Advances in Applied Mathematics 应用数学进展, 2020, 9(8), 1206-1212
Published Online August 2020 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2020.98141

Hans汉斯

Short-Term Consumer Index Prediction Based on ARIMA-Holt Index Smoothing Model

Han Zhao, Deshan Sun

School of Mathematics, Liaoning Normal University, Dalian Liaoning
Email: 1094008976@qq.com

Received: Jul. 25th, 2020; accepted: Aug. 10th, 2020; published: Aug. 17th, 2020

Abstract

Taking the monthly historical data of Shanghai Consumer Price Index (CPI) for the 10-year period from 2008 to 2017 as a sample, the time series test method was used to analyze the correlation and establish the ARIMA model. At the same time, a variety of forecasting methods are used to forecast the level of Shanghai Residents' consumption index in the first quarter of 2018. The results show that the Shanghai consumer index has a clear trend, and the Holt index smooth forecasting method has better forecasting ability, and the effect is ideal, which provides a certain reference for short-term forecasting.

Keywords

CPI, ARIMA Model, Holt Exponential Smoothing, Short-Term Forecast

基于ARIMA-Holt指数平滑模型的短期居民消费指数预测

赵　晗，孙德山

辽宁师范大学数学学院，辽宁　大连
Email: 1094008976@qq.com

收稿日期：2020年7月25日；录用日期：2020年8月10日；发布日期：2020年8月17日

摘　要

以上海市2008~2017年10年间居民消费指数(CPI)的月度历史数据为样本，采取时间序列检验方法对其

文章引用: 赵晗, 孙德山. 基于 ARIMA-Holt 指数平滑模型的短期居民消费指数预测[J]. 应用数学进展, 2020, 9(8): 1206-1212. DOI: 10.12677/aam.2020.98141

进行了相关分析，建立了ARIMA模型。同时利用多种不同预测方法对2018年第一季度上海市居民消费指数水平进行预测，结果表明：上海市居民消费指数具有明显的趋势性，且Holt指数平滑预测方法具有更优的预测能力，效果较为理想，为短期预测提供一定的借鉴。

关键词

CPI，ARIMA模型，Holt指数平滑，短期预测

1. 引言

居民消费指数(CPI)是用来反映居民家庭购买消费商品及服务的价格水平的变动情况，CPI 能在特定时段内度量一组代表性消费商品及服务项目的价格水平随时间而变动的情况，是对中国经济增速最稳定的拉动。所以分析居民消费指数水平的波动情况对于总体经济发展具有很重要的意义。

一般而言，可以根据已知数据建立多元线性回归模型预测 CPI，但该种方法过于传统往往预测效果不太理想。时间序列分析 ARIMA 模型是由伯克斯和詹金斯(Box-Jenkins)于上个世纪 70 年代系统提出的不同于回归模型的根据观测的时间序列数据，通过曲线拟合和参数估计建立数学模型的一种分析方法，所以现在大多采用时间序列分析模型进行预测。李菊梅建立了 ARMA 模型对我国 1984～2005 年的年度进出口数据进行分析，发现其短期预测效果比长期预测效果好[1]；张立杰等人基于自回归移动平均及支持向量机方法对中国棉花价格预测，得到了很好的效果[2]；敬久旺利用 ARIMA 乘积季节模型，对我国海关进出口商品对总值进行时间序列分析，发现该种方法预测能力更优，能充分反映我国海关进出口商品总值的时间序列变化规律[3]。本文在宏观经济理论的基础上，以上海市 2008~2017 年月度 CPI 变化的时间序列为研究切入点，借助 R 语言统计软件，建立相应的 CPI 时间序列 ARIMA 模型，并用多种方法对 2018 年第一季度上海市 CPI 进行有效预测分析，得到了较好的结果。

2. ARIMA 模型和 Holt 指数平滑

2.1. ARIMA 模型

一般来讲，随机时间序列模型包括移动平均模型(MA)、自回归模型(AR)和自回归移动平均模型(ARMA)，对于非平稳时间序列通常采取自回归综合移动平均模型(ARIMA)。

在一个 p 阶自回归模型中，序列的每一个值都可以用它之前 p 个值的线性组合来表示：

$$\mathrm{AR}(p): Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t \tag{1}$$

其中 Y_t 是时序中的任一观测值，μ 是序列均值，β 是权重，ε_t 是随机扰动。在一个 q 阶移动平均模型中，序列的每一个观测值都可以用它之前的 q 个残差的线性组合来表示：

$$\mathrm{MA}(q): Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} - \varepsilon_t \tag{2}$$

其中 ε_t 是预测残差，θ 是权重。这两种方法的混合即 $\mathrm{ARMA}(p,q)$ 模型，其表达式如下：

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3}$$

此时，序列中的每个观测值用过去的 p 个观测值和 q 个残差的线性组合来表示。

在一般的自回归移动平均模型中，序列仅有趋势性，假设 x_t 表示随机序列，并假设 $Lx_t = x_{t-1}$，其中 L 是滞后算子。如果存在非负整数 d 满足：

$$\phi(L)\nabla^d x_t = \theta(L)\varepsilon_t \tag{4}$$

式中函数表示为：

$$\begin{cases} \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \\ \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \\ \nabla^d = (1-L)^d \end{cases} \tag{5}$$

且 $|L| \le 1$，$\phi(L)$ 与 $\theta(L)$ 互素，$\phi_p \theta_q \ne 0$，$\{\varepsilon_t\}$ 是白噪声序列，存在 $E(\varepsilon_t) = 0$，$E(\varepsilon_t)^2 < \infty$。ARIMA$(p,d,q)$ 模型意味着序列被差分了 d 次，且序列中的每个观测值都是用过去的 p 个观测值和 q 个残差的线性组合来表示[4]。

2.2. Holt 指数平滑

Holt 指数平滑可以对有水平项和趋势项的时序进行拟合，时刻 t 的观测值可以表示为：

$$Y_t = \text{level} + \text{slope} \times t + \text{irregualr} \tag{6}$$

其中平滑参数 α 控制水平项的指数型下降，β 控制斜率的指数型下降，参数取值越大意味着越近的观测值的权重越大。

3. 实证分析

3.1. 数据的选取与说明

选取 2008~2017 年 10 年间上海市居民消费价格指数为实验样本，数据取自上海市统计年鉴，每个月度的 CPI 值如表 1 所示。基于上述理论，对时间序列数据进行建模和分析，依此采用多种方法对 2018 上半年 CPI 进行监控预测并与实际值进行比较分析，从而评价模型的预测能力。

Table 1. Consumer Price Index (CPI) in Shanghai from 2008 to 2017
表 1. 上海市 2008~2017 居民消费价格指数

年份月份	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
1	105.9	101.7	101.1	104.3	104.9	102.1	102.0	101.8	102.5	103.7
2	105.9	100.8	101.2	104.5	104.4	102.3	102.8	102.2	102.7	102.6
3	106.8	100.4	101.5	104.6	104.2	102.3	102.7	102.3	103.0	102.2
4	107.1	99.9	101.8	104.7	104.0	102.2	102.6	102.4	103.1	102.0
5	107.1	99.7	102.0	104.6	103.8	102.2	102.7	102.4	103.0	102.0
6	107.1	99.5	102.2	105.0	103.6	102.2	102.7	102.4	103.1	101.9
7	107.1	99.3	102.5	105.1	103.4	102.2	102.7	102.4	103.1	101.8
8	106.9	99.3	102.6	105.2	103.2	102.2	102.7	102.5	103.1	101.8
9	106.7	99.3	102.7	105.2	103.1	102.2	102.7	102.4	103.2	101.8
10	106.4	99.4	102.8	105.3	103.0	102.3	102.7	102.4	103.2	101.7
11	106.1	99.5	103.0	105.2	102.9	102.3	102.7	102.4	103.2	101.7
12	105.8	99.6	103.1	105.2	102.8	102.3	102.7	102.4	103.2	101.7

3.2. 数据处理与模型构建

3.2.1. 数据的处理与分析

本文使用的统计软件为 R 语言，首先验证序列的平稳性。画出该序列的折线图，从上表 1 和序列的折线图(图 1)可以发现，序列受时间趋势影响明显，属于非平稳序列，因此有必要对原对数后的时间序列作差分处理，发现 2 阶差分效果最好，处理后的时序如图 2 所示。

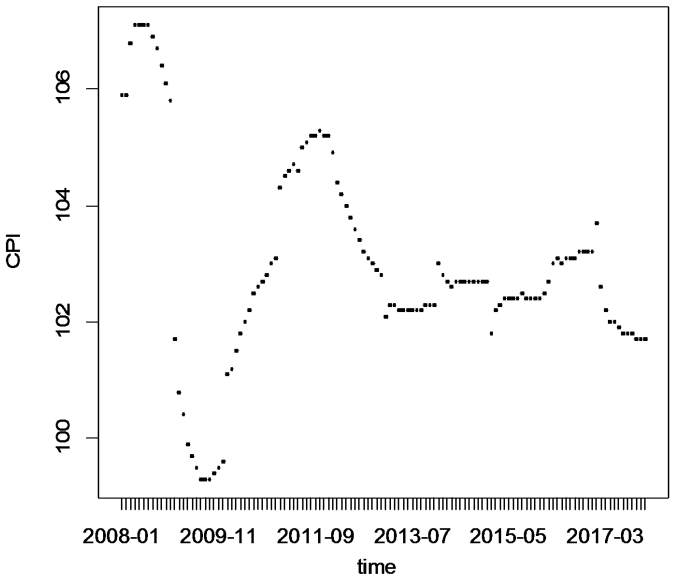

Figure 1. The original sequence
图 1. 原始序列

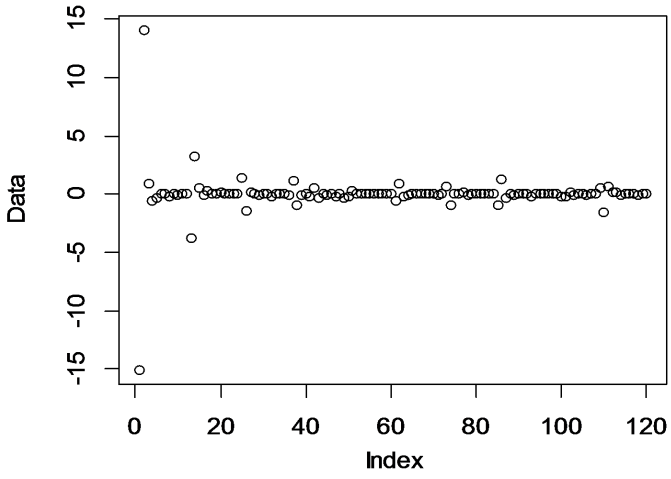

Figure 2. Post-difference sequence
图 2. 差分后序列

3.2.2. 模型确立

对差分后的时序进行 ADF 检验显示此时序列平稳，接下来要进行模型的选取与检验，一般来说，建立 ARIMA 模型的步骤包括：

(1) 确保时序的平稳性；

(2) 找到一个(或几个)合理的模型(即选定可能的p值和q值)；

(3) 拟合模型；

(4) 从统计假设和预测准确性等角度评估模型；

(5) 预测[5]。

我们需要为模型选定参数p，d和q，差分次数为2，d值已经确定，下面通过ACF和PACF图来选择备选模型的p和q [6]，从CPI时序的ACF与PACF结果呈现的特点中可发现：一方面，从ACF图中可以发现序列自1阶开始逐渐增大，自相关数值迅速趋于0，这说明该序列表现出上升态势；另一方面，从PACF图中，偏自相关函数呈明显的下降趋势，逐渐减小到0，虽然11阶偏自相关系数值超出边界，很可能属于偶然出现的，尽管如此，1阶到5阶相关函数显著不为零，也说明该时间序列变动具有趋势性，检验结果如图3所示。

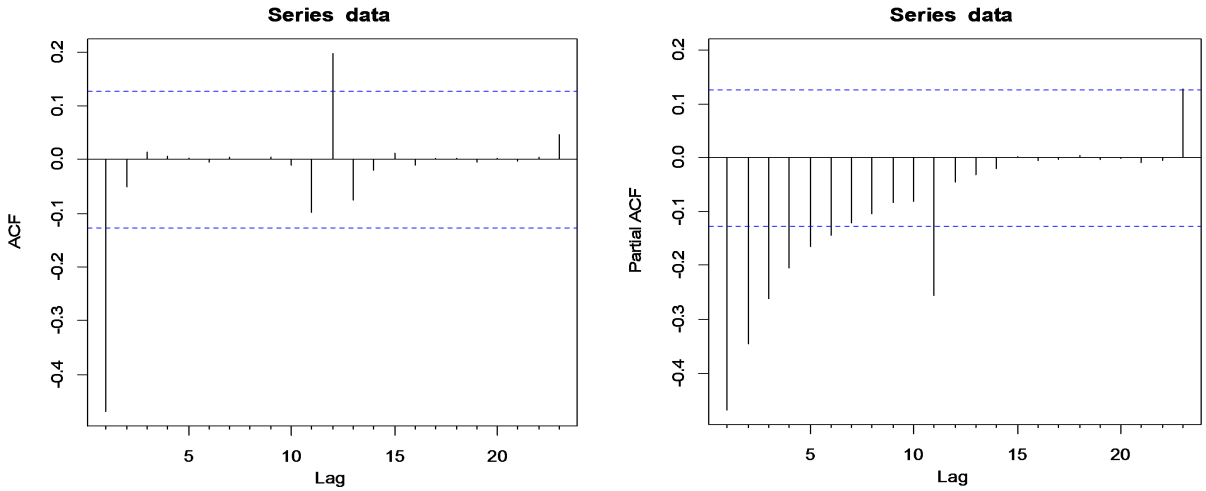

Figure 3. Autocorrelation and partial autocorrelation of CPI sequence after quadratic difference
图3. 二次差分后的CPI序列自相关和偏自相关图

由上述大概可以确定模型，观察自相关函数图与偏自相关函数图可得出相应结论：

(1) 自相关系数呈现1阶截尾，偏自相关系数呈现拖尾状态。自相关函数图中1阶自相关系数显著不为0，1阶后显著为0，可取移动平均阶数$q=1$。偏自相关函数中偏自相关系数前6阶显著不为0，呈线性增长趋于0，可取自回归阶数$p=0$。

(2) 时序不具有季节性。在上述序列的平稳性分析中，对时序进行季节性检验，发现该序列不具有季节性，同时非季节性差分结果$d=2$，序列具有趋势性。

根据上述结论进行模型确认和参数估计得到最终预测模型为ARIMA (0, 2, 1)，参数估计$\delta^2 \approx 1.083$，故拟合后的ARIMA模型为：

$$Y_t = 0.9189Y_{t-1} + 0.0284Y_{t-2} + \varepsilon_t - 1.083\varepsilon_{t-1} \tag{7}$$

3.2.3. 模型拟合与评估

综上已经确立了ARIMA模型，下面我们进行模型的拟合与评估。模型拟合效果以对百分比误差的绝对值做平均，即MAPE的值为指标，本案例MAPE的值为0.22%，证明拟合效果比较理想。

一般来说，一个模型如果合适，则模型的残差应该满足均值为0的正态分布，换句话说，模型的残差应该满足独立正态分布，本文利用Ljung-Box test评价模型，Box.test函数可以检验残差的自相关系数

是否都为 0，在本实验中，模型残差的 p 值为 0.9339 远大于显著性水平，接受原假设，即可以认为残差的自相关系数为 0，是白噪声序列。但残差并不符合正态分布，前后存在较大的波动，中间比较聚集。拟合结果(图 4)与 QQ 图检验(图 5)如下：

```
                      ME      RMSE       MAE           MPE      MAPE    MASE
Training set -0.008614664 0.4798109 0.2285154 -0.006790934 0.2227495 1.0791
                     ACF1
Training set 0.007478567
```

Figure 4. The fitting results
图 4. 拟合结果

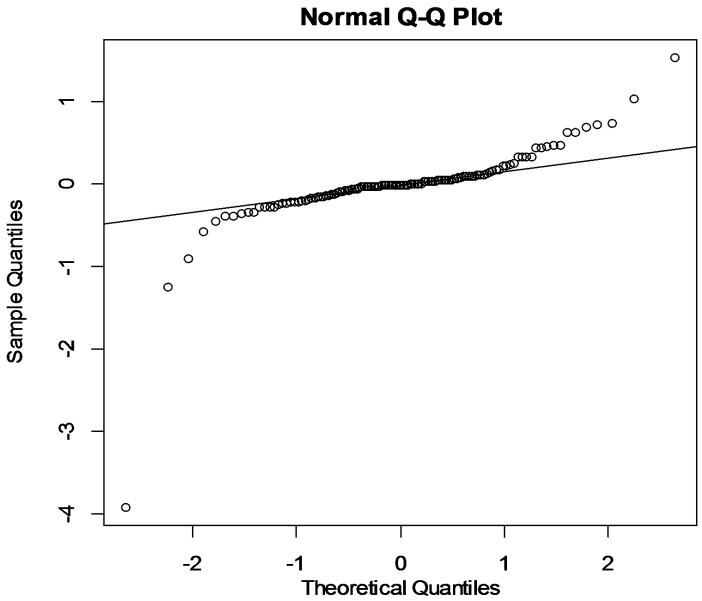

Figure 5. Residual normal Q-Q graph
图 5. 残差正态 Q-Q 图

4. CPI 短期预测与分析

以上建立了基于 ARIMA 的时间序列模型，为了验证模型的有效性，对 2018 年第一季度上海市居民消费水平指数进行预测，采取自动预测，ARIMA 模型直接预测和 Holt 指数平滑预测三种方法，并与实际值进行对比发现 Holt 指数平滑预测效果最好，平均相对误差为 $e = 0.1801$，但随着预测时间的增长，预测精度有所下降，也说明了该模型更适用于短期预测。

预测结果如表 2、表 3 所示。

Table 2. Comparison of prediction accuracy results
表 2. 预测精度结果对比

预测方法	MAPE (平均相对误差)
自动预测	0.2103
ARIMA 模型预测	0.2227
Holt 指数平滑	0.1801

Table 3. Model prediction versus real value
表 3. 模型预测与真实值对比

预测值	真实值	预测精度
100.9	101.1	0.9981
101.6	101.8	0.9980
101.9	101.8	0.9990

5. 总结

本文以上海市 2008~2017 年间的居民消费指数的月数据作为样本，进行差分选阶，建立了 ARIMA 模型并利用 Holt 指数平滑进行短期预测，结果表明该模型对于短期预测具有很好的性能，效果较为理想。该模型易于理解，在实际生活中有广泛的应用，能为短期预测提供很好的借鉴。

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] 王谦, 管河山. 中国进出口总额时间序列 SARIMA 模型的实证[J]. 经济论坛, 2018(12): 78-83.

[2] 常月, 冯宇旭, 曹显兵. 基于非线性时间序列模型的股票分析与预测[J]. 数学的实践与认识, 2018, 48(22): 21-26.

[3] 李欣阳, 李素娟, 刘晓迪, 樊安彤, 闫萍, 刘洪庆. 自回归移动平均乘积季节模型在甲型肝炎发病数中的应用[J]. 山东大学学报(医学版), 2018, 56(12): 103-108.

[4] 孙皖宁, 杨静, 杨依依, 刘桐同, 白晓东. 基于 SARIMA 模型对中国 GDP 分析及预测[J]. 中国集体经济, 2018(36): 78-80.

[5] 张春露, 白艳萍. ARIMA 时间序列模型和 BP 神经网络组合预测在铁路客座率中的应用[J]. 数学的实践与认识, 2018, 48(21): 105-113.

[6] 杨进, 陈亮. 基于小波神经网络与 ARIMA 组合模型在股票预测中的应用[J]. 经济数学, 2018, 35(2): 62-67.