

[文章编号] 1003-4684(2020)01-0106-04

中国消费者价格指数应用研究

——基于调和分整自回归移动平均模型

王 锦, 商 豪

(湖北工业大学理学院, 湖北 武汉 430068)

[摘 要] 针对中国居民消费者价格指数时间序列数据呈现出的记忆特征,采用描述性统计分析和经典时间序列分析相结合的方法,建立两种不同的长记忆模型,分析比较两种模型的谱密度函数拟合图和三期预测值误差,结果显示 ARFIMA 模型的平均预测绝对误差为 37.34%,而 ARTFIMA 模型的平均预测绝对误差为 28.57%,因此 ARTFIMA 模型的拟合预测效果更好。由此得出中国居民消费者价格指数序列更加符合半长期记忆性的结论。

[关键词] 时间序列; 长记忆性; ARTFIMA 模型; ARFIMA 模型

[中图分类号] O212 **[文献标识码]** A

近年来用于刻画时间序列长记忆性特征的模型层出不穷,其中,Granger^[1]和 Hosking^[2]提出的 ARFIMA 模型成为描绘时间序列长期记忆性常用的工具。ARFIMA 模型克服了传统时序模型只能描绘短期记忆性的缺陷,通过引入分数维差分参数,使其具备拟合时间序列过程长记忆性的能力。ARFIMA 模型得到广泛应用,但有学者指出该模型在拟合长期记忆性较弱的时间序列过程中是失效的^[3-4]。基于此问题,Meerschaert^[5]等人发现调和分整过程展现出的半长期记忆性(semi-long range dependence)能够很好地刻画长期记忆性较弱的时间序列过程。这种半长期记忆性具体表现为序列过程的自相关函数在前期缓慢下降,呈现双曲线形式,后期则以指数率迅速衰减。由此 Meerschaert^[6]在研究大气湍流速度数据的过程中首次在 ARFIMA 模型的基础上加入调和参数,提出了调和分整自回归移动平均模型(ARTFIMA),用以拟合具有半长期记忆性的平稳时间序列过程。

本文通过分析中国居民消费者价格指数的记忆特征,针对过程中存在的长期记忆性特征建立 ARTFIMA 模型,并与 ARFIMA 模型进行对比,最后通过谱密度拟合函数和预测效果验证 ARTFIMA 模型在中国居民消费者价格指数的拟合和预测方面比 ARFIMA 模型更具优越性。

1 模型说明

1.1 分整自回归移动平均模型

分整自回归移动平均模型 ARFIMA(p, d, q)可以定义为整数阶差分 ARIMA(p, D, q)模型的推广形式,其差分参数拓展到分数维^[7]。ARFIMA(p, d, q)表达形式如下:

$$\begin{cases} \Phi(B)(1-B)^d(x_t - \mu_t) = \Theta(B)\epsilon_t \\ \mu_t = E(x_t) \\ \epsilon_t \sim NID(0, \sigma_\epsilon^2) \end{cases}$$

式中: $d \in R$ 且 $d < 0.5$, $p, q \in Z$; $\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2, \dots, \varphi_p B^p$, $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ 分别为 p 阶自回归多项式与 q 阶移动平均多项式, $\Phi(B) = 0$, $\Theta(B) = 0$ 的所有根都在单位圆外,且没有重根; B 为滞后算子; $(1-B)^d$ 为分数维差分算子,即

$$(1-B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j B^j$$

则称 $y_t = x_t - \mu_t$ 为分整自回归移动平均模型。当 $d < 0.5$,零均值的 ARFIMA(p, d, q)过程是协方差平稳的;如果 $0 < d < 0.5$,则认为过程具有长记忆性。

1.2 调和分整自回归移动平均模型

调和分整自回归移动平均模型 ARTFIMA(p, d, λ, q)在上述 ARFIMA(p, d, q)模型的基础上加入了调和参数 λ 。具体表达式如下:

[收稿日期] 2019-08-12

[基金项目] 湖北省教育厅人文社会科学研究项目(14G191)

[第一作者] 王 锦(1995-),女,安徽亳州人,湖北工业大学硕士研究生,研究方向为金融统计

[通信作者] 商 豪(1982-),女,湖北罗田人,湖北工业大学副教授,研究方向为金融计量

$$\begin{cases} \Phi(B)(1-e^{-\lambda}B)^d(x_t-\mu_t)=\Theta(B)\varepsilon_t \\ \mu_t=E(x_t) \\ \varepsilon_t\sim NID(0,\sigma_\varepsilon^2) \end{cases}$$

式中： $d \in R, \lambda > 0, p, q \in Z; \Phi(B)=1-\varphi_1 B-\varphi_2 B^2, \cdots, \varphi_p B^p, \Theta(B)=1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q$ 分别为 p 阶自回归滞后算子与 q 阶移动平均滞后算子， $\Phi(B)=0, \Theta(B)=0$ 的所有特征根都在单位圆内，且没有公因子； B 为后移算子； $(1-e^{-\lambda}B)^d$ 是加入调和参数的分数维差分算子，即

$$(1-e^{-\lambda}B)^d f(t)=\sum_{j=0}^{\infty}(-1)^j e^{-\lambda j} \frac{\Gamma(1+d)}{j! \Gamma(1+d-j)} f(t-j)$$

其中 $\Gamma(d)=\int_0^{\infty} e^{-x} x^{d-1} dx$ 为伽马函数。称 $y_t=x_t-\mu_t$ 为调和分整自回归移动平均模型。与 ARFIMA(p, d, q) 模型不同,对于 $\forall d \in R$, 零均值 ARTFIMA(p, d, λ, q) 过程的协方差都是平稳的。ARTFIMA 模型的参数估计方法参考文献 [8]。

2 数据说明及预处理

本文的实证分析对象为中国居民消费者价格指数(CPI),数据来源于中国统计年鉴。选取从 2000 年 2 月到 2019 年 3 月共 230 个月度数据,数据无缺失和异常情况。初步的描述性统计分析结果显示样本数据呈现轻微右偏和尖峰分布。运用 Jarque-Bera 统计量检验得出 CPI 序列不满足正态分布。

表 1 CPI 序列的基本统计特征

样本量	230	最大值	108.70
均值	102.21	中位数	101.90
标准差	2.01	偏度(skew)	0.77
最小值	98.20	峰度(kurt)	0.79

从 CPI 序列的时间序列图(图 1)可以看出,CPI 序列没有明显的趋势性,但不具备平稳性特征。利用 ADF 方法检验 CPI 序列的单位根情况,结果显示序列非平稳。故对 CPI 进行一阶差分后得到一组新的时间序列,记为 CPID1。一阶差分后的 CPI 序列的时序图如图 2 所示。CPI 和 CPID1 的 ADF 检验结果见表 2,从检验结果可以看出 CPID1 为平稳的时间序列过程,具备经典时间序列模型平稳性建模的前提条件。

表 2 CPI 和 CPID1 样本序列的单位根检验结果

	D-F 统计量值	P 值	结果
CPI	-2.552	0.107	接受原假设,存在单位根
CPID1	-9.117	<0.01	拒绝原假设,不存在单位根

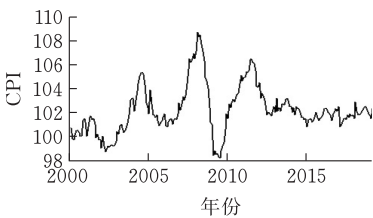


图 1 CPI 原始数据时间序列图

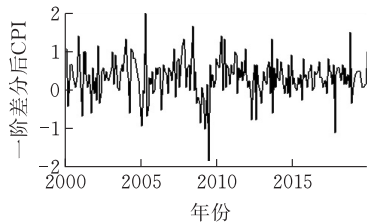


图 2 一阶差分后 CPID1 时间序列图

3 实证分析

3.1 CPI 序列的长记忆性检验

Hurst 提出的重标极差分析法(R/S 方法)是分析时间序列是否具有长期记忆性的有力工具,此方法通过计算 Hurst 指数判断时间序列数据存在的记忆特征。当 Hurst 指数介于数值 0.5 与 1 之间时,说明该序列存在持续性状态,即具有长期记忆性,且当 Hurst 指数越接近 1,相关性越强,持续性状态越明显;Hurst 指数越接近 0.5,说明存在越大的噪声,持续性状态越微弱。Hurst 指数 H 的计算原理如下^[9]:

1)定义待研究的时间序列数据为 $\{x_t, t > 0\}$, 将其划分为长度为 n 的 A 个连续子序列区间,每一个子序列区间记为 $I_a, a=1, \cdots, A$;

2)对各个子序列区间计算其均值

$$e_a=\frac{1}{n}\sum_{i=1}^n x_{i,a}$$

3)对每个子序列区间计算累积均值离差

$$y_{k,a}=\sum_{i=1}^k(x_{i,a}-e_a),k=1,2,\cdots,n$$

4)计算每个子序列区间的标准差

$$s_a=\sqrt{\frac{1}{n}\sum_{k=1}^n(x_{k,a}-e_a)^2}$$

5)计算 A 个子序列区间的平均重标度极差

$$(R/S)_n=\frac{1}{A}\sum_{a=1}^A\frac{R_a}{S_a}$$

其中 $R_a=\max(y_{k,a})-\min(y_{k,a}), k=1, \cdots, n$ 是单个子序列区间的极差;

6)利用 Hurst 提出的模型,建立关系式

$$(R/S)_n=(c\times n)^H$$

两边取对数可得

$$\log(R/S)_n=H\times\log n+\log c$$

通过最小二乘回归求解 H 值。

利用上述 R/S 方法计算 CPI 序列的 H 指数, 结果见表 3。

表 3 R/S 长记忆性检验结果

序列	观测值/个	H 指数值	R^2
CPI	230	0.621	0.95

根据 R/S 长记忆检验方法计算步骤, 通过线性回归计算得出 $H = 0.621$, 结合拟合优度 R^2 统计量可知线性拟合效果较好。从 H 指数的数值来看, 中国消费者价格指数序列存在长记忆性, 但长记忆性较弱。

3.2 建立 ARTFIMA 模型

根据长记忆性检验结果, 建立可以拟合时间序列长记忆性的 ARTFIMA 模型。首先针对一阶差分后的 CPID1 样本序列建立 ARIMA 模型, 用以消除原始序列中存在的短期相关性。结合最小 AIC 和 BIC 准则(表 4)选取 $p = 2, q = 2$ 作为相对最优阶数。运用 R 软件和极大似然估计法^[8]对 CPI 序列建立 ARTFIMA(2, $d, \lambda, 2$)模型并进行参数估计, 参数估计结果见表 4。ARTFIMA(2, $d, \lambda, 2$)模型的参数估计结果见表 5。

表 4 ARIMA 模型信息准则表(CPID1)

p	q	AIC	BIC
0	0	425.76	432.64
0	1	427.61	437.92
0	2	426.36	440.11
1	0	427.57	437.88
2	0	425.57	439.32
1	1	425.10	438.85
1	2	423.99	441.18
2	1	424.15	441.34
2	2	402.43	423.06

表 5 ARTFIMA 模型参数估计结果

参数	估计值	参数	估计值
d	1.209	φ_2	-0.773
λ	0.210	θ_1	1.807
φ_1	1.673	θ_2	-0.999

模型表达式为:

$$\begin{cases} (1 + 1.673B - 0.773 B^2)(1 - e^{-0.210} B)^{1.209} x_t = \\ \epsilon_t + 1.807 \epsilon_{t-1} - 0.999 \epsilon_{t-2} \\ \epsilon_t \sim NID(0, \sigma_\epsilon^2) \end{cases}$$

3.3 建立 ARFIMA 模型

结合前人的研究^[10]和上述长记忆性检验的结果, 建立长记忆 ARFIMA(2, $d, 2$)模型。利用极大似然估计法得出参数估计结果(表 6)。

表 6 ARFIMA 模型参数估计结果

参数	估计值	参数	估计值
d	0.058	θ_1	1.029
φ_1	1.888	θ_2	-0.181
φ_2	-0.908		

模型表达式如下:

$$\begin{cases} (1 + 1.888B - 0.908 B^2)(1 - B)^{0.058} x_t = \\ \epsilon_t + 1.029 \epsilon_{t-1} - 0.181 \epsilon_{t-2} \\ \epsilon_t \sim NID(0, \sigma_\epsilon^2) \end{cases}$$

3.4 模型对比

3.4.1 谱密度拟合函数对比 运用 R 软件画出上述建立的 ARFIMA 模型(图 3)与 ARTFIMA 模型(图 4)的谱密度函数拟合图, 该图采用对数刻度反映频率与谱密度之间的幂律关系, 其中黑色实心曲线为模型的光谱拟合情况。从谱密度函数拟合图的对比中可以看出, ARTFIMA 模型在调和参数的作用下, 对低频处的数据进行了更好的拟合。

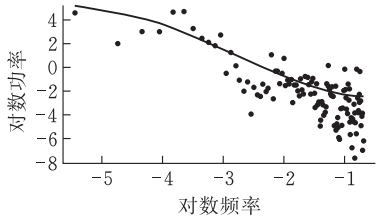


图 3 ARFIMA 模型谱密度函数拟合图

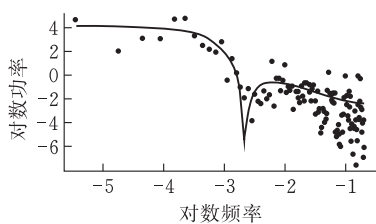


图 4 ARTFIMA 模型谱密度函数拟合图

3.4.2 预测效果对比 运用上述建立的 ARFIMA 和 ARTFIMA 模型对中国居民消费者价格指数进行 3 期预测, 结果如表 7 所示。

表 7 预测效果对比

	实际值	ARFIMA 预测值	ARTFIMA 预测值
2019 年 4 月	102.5	102.21	102.32
2019 年 5 月	102.7	102.25	102.33
2019 年 6 月	102.7	102.29	102.37
平均预测绝对误差/%		37.34	28.57

$$\text{预测绝对误差} = |(\text{预测值} - \text{实际值}) / (\text{实际值} / 100)|$$

从预测结果可以看出, ARTFIMA 模型预测误差为 28.57%, 而 ARFIMA 模型的预测误差达到 37.34%, 说明 ARTFIMA 模型在中国居民消费者价格指数序列的预测方面比 ARFIMA 模型更具优越性。

3.5 模型有效性检验

对上述两个模型产生的残差进行白噪声检验，用于验证建立模型的有效性。具体方法为：分别对残差序列和残差平方序列进行 Ljung-Box 纯随机性检验,仅当残差和残差平方序列同时通过 Ljung-Box 检验,才认为此残差是白噪声序列。从检验结果可以看出,各残差序列均是白噪声序列,说明模型拟合是合理的。

表 8 残差白噪声检验

	<i>p</i> 值
ARTFIMA 模型残差	0.8383
ARTFIMA 模型残差值平方	0.1597
ARFIMA 模型残差	0.8275
ARFIMA 模型残差值平方	0.5772

4 结束语

通过上述实证分析发现,中国居民消费者价格指数虽具有长期记忆性特征,但序列的相关性较弱,说明中国的消费者市场没有表现出随机游走的特性,并不是一个完全独立的过程,具有非线性、依赖性和潜在的可预测性。针对该序列表现出的长期记忆性建立 ARTFIMA 模型,取得了较好的拟合和预测效果,为针对时间序列过程记忆性的建模方法提供了新的视角和可能。

[参 考 文 献]

[1] Granger C W J . Long Memory Relationship and the

Aggregation of Dynamic Models[J]. Journal of Econometrics, 1980, 14(2):227-238.

[2] Jonathan R. M. Hosking. Fractional Differencing[J]. Biometrika, 1981, 68(1):165-176.

[3] 刘文财,刘豹,张维.ARFIMA 模型在中国股票市场预测的失效[J].系统工程理论方法应用,2002(2):94-100.

[4] 金秀,姚瑾,庄新田.基于分数阶差分的 ARFIMA 模型及预测效果研究[J].数理统计与管理,2007(5):896-907.

[5] Meerschaert M M, Sabzikar F. Stochastic integration for tempered fractional Brownian motion[J]. Stochastic Processes and their Applications, 2014, 124: 2363-2387.

[6] Meerschaert M M, Sabzikar F, Phanikumar M S, et al. Tempered fractional time series model for turbulence in geophysical flows [J]. Journal of Statistical Mechanics: Theory and Experiment, 2014, 10, 1088/1742-5468.

[7] Watkins N, Graves T, Franzke C, et al. Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models[J]. Statistical Mechanics and its Applications, 2017, (473): 60-71.

[8] Sabzikar F, Mcleod I, Meerschaert M M. Parameter estimation for ARTFIMA time series [J]. Journal of Statistical Planning and Inference, 2019, 200: 129-145.

[9] 黄志平.上证综指的长记忆性检验及其预测研究[D].南昌:江西财经大学,2019.

[10] 张世英,刘菁.长记忆性时间序列及其预测[J].预测,1999(3):50-51.

Application Research of Chinese Consumer Price Index Based on ARTFIMA Model

WANG Jin, SHANG Hao

(School of Sciences , Hubei Univ. of Tech., Wuhan 430068, China)

Abstract: The Autoregressive Tempered Fractionally Integrated Moving Average(ARTFIMA) time series model is applied in this paper. Aiming at the memory characteristics of the Chinese Consumer Price Index data, a combination of descriptive statistical analysis and classical time series analysis is used to establish two different long memory models, and the spectral density functions of the two models are compared. The combined graph and predicted value error show that the ARTFIMA model has better fitting effect and prediction effect. This concludes that the Chinese consumer price index sequence is more consistent with the semi-long-term correlation.

Keywords: time series; long memory; ARTFIMA model; ARFIMA model