

Forecast and Research on GDP of Fujian Province Based on Time Series Analysis

Zhiyi Zhou, Qitong Ou

School of Applied Mathematics, Xiamen University of Technology, Xiamen Fujian
Email: 942830176@qq.com, ouqitong@xmut.edu.cn

Received: May 27th, 2020; accepted: Jun. 10th, 2020; published: Jun. 17th, 2020

Abstract

Firstly, based on the time series analysis theory, the GDP of Fujian province from 1978 to 2015 was fitted and analyzed by Eviews software; the ARIMA(4,2,2) model and ARIMA(4,3,2) model were established. Secondly, using the two models, the GDP of Fujian province in 2016~2018 was used to test. We found the prediction error of ARIMA(4,3,2) is smaller, so it is considered that the model has a good effect. Finally using ARIMA(4,3,2) model, the GDP of Fujian province in 2019~2023 was predicted.

Keywords

Time Series Analysis, ARIMA Model, GDP of Fujian Province, Eviews Software

基于时间序列分析的福建省GDP预测研究

周芷伊, 欧启通

厦门理工学院应用数学学院, 福建 厦门
Email: 942830176@qq.com, ouqitong@xmut.edu.cn

收稿日期: 2020年5月27日; 录用日期: 2020年6月10日; 发布日期: 2020年6月17日

摘要

首先基于时间序列分析原理, 通过Eviews软件对福建省1978~2015年的GDP数据进行拟合分析, 建立了ARIMA(4,2,2)和ARIMA(4,3,2)模型。然后应用两模型对福建省2016~2018年的GDP进行检验, 发现ARIMA(4,3,2)模型的预测误差较小, 所以认为该模型拟合效果较好。最后应用ARIMA(4,3,2)模型对福建省2019~2023年的GDP进行预测。

关键词

时间序列分析, ARIMA模型, 福建省GDP, Eviews软件

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

福建省是我国沿海经济带及海峡西岸经济区的重要省份, 近些年随着“一带一路”等政策的逐步实施, 经济得到持续提升, 但是福建省经济水平和其他先进省份相比还是存在一些差距的。因此对福建省GDP的趋势做出预测, 将有助于政府对未来做出更好的决策, 这将有益于政府对宏观经济的调控, 从而使宏观经济得到良好的发展[1]。目前GDP的预测模型主要有多元回归模型、生产函数模型、灰色理论模型、神经网络模型、时间序列模型等[2]。其中时间序列模型较为常用, 近年来我国也有许多应用时间序列模型对GDP进行预测研究的例子[3][4][5]。常用的时间序列模型包括自回归(AR)模型、移动平均(MA)模型、自回归移动平均(ARMA)模型、求和自回归移动平均(ARIMA)模型[6]。本文主要使用时间序列模型中的ARIMA模型, 通过Eviews软件, 结合1978~2015年福建省GDP数据建立时间序列模型, 并应用建立的模型对2016~2018年的福建省GDP数据进行预测, 通过对比预测值与真实值的差值计算出相对误差, 从而来考量模型预测的准确性, 选择较准确的ARIMA模型对福建省未来GDP进行预测。

2. 数据收集

为了分析和预测福建省GDP, 我们在福建省统计局官网的统计年鉴收集到了1978~2018年的福建省GDP数据[7], 如表1。

Table 1. GDP of Fujian province from 1978 to 2018

表 1. 1978~2018 年福建省 GDP

年份	生产总值(亿元)	年份	生产总值(亿元)	年份	生产总值(亿元)
1978	66.37	1992	784.68	2006	7583.85
1979	74.11	1993	1114.2	2007	9248.53
1980	87.06	1994	1644.39	2008	10,823.01
1981	105.62	1995	2094.9	2009	12,236.53
1982	117.81	1996	2484.25	2010	14,737.12
1983	127.76	1997	2870.9	2011	17,560.18
1984	157.06	1998	3159.91	2012	19,701.78
1985	200.48	1999	3414.19	2013	21,868.49
1986	222.54	2000	3764.54	2014	24,055.76
1987	279.24	2001	4072.85	2015	25,979.82
1988	383.21	2002	4467.55	2016	28,519.15
1989	458.4	2003	4983.67	2017	32,182.09
1990	522.28	2004	5763.35	2018	35,804.04
1991	619.87	2005	6554.69		

3. ARIMA(4,2,2)模型的建立

3.1. 平稳性检验

用 Excel 绘制出福建省 1978~2015 年的 GDP 时序图, 如图 1。

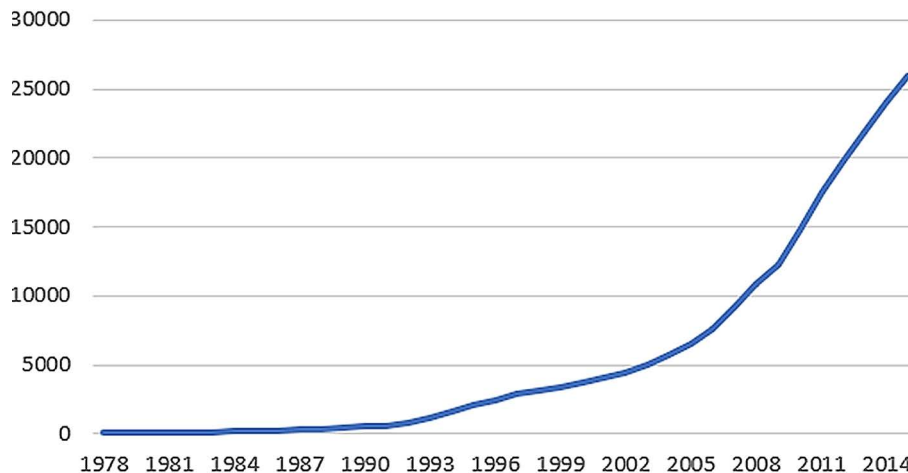


Figure 1. Sequence diagram of Fujian province's GDP from 1978~2015

图 1. 福建省 1978~2015 年 GDP 时序图

由时序图可以看出, 大概从 2002 年开始, 福建省 GDP 增长速度较快, 从总体趋势来看, 该时间序列呈明显的上升趋势, 因此可认为该时间序列为非平稳序列。接下来的模型建立过程中, 我们可以通过单位根检验的方法进一步对序列的平稳性进行检验, 在 Eviews 中将福建省 GDP 的时间序列命名为 {X} 并进行平稳性检验, 得到检验结果如图 2。

由运行结果可知, 显著水平为 10% 的 t 统计量值为 -3.215267, 而时间序列 {X} 的 t 值为 1.419408, 大于显著水平 10% 下的 t 值, 所以不能拒绝原假设, 序列存在单位根, 即该时间序列是非平稳的。为了使非平稳序列变为平稳序列, 我们应对其进行差分, 但由于原始序列呈指数增长趋势, 所以我们首先要对原始序列取对数, 来消除可能存在的异方差问题。将取对数后的序列 {LN X} 进行 ADF 检验, 结果如图 3。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	1.419408	1.0000
Test critical values:		
1% level	-4.284580	
5% level	-3.562882	
10% level	-3.215267	

Figure 2. Unit root test of {X}

图 2. {X} 的单位根检验

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.531493	0.7971
Test critical values:		
1% level	-4.273277	
5% level	-3.557759	
10% level	-3.212361	

Figure 3. Unit root test of {LN X}

图 3. {LN X} 的单位根检验

由运行结果可知, 取对数后的时间序列依然不平稳, 因此我们应再对其进行一阶差分, 得到序列 {D(LNX)}, 其单位根检验结果如图 4。由运行结果可知, 一阶差分后显著水平为 10%时的 t 统计量值为 -3.215267, 而时间序列 {D(LNGX)} 的 t 值为 -2.289593, 大于 -3.215267, 所以不能拒绝原假设, 该时间序列依然是非平稳的, 所以接下来我们要继续对序列进行二阶差分, 得到序列 {D(LNX,2)}, 其单位根检验结果如图 5。

由运行结果可知, 二阶差分后时间序列 {D(LNX,2)} 的 t 统计量值 -7.130786 小于显著水平 1% 的临界值 -4.252879, 所以不存在单位根, 即序列平稳。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.289593	0.4269
Test critical values: 1% level	-4.284580	
5% level	-3.562882	
10% level	-3.215267	

Figure 4. Unit root test of {D(LNX)}
图 4. {D(LNX)} 的单位根检验

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.130786	0.0000
Test critical values: 1% level	-4.252879	
5% level	-3.548490	
10% level	-3.207094	

Figure 5. Unit root test of {D(LNX,2)}
图 5. {D(LNX,2)} 的单位根检验

3.2. 模型识别

二阶差分后的时间序列 {D(LNX,2)} 已经平稳, 接下来我们将通过自相关图和偏自相关图来对模型进行识别, 选取合适的模型并判断 p,q 的取值。使用 Eviews 绘制出序列的自相关图和偏自相关图如图 6。

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.028	0.028	0.0315	0.859
		2	-0.533	-0.534	11.472	0.003
		3	0.151	0.265	12.414	0.006
		4	0.072	-0.384	12.635	0.013
		5	-0.257	0.046	15.560	0.008
		6	0.095	-0.071	15.970	0.014
		7	0.088	-0.123	16.337	0.022
		8	-0.116	0.071	16.999	0.030
		9	0.068	-0.062	17.231	0.045
		10	-0.001	-0.097	17.231	0.069
		11	-0.208	-0.143	19.590	0.051
		12	-0.002	-0.121	19.591	0.075
		13	0.185	0.064	21.616	0.062
		14	-0.013	-0.095	21.627	0.087
		15	-0.094	-0.026	22.199	0.103
		16	0.072	-0.002	22.555	0.126

Figure 6. Autocorrelation diagram and partial correlation diagram of {D(LNX,2)}
图 6. {D(LNX,2)} 的自相关图和偏自相关图

时间序列{D(LNX,2)}的自相关图系数在 $k=2$ 以后进入二倍标准差内, 偏自相关图系数在 $k=4$ 后进入二倍标准差内, 且进入二倍标准差内的自相关系数和偏自相关图系数没有快速趋于 0, 还存在着较大的波动数值, 所以认为该序列的自相关系数和偏自相关系数均为拖尾, 可以尝试建立 $p < 4, q < 2$ 的 ARIMA(p,2,q)模型, 选择符合条件的最优模型进行预测。

3.3. 参数估计

经过尝试发现, 只有 ARIMA(4,2,2)模型通过了检验, 所以这里以 ARIMA(4,2,2)模型的建模过程为例。在 Eviews 的运行结果中, AR(1)和 AR(3)的 Prob(伴随概率)值分别为 0.1287 和 0.6662, 均大于 0.05, 故没有通过 t 检验, 参数不显著。剔除不显著的项再进行估计, 结果发现 MA(1)项的伴随概率为 0.5855 大于 0.05, 还是没有通过 t 检验, 将其剔除再次进行估计, 最终结果如图 7。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002500	0.005357	-0.466704	0.6443
AR(2)	-1.155004	0.131887	-8.757538	0.0000
AR(4)	-0.493897	0.113724	-4.342954	0.0002
MA(2)	0.999939	0.061040	16.38166	0.0000
R-squared	0.597533	Mean dependent var		-0.000129
Adjusted R-squared	0.554411	S.D. dependent var		0.069243
S.E. of regression	0.046221	Akaike info criterion		-3.194289
Sum squared resid	0.059819	Schwarz criterion		-3.011072
Log likelihood	55.10862	Hannan-Quinn criter.		-3.133557
F-statistic	13.85696	Durbin-Watson stat		1.383367
Prob(F-statistic)	0.000010			
Inverted AR Roots	.25+.80i	.25-.80i	-.25+.80i	-.25-.80i

Figure 7. Modeling result of ARIMA(4,2,2) after the AR(1), AR(3), MA(1) are eliminated

图 7. ARIMA(4,2,2)剔除 AR(1)、AR(3)、MA(1)项的建模结果

此时 Prob 值均小于 0.05, 参数显著, 且根均在单位圆内, 所以可以建立该模型, 模型方程为:

$$x_t = -0.0025 - 1.155004x_{t-2} - 0.493897x_{t-4} + \varepsilon_t - 0.999939\varepsilon_{t-2}$$

3.4. 模型检验

接下来对建立的模型进行白噪声检验, 检验模型的适应性。判断方法有两种, 一种是观察自相关图的系数是否在二倍标准差内, 如果是则为白噪声; 另外一种看伴随概率是否大于 0.005, 如果均大于 0.005 则为白噪声。在 Eviews 中进行白噪声检验, 结果如图 8。

ARIMA(4,2,2)模型残差的自相关系数均在二倍标准差内, P 值也均大于 0.005, 所以可以认为序列之间没有任何关联, 即为白噪声序列, 可以用该模型进行预测。

4. ARIMA(4,3,2)模型的建立

4.1. 平稳性检验

之前我们对取对数后的时间序列{LNX}进行了二阶差分并建立了模型, 现在我们来尝试进行三阶差分并建模, 对比二阶差分和三阶差分的效果, 选出预测效果最好的模型。在 Eviews 中对取对数后的时间序列{LNX}进行三阶差分得到序列{D(LNX,3)}, 进行单位根检验, 结果如图 9。

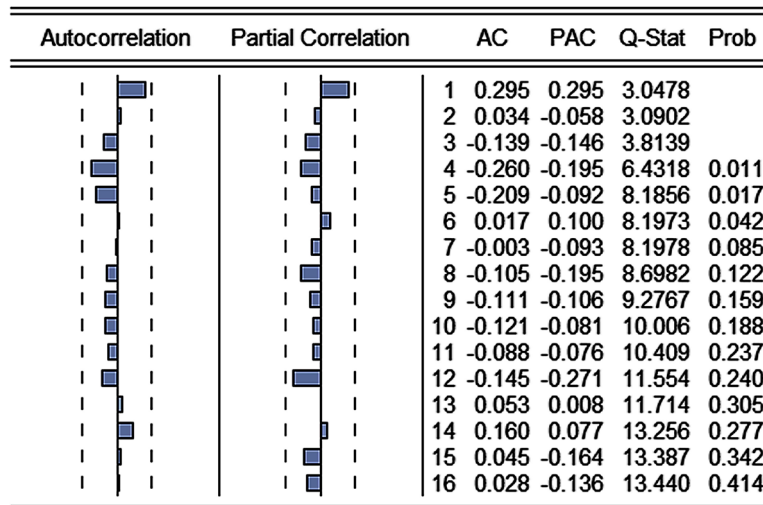


Figure 8. Residual correlation diagram of ARIMA(4,2,2) model
 图 8. ARIMA(4,2,2)模型的残差相关图

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.928516	0.0021
Test critical values:		
1% level	-4.284580	
5% level	-3.562882	
10% level	-3.215267	

Figure 9. Unit root test of {D(LNX,3)}
 图 9. {D(LNX,3)}的单位根检验

由单位根检验结果可知, 三阶差分后序列 {D(LNX,3)} 的 t 统计量值为 -4.928516 小于显著水平为 1% 的临界值 -4.284580, 所以不存在单位根, 该序列平稳。

4.2. 模型识别

在 Eviews 中做三阶差分的自相关图和偏自相关图如图 10。时间序列 {D(LNX,3)} 的自相关图系数在 k = 5 以后进入二倍标准差内, 偏自相关图系数在 k = 4 后进入二倍标准差内, 且进入二倍标准差内的自相关系数和偏自相关图系数没有快速趋于 0, 还存在着较大的波动数值, 所以认为该序列的自相关系数和偏自相关系数均为拖尾, 可以尝试建立 p < 4, q < 5 的几个 ARIMA(p,3,q)模型, 选出通过检验的最优模型。

4.3. 参数估计

经过尝试, 只有 ARIMA(4,3,2)模型通过了检验, 所以这里以建立 ARIMA(4,3,2)模型为例进行操作。在 Eviews 的几次运行结果中, AR(1)、AR(3)和 MA(1)的伴随概率均大于 0.05, 没有通过 t 检验, 参数不显著, 剔除不显著项重新估计, 最终得到结果如图 11。

此时伴随概率均小于 0.05, 参数显著且根均在单位圆内, 所以该模型可建模, 模型方程为:

$$x_t = 0.004544 - 1.272193x_{t-2} - 0.605696x_{t-4} + \varepsilon_t - 0.999971\varepsilon_{t-2}$$

4.4. 模型检验

利用 Eviews 对 ARIMA(4,3,2)模型进行白噪声检验, 结果如图 12。

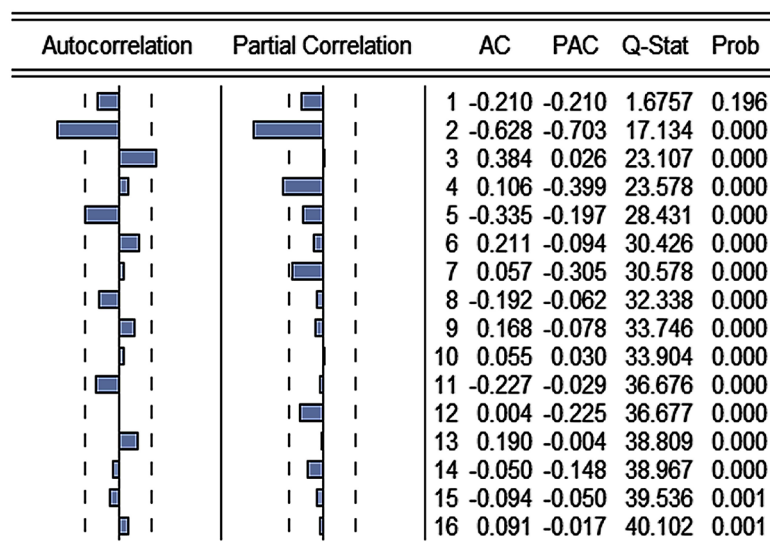


Figure 10. Autocorrelation diagram and partial correlation diagram of $\{D(LNX,3)\}$

图 10. $\{D(LNX,3)\}$ 的自相关图和偏自相关图

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004544	0.006755	0.672727	0.5068
AR(2)	-1.272193	0.093094	-13.66574	0.0000
AR(4)	-0.605696	0.091160	-6.644336	0.0000
MA(2)	0.999971	0.050818	19.67734	0.0000
R-squared	0.685989	Mean dependent var		-0.004638
Adjusted R-squared	0.651099	S.D. dependent var		0.094313
S.E. of regression	0.055709	Akaike info criterion		-2.817440
Sum squared resid	0.083794	Schwarz criterion		-2.632410
Log likelihood	47.67032	Hannan-Quinn criter.		-2.757125
F-statistic	19.66142	Durbin-Watson stat		2.443716
Prob(F-statistic)	0.000001			
Inverted AR Roots	.27+.84i	.27-.84i	-.27+.84i	-.27-.84i

Figure 11. Modeling result of ARIMA(4,3,2) after the AR(1), AR(3), MA(1) are eliminated

图 11. ARIMA(4,3,2)剔除 AR(1)、AR(3)、MA(1)项的建模结果

ARIMA(4,3,2)模型残差的自相关系数均在二倍标准差内, P 值也均大于 0.005, 所以可以认为该模型序列之间没有任何关联, 即为白噪声序列, 所以可以用于预测。

5. 预测结果检验

利用 Eviews 对福建省 2016~2018 年的 GDP 进行预测, 得到预测值, 但此时预测后的结果是取对数后的预测值, 我们需要将其还原为原序列值的形式, 还原后得到的数据如表 2。

通过对比可以发现, ARIMA(4,3,2)模型的平均相对误差更小, 预测效果更好, 所以应选择 ARIMA(4,3,2)模型进行预测, 得到福建省 GDP 预测值如表 3。

预测出 2019~2023 年的 GDP 值分别为 39,077.03 亿元、44,903.9 亿元、52,959.35 亿元、63,306.92 亿元、78,094.12 亿元。

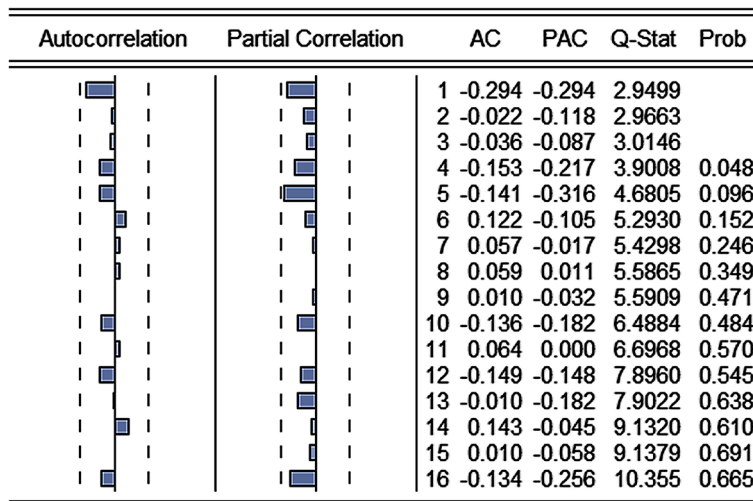


Figure 12. Residual correlation diagram of ARIMA(4,3,2) model
 图 12. ARIMA(4,3,2)模型的残差相关图

Table 2. Comparison of model results
 表 2. 模型结果比较

年份	真实值	ARIMA(4,2,2)模型	差值	相对误差
2016	28,519.15	29,013.31	494.16	1.7327%
2017	32,182.09	33,052.75	870.66	2.7054%
2018	35,804.04	36,147.03	342.99	0.958%
年份	真实值	ARIMA(4,3,2)模型	差值	相对误差
2016	28,519.15	28,696.2	117.05	0.4104%
2017	32,182.09	32,257.63	75.54	0.2347%
2018	35,804.04	35,457.49	346.55	0.9679%

Table 3. GDP forecast of Fujian province
 表 3. 福建省 GDP 预测

年份	2019	2020	2021	2022	2023
预测值	39,077.03	44,903.9	52,959.35	63,306.92	78,094.12

6. 总结与展望

本文在对福建省 GDP 进行预测研究过程中, 首先阐述了研究现状和研究方法, 并介绍了研究思路 and 运用的几种模型。然后在福建省统计局官网上收集到了福建省的 GDP 数据, 运用 Eviews 软件, 基于时间序列分析原理对数据进行了平稳性检验、模型识别、参数估计、白噪声检验等一系列处理, 最终建立了 ARIMA(4,2,2)及 ARIMA(4,3,2)两个模型, 分别用它们进行预测, 并对比预测结果, 计算出预测误差, 最终发现 ARIMA(4,3,2)模型的预测效果相对于 ARIMA(4,2,2)较好, 所以确定 ARIMA(4,3,2)是最优模型, 可以对福建省未来 GDP 进行有效预测。

在经济全球化的今天, 某个国家或地区的经济情况很容易受到自然灾害或国际形势等突发情况的影响, 面对这些情况需要具体问题具体分析, 预测的准确程度难免会有所降低。此外, ARIMA 模型也只在短期预测中较为精确, 并不适用于长期预测。即便如此, 我们仍然可以使用 ARIMA 模型进行粗略预测,

判断出未来短期内的经济走势, 对已存在的问题作出适当调整, 对将来有可能发生的问题做好准备, 这将会有利于政府经济计划的改良并促进经济发展。

参考文献

- [1] 宋喆磊. 改革开放以来福建省经济增长因素的实证分析[J]. 企业导报, 2009(6): 112-113.
- [2] 李守丽. 时间序列模型在地级市 GDP 预测中的应用[D]: [硕士学位论文]. 郑州: 郑州大学, 2013.
- [3] 戴琳琳. 基于 ARIMA 模型的青岛市 GDP 预测分析[J]. 河北能源职业技术学院学报, 2019(3): 60-62.
- [4] 张旭昌. 基于时间序列分析的河北省 GDP 预测模型研究[J]. 改革与开放, 2018(21): 16-18.
- [5] 王鄂, 张霆. 时间序列在湖南省 GDP 预测中的应用[J]. 青岛大学学报(自然科学版), 2019, 32(3): 136-140.
- [6] 孙铁轩, 邵春福. 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): 348-359.
- [7] 福建统计局官网. 1978-2018 年我国生产总值[EB/OL]. <http://tjj.fujian.gov.cn/>, 2020-02-01.