

# Forecast of the Total Sales of Wholesale and Retail in Shandong Province—Based on ARIMA Model

Yongjuan Ma

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan  
Email: milio727@outlook.com

Received: Aug. 10<sup>th</sup>, 2016; accepted: Aug. 24<sup>th</sup>, 2016; published: Aug. 31<sup>st</sup>, 2016

Copyright © 2016 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The sales price of wholesale and retail of commodities is one of the important factors that affect people's living standard. It has positive significance to forecast the total sales of the wholesale and retail of commodities. Time series analysis provides a suit of methods dealing with dynamic data with the scientific basis, namely with the analysis and investigation, we can constitutionally know the structure and complex character of the data, so we can achieve the purpose of forecasting its development trend and putting up essential control. In this paper, using the method of Box-Jenkins ARIMA model, the date sequence of wholesale and retail sales from 1979 to 2014 of Shandong province is analyzed, the model of auto regressive integrated moving average ARIMA (0,1,6) was established. The results show that the ARIMA model has good fitting effects on the original data sequence. The prediction effect can be used for short-term prediction of the total sales of wholesale and retail in Shandong province. It provides a basis and reference for government departments to formulate economic plans. According to the results of the prediction model established, in Shandong province the wholesale and retail sales remain high growth trend in the coming years.

## Keywords

Auto Regressive Integrated Moving Average Model, Wholesale and Retail Industry, Prediction of Total Sales, Time Series Analysis

---

# 山东省批发和零售业销售总额的预测——基于ARIMA模型

马永娟

云南财经大学统计与数学学院, 云南 昆明  
Email: milio727@outlook.com

收稿日期: 2016年8月10日; 录用日期: 2016年8月24日; 发布日期: 2016年8月31日

## 摘要

商品批发与零售业销售价格是影响人们生活水平的重要因素之一, 因此对商品批发和零售业销售总额进行预测有着积极的意义, 时间序列分析提供了一套具有科学依据的动态数据处理方法, 就是通过对模型的分析研究去了解数据的内在结构和复杂特性, 从而达到预测其发展趋势并进行必要的控制的目的。本文利用Box-Jenkins法的ARIMA模型, 对1979~2014年山东省批发和零售业销售总额数据序列进行分析, 建立了1979~2014年山东省批发和零售业销售总额的自回归移动平均模型ARIMA(0,1,6)。检验结果显示, ARIMA(0,1,6)模型对原始数据序列有着较好的拟合效果, 可用于短期内山东省批发和零售业销售总额的预测, 为政府部门制定经济计划提供依据和参考。根据建立的模型预测结果, 山东省批发和零售业销售总额在未来几年仍将保持较高的增长趋势。

## 关键词

ARIMA模型, 批发零售业, 销售总额预测, 时间序列分析

## 1. 引言

改革开放以来, 我国的批发和零售业得到了较快发展, 促使社会消费品零售总额持续较快增长。我国批发和零售业无论从繁荣发展的程度看, 还是从对国民经济的贡献率看, 都是我国国民经济十分重要的行业, 对拉动和扩大我国内需, 促进我国经济快速平稳发展起到了越来越大的作用[1]-[5]。国外大型批发和零售业全面进入我国, 对我国本土批发和零售业带来了强大冲击和市场压力, 不仅使我国批发和零售业的国际化程度越来越高, 还使我国境内的批发和零售业的市场竞争变得越来越激烈, 批发和零售业的演变和发展趋势越来越复杂[6] [7]。面对国际上批发和零售业在我国零售市场上的规模化扩张, 我国的批发和零售行业能否在“同台竞争”中不断发展壮大自己, 还有我国零售和批发业未来的发展趋势, 都是我们长久以来十分重视的问题。

本文以山东省批发和零售业为研究对象, 以此来研究山东省批发和零售业未来的发展趋势。批发和零售业销售总额数据序列受到许多因素的影响, 这些因素之间存在着很多复杂的联系, 往往会表现出某种随机性, 但是它们彼此之间也存在着统计上的依赖关系。本文把山东省批发和零售业的销售总额作为一个时间序列, 根据时间序列的性质[8], 再以以前的资料为依托得到它的变化规律, 以此来预测批发和零售业未来发展的趋势。为了准确的预测山东省批发和零售业的发展趋势, 要求所建模型必须符合实际情况, 并且满足统计理论的要求, 这就需反复使用自回归单程移动平均模型 ARIMA(p,d,q)调整模型[9]-[12], 选择出最适合山东省批发和零售业销售总额预测的模型, 并应用此模型为山东省批发与零售业销售总额进行预测, 预测结果可以为相关政府部门就此行业发展计划的制定提供一定的依据和参考价值。

## 2. 山东省批发和零售业销售总额模型方法

### 2.1. 数据查询

首先从山东省统计局网站查询数据, 得到山东省从 1979 年到 2014 年的批发和零售业的销售总额数

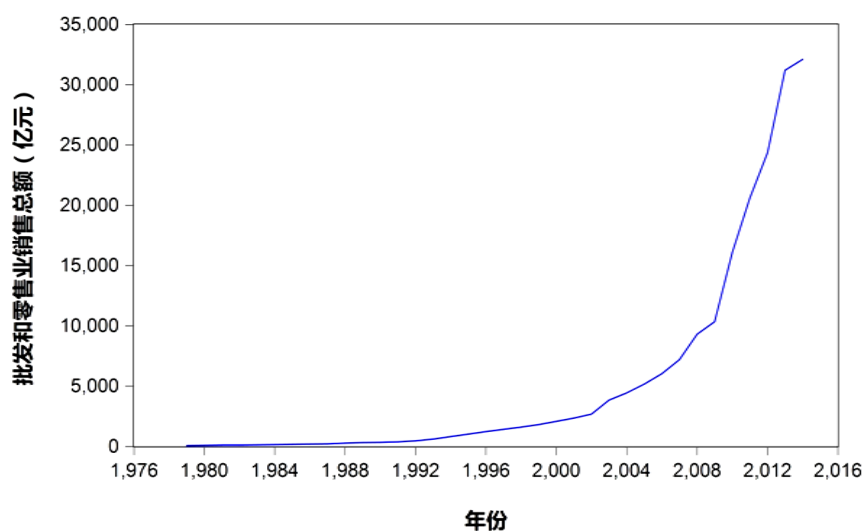
值(用  $y_t$  表示)。随着时间的推移,山东省批发和零售业的销售总额是逐年增大的。具体数值如表 1 所示。

按此表 1,山东省批发和零售业销售总额数据作散点图,如图 1,从图中可以清晰的看出,批发和零售业销售总额序列  $y_t$  整体呈现一定的上升趋势,即山东省批发和零售业销售总额是逐年递增的。特别是 2010 年以后,呈现出强劲的增长势头。该序列为非平稳序列[13],有明显的上升趋势。

**Table 1.** Total sales in wholesale and retail industry (unit: one hundred million yuan)

**表 1.** 批发和零售业销售总额 (单位: 亿元)

时间	销售总额	时间	销售总额
1979	78.38	1997	1425.50
1980	94.61	1998	1600.27
1981	107.10	1999	1807.00
1982	112.90	2000	2075.94
1983	127.67	2001	2340.68
1984	147.25	2002	2691.49
1985	173.62	2003	3836.66
1986	194.85	2004	4444.04
1987	217.34	2005	5173.89
1988	287.09	2006	6044.60
1989	315.80	2007	7205.96
1990	338.02	2008	9314.97
1991	392.19	2009	10,348.40
1992	471.87	2010	16,105.60
1993	617.53	2011	20,586.70
1994	813.17	2012	24,366.90
1995	1024.82	2013	31,193.4
1996	1226.57	2014	32,112



**Figure 1.** Trend chart of the total wholesale and retail sales

**图 1.** 批发和零售业销售总额趋势

## 2.2. ARIMA 模型原理

ARIMA( $p,d,q$ )模型, 即自回归单程移动平均模型具有如下结构

$$\begin{cases} \Phi(L)\Delta_d y_t = c + \Theta(L)\varepsilon_t, \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t, \\ Ey_t \varepsilon_t = 0, s < t. \end{cases}$$

上式中,  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  为 ARMA( $p,q$ )模型的自回归系数多项式,  $\Delta_d = (1-L)^d$ ,  $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  为 ARMA( $p,q$ )模型的移动平均系数多项式, 并且上式可以简记为

$$\Delta_d y_t = u + \frac{\Theta(L)}{\Delta_d y_t} \varepsilon_t.$$

式中  $u$  是时间序列  $y_t$  的均值,  $\{\varepsilon_t\}$  为零均值的白噪声序列。由此可知, ARIMA1 模型实质上就是差分运算与 ARMA 模型的组合。

以时间序列的自相关分析为基础, ARIMA( $p,d,q$ )模型在经济预测的过程中既考虑了经济现象在时间序列上的依存性, 又考虑了其随机波动的干扰性, 此模型对于经济运行短期趋势的预测准确率较高, 是应用比较广泛的时间序列分析方法之一。

若时间序列  $y_t$  通过  $d$  次差分后可转换成平稳时间序列, 即  $y_t \sim I(d)$ , 则

$$\mu_t = \Delta^d y_t = (1-B)^d y_t,$$

$\mu_t$  为平稳序列, 即  $\mu_t \sim I(0)$ , 于是可建立 ARMA( $p,q$ )模型

$$\mu_t = c + \phi_1 \mu_{t-1} + \dots + \phi_p \mu_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

经  $d$  阶差分后的 ARMA( $p,q$ )模型称为 ARIMA( $p,d,q$ )模型, 即自回归单程移动平均模型, 其中  $p$  为自回归模型的阶数,  $q$  为移动平均的阶数,  $\varepsilon_t$  为白噪声过程。模型参数估计可采用条件最小二乘估计(CLS)或最大似然估计(ML)。

而本文对山东省批发和零售业销售总额的研究, 通过往年的历史数据随着时间的变化是如何变化的, 进而建立一个模型, 使得此模型有效的反映出销售总额随时间变化规律。从而进行未来发展趋势以及销售总额的预测。ARIMA 模型法完全符合我们的研究对象, 所以我们本文选用 ARIMA 模型法进行建模分析与预测。

## 3. 山东省批发和零售业销售总额 ARIMA 模型

### 3.1. 数据处理

山东省批发和零售业的销售总额数据变动幅度比较大, 是非平稳序列。为了消除原始数据序列的非平稳性, 使数据更为平稳, 本文采用对山东省批发和零售业销售总额取对数形式(记为  $\ln y_t$ ), 并利用 EXCEL 对  $\ln y_t$  序列作一阶差分(记为  $\Delta \ln y_t$ ), 对  $\ln y_t$  序列作二阶差分(记为  $\Delta^2 \ln y_t$ ), 所得数据对一阶差分后的数据  $\Delta \ln y_t$  作序列趋势图, 见图 2, 从图中可以看出, 时间趋势基本消除, 可认为是平稳性序列。但序列的趋势图只能粗略地判断序列具有平稳性, 理论上应用单位根检验法检验。

序列的趋势图只能粗略地判断序列是否具平稳性不能准确地判断该序列是否为平稳性序列, 为了更加准确的判断批发和零售业销售总额数据是否为平稳序列, 我们对  $\Delta \ln y_t$  进行平稳性检验(ADF 检验),  $\Delta \ln y_t$  序列检验  $ADF = -4.876862$ , 分别小于三个检验水平不同的临界值 ( $t_{1\%} = -3.653730, t_{5\%} = -2.957110, t_{10\%} = -2.617434$ ), 拒绝了存在单位根的零假设,  $\Delta \ln y_t$  是平稳时间序列。

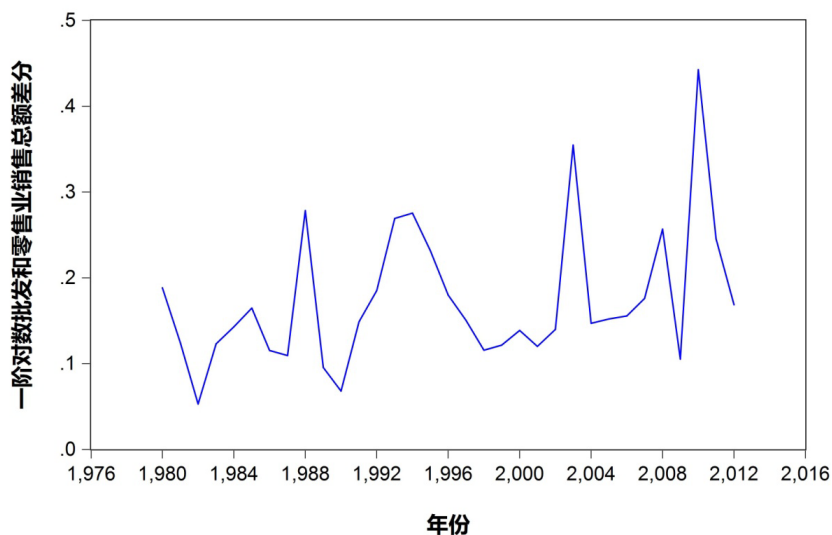


Figure 2.  $\Delta \ln y_t$  sequence trend figure

图 2.  $\Delta \ln y_t$  的序列趋势图

在建模过程中，没有必要进行二阶差分。取对数的山东省批发和零售业销售总额序列为 1 阶单整序列，即  $\Delta \ln y_t \sim I(1)$ 。

### 3.2. 模型建立与检验

由以上对序列  $\Delta \ln y_t$  的 ADF 检验，可确定 ARIMA( $p, d, q$ ) 模型中  $d$  的取值为 1。同时序列  $\Delta \ln y_t$  的  $DW = 1.985865$ ，说明序列  $\Delta \ln y_t$  不存在自相关。为了确定 ARIMA 模型中的  $p$  和  $q$  的值，我们作序列  $\Delta \ln y_t$  的自相关(ACF)和偏自相关(PACF)图，见图 3。

由图 3  $\Delta \ln y_t$  序列的自相关与偏自相关图可以看出，ACF 表现截尾，但 PACF 表现拖尾，故选用模型 ARIMA(0,1, $q$ ) 型。自相关系数在滞后阶数为 6 时显著不为零，在滞后阶数大于 6 时基本都处于置信带内，所以可取  $q = 6$ 。我们对时间序列  $\Delta \ln y_t$  反复拟合，进而得到一个较好的时间序列模型。回归结果如表 2。

故估计的 ARIMA(0,1,6) 模型如下：

$$\Delta \ln y_t = 0.181642 + \varepsilon_t + 0.820629\varepsilon_{(t-6)}.$$

山东省批发和零售业销售总额随时间演化的模型即得到。对模型的 Q 统计量进行白噪声检验，即模型检验如图 4。

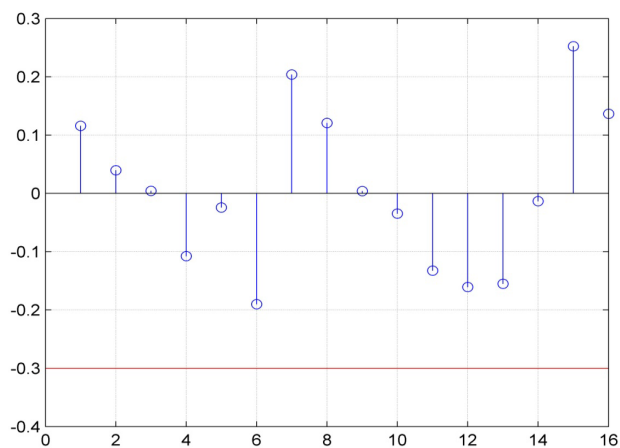
对残差序列进行白噪声检验，可以看出 ACF 和 PACF 都没有显著异于零，Q 统计量的 P 值都远远大于 0.05，因此可以认为残差序列为白噪声序列，模型信息提取比较充分。常数和滞后六阶参数的 P 值都很小，参数显著；因此整个模型比较精简，山东省批发和零售业销售总额的 ARIMA 模型较优。能有效地反应批发与零售业销售总额随时间的变化趋势。

山东省批发和零售业销售总额随时间演化的模型即得到。对模型的 Q 统计量进行白噪声检验，即模型检验如图 5 和图 6。

对残差序列进行白噪声检验，可以看出 ACF 和 PACF 都没有显著异于零，Q 统计量的 P 值都远远大于 0.05，因此可以认为残差序列为白噪声序列，模型信息提取比较充分。常数和滞后六阶参数的 P 值都很小，参数显著；因此整个模型比较精简，山东省批发和零售业销售总额的 ARIMA 模型较优。能有效地反应批发与零售业销售总额随时间的变化趋势。

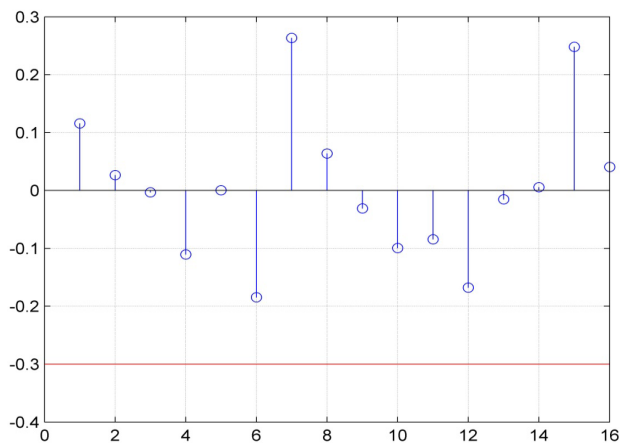
**Table 2.** Parameter estimation  
**表 2.** 参数估计

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.181642	0.007415	24.49518	0.0000
MA(6)	-0.860269	0.058949	-14.59342	0.0000
R-squared	0.327539	Mean dependent var		0.173912
Adjusted R-squared	0.305847	S.D. dependent var		0.082022
S.E. of regression	0.068337	Akaike info criterion		-2.470036
Sum squared resid	0.144769	Schwarz criterion		-2.379339
Log likelihood	42.75559	Hannan-Quinn criter.		-2.439519
F-statistic	15.09936	Durbin-Watson stat		1.749398
Prob (F-statistic)	0.000501			
Inverted MA Roots	0.98	0.49 - 0.84i	0.49 + 0.84i	-0.49 - 0.84i
	-0.49 + 0.84i		-0.98	



**Figure 3.**  $\Delta \ln y_t$  autocorrelation figure

**图 3.**  $\Delta \ln y_t$  的自相关图



**Figure 4.**  $\Delta \ln y_t$  partial autocorrelation figure

**图 4.**  $\Delta \ln y_t$  的偏自相关图

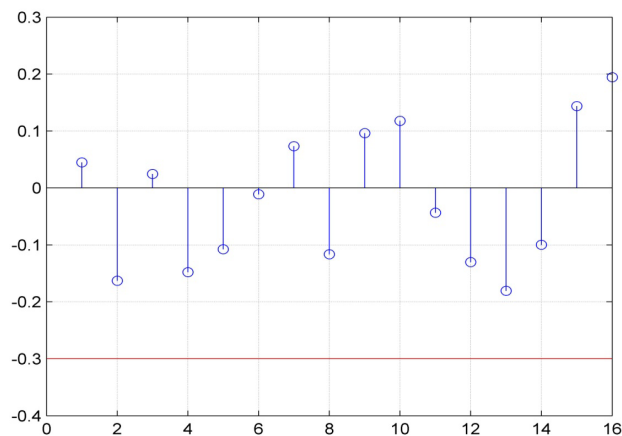


Figure 5. White noise autocorrelation plot

图 5. 白噪声的自相关图

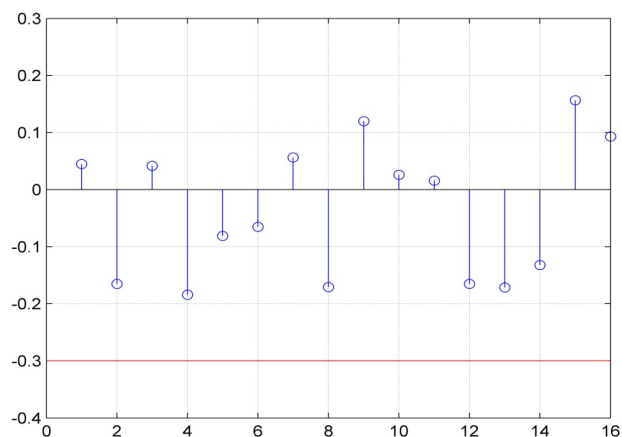


Figure 6. White noise partial autocorrelation plot

图 6. 白噪声的偏自相关图

### 3.3. 模型预测

根据检验和比较的结果,选择最后的方程模型,使用 EViews 软件中的预测功能对模型进行预测时,如果预测误差较小,就可以用这个模型预测,即可得到原时间序列的将来走势。但是如果预测误差较大,我们就需要重新调整模型,直到选择出合理的预测模型。EViews 软件中,预测 2007 年至 2012 年的  $\Delta \ln y_t$  序列值,预测情况如图 7 所示。

图中左边是预测值与置信区间,右边是预测的误差。Root mean squared error 代表均方误差方,MAE 表示平均绝对误差,MAPE 表示平均绝对误差百分比。Theil inequality coefficient 表示不等系数,bias proportion 表示偏误,即预测均值与真实均值的偏离程度;variance proportion 表示方差误,用来反映预测波动与真实波动之间的差异;covariance proportion 表示协方差误,反映残存非系统性预测误差,该误差占比越大,预测效果越好。本例中的协方差误(0.552111)远远大于方差误(0.367295),因此预测效果较好。

利用 ARIMA(0,1,6)模型对山东省 2011~2014 年批发和零售业的销售总额作预测,并与真实值比较,进一步检验模型的适应性。具体结果见表 3。

由表 3 可知,实际值和预测值相差较小,预测的相对误差都不超过 5%,说明模型非常适合短期预测。通过模型还预测出山东省 2015 年至 2017 年批发和零售业销售总额,具体结果见表 4。



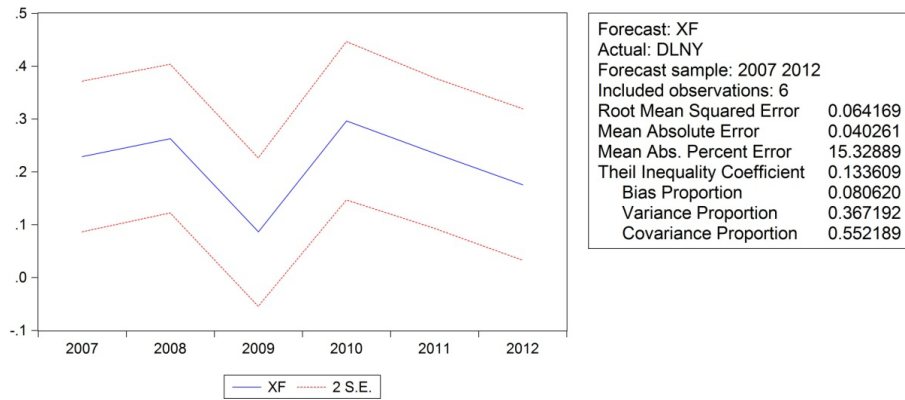


Figure 7. Predictive diagram and related statistics of  $\Delta \ln y_t$  sequence

图 7.  $\Delta \ln y_t$  序列的预测图及相关统计量

Table 3. Comparison between real and predictive values of total sales in the wholesale and retail industry of Shandong province

表 3. 山东省批发和零售业销售总额真实值和预测值的对照

年份	真实值	预测值	相对误差%
2011	20,586.70	20,372.03	1.05
2012	24,366.90	24,540.27	0.71
2013	31,193.4	31,586.6421	1.26
2014	32,112	31,868.4486	0.76

Table 4. 2015-2017 predictive value of the total amount of wholesale and retail sales of Shandong province (unit: one hundred million yuan)

表 4. 2015 年~2017 年山东省批发和零售业销售总额的预测值 (单位: 亿元)

年份	预测值
2015	43,498.9005
2016	46,002.4426
2017	54,669.2012

#### 4. 结束语

ARIMA(0,1,6)模型对于山东省批发与零售业销售总额的预测结果显示其预测值与真实值相对误差小于 5%，从而可以判断此模型对原序列有较好的预测效果。同时预测结果表明，山东省批发和零售业年度销售总额保持较快增长；其中批发和零售业销售总额是社会消费品零售总额的主要组成部分，故社会消费品零售总额也保持较快增长。社会消费品零售总额可以很好的反映人民物质文化生活水平，还能够反映出商品购买力的实现程度和市场规模，批发和零售业销售总额是社会消费品零售总额的主要组成部分，故山东省未来几年的经济将处于快速发展的阶段。

#### 参考文献 (References)

- [1] 赵蕾, 陈美英. ARIMA 模型在福建省 GDP 预测中的应用[J]. 科技和产业, 2007, 7(1): 45-47.
- [2] 李子奈, 潘文卿. 计量经济学[M]. 北京: 高等教育出版社, 2011: 261-305.
- [3] 李智勇, 孙小英. ARMA 模型在批发和零售贸易餐饮业预测中的应用[J]. 北京市财贸管理学院学报, 2006, 5(3):



25-28.

- [4] 安鸿志. 时间序列分析[M]. 上海: 华东师范大学出版社, 1992.
- [5] 易丹辉. 数据分析与 Eviews 应用[M]. 北京: 中国人民大学出版社, 2008: 146-154.
- [6] 全福生, 彭白玉. ARMA 模型在中国股市中的应用[J]. 衡阳师范学院学报, 2009, 30(3): 26-28.
- [7] 殷英, 谭元发. 我国新能源 ARMA 模型分析与预测[J]. 数学理论与应用, 2009, 29(4): 72-77.
- [8] 王振龙. 时间序列分析[M]. 北京: 中国统计出版社, 1999: 89-125.
- [9] 范玉妹, 玄婧. ARMA 算法在 GDP 预测中的应用[J]. 江南大学学报: 自然科学版, 2010, 9(6): 736-740.
- [10] 王若羽. 基于 ARMA 模型的我国国内生产总值的预测研究[J]. 商场现代化, 2011, 649(16): 152.
- [11] 李先孝. 时间序列分析基础[M]. 武汉: 华中理工大学出版社, 1991: 113-116.
- [12] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2005.
- [13] Badescu, V. (2001) Time-Series Analysis and Forecasting of Daily Averaged Surface Pressure on Mars. *Meteorology and Atmospheric Physics*, 78, 195-204. <http://dx.doi.org/10.1007/s703-001-8173-x>

#### 期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>