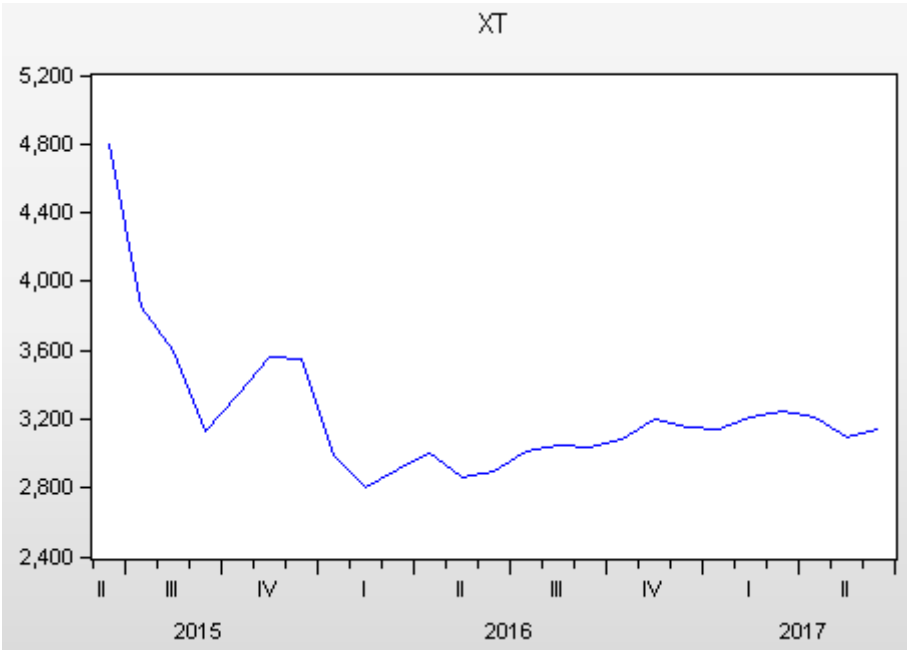


图1 上证 A 股指数时序图

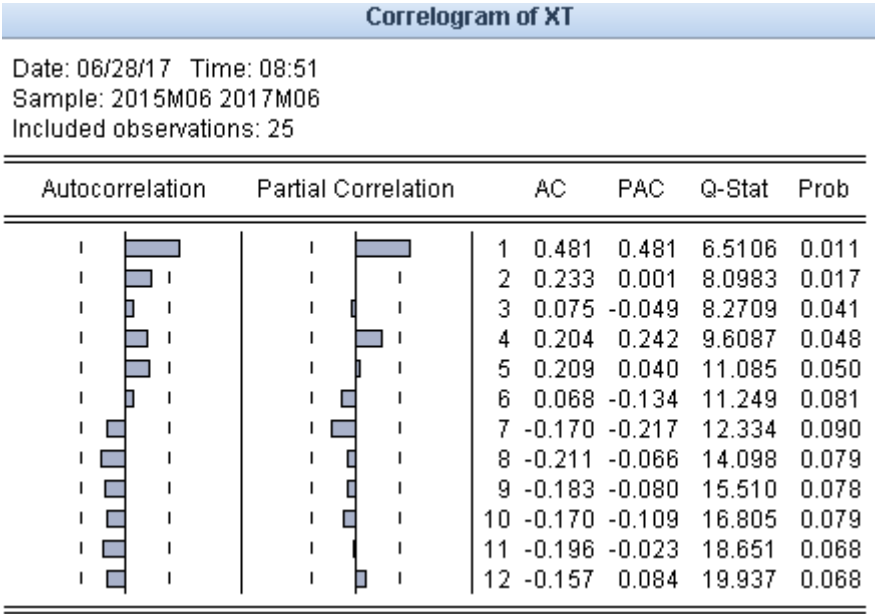


图2 序列 X_t 的自相关图和偏自相关图

从图 1 曲线的走势可以看出 2015 年 6 月之后有一个骤降，之后的走势都很平稳。

2、ADF 检验

下面，通过 ADF 检验来验证它的平稳性。如图 3 所示。

Augmented Dickey-Fuller Unit Root Test on XT		
Null Hypothesis: XT has a unit root		
Exogenous: Constant		
Lag Length: 5 (Automatic - based on SIC, maxlag=5)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.106059	0.0056
Test critical values: 1% level	-3.831511	
5% level	-3.029970	
10% level	-2.655194	

*Mackinnon (1996) one-sided p-values.

图3 序列 X_t 的 ADF 单位根检验结果

通过图3所示的单位根检验结果可以看出, ADF 检验的 t 值为-4.106 小于 1% 的临界水平-3.83, 相应的概率 P 值为 0.0056, 小于 0.01, 可以拒绝序列非平稳的原假设, 因此上证 A 股指数的时间序列数据是零阶单整的, 即序列 X_t 是平稳的, 所以可以直接建立 ARMA 模型。

(三) 模型的识别与选择

模型的识别可以通过自相关和偏自相关图来判别。为了更好地拟合模型, 本文首先采用 Schwarz criterion 施瓦茨准则和 Akaike info criterion 赤池信息准则及模型拟合优度检验进行 ARMA (p, q) 的选择。选取拟合效果较好并且不太繁杂的模型为最优模型。AIC 准则可以再模型极大似然的基础上, 对模型的阶数和相应参数同时给出一种最佳估计。

由图2中 ACF 和 PACF 拖尾的性质可以初步判断建立 ARMA 模型。通过对 ARMA (1, 1)、ARMA (2, 1)、ARMA (3, 1)、ARMA (1, 2)、ARMA (2, 2)、ARMA (3, 2)、ARMA (3, 3) 模型的比较可以得到最优模型, 模型比较结果如表2所示。

表2 模型比较结果

模型	拟合优度	调整后的拟合优度	AIC 值	SC 值
ARMA (1, 1)	0.6690	0.6375	13.0119	13.1592
ARMA (2, 1)	0.8172	0.7883	12.1649	12.3624
ARMA (3, 1)	0.6304	0.5434	12.7762	13.0241
ARMA (1, 2)	0.6693	0.6197	13.0941	13.2905
ARMA (2, 2)	0.5906	0.4996	13.0581	13.3050
ARMA (3, 2)	0.7213	0.6342	12.5849	12.8825
ARMA (3, 3)	0.8499	0.7898	12.0569	12.4040

从表2模型比较结果中可以发现 ARMA (3, 3) 的拟合优度和调整后的拟合优度是最大的, 分别为 0.8499、0.7898; 它的 AIC 值和 SC 值是最小的, 分别为 12.0569、12.4040。因此, 可以得出 ARMA (3, 3) 为最优模型。模型的估计结果如图4所示。

Dependent Variable: XT
Method: Least Squares
Date: 06/28/17 Time: 10:30
Sample (adjusted): 2015M09 2017M06
Included observations: 22 after adjustments
Convergence achieved after 204 iterations
MA Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3092.705	357.8033	8.643588	0.0000
AR(1)	0.616093	0.404787	1.522015	0.1488
AR(2)	0.527331	0.286406	1.841203	0.0855
AR(3)	-0.286535	0.182613	-1.569087	0.1375
MA(1)	-0.134554	0.631071	-0.213214	0.8340
MA(2)	-1.900983	0.975893	-1.947942	0.0704
MA(3)	-1.498199	0.601573	-2.490468	0.0250
R-squared	0.849917	Mean dependent var	3118.674	
Adjusted R-squared	0.789884	S.D. dependent var	193.0359	
S.E. of regression	88.48449	Akaike info criterion	12.05690	
Sum squared resid	117442.6	Schwarz criterion	12.40405	
Log likelihood	-125.6259	Hannan-Quinn criter.	12.13868	
F-statistic	14.15750	Durbin-Watson stat	2.052691	
Prob(F-statistic)	0.000020			
Inverted AR Roots	.84	.48	-.71	
Inverted MA Roots	1.73	-.80-.48i	-.80+.48i	
Estimated MA process is noninvertible				

图 4 ARMA (3, 3) 的参数估计结果

(四) 模型的检验与确定

由图 4 中的 t 统计量的 P 值可以看出, ARMA (3, 3) 未通过显著性检验。所以需要剔除不通过检验的参数, 最后得到通过了显著性检验的 AR (1) 模型, 如图 5 所示, t 统计量的 P 值均小于 1%。

Dependent Variable: XT
Method: Least Squares
Date: 06/28/17 Time: 14:39
Sample (adjusted): 2015M07 2017M06
Included observations: 24 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3103.983	65.16882	47.62987	0.0000
AR(1)	0.483889	0.078110	6.194960	0.0000
R-squared	0.635626	Mean dependent var	3168.879	
Adjusted R-squared	0.619064	S.D. dependent var	253.6909	
S.E. of regression	156.5781	Akaike info criterion	13.02464	
Sum squared resid	539367.6	Schwarz criterion	13.12281	
Log likelihood	-154.2957	Hannan-Quinn criter.	13.05069	
F-statistic	38.37753	Durbin-Watson stat	1.569141	
Prob(F-statistic)	0.000003			
Inverted AR Roots	.48			

图 5 AR (1) 模型的参数估计结果

接下来，需要对拟合模型 AR (1) 的残差序列进行白噪声检验，若属于白噪声序列，则认为拟合模型显著有效。如图 6 所示为 AR (1) 模型的残差序列的 ACF 和 PACF 图。

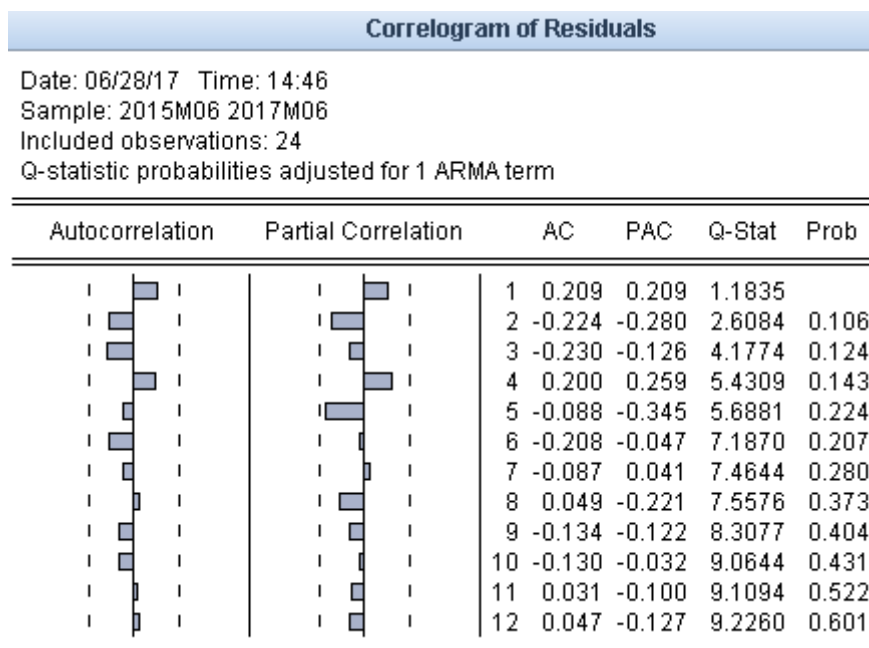


图 6 AR (1) 模型的残差序列的 ACF 和 PACF 图

从图 6 可以得出序列的相关分析结果：

- (1) 可以看出自相关系数始终在零周围波动，判定该序列为平稳时间序列。
- (2) 看 Q 统计量的 P 值：该统计量的原假设为残差的 1 期、2 期……k 期的自相关系数均近似为 0，备择假设为自相关系数中至少有一个不等于 0，因

此如图可知，该 P 值都>1%的显著性水平，所以接受原假设，即序列是纯随机序列，即白噪声序列。因此，模型 AR（1）是合理的。

可以得出 AR（1）模型的表达式为：

$$X_t = 3103.983 + 0.4839 * X_{t-1}$$

四、模型的拟合与预测

静态预测是采用滞后因子变量的实际值而不是预测值来计算一步向前的结果，所以 Eviews 软件的静态预测所得的结果比动态预测更加准确。因此本文采用的是静态预测，根据所建立的 AR（1）模型对 2017 年 1 月到 2017 年 6 月的上证 A 股指数的平均值进行预测，如图 7 所示。

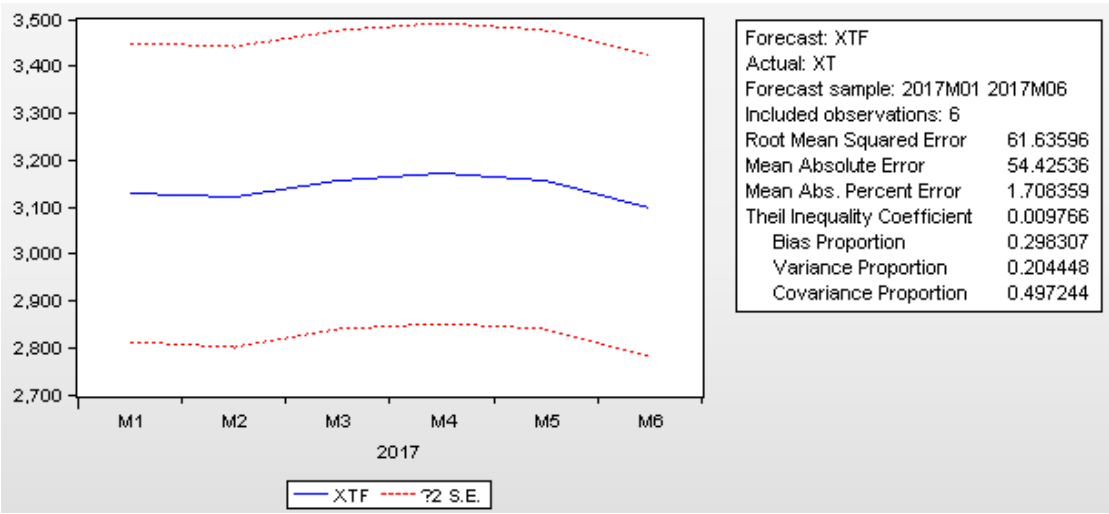


图 7 AR（1）模型的预测结果

在图 7 中，虚线是预测置信区间，实线代表真实值的变化趋势。可以看到随着向后预期的增加，预测置信区间变化程度不明显，从而表明预测期越往后，模型的预测精度无明显变化。这是因为静态模型中每一项预测值都是由前一项真实值来进行的预测，相比动态模型，不会积累误差，所以静态模型每一项预测值都更为准确。

另一方面，图 7 中的评价指标也显示了此次预测精度非常高。如 Theil 不等系数在 0-1 之间，Theil 不等系数等于 0 表示预测与真值完全相同，图 7 中的 Theil 不等系数为 0.0098，表明 AR（1）模型对 2017 年 1 月到 2017 年 6 月的上证 A 股指数平均值的预测准确度极高。下面对 2017 年 1 月到 2017 年 6 月的上证 A 股指数平均值的真实值、预测值和预测误差进行整理，如下表 3 所示。

表 3 上证 A 股指数平均值的真实值、预测值和预测误差

	2017 年 1 月	2017 年 2 月	2017 年 3 月	2017 年 4 月	2017 年 5 月	2017 年 6 月
实际值	3136.3257	3211.4586	3240.6875	3213.2640	3094.5787	3136.7955
预测值	3129.0763	3119.6330	3155.9891	3170.1326	3156.8627	3099.4321

预测误差	0.2311%	2.8593%	2.6136%	1.3423%	2.0127%	1.1911%
------	---------	---------	---------	---------	---------	---------

从表 3 中可以发现，2017 年 1 月到 2017 年 6 月的实际值与预测值之间的差距非常小，预测偏差都在 3%之内，因此该模型预测的精度很高，同时也表明本研究建立的 AR（1）模型非常优良。

根据所建立的 AR(1)模型对未来三个月的上证 A 股指数的平均值进行预测，如下表 4 所示。

表 4 上证 A 股指数平均值的未来三个月的预测值			
	2017 年 7 月	2017 年 8 月	2017 年 9 月
预测值	3119.8604	3111.6657	3107.7003

从表 4 中可以看出，上证 A 股指数平均值的未来三个月的预测值有略微下降趋势，基本可以判断出未来三个月的上证 A 股指数具有下降的趋势。

五、结论及缺陷分析

（一）研究结论

1、建模结果

对于上证 A 股指数时间序列的波动性进行建模，引入了关于 ARIMA 模型的建模理论，通过模型识别与参数估计得到最优模型 ARMA（3,3）。然后进行模型检验，剔除未能通过显著性检验的参数项，最后得到通过检验的 AR（1）模型，结果如下：

$$X_t = 3103.983 + 0.4839 * X_{t-1}$$

2、预测结果

用所得到的模型进行预测，本文采用静态预测的方法对上证 A 股指数的平均值进行预测，每一项预测值都是基于前一项的实际值进行预测所得到的。相比动态预测中的由初始实际值进行的递推预测，无形地积累了预测过程中的每一项预测误差。这种静态预测的方法消除了时间序列预测值的统计误差，本文对 2017 年 1 月到 2017 年 6 月的上证 A 股指数平均值进行静态预测，再与真实值做相对误差计算，得到较小的预测误差，即每个月的误差都不超过 3%，说明预测精度较高，因此，本文采用静态预测方法进行对上证 A 股指数平均值的未来三个月的上证 A 股指数进行预测，得到的预测值如下：

上证 A 股指数平均值的未来三个月的预测结果			
	2017 年 7 月	2017 年 8 月	2017 年 9 月
预测值	3119.8604	3111.6657	3107.7003

从预测结果中可以看出，上证 A 股指数平均值的未来三个月的预测值有略微下降趋势。

(二)对策与不足

本文利用 ARMA 模型，采用 eviews 软件对月度上证 A 股指数进行预测分析，该方法可以对现实生活中具有明显趋势的经济数据进行建模，研究其内在规律，更好地把握未来的发展趋势。从结果可知，对于短期预测来说，运用 ARMA 模型具有一定的可行性，但其只限于短期预测，对于长期趋势或者突发情形，就会表现出局限性，预测的偏差会比较大，因为影响股票价格波动的因素很复杂，有些也很难量化，包括如宏观经济政策因素、政府政策、国际环境影响等。这些因素在 ARMA 模型中只能以随机扰动项表示，而在预期的期望值中却无法表现出来。另外，这种传统的多元回归模型均为齐方差模型，且假定模型残差的均值为零、方差为常数，但实际上我国股价指数序列往往存在异方差现象。

鉴于此，我们可以采用其他用途更加广泛的模型，例如 ARCH 模型、GARCH 模型、EGARCH 模型等，这些模型考虑的综合因素更多更全面，从而得到的相关的预测结果更加贴近实际，可以得到更好的参考。

参考文献

- [1]Brown, Rozeff. Arbitrage, Short sales, and Financial Innovation[J]. Econometrica, 59:1041-1068, 1978.
- [2]Brown et al. The Microstructure of Stock Markets. CEPR Discussion Papers, 1987.
- [3]Barefield, Comiskey. Efficient capital markets II. Journal of Finance 46, 1975-1617.
- [4]Fried, Givoly. Economic Forces and the Stock Market[J]. Journal of Business, 1982, 59(3): 383-403.
- [5]O' Brien. New facts in finance[J]. Economic Perspectives Federal Reserve Bank of Chicago, 1988, 23(3):36-58.
- [6]Calderon. Investor sentiment and the near-term stock market[J]. Journal of Empirical Finance, 1993.
- [7]孙赞生. SV-MAE模型的构建及在上证指数中的应用[D]. 苏州大学, 2016.
- [8]崔远远、文忠桥等. 基于ARIMA模型对上证指数月度时间序列的分析和预测[J]. 安徽财经大学, 2015.
- [9]卢金长. 数据挖掘技术在上证指数预测中的应用研究[D]. 暨南财经大学, 2014.
- [10]纪华东. 中国股票市场有效性研究综述机. 市场周刊: 理论研巧, 2014(1): 85-86.