

基于 ARIMA 模型的 IT 行业薪酬变化分析

刘量

摘要 在 21 世纪的今天,随着互联网行业的高速发展,对于互联网行业发展趋势的研究也变得愈发重要。而在此类型的研究中,对 IT 行业的薪酬变化的研究既可以反映出互联网行业市场的当前发展状态,又可以体现出互联网行业的发展模式和发展的前景。为有效研究 IT 行业的薪酬变化,IT 行业的劳动者报酬和从业人员数量是极具代表性的数据。采用 SPSS 对数据进行拟合,虽然能够反映出数据的增减性,但对数据的特征挖掘不足,且无法有效预测其变化趋势。基于此,使用自回归积分滑动平均模型进行研究能够有效解决以上问题。最终得到结论,在未来十年间,IT 行业的薪酬还会进一步的稳定提升,市场发展前景良好,市场目前的发展状态良好。

关键词: IT 行业;薪酬变化;ARIMA 模型;R 语言

1 引言

互联网是在 1969 年由美国最早提出的,最初的互联网源于美国国防部高级研究计划署,是专门应用于军事连接的。1983 年,美国国防部将 APA 网络划分为军事网络和民用网络,同时局域网和广域网的产生和蓬勃发展进一步促进了互联网的发展。经过了 20 多年互联网的蓬勃发展,20 世纪 90 年代,互联网全面进入高速发展的时代^[1]。直到 1994 年,中国最早的互联网也应运而生。时至今日,中国的互联网发展已经相当成熟,据中国互联网络信息中心(CNNIC)发布的第 46 次《中国互联网络发展状况统计报告》显示,截至 2020 年 6 月,中国网民规模达 9.40 亿,较今年 3 月增长 3625 万,互联网普及率达 67.0%。

借助飞速发展的互联网催生出的新的经济增长点,出现了许多的互联网类型企业,主要涉及门户网站、电子商务、网络游戏、即时通讯等多个领域。据国研网发布的数据显示,互联网企业在 2018 年的软件业务收入已经达到 619087338 万元,利润总额达到 89615917 万元,其变化趋势如图 1 所示。

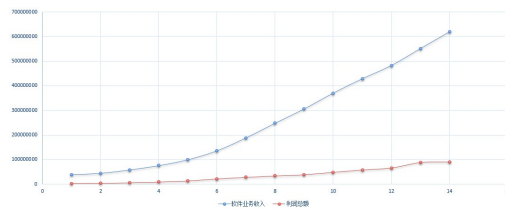


图 1 数据变化趋势图

尤其是近几年来,智能手机的出现,使得互联网不再局限于依托 PC 端工具,可以通过智能手机实现各种互联网服务,出现爆发式的发展。互联网的出现颠覆了人们传统的生活方式,使用互联网学习娱乐、社交、购买理财产品、电子商务等等已经司空见惯,对于互联网行业发展趋势的研究变得尤为重要^[1]。而在此类型的研究中,对 IT 行业薪酬变化的研究不仅能够反映出互联网行业市场的当前发展状态,更可以体现出互联网行业的发展模式和发展的前景。

2 分析方法

为了研究 IT 行业薪酬的变化,我们选取了国研网的劳动者报酬和从业人员数量这两个指标,通过这两个关键指标来评估 IT 行业的市场发展现状和 market 发展前景,以此研究 IT 行业薪酬变化的过程和趋势。

我们首先计算这两个关键指标的增长率,得到的结果如表 1 所示。

表 1 增长率结果

年份	劳动者报酬增长率/%	从业人员数量增长率/%
2006	-	46.104
2007	4.346	18.569
2008	35.8	1.05
2009	41.566	37.984
2010	46.942	27.8
2011	36.718	26.232
2012	31.444	21.655
2013	25.52	12.389
2014	22.249	16.064
2015	3.242	5.22
2016	16.436	2.011
2017	21.476	5.417
2018	21.818	4.367

从中我们不难发现,虽然增长率都是正数,表示劳动者薪酬和从业人员数量都是连年增长,但是其增长率却存在着变化特征。

2.1 SPSS 拟合分析

首先,我们通过 SPSS,分别对劳动者报酬和从业人员数据这两项关键数据进行拟合分析。

劳动者报酬的 SPSS 拟合结果如图 2 所示。

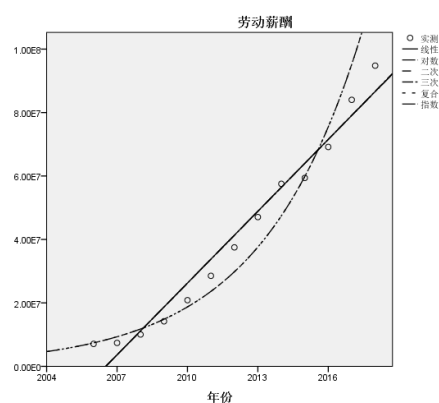


图2 劳动者报酬 SPSS 拟合结果

从业人员数量的 SPSS 拟合结果如图 3 所示。

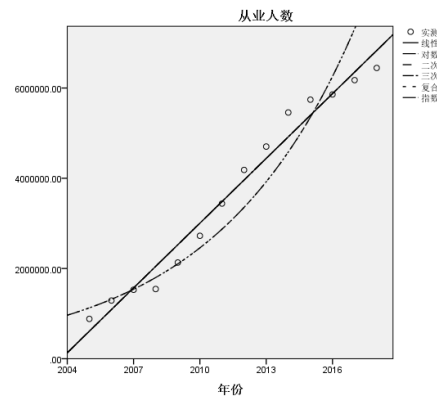


图3 从业人员数量 SPSS 拟合结果

通过 SPSS 的拟合,我们不难发现,无论是劳动者报酬还是从业人员数量,从 2005 年至 2018 年均稳步增长。但受限于 SPSS 的拟合方法有限,其拟合函数类型有限,SPSS 拟合只能有效反应出数据的增长性,但其他数据特征无法进行有效发掘,对未来数据变化也无法进行有效的预测。基于此,我们提出使用 ARIMA 时间序列模型,对劳动者报酬和从业人员数量进行进一步的分析,并预测其未来的发展趋势。

2.2 ARIMA 模型预测分析

对于较少的时间段的时间预测,因为数据量较少,所以直接使用神经网络是不现实的。因此对于无症状感染者的分布预测模型采用 ARIMA (Autoregressive Integrated Moving Average Model) 模型实现^[2]。ARIMA 模型的算法流程图如图 4 所示。

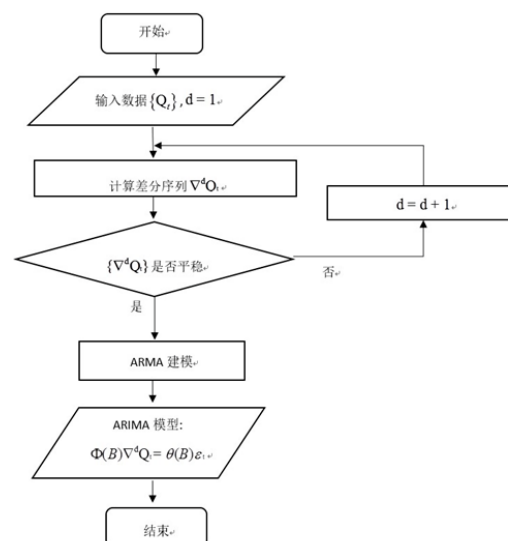


图4 ARIMA 模型算法流程图

ARIMA 模型是由 Box 和 Jenkins 提出的用于时间序列分析和预测的一种线性模型。在时间序列分析中，需要设置 3 个参数，分别为自回归阶数 p ，差分阶数 d ，移动平均阶数 q 。

$ARIMA(p,d,q)$ 的一般形式如下^[3]：

$$\begin{cases} \phi(B)\nabla^d y_t = \theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0 \quad var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_s) = 0 \quad s < t \end{cases}$$
$$\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

其中， B 为后移算子， ∇^d 为高阶差分， φ_p 和 θ_q 为自回归参数和移动平均参数。

通过 R 语言，首先我们对劳动者报酬数据绘制图像，得到的结果如图 5 所示。

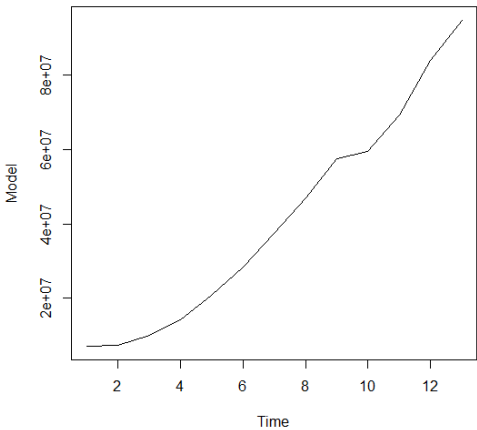


图5 劳动者报酬时间序列图

在此基础上，我们对劳动者报酬进行差分运算，并得到结果。我们发现，进行 2 阶差分运算，数据趋于平稳，如图 6 所示。

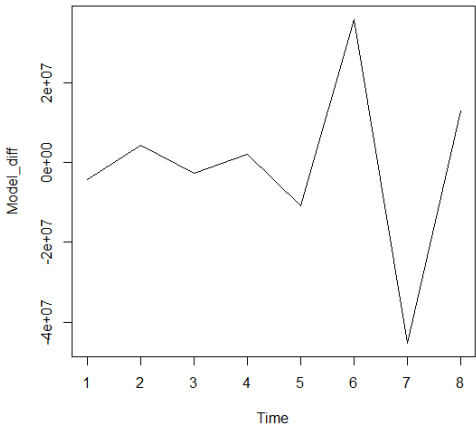


图6 劳动者报酬差分图

在此基础上，我们使用 R 语言绘制自相关图和偏自相关图，并分析得到的自相关图和偏自相关图的数据特征。依据这些特征，来决定我们的 ARIMA 模型的三个参数的取值，以此来建立模型。我们得到的结果如图 7 和图 8 所示。

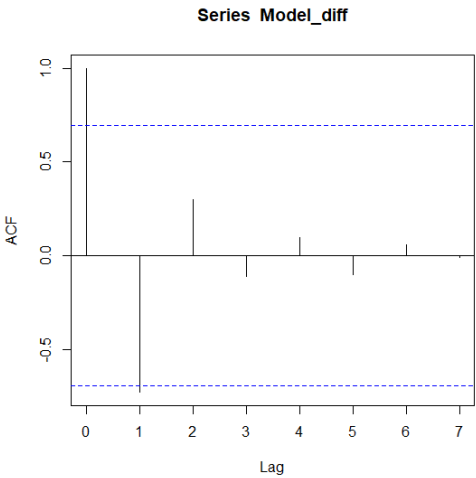


图7 自相关图

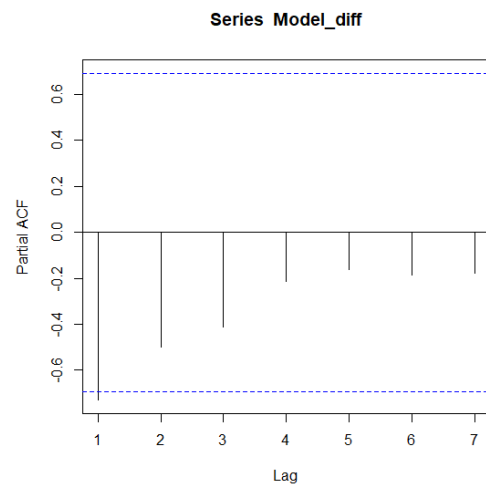


图8 偏自相关图

通过观察自相关图和偏自相关图,我们发现,自相关图在1阶之后拖尾,偏自相关图在1阶之后拖尾。所以我们应该构造 ARIMA(1, 2, 1) 模型。

使用 R 语言用构建完成的模型对未来进行预测,如图9所示。

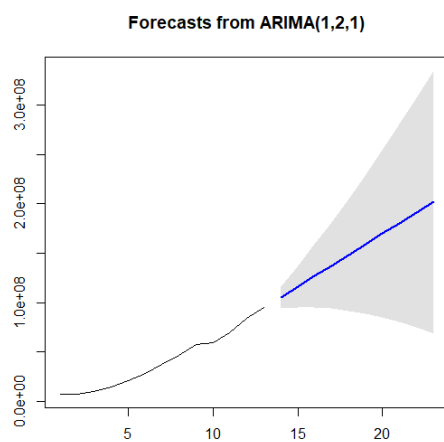


图9 劳动者报酬预测

基于同样的方法,我们对从业人员数量同样构建 ARIMA 模型进行预测,预测的结果图如图10所示。

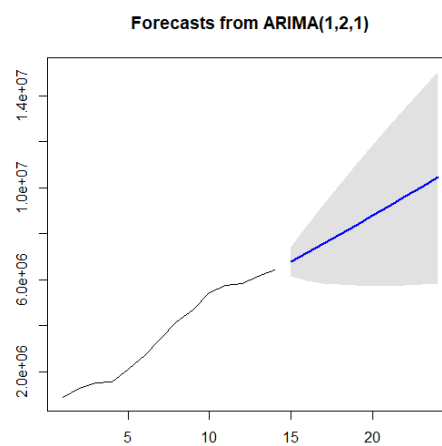


图10 从业人员数量预测

我们统计预测劳动者报酬和从业人员数量得到的数据,结果如表2所示。

表2 预测结果

年份	劳动者报酬	从业人员数量
2019	105511941	6788253
2020	116236619	7167822
2021	126959774	7565641
2022	137682615	7972566
2023	148405392	8384034
2024	159128156	8797770
2025	169850917	9212636
2026	180573678	9628068
2027	191296439	10043781
2028	202019199	10459634

3 分析结果

通过对劳动者报酬和从业人员数量构造 ARIMA 模型,并预测数据在未来十年的变化,我们发现,在未来十年间,无论是劳动者报酬还是从业人员数量,均会保持进一步的增长。通过对数据的分析,我们预测在未来十年间,IT 行业的薪酬还会进一步的稳定提升,市场发展前景良好,市场目前的发展状态良好。

4 结论

我们通过对劳动者报酬数据和从业人员数量数据的分析,预测在未来 IT 行业薪酬还会进一

步的稳步提升，预期在 2027 年，IT 行业从业人
员数量将突破千万。

参考文献

- [1]张璐. 我国现代互联网企业核心竞争力提升研究[D]. 西北农林科技大学, 2020.
- [2]金旗, 裴昌幸, 朱畅华. ARIMA 模型法分析网络流量[J]. 西安电子科技大学学报, 2003(01):6-10.
- [3]王英伟, 马树才. 基于 ARIMA 和 LSTM 混合模型的时间序列预测[J]. 计算机应用与软件, 2021, 38(02):291-298.