

文章编号: 2095-2163(2020)05-0240-05

中图分类号: C81

文献标志码: A

基于 ARIMA 模型职工平均工资的分析与预测

王宁宁

(上海工程技术大学 管理学院, 上海 2016020)

摘要:以 1978—2017 年上海市职工平均工资统计数据为研究对象,采用基于 ARIMA 模型的时间序列方法,对平均工资进行预测,借助 SAS 软件对数据进行拟合分析,建立了 ARIMA(1,0,1) 模型,并在此基础上对 2018—2027 年上海市职工平均工资进行预测。通过对职工工资数据序列检验,得出最终模型。从实用层面评估了预测模型的建模方法和预测数据的可信度。结果显示,ARIMA(1,0,1) 模型能对上海市职工平均工资起到较好的预测效果。

关键词: ARIMA 模型; 职工平均工资; 预测

The analysis and prediction of the average salary of Shanghai employees based on ARIMA model

WANG Ningning

(School of Management, Shanghai University of Engineering technology, Shanghai 2016020, China)

[Abstract] This paper takes the statistical data of the average salary of employees in Shanghai from 1978 to 2017 as the research object, adopts the time series method based on ARIMA model to predict the average salary, and establishes the ARIMA(1,0,1) model with the help of SAS software to fit and analyze the data. On this basis, it predicts the average salary of employees in Shanghai from 2018 to 2027. The final model is obtained by examining the salary data sequence of employees. The modeling method of the prediction model and the reliability of the prediction data are evaluated from the practical level. The results show that ARIMA(1,0,1) model has a good prediction effect on the average salary of employees in Shanghai.

[Key words] ARIMA model; Average wages of employees; predict

0 引言

上海作为世界著名的金融中心之一,有着庞大的人口规模和发达的经济水平^[1]。所以研究上海的经济水平具有一定的代表意义。劳动工资是收入的重要组成部分,是衡量收入、分配与劳动力发展水平的重要指标。职工工资水平可以由职工平均工资体现,而和职工平均工资相关的指标与职工的利益息息相关,这项指标对未来人口的影响和引导就业非常重要。除此之外,员工的平均工资水平对社会保障缴费的基数、对公积金的缴费基数及退休员工的基本养老金都存在一定影响。因为职工的平均工资水平关系到每个员工的利益,所以对职工给的平均工资进行预测具有深刻的意义。比如,职工工资的最低标准等都要依据参考社会平均工资水平进行制定^[2]。目前,很多专家学者运用多种方法对工资进行预测,如线性回归法、指数平滑法、Logistic 模型等。将职工的平均工资按年份排序,就构成了一个时间序列,可以用时间序列分析的方法对职工平均工资的时间序列进行建模和预测^[3]。

时间序列预测方法主要是根据等时间内随机变

量的变化,推测潜在的趋势,并对未来的变化做出合理的估计^[4]。因为时间单位存在差异,序列中关于时间的变量可以是各种形式,并不局限于日月年等^[5]。在对涉及时间序列问题的分析中,一般通过建立有关时间序列的模型对数据随时间的变化趋势进行量化^[6]。本文尝试运用时间序列数据分析中常用的 ARIMA 模型对上海市职工平均工资进行分析,同时对未来职工工资水平做出一定的预测。

1 实证研究

(1) ARIMA 模型。ARIMA 模型是通过自身的变动情况,结合差分方程,对未来的变化趋势进行线性预测的一种精度较高、拟合性较强的自回归求和滑动平均模型^[7]。

(2) 数据来源及预处理。通过查阅统计年鉴,可以得到上海市 1978 至 2017 年度职工平均工资统计表,利用 SAS 处理得到上海职工平均工资时序,如图 1 所示。

从时序图可以看出,上海市职工平均工资在过去的 38 年中,总体呈指数上升趋势,尤其是在后面几年增速突飞猛进。由此可以判断,这个时间序列

具有非平稳性,是一个非平稳时间序列。

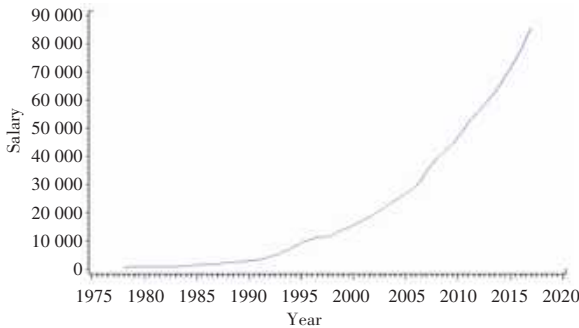


图 1 上海历年职工平均工资时序图

Fig. 1 Timing chart of average salary of employees over the years in Shanghai

(3)模型的初步建立与检验。由图 1 可以看出,时间序列是非平稳的。而非平稳的时间序列不能直接用来建立时间序列模型。所以,需要对上海市历年职工平均工资序列进行平稳化处理,对原始数据进行一定的变换是必要的。通过对这个序列进行取对数,把它的指数趋势转变为线性的趋势,再对这个序列进行差分处理^[8]。通过对原始数据进行

取对数,就可以获得如图 2 所示的时序图。

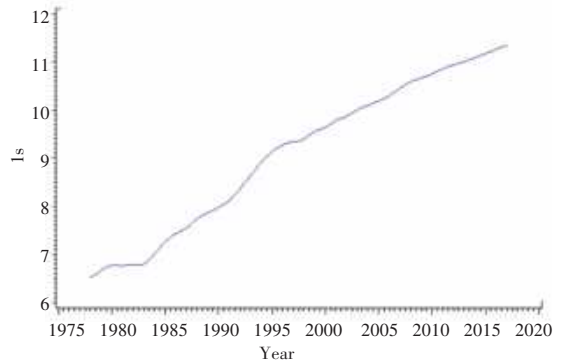


图 2 上海平均工资取对数后的时序图

Fig. 2 Timing chart of average wages in Shanghai after taking logarithm

由图 2 可见,取对数后的序列还是非平稳的,但可通过该序列的原始数据趋势图大致了解这个序列的平稳性。为了进一步了解上海平均工资取对数后序列的性质,下面给出 1s 序列的自相关系数。自相关图表现为自相关函数并未快速下降到零,也没有在零附近收敛。

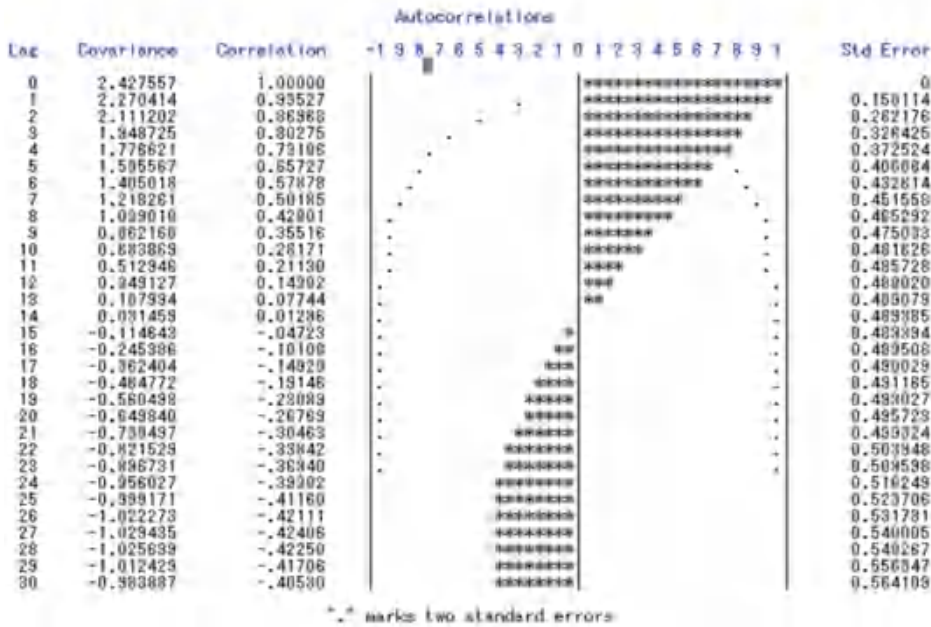


图 3 上海平均工资取对数后的自相关系数

Fig. 3 Autocorrelation coefficient of the logarithm of the average wage in Shanghai

由图 3 可知,由于自相关函数的衰减速度明显变得缓慢,从而可以推断 1s 序列具有非平稳的性质,这个结论与 1s 序列图上观察到的结果相同,对其取完对数后的时间序列依旧具有线性趋势,但也不平稳。因此,可以利用一阶差分运算对存在的这种线性趋势进行消除,使该序列平稳。所以要对 1s

变量差分操作后的序列展开平稳性识别检验。进行一阶差分后,检验过程如下:

(1)观察样本自相关系数。如果一个平稳序列的自相关函数呈现截尾性质,而呈现偏自相关函数拖尾性,就可用 MA 模型拟合该序列。从图 4 可以看出:对其完成一阶差分后的样本自相关系数在 1

步以后具有截尾的性质,由此可以初步判定为 MA (1) 模型,然后再对其展开参数估计,同时对参数展开 T 检验。之后,对进一步得到的结果展开白噪声检验。因为纯随机序列指的是对以前值没有记忆,

对将来值也没有影响的序列,所以这种序列不存在分析价值。因此,为了确定一个序列是否具有继续研究的价值和必要,在决定对其建立模型前要进行白噪声检验,并进行参数估计^[9]。



图4 进行一阶差分后的样本自相关系数

Fig. 4 Sample autocorrelation coefficient after first difference

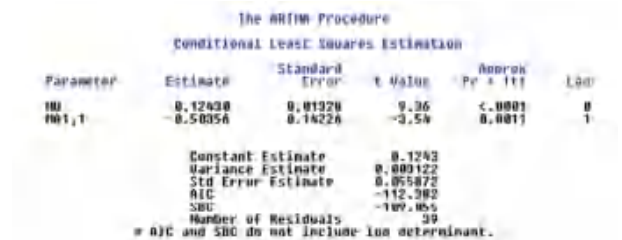


图5 MA(1)模型参数估计和检验结果

Fig. 5 MA(1) model parameter estimation and test results



图6 MA(1)模型白噪声检验结果

Fig. 6 White noise test results of MA(1) model

从图5、图6可以看出:在5%的显著性水平下,MA(1)的模型通过了白噪声检验,说明拟合比较充分,参数也通过了T检验,模型是比较理想的。

(2)观察样本偏向关系数。如果AR(p)模型是平稳的,那么它的偏自相关函数就有p步截尾的性质。从图7可以看出:经过一阶差分之后样本的偏相关系数在1步后具有截尾的性质,所以可以将它初始认定为AR(1)模型。然后再对其进行参数估计,同时对参数展开T检验,对检验所得结果再一次进行白噪声检验和参数估计。

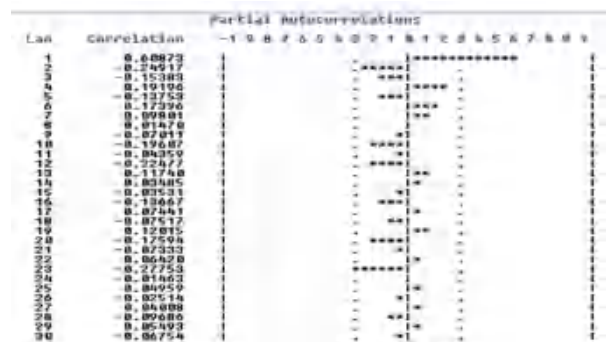


图7 进行一阶差分后的样本偏相关系数

Fig. 7 Partial correlation coefficient of samples after first difference

从图8、图9可以看出:在显著性水平为5%的情况下,AR(1)模型也通过了白噪声检验,说明拟合的效果是比较充分的;参数均通过了T检验,此模型是比较理想的。

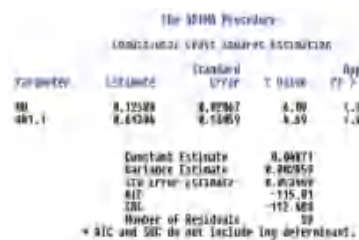


图8 AR(1)模型参数估计和检验结果

Fig. 8 AR(1) model parameter estimation and test results

Lag	Chi-Square	df	Pr > ChiSq	Autocorrelation									
4	6.29	5	0.2782	0.152	-0.045	-0.215	-0.007	0.138	0.049				
12	16.82	11	0.1483	0.067	0.219	0.125	-0.219	0.122	-0.298				
18	23.13	13	0.1451	-0.198	0.018	-0.298	-0.224	0.098	-0.172				
24	26.79	23	0.2871	0.088	-0.091	-0.278	0.019	0.088	-0.074				
30	27.92	29	0.5273	-0.091	-0.053	0.018	-0.014	-0.014	-0.011				

图9 AR(1)模型白噪声检验结果

Fig. 9 White noise test results of AR (1) model

2 模型的优选和数据预测

从上述分析中可以看到：经过对原始数据取对数，变换之后数据相对比较理想的模型是 MA(1)模型和 AR(1)模型。接着，再依据 AIC 信息准则对比较理想的这两个模型进行优选。根据 AIC 准则，可以确定模型拟合度是否优良。尽管根据 AIC，数据拟合度越优良越好，但还是要避免拟合度过犹不及的情况。因此选择模型时，应优先考虑 AIC 值最小的。

从图 5、图 8 可以看到：MA(1)模型的 AIC 值为 -112.382，AR(1)模型的 AIC 值为 -115.81，即 AR(1)模型的 AIC 值较小。因此把 AR(1)模型选为拟合原始数据经过变换后数据的最终模型，进一步选 ARIMA(1,0,1)作为拟合原始数据最终模型。

结合前述 AR(1)模型的分析过程，AR(1)模型的分析结果如下：

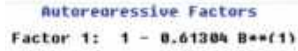


图10 AR(1)模型的结果

Fig. 10 Results of AR (1) model

3 结束语

根据此模型，利用 ARIMA(1,0,1)模型对上海市 2018—2027 年职工平均工资进行预测，同时再对预测值进行一定的改变，目的是为了获得在原始度量下的预测值。预测结果如图 11 所示。

Year	Y _t	Forecast	Conf. Interval	Lower 95%	Upper 95%	Actual	Residual
1	11976	11976	11976	11976	11976	11976	0
2	12179	12179	12179	12179	12179	12179	0
3	12382	12382	12382	12382	12382	12382	0
4	12585	12585	12585	12585	12585	12585	0
5	12788	12788	12788	12788	12788	12788	0
6	12991	12991	12991	12991	12991	12991	0
7	13194	13194	13194	13194	13194	13194	0
8	13397	13397	13397	13397	13397	13397	0
9	13600	13600	13600	13600	13600	13600	0
10	13803	13803	13803	13803	13803	13803	0
11	14006	14006	14006	14006	14006	14006	0
12	14209	14209	14209	14209	14209	14209	0
13	14412	14412	14412	14412	14412	14412	0
14	14615	14615	14615	14615	14615	14615	0
15	14818	14818	14818	14818	14818	14818	0
16	15021	15021	15021	15021	15021	15021	0
17	15224	15224	15224	15224	15224	15224	0
18	15427	15427	15427	15427	15427	15427	0
19	15630	15630	15630	15630	15630	15630	0
20	15833	15833	15833	15833	15833	15833	0
21	16036	16036	16036	16036	16036	16036	0
22	16239	16239	16239	16239	16239	16239	0
23	16442	16442	16442	16442	16442	16442	0
24	16645	16645	16645	16645	16645	16645	0
25	16848	16848	16848	16848	16848	16848	0
26	17051	17051	17051	17051	17051	17051	0
27	17254	17254	17254	17254	17254	17254	0
28	17457	17457	17457	17457	17457	17457	0
29	17660	17660	17660	17660	17660	17660	0
30	17863	17863	17863	17863	17863	17863	0
31	18066	18066	18066	18066	18066	18066	0
32	18269	18269	18269	18269	18269	18269	0
33	18472	18472	18472	18472	18472	18472	0
34	18675	18675	18675	18675	18675	18675	0
35	18878	18878	18878	18878	18878	18878	0
36	19081	19081	19081	19081	19081	19081	0
37	19284	19284	19284	19284	19284	19284	0
38	19487	19487	19487	19487	19487	19487	0
39	19690	19690	19690	19690	19690	19690	0
40	19893	19893	19893	19893	19893	19893	0
41	20096	20096	20096	20096	20096	20096	0
42	20299	20299	20299	20299	20299	20299	0
43	20502	20502	20502	20502	20502	20502	0
44	20705	20705	20705	20705	20705	20705	0
45	20908	20908	20908	20908	20908	20908	0
46	21111	21111	21111	21111	21111	21111	0
47	21314	21314	21314	21314	21314	21314	0
48	21517	21517	21517	21517	21517	21517	0
49	21720	21720	21720	21720	21720	21720	0
50	21923	21923	21923	21923	21923	21923	0

图 11 预测结果

Fig. 11 Forecast

从图 11 中可知，2018—2027 年上海市职工平均工资预测值增长符合社会发展规律，说明建立的 ARIMA(1,0,1)是科学合理的。除此之外，从表中可以看出上海市的经济状况良好，上海市职工的生活水平将会越来越高，还将继续以较快的速度发展。本文通过这个模型对上海市职工工资进行预测，同时预测结果表明精度较好。因此，可以根据这门技术为有关部门制定或者相应调整社保缴费基数、公积金缴费基数、以及与退休人员利益相关的政

策的制定提供一定的科学依据和参考。同时有助于推动上海市职工与雇主之间的关系良好健康发展，对维持社会的健康稳定具有一定意义。

当使用 ARIMA 模型对有时间相关性的数据进行预测时，要依据所面临问题的特点，对其量体裁衣并选择最适合的模型。同时，需要注意这种模型属于精度较高的时间序列短期预测模型，对预测短期的数据比较精确，如果延长预测期，预测误差也会随之相应的增大。（下转第 247 页）