

基于时间序列方法的湖北省居民 消费价格指数预测分析

杜文晟¹

(湖北师范大学 经济管理与法学院, 湖北 黄石 435002)

【摘要】: 以湖北省的居民消费价格指数数据为研究对象, 使用 ARIMA 模型对数据进行拟合, 并据此作出预测。

【关键词】: 湖北省 消费价格指数 时间序列 ARIMA

【中图分类号】 F222 **【文献标识码】** A **【文章编号】** 2096-3130(2019)06-0045-04

一、引言

居民消费价格指数(CPI)是根据居民用于与生活资料消费密切相关的的一篮子商品和劳务价格制定的指标体系, 用于反映消费品的物价变动。CPI 作为衡量通货膨胀或者紧缩程度的重要指标的同时, 也是国家宏观调整宏观经济政策的重要依据。尤其是最近十几年, CPI 成为了大众新闻媒体报道里常见的话题, 甚至有不少网民将其看做衡量通胀指标的唯一因素。

从宏观经理理论上来说, CPI 和国内生产总值平减指数(GDP Deflator)都是作为衡量通胀指标的重要依据, 然而 CPI 倾向于夸大通胀率, 而 GDP 平减指数则相反。所以不能简单用 CPI 作为通胀率指标。但是研究国家或者一个区域的 CPI 指数的变动, 可以从一个侧面反映国家或者地区的民生情况。CPI 的统计数据是一个典型的时间序列, 一般可以按照月份、季度和年度来统计数据, 按照一篮子固定商品价格计算当期或者累计的价格指数。

二、时间序列模型概述

时间序列模型可分为确定性模型和随机性模型, 一般通过观察时序图和相关函数图的方法来判断区分。主流的随机性时间序列分析模型由 Box 和 Jackins 创立, 包括自回归模型(Auto Regression, AR)、移动平均模型(Moving Average, MA)、自回归移动平均模型(ARMA)和自回归求和移动平均模型(ARIMA)。

AR 模型基于一个非常简单但也非常普遍的假设: 研究对象的历史状态对当前状态是有影响的, 有影响历史状态可能是以前某一个时刻或者几个时刻的, 这些时刻可能和本时刻相连也可能相差几个时间间隔。假设第 t 时刻状态为 Y_t , 那么 p 阶 AR 模型(或者简称为 AR(p) 模型)可以表示为 $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$, 其中序列 $\{\varepsilon_t\}$ 是独立同分布的纯随机项, 而 μ 则是 t 时刻的序列均值, 可能是固定也可能是不固定的。

MA 模型是基于纯随机项生成的, 也就是 AR 模型中的 $\{\varepsilon_t\}$ 序列。通常假设该随机项是一个白噪声序列, 即该序列服从均值为

¹**作者简介:** 杜文晟, 男, 湖北武汉人, 高级实验师, 研究兴趣为经济统计与数据分析年生, 湖北武汉人, 高级实验师, 研究兴趣为经济统计与数据分析。

0, 方差恒定的同一个分布函数, 并且协方差矩阵是一个对角阵, 对角线以外元素为 0。更强的假设是该序列服从正态分布(高斯分布), 这就是高斯白噪声假设。MA(q) 模型的形式为 $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$ 。ARMA 模型则是综合了上述两个模型的内容, ARMA(p, q) 模型形式为 $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$ 。

使用以上模型来建模的时候, 都追求平稳模型, 如果模型不平稳, 则使用差分方法进行处理, 构成自回归求和移动平均模型 ARIMA(p, d, q), 其中 d 为差分次数, 大多数情况下 d=1 或者最多等于 2。当然序列也可能是有确定性趋势, 这时候就需要用趋势外推方程了, 类似于回归分析。

三、模型构建和预测

我们选择了湖北省 1994 年 1 月至 2019 年 2 月一共 302 个数据点(数据来源: 凤凰财经网), CPI 数据为当月值(上个月为 100)。数据的一部分(1994.01—2014.12 共 252 个数据点)被用于建模, 其余部分被用于查看模型预测效果。下面的工作全部在 R 语言统计软件的环境下完成。

首先绘制时序图和相关函数图, 以便观察时间序列的基本特征。如图 1, 图 2。

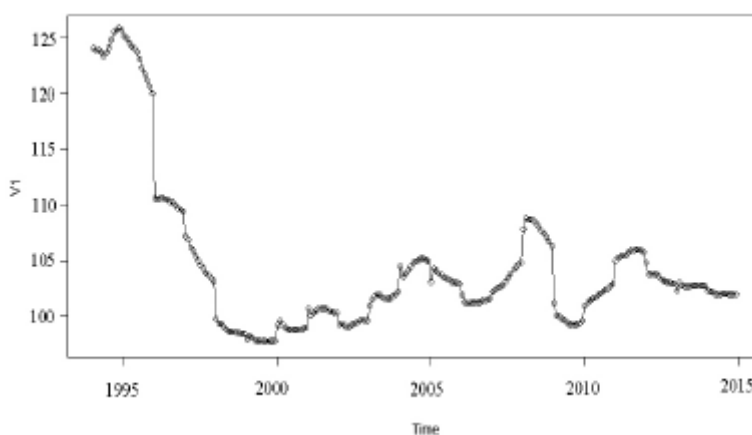


图 1 湖北省居民消费价格指数时序图(1994 年 1 月-2014 年 12 月)

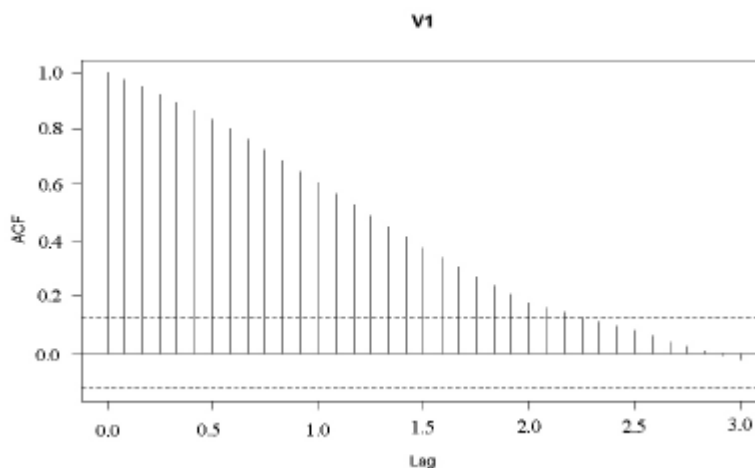


图 2 数据自相关函数图

从时序图来看,1994 年的 CPI 指数明显比后面高出很多,实际上数据显示当年湖北省 CPI 达到 124 以上(相当于 CPI 增长率 24%)。在 90 年代初期,中国经济一度过度扩张,广义货币 M2 发行量年年激增。1993 年中国经济冲上新一轮高峰,基建投资迅猛增加,社会需求强力扩张,货币投放增长同比达到 34%的惊人水平。时任国务院副总理朱镕基采取一系列措施控制经济过热,宁可经济增长慢一点,也要降低高通胀率。此后 CPI 一路下降,到了本世纪初已经在 100 附近徘徊,此后 CPI 曾在 2008 年再次达到顶峰,次年一月迅速下跌,这是受到了 2008 年国际金融危机影响。所有这些情况在湖北省的 CPI 数据上都有所反映。

按照建模的需要,过高和过低的数据应当被当做异常点除去。但是考虑到中国经济的延续性,我们仍然保留了这一部分数据。当然无论是否去除,从时序图上都看不出来明显的长期趋势,故初步判断序列为随机性时间序列(无确定性趋势)。所以我们考虑使用 ARIMA 模型建模。继续绘制自相关函数(ACF)图和偏相关函数(PACF)图。从自相关函数图可以清晰看出,相关系数随着滞后增加而下降,但是下降速度非常慢,知道 20 阶还未降到标准线以下,且呈现线性递减趋势。这种情况表明时间序列存在单位根,进一步的 ADF 检验也证实了这一点,结果如下:

```
Value of test - statistic is: -1.576
Critical values for test statistics:
      1pct   5pct   10pct
taul -2.58  -1.95  -1.62
```

ADF 检验 Q 统计量计算为-1.576,因为 ADF 是左侧检验,所以结果落在了 10%置信区间内,即应当接受原假设(序列存在单位根)。而对序列的一次一阶差分的时序图和 ACF 如图 3 和图 4 所示。

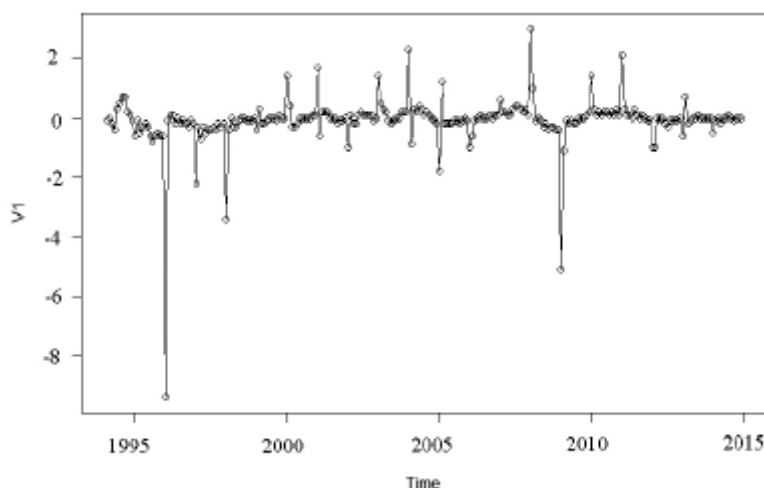


图 3 一次差分时序图

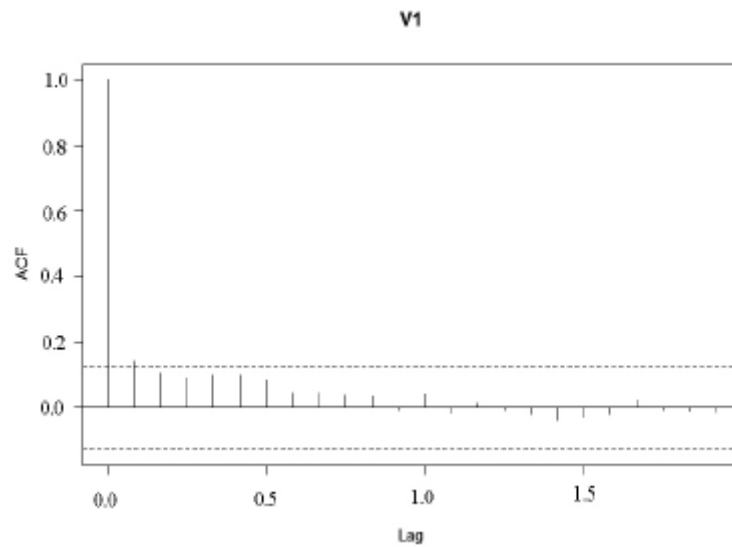


图 4 一次差分后的自相关函数图

从图形上看, 序列已经平稳, 再次做 ADF 检测结果如下:

```
Value of test - statistic is: -9.3054
Critical values for test statistics:
      1pet   5pet   10pet
taul  -2.58  -1.95  -1.62
```

很明显, 已经可以拒绝原假设了, 序列中不再存在单位根。

查看差分后的相关函数如图 4 和图 5 所示。

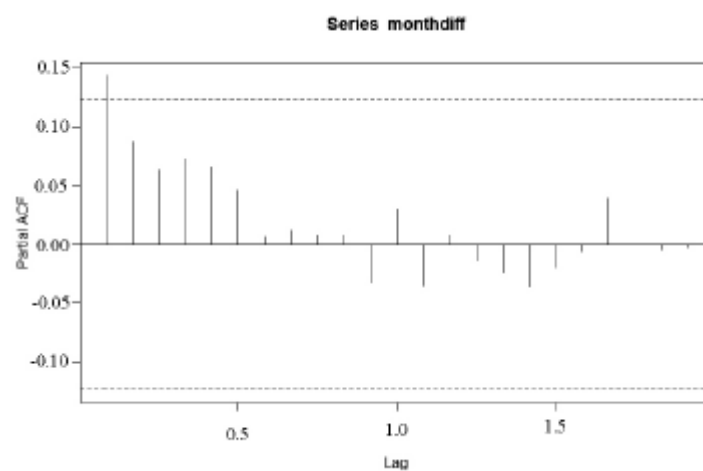


图 5 差分序列的偏自相关函数

从两个图中可以发现很明显的拖尾特性, 并且季节性不明显。根据一般的阶数识别标准, 我们有如下规则:

	自相关函数图	偏相关函数图	模型类型
截尾/拖尾	q 阶截尾	拖尾	MA (q)
截尾/拖尾	拖尾	P 阶截尾	AR (p)
截尾/拖尾	拖尾	拖尾	ARMA

判断模型为非季节性 ARMA (p, q), 需要确定 p 和 q 的值。我们使用主流的 AIC 原则来判断, 阶数从低到高依次拟合 ARMA 模型, 当 p=1 且 q=1 时, AIC 达到最小 (629.722), 从而确定差分序列为 ARMA (1, 1) 模型。

使用条件最小二乘法估计模型参数, 结果为:

Coefficient (s) :

arl	mal	intercept
0.84675	-0.73628	-0.01356

因此差分序列模型为 $Y_t = -0.01356 + 0.84675Y_{t-1} + \varepsilon_t + 0.73628 \varepsilon_{t-1}$, 如果还原为原始 (未差分) 序列, 那么模型为:

$$Y_t = -0.01356 + 1.84675Y_{t-1} - 0.84675Y_{t-2} + \varepsilon_t + 0.73628 \varepsilon_{t-1}$$

所以从形式上看, ARIMA (p, d, q) 模型可以转换为一个 (非平稳) ARMA (p+d, q) 模型。当然如果用这个模型来估计参数, 那么会和 ARIMA (p, d, q) 的不一样。

LBQ 检验模型残差结果如下:

X-squared=0.0029943, df=1, p-value=0.9564

实际上 LBQ 检验的 P 值随着滞后数增加递增, 所以我们可以拒绝原假设, 认为残差序列为纯随机序列, 因此模型通过白噪声检验。

使用上述模型对未来 24 个月的预测结果和真实值如下:

	预测均值	Lo80	Hi80	Lo95	Hi95	真实值
Jan-15	101.9926	100.92103	103.0642	100.35376	103.6315	101
Feb-15	101.9863	100.38382	103.5887	99.53554	104.437	101.3
Mar-15	101.9808	99.92057	104.041	98.82997	105.1316	101.5
Apr-15	101.976	99.49409	104.458	98.18022	105.7719	101.6
May-15	101.9719	99.09187	104.852	97.56725	106.3766	101.6
Jun-15	101.9684	98.70833	105.2285	96.98255	106.9543	101.6

Jul-15	101.9654	98.34046	105.5903	96.42155	107.5092	101.6
Aug-15	101.9627	97.98642	105.9391	95.88149	108.044	101.7
Sep-15	101.9605	97.64496	106.276	95.36046	108.5605	101.6
Oct-15	101.9585	97.31512	106.6019	94.85705	109.06	101.6
Nov-15	101.9568	96.99611	106.9176	94.37007	109.5436	101.5
Dec-15	101.9554	96.68728	107.2235	93.89852	110.0122	101.5
Jan-16	101.9541	96.38802	107.5202	93.4415	110.4667	101.3
Feb-16	101.953	96.09778	107.8083	92.99819	110.9079	101.8
Mar-16	101.9521	95.81606	108.0881	92.56784	111.3364	102
Apr-16	101.9513	95.54239	108.3602	92.14973	111.7528	102.1
May-16	101.9506	95.27633	108.6248	91.74319	112.158	102.2
Jun-16	101.95	95.01747	108.8825	91.34761	112.5524	102.1
Jul-16	101.9495	94.76541	109.1335	90.96239	112.9365	102.2
Aug-16	101.949	94.51979	109.3783	90.58699	113.3111	102.1
Sep-16	101.9486	94.28026	109.617	90.22087	113.6764	102.1
Oct-16	101.9483	94.04651	109.8501	89.86356	114.033	102.2
Nov-16	101.948	93.81823	110.0778	89.51459	114.3814	102.2
Dec-16	101.9478	93.59514	110.3004	89.17353	114.722	102.2

可以看到, 预测误差始终在 5%以内, 但是方差范围(Low 数据和 Hi 数据之间)随着时间推移不断扩大, 这也是 ARIMA 模型的特点。实际上, ARIMA 模型的预测方差是时间的线性函数。

四、结论与讨论

从上文利用 ARIMA 模型对 2015 年 1 月—2016 年 12 月的 CPI 预测结果来看, 绝对误差在 0.5 以内, 相对误差小于 3.5%, 可以说还是比较好的。建模过程中发现判定系数为 0.7, 应该说不是特别高, 究其原因应当是 1994—1997 我国宏观经济政策调整(即著名的“经济软着陆”)使得这几年的 CPI 在这 302 个数据点中偏离较大, 几乎成为异常点群。同时我们也发现, 就湖北省而言, 其当月 CPI 的时间序列数据并未表现出明显的季节性(我们另外做的数据分析表明, CPI 累计值的季节性也来自于统计指标本身), 这说明湖北省的居民消费情况比较平稳, 近 20 年并无大起大落或者趋势跟风之类情况。不过从模型中来看, 其走势是微弱上扬的(存在单位根), 这可能是因为近年来人力成本、原材料成本和能源成本都在上升, 进而导致各类产品和服务价格逐步上涨造成的。

宏观经济调控的目标应该是使得 CPI 保持一个温和的上升幅度, 既能够刺激经济发展, 又不至于因物价过快上涨导致民生问题, 影响老百姓的幸福指数。至少目前来看, 湖北省的 CPI 上升幅度是温和的, 有利于经济健康发展的。在供给侧结构性改革的大背景下, 政府的宏观调控可以集中在引导企业产业结构转型, 鼓励发展低能耗、高附加值产业和高科技产业, 同时实行稳健的货币政策, 控制好货币增量。企业也应当积极消除原材料和燃料动力采购当中的浪费情况, 优化管理, 节约成本, 从而使得产品更加具有市场竞争力。

参考文献:

[1]Box G, Jenkins G. Time Series Analysis: Forecasting and Control. John Wiley & Sons. Inc, 2008.

[2]Cryer J, Chan K. Time Series Analysis with Applications in R. Springer Science Business Media LLC, 2008.

[3]董大勇,刘珂言.基于 ARIMA 模型的工业品出厂价格指数预测与分析[J].统计与决策,2016,445(1):179-181.

注释:

1R 语言统计软件和 Eviews 或者 stata 不同,其 AIC 值相当于 Eviews 计算结果乘以样本容量,故而非常大,本文中减去 2 并除以 252 后 AIC 值和 Eviews 相同,为 2.49。