

Research on Financial Revenue Forecast Based on Optimal ARIMA Model

Lanling Liu, Deshan Sun

School of Mathematics, Liaoning Normal University, Dalian Liaoning
Email: 1171911810@qq.com

Received: Mar. 1st, 2020; accepted: Mar. 16th, 2020; published: Mar. 24th, 2020

Abstract

Fiscal revenue is an important economic lever for the state to implement macroeconomic regulation and control. Accurately predicting fiscal revenue is of great significance to it, and the ARIMA model is an effective tool for short-term forecasting of fiscal revenue. Taking the 1950~2016 fiscal revenue as a sample, multiple sets of ARIMA models were established, and then multiple screening criteria were applied to find the optimal lag order p and q . Finally, the optimal ARIMA(3,2,8) model was determined. The model has passed a number of hypothesis tests to predict the fiscal revenue of 2017~2018, and the conclusions indicate that the prediction accuracy is high. At the same time, this model is also used to forecast the fiscal revenue of 2019~2022 in the next few years, providing certain reference value for the country and the government.

Keywords

ARIMA Model, Fiscal Revenue, Forecast

基于最优ARIMA模型财政收入预测研究

刘兰苓, 孙德山

辽宁师范大学数学学院, 辽宁 大连
Email: 1171911810@qq.com

收稿日期: 2020年3月1日; 录用日期: 2020年3月16日; 发布日期: 2020年3月24日

摘要

财政收入是国家对经济实行宏观调控的重要经济杠杆, 准确预测财政收入对其意义重大, 而ARIMA模型是短期预测财政收入的有效工具。以1950~2016年财政收入为样本, 建立多组ARIMA模型, 进而运用重

筛选准则, 找到最优滞后阶数 p 和 q , 最后确定了最优ARIMA(3,2,8)模型。该模型通过了多项假设检验, 对2017~2018年的财政收入进行预测, 结论表明预测精度高。同时也利用该模型对2019~2022年的财政收入进行了预测, 为国家及政府提供一定的参考价值。

关键词

ARIMA模型, 财政收入, 预测

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

财政收入是衡量一个国家经济实力的标志, 也是一个国家对经济实行宏观调控的重要手段, 财政收入的增长关系着一个国家和社会的经济发展和进步。因此, 准确了解财政收入的预测值具有重要的现实意义, 能够为政府提供一定的决策[1]。

2. ARIMA(p,d,q)模型

时间序列分析主要分为三种模型: 包括移动平均模型(MA)、自回归模型(AR)、以及自回归移动平均模型(ARMA)。其中 AR 模型和 MA 模型是 ARMA 的一种特殊情况, 本文研究的 ARIMA 模型也是 ARMA 模型的推广[2], ARIMA 模型是经过差分后能够将非平稳时间序列转换成平稳的序列, 依然适合 ARMA 模型的称为 ARIMA 模型。ARIMA 模型又称为差分自回归移动平均模型, 是著名的时间序列预测分析方法之一, 由美国统计学家 Box 和英国统计学家 Jenkins 于 20 世纪 70 年代初所提出[3]。原理是将因变量仅对它的滞后值, 以及随机误差项的现值和滞后值进行回归, 进而建立模型, 再次按照 ARMA 模型的方法进行建立。

3. 建立模型

若 X_t 是 d 阶单整序列, 其模型结构如下:

$$X_t = C + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q}$$

其中 C 是常数, ε_t 是白噪声, α 和 β 是待估参数, p 是自回归模型的最大阶数, 而 q 是移动回归模型的最大阶数。

3.1. ARIMA 模型预测步骤

- › 步骤 1 识别
利用相关图和偏相关图, 找出适当的 p , d 和 q 值
- › 步骤 2 估计
估计回归模型中的自回归和移动平均项的参数
- › 步骤 3 诊断
对回归模型进行检验, 看残差是不是白噪声, 如果是就接受拟合模型, 反之则从步骤 1 重新开始做[4]
- › 步骤 4 预测

利用回归模型进行预测

3.2. 数据来源

数据选自《中国统计年鉴》，本文采用 1950~2018 年的财政收入总值，其中无国内外债务收入。具体财政收入的时间走势图见图 1。

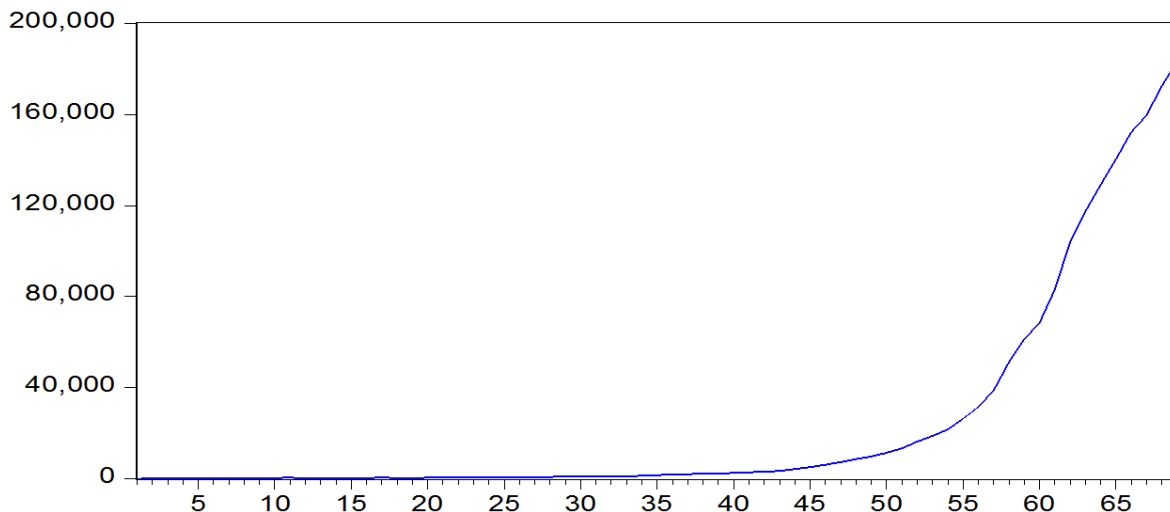


Figure 1. 1950~2018 annual fiscal revenue (100 million yuan)

图 1. 1950~2018 年度财政收入(亿元)

4. ARIMA 模型对财政收入建模

4.1. 财政收入序列平稳性检验

本文采用 Eview8.0 软件对数据进行统计分析，1950~2018 年的财政收入总值数据来源《中国统计年鉴》，财政收入中不包括国内外债务收入。首先对导入的数据财政收入 Y 取对数，会得到一个新的序列 Y_1 ，然后再对 Y_1 序列进行单位根测验，得知 Y_1 序列是不平稳的[5]，因此对 Y_1 进行一阶差分，得到结果如图 2，依然不平稳，再进行二阶差分，得到结果如图 3， p 值等于 0，所以财政收入 Y 是二阶单整。因此建立 ARIMA($p,2,q$) 模型。

还应该确定一下滞后阶数 p 和 q ，然后再进行自相关系数检验，最终确定估计方程。

Null Hypothesis: Y2 has a unit root
 Exogenous: Constant
 Lag Length: 5 (Automatic - based on SIC, maxlag=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.432259	0.1373
Test critical values:		
1% level	-3.540198	
5% level	-2.909206	
10% level	-2.592215	

Figure 2. Sequence stationary test results after first order difference

图 2. 一阶差分后序列平稳性检验结果

Null Hypothesis: Y3 has a unit root
 Exogenous: Constant
 Lag Length: 4 (Automatic - based on SIC, maxlag=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.521561	0.0000
Test critical values:		
1% level	-3.540198	
5% level	-2.909206	
10% level	-2.592215	

Figure 3. Sequence stationary test results after second-order difference

图 3. 二阶差分后序列平稳性检验结果

4.2. 建立最优的 ARIMA 模型

从图 3 中可以看出, 自相关系数(AC)和偏相关系数(PAC)的变动没有一个很明显的规则, 但可以初步判别从第 6 阶后逐渐下降。因此根据 Akaike info、Schwarz 以及 Hannan-Quinn 准则筛选, 进行多种 p 和 q 的建模尝试, 通过 LM 检验法、JB 统计量检验法、拟合优度检验[6], 找到最优的 p, q 值, 分别为 $p=3, q=8$, 所以建立 ARIMA(3,2,8) 模型[7]。如下图 4:

Dependent Variable: Y3
 Method: Least Squares
 Date: 08/06/19 Time: 15:56
 Sample (adjusted): 6 69
 Included observations: 64 after adjustments
 Convergence achieved after 28 iterations
 MA Backcast: -2 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.002514	0.001270	1.979769	0.0530
AR(1)	-0.225613	0.134472	-1.677766	0.0994
AR(2)	-0.296851	0.116901	-2.539341	0.0141
AR(3)	-0.513466	0.124260	-4.132177	0.0001
MA(1)	-0.394289	0.088692	-4.445616	0.0000
MA(2)	-0.420484	0.100807	-4.171165	0.0001
MA(3)	0.359424	0.102011	3.523368	0.0009
MA(4)	-0.100190	0.082428	-1.215488	0.2297
MA(5)	-0.238780	0.095343	-2.504434	0.0154
MA(6)	0.481708	0.083327	5.780908	0.0000
MA(7)	0.221149	0.089200	2.479237	0.0164
MA(8)	-0.866370	0.063266	-13.69418	0.0000
R-squared	0.728334	Mean dependent var		-0.001235
Adjusted R-squared	0.670866	S.D. dependent var		0.149806
S.E. of regression	0.085944	Akaike info criterion		-1.902878
Sum squared resid	0.384092	Schwarz criterion		-1.498087
Log likelihood	72.89209	Hannan-Quinn criter.		-1.743410
F-statistic	12.67378	Durbin-Watson stat		2.034195
Prob(F-statistic)	0.000000			
Inverted AR Roots	.26+.79i	.26-.79i	-.75	
Inverted MA Roots	.99	.80-.58i	.80+.58i	.13+.99i
	.13-.99i	-.74-.60i	-.74+.60i	-.98

Figure 4. ARIMA(3,2,8) model output

图 4. ARIMA(3,2,8) 模型的输出结果

残差拟合图见图 5。

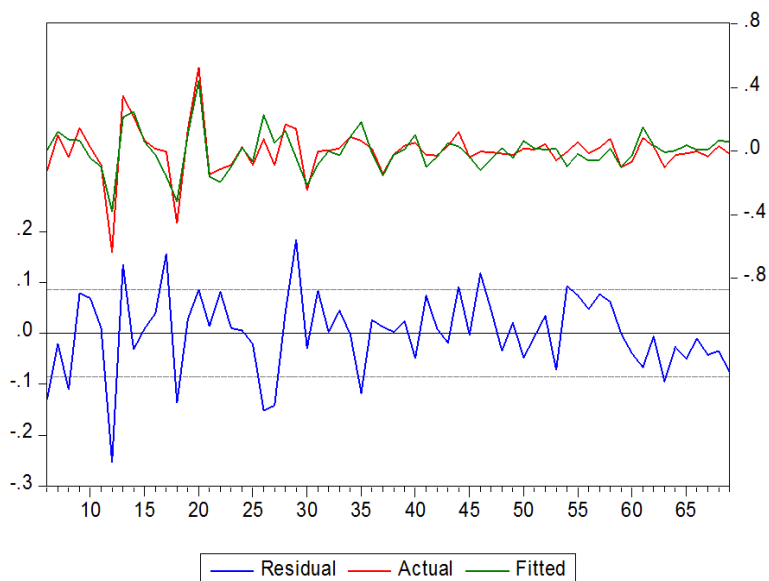


Figure 5. ARIMA(3,2,8) residual fit of the model

图 5. ARIMA(3,2,8) 模型的残差拟合图

4.3. 拟合优度检验

从图 4 中可以看出 Adjusted R-square 值为 0.670866，残差拟合图见图 5，能够看出残差拟合效果不错，尤其是第 57 个数据，也就是 2005 年之后的预测精度是比较准确的。

4.4. 残差正态性检验

对残差进行正态性检验，结果如图 6，图中可以看出 JB 统计量所对应的概率值是大于 0.1 的置信水平，所以该模型得到的回归方程的残差是服从正态分布。

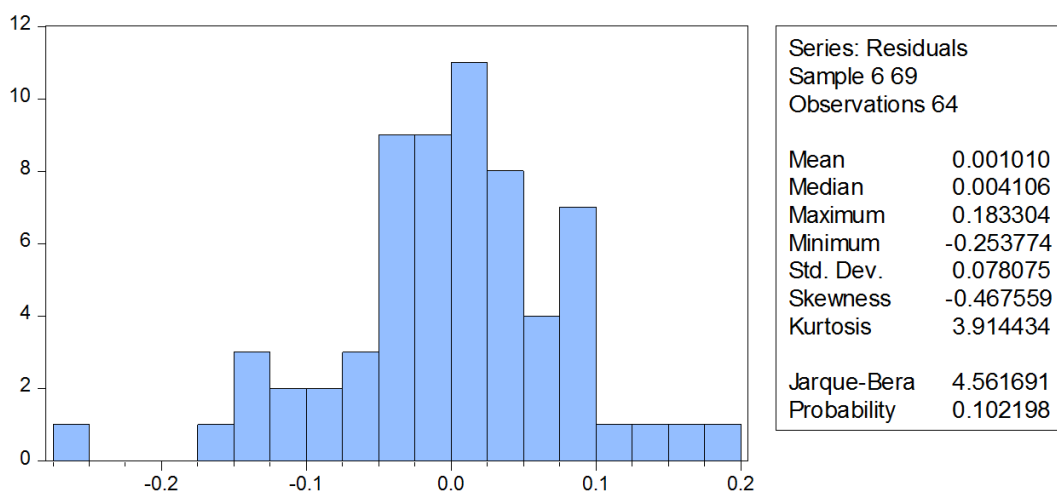


Figure 6. ARIMA(3,2,8) results of the model's residual normality test

图 6. ARIMA(3,2,8) 模型的残差正态性检验结果

4.5. 方程显著性检验

从图 4 能够看出 F -statistic 的值为 12.67378, 所对应的概率值 p 为 0, 说明该模型通过了 F 检验。

4.6. 变量显著性检验

从图 4 中能够看出, 在 0.05 置信水平下, 有 AR(2)、MA(1)、MA(2)、MA(3)、MA(5)、MA(6)、MA(7)、MA(8)通过了系数显著性 t 测验, 其它的应该从模型方程中剔除。

4.7. 对 ARIMA(3,2,8) 模型进行残差自相关检验(LM 检验法)

根据模型的特点, 采用 LM 检验法, 该模型自相关检验的结果如下图 7。从图 7 中可以看出 F -statistic 和 Obs*R-square 所对应的概率 p 值分别是 0.6735 和 0.6030, 都是大于 0.1 的显著性水平, 因此说明残差不存在序列相关性, 为白噪声[8]。

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.398546	Prob. F(2,48)	0.6735
Obs*R-squared	1.011754	Prob. Chi-Square(2)	0.6030

Figure 7. Residual series autocorrelation test of the model

图 7. ARIMA(3,2,8) 模型的残差序列自相关检验

因此, 得到 ARIMA(3,2,8) 模型的最终的回归方程:

$$Y_3 = -0.296851Y_{3,t-2} - 0.394289V_{t-1} - 0.420484V_{t-2} + 0.359424Y_{3,t-3} - 0.238780V_{t-5} + 0.481708V_{t-6} + 0.221149V_{t-7} - 0.866370V_{t-8}$$

其中 $F = 12.67378, AIC = -1.902878, SC = -1.498087$ 。

5. 模型的预测

利用建立的 ARIMA(3,2,8) 模型对 2017~2020 年的财政收入进行预测, 根据差分反推计算, 财政收入 $_t =$ 财政收入 $_{t-1} \times e^{Y_3}$ (Y_3 表示取对数后的财政收入)进而得到预测[9]。2017~2018 年的财政收入是已知, 下表 1 将年财政收入的真实值与预测值进行对比分析, 计算出预测误差率。见表 1:

Table 1. Forecasting fiscal revenue forecast for 2016~2022 using models

表 1. 利用 ARIMA(3,2,8) 模型预测 2016~2022 年财政收入预测

	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年
Y_3	0.0445856	-0.0008915	-0.0263371	0.0318133	0.0104155	0.0505950
实际值(亿元)	172,592.77	183,351.84	无	无	无	无
预测值(亿元)	178,210.09	197,844.35	205,592.43	224,969.65	255,683.39	284,048.39
增长率%	10.16	11.02	3.92	9.43	13.65	11.09
预测误差率%	-3.25	-7.90	无	无	无	无

6. 结论

从表 1 可以看出, 对 2017~2018 年财政收入模型预测的误差率分别是-3.25%和-7.90%, 预测精度比较高。因此对 2019~2022 年财政收入预测, 也是有比较好的预测精度[10]。

本文研究的模型预测精度较高主要原因有三: 第一, 选取的数据样本是国家统计局网站上的1950~2018年的财政收入总值, 总计样本个数为69个, 满足ARIMA模型所需的样本数量; 第二, 模型分析时也进行了多组试验, 并根据AIC和SC准则、LM检验法、JB统计量检验法及拟合优度检验等多重筛选原则, 找到适合模型的最优的 p, q 值, 进而建立了最优的财政收入预测模型, 是ARIMA(3,2,8)模型[11]; 第三, 模型已经通过了多种检验, 并且拟合优度较好, 残差高度也符合正态分布, 且为白噪声。

基金项目

辽宁省自然科学基金指导计划项目(编号: 2019-ZD-0471)。

参考文献

- [1] 郑鹏辉. 基于ARIMA模型的组合模型研究[D]: [硕士学位论文]. 秦皇岛: 燕山大学, 2009.
- [2] 熊志斌. ARIMA融合神经网络的人民币汇率预测模型研究[J]. 数量经济技术经济研究, 2011, 28(6): 64-76.
- [3] 李娜, 薛俊强. 基于最优ARIMA模型的我国GDP增长预测[J]. 统计与决策, 2013(9): 23-26.
- [4] 章博伟. 北京市财政收入预测[D]: [硕士学位论文]. 大连: 东北财经大学, 2017.
- [5] 吕宁. 地方财政一般预算收入预测模型研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2006.
- [6] 郑鹏辉, 单锐, 陈静. 时间序列分析在我国财政收入预测中的应用[J]. 重庆文理学院学报(自然科学版), 2008, 27(2): 15-18.
- [7] 徐向辉. 基于ARIMAX模型的全国财政收入的预测与分析[D]: [硕士学位论文]. 大连: 大连理工大学, 2019.
- [8] 孙轶轩, 邵春福, 计寻, 朱亮. 基于ARIMA与信息粒化SVR组合模型的交通事故时序预测[J]. 清华大学学报(自然科学版), 2014, 54(3): 348-353+359.
- [9] 成云, 成孝刚, 谈苗苗, 周凯, 李海波. 基于ARIMA和小波神经网络组合模型的交通流预测[J]. 计算机技术与发展, 2017, 27(1): 169-172.
- [10] 蒋泽迪, 程毛林. ARIMA乘积季节模型在财政收入预测中的应用[J]. 苏州科技大学学报(自然科学版), 2018, 35(1): 28-32.
- [11] 薛俊强. 宁波港集装箱吞吐量预测模型的选择[J]. 华东经济管理, 2013, 27(5): 169-172.