中国农村CPI预测研究

魏铭宏

北方工业大学, 北京

收稿日期: 2021年11月23日; 录用日期: 2021年12月8日; 发布日期: 2021年12月24日

摘要

本文研究了农村CPI各分类价格指数在1998~2018年变化趋势,对总分类与各分类价格指数进行逐步回归,构建因变量与自变量的多元线性回归关系模型,运用逐步回归方法建立MLR模型,得出影响农村CPI的分类价格指数,确定其相关性变量,为进一步优化相关性系数使模型预测效果更吻合,进一步构建GM(1,N)模型进行预测;再对数据进行光滑性处理,在分析原模型背景值构造形式的基础上对其进行优化,重新构建GM(1,N)模型的背景值,提出了改进后的灰色GM(1,N)模型,通过运算和模型对比分析,发现该改进模型能够有效地提高预测精度,在对农村CPI预测中具有较强的实用性。

关键词

居民消费价格指数,GM(1,N)模型,MLR模型,改进的GM(1,N)模型

Study on the Prediction of Rural CPI in China

Minghong Wei

North China University of Technology, Beijing

Received: Nov. 23rd, 2021; accepted: Dec. 8th, 2021; published: Dec. 24th, 2021

Abstract

This paper studies the change trend of each classified price index of rural CPI from 1998 to 2018, makes a stepwise regression between the general classification and each classified price index, constructs a multiple linear regression relationship model between dependent variables and independent variables, establishes an MLR model by using the stepwise regression method, obtains the classified price index affecting rural CPI, and determines its correlation variables. In order to further optimize the correlation coefficient and make the prediction effect of the model more consistent, the GM(1,N) model is further constructed for prediction; Then the data is smoothed, the background value of the original model is optimized based on the analysis of the structural form of the background value of the original model, the background value of GM(1,N) model is reconstructed, and the improved grey GM(1,N) model is proposed. Through operation and model comparative analysis, it is found that the improved model can effectively improve the prediction accuracy and

文章引用: 魏铭宏. 中国农村 CPI 预测研究[J]. 统计学与应用, 2021, 10(6): 1045-1052. DOI: 10.12677/sa.2021.106111

has strong practicability in rural CPI prediction.

Keywords

Consumer Price Index, GM(1,N) Model, MLR Model, Improved GM(1,N) Model

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/



Open Access

1. 引言

1.1. 背景与意义

构成 CPI 的指数有食品、烟酒用品及居住、交通通讯、医疗保健个人用品等多个方面。CPI 作为一个补偿指数,以部分或全部补偿消费品和服务价格的变化补偿生活费用的变化。CPI 也是判断是否通货膨胀的一个重要指标,CPI 的高低可以在一定水平上说明通货膨胀的严重程度,CPI 上升幅度很大说明通货膨胀程度越大,市场经济不稳定,央行会有缩紧货币政策的可能,因此 CPI 指数升幅过高往往不被人们所欢迎,一般 CPI 大于 3%的增幅认为是通货膨胀。CPI 过高不仅给居民的生活带来更重的负担,同时还阻碍国民经济的全面协调可持续发展。

当前国内学术界对我国农村 CPI 波动的研究较少,目前的研究现状主要是定性分析我国农村物价涨幅高于城市的现状及原因。本文主要从定量分析我国农村总 CPI 与各分类 CPI 之间的相关性,首先通过利用 SPSS 软件对农村总 CPI 与食品烟酒、居住交通与通信等分类价格指数进行多元线性回归分析,用逐步回归方法建立模型 MLR,在此基础上建立总 CPI 指数与相关消费价格指数的灰色系统模型 GM(1,3),及通过灰度模型进行改进,以便不同的模型处理数据的过程中可以起到互补的作用。同时对比分析三种模型的预测效果,确定更实效的模型。

1.2. 文献综述

近年来许多国内外学着研究 CPI 指数,邓海云研究 CPI 八大类着手比较城乡差异,分析差异产生的结构性原因,并试图提出缓解城乡物价差异的措施,以促进中国城乡物价发展的平衡[1]。也有学者对比分析 MLR 和 GM(1,N)模型的预测效果,确定更实效的模型,从中可以分析得出我国农村各类 CPI 变动对总 CPI 的影响情况[2]。也有基于 SARIMA 模型,预测居民消费价格指数[3]。还有综合以往文献的研究,通过传统经济预测方法和机器学习方法的对比,基于新常态背景对 CPI 进行合理的预测和分析,比较不同预测方法的优劣及适用性[4]。

本文既从 CPI 各项分类指标着手研究,也通过建立预测模型的方法研究 CPI 总指数与分类指标的相关性,通过构建模型 MLR 研究哪几个因素与城乡总 CPI 具有较强相关性。由于 GM(1,N)模型在经济及空情报雷达费用等方面有着良好的预测效果[5] [6],本文运用 GM(1,N)模型进一步改善运用 MLR 模型预测 CPI 指数的精度及实效性,再通过改进的 GM(1,N)模型进一步改善预测准确率,寻找最适合预测 CPI 的模型及着重需要改善的方面,提出改善农村 CPI 的建议。

2. 基于 MLR 模型的 CPI 研究

根据上文研究发现,各大类居民消费价格指数与总指数变化有时存在相同趋势,因此我们假设各大

类 CPI 与总 CPI 存在相关性,并通过建立多元线性回归模型来定量分析他们之间的关系。

2.1. 模型建立

农村:
$$y_1 = \beta_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \dots + \beta_{17}x_{17}$$
 公式(1)

其中, β_{ii} 代表各变量的系数。模型其他符号的含义如表1所示。

Table 1. Symbol meaning in the model

表 1. 模型中符号含义

y_1	x_{11}	x_{12}	x_{13}	x_{14}	<i>x</i> ₁₅	<i>x</i> ₁₆	<i>x</i> ₁₇
农村总	农村	农村	农村生活	农村	农村交通	农村教育	农村
CPI	食品烟酒	衣着	用品及服务	医疗保健	和通信	文化和娱乐	居住

我们选取 1998 年~2016 年的数据用来求取模型参数,2017 年与 2018 年的数据用作预测。利用 SPSS 软件,采用逐步回归法,依据当 F的概率小于 0.05 时进入模型、大于 0.10 时剔除模型的规则,选择变量,得出结果。

2.2. 农村 CPI 研究

Table 2. Summary of rural models

表 2. 农村模型摘要

模型	R	\mathbb{R}^2	调整后的 R ²	标准偏斜度错误
1	0.907	0.823	0.813	1.0537
2	0.944	0.891	0.877	0.8531
3	0.967	0.934	0.921	0.6834

a.预测值: (常数), 食品烟酒; b.预测值: (常数), 食品烟酒, 居住; c.预测值: (常数), 食品烟酒, 居住, 交通和通信; d.应变数: 农村总 CPI。

根据上表2回归方程的拟合效果可知,选择模型3,即当引入食品烟酒、居住、交通和通信时,复相关系数、判定系数和调整的判定系数等取值都很大可知,这三个变量进入方程时,回归方程的拟合效果最好。

Table 3. Rural ANOVA table 表 3. 农村方差分析表

模型		平方和	df	均值平方	F	显著性
	回归	87.746	1	87.716	72.029	0.000
1	残差	18.875	17	1.110		
	总计	106.621	18			
2	回归	94.977	2	47.488	65.253	0.000
	残差	11.644	16	0.728		
	总计	106.621	18			
3	回归	99.615	3	33.205	71.088	0.000
	残差	7.006	15	0.467		
	总计	106.621	18			

由方差分析表 3 对比发现,虽然三个模型均通过了显著性检验,但模型 3 的拟合效果最好。综上, 选择模型三作为农村 CPI 的最终模型。

Table 4. Rural coefficient 表 4. 农村系数

模型 -		非标准化系数		标准化系数	Т	显著性 -	置信区间	
		В	标准误	Beta	1	业者性	下限	上限
1	(常数)	61.085	4.639		13.168	0.000	51.298	70.872
1	食品烟酒	0.396	0.045	0.907	8.890	0.000	0.302	0.490
	(常数)	37.824	8.280		4.568	0.000	20.271	55.377
2	食品烟酒	0.254	0.058	0.581	4.391	0.000	0.131	0.377
	居住	0.372	0.118	0.417	3.152	0.006	0.122	0.621
	(常数)	14.985	9.826		1.525	0.148	-5.958	35.928
2	食品烟酒	0.161	0.055	0.369	2.935	0.010	0.044	0.278
3	居住	0.363	0.094	0.407	3.837	0.002	0.161	0.564
	交通和通信	0.338	0.107	0.303	3.151	0.007	0.109	0.566

根据模型 3 的分析结果可知,首先 3 个变量均通过了回归系数 T 检验,说明线性关系显著。其次,根据 Beta 值可以得出影响农村总 CPI 的各大类指数由强到弱依次为居住类、食品烟酒类和交通和通信类。最后,根据表 4 我们可以建立农村的预测模型为:

$$y_1 = 14.985 + 0.161x_{11} + 0.338x_{15} + 0.363x_{17}$$

通过上述多元线性关系可以对某一年农村 CPI 进行预测,即将食品烟酒、居住与交通和通信代入公式,得到农村总 CPI。

现在将 2017 年和 2018 年统计数据作为预测数据,可得: 2017 年预测值为 102.45,实测值为 101.3,相对误差为 1.352%; 2018 年预测值为 103.17,实测值为 102.1,相对误差为 1.05%;两次误差的绝对值都控制在 1.2%以内,说明拟合效果较好。

3. 基于 GM(1.N)模型的 CPI 研究

3.1. GM(1,N)模型构建机理

GM(1,N)模型是一个基于一系列相互关联的系统状态模型,它不但可以了解整个系统的变化,还可以了解系统中各个环节的发展变化,是全方位描述系统特征的理想方法。

设
$$X_1^{(0)} = (X_1^{(0)}(1), X_1^{(0)}(2), \cdots, X_1^{(0)}(n))$$
 为行为变量数据序列,
$$X_i^{(0)} = (X_i^{(0)}(1), X_i^{(0)}(2), \cdots, X_i^{(0)}(n))$$
 为因子变量序列 $X_i^{(1)}$ 为 $X_i^{(0)}$ 的 1-AGO 序列,

 $Z_1^{(1)}$ 为 $X_i^{(1)}$ 的紧邻均值序列,则称 $X_1^{(0)}(k) + aZ_1^{(0)}(k) = \sum_{i=2}^{N} b_i X_i^{(0)}(k)$ 为 GM(1,N)灰色微分方程。在 GM(1,N)

灰色微分方程中-a 为发展系数, $b_i X_i^{(0)}(k)$ 为驱动项,b 为驱动系数, $u = \left[a, b_2, b_3, \cdots, b_N\right]^{\mathsf{T}}$ 为参数向量。

$$B = \begin{bmatrix} -Z_1^{(1)}(2) & -X_2^{(1)}(2) & \cdots & -X_N^{(1)}(2) \\ -Z_1^{(1)}(3) & -X_2^{(1)}(3) & \cdots & -X_N^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -Z_1^{(1)}(n) & -X_2^{(1)}(n) & \cdots & -X_N^{(1)}(n) \end{bmatrix}, \quad Y = \begin{bmatrix} X^0(2) \\ X^0(3) \\ \vdots \\ X^0(n) \end{bmatrix}$$
则参数向量 u 的最小二乘估计满足:

$$u = (B^{T}B)^{-1}B^{T}Y$$
, 影子矩阵为 $\frac{dx_{1}^{(1)}}{dt} + ax_{1}^{(1)}(t) = \sum_{i=2}^{N} b_{i}x_{i}^{(1)}(t)$ 。

3.2. GM(1,N)模型在农村 CPI 方面实证分析

根据逐步回归模型得知农村总 CPI 与食品烟酒、居住、交通和通信这三个因素有较好的线性相关性,拟合度 R 方属于较拟合,效果比较好,绝对误差控制在 1.05%以内,为了准确的预测农村总 CPI,使模型更具优化型,构建 GM(1,3)模型。根据逐步回归模型可知 GM(1,3)模型的特征变量为食品烟酒、居住、交通和通信这三个消费价格指数。 X_1 , X_2 , X_3 , X_4 分别代表农村总 CPI、食品烟酒、居住、交通和通信的消费价格指数。其中农村居民总消费价格指数 $X_1^{(0)}$ 有 21 组数据作为相关因素序列,即:

$$X_1^{(0)} = (99.0, 98.5, 99.9, 100.8, 99.6, 101.6, 104.8, 102.2, 105.4, 105.4, 106.5, 99.7, 103.6, 105.8, 102.5, 102.8, 101.8, 101.3, 101.9, 101.3, 102.1)$$

$$X_2^{(0)} = (96.6, 96.0, 97.5, 99.8, 99.3, 103.4, 111.5, 102.5, 102.1, 113.6, 114.0, 100.1, 107.5, 112.4, 104.0, 104.9, 102.6, 102.4, 104.0, 98.9, 98.9, 101.1)$$

$$X_3^{(0)} = (98.8, 99.9, 102.8, 100.3, 100.1, 101.0, 105.8, 105.2, 104.6, 104.4, 108.2, 98.5, 104.5, 105.7, 101.9, 102.3, 101.9, 99.7, 100.6, 102.7102.7, 103.3)$$

$$X_4^{(0)} = (95.9, 94.5, 93.6, 98.7, 98.2, 98.6, 99.8, 100.3, 101.3, 100.6, 100.7, 98.2, 100.3, 101.3, 100.6, 100.1, 100.0, 98.1, 98.9, 101.4, 101.4, 101.8)$$

把数据分别做一次累加,建立灰色微分方程,参数向量 u 的最小二乘估计为: $u = \left(B^{\mathsf{T}}B\right)^{-1}B^{\mathsf{T}}Y$,用 python 进行测算可得到进一步的结果。可知道矩阵 B 及矩阵 Y 带入公式 $\hat{u} = \left(B^{\mathsf{T}}B\right)^{-1}B^{\mathsf{T}}Y$,可知道 u 的最小二乘估计为 $\hat{u} = \begin{bmatrix} 2.09, 0.33, 0.94, 0.83 \end{bmatrix}$,从而求出灰微分方程的系数。

得出 2017 年预测值为 101.10,实际值为 101.3,相对误差为 0.20%;同理可知 2018 年预测值为 103.09,而实测值为 102.1,相对误差为 0.97%;预测的两年数据误差绝对值都控制在很小数以内,拟合效果相对逐步回归更好,从而说明对相关变量的贡献系数起到了进一步的优化作用。

4. GM(1,N)模型的改进

4.1. 灰色关联分析

采用 GM(1,N)模型进行费用预测 CPI 时,需明确费用与各个因素之间的相互关系。模型显著因素的选择将直接影响到最后的预测精度。预测模型显著因素选择过少,无法有效地反映外部因素对装备费用的影响,显著因素选择过多,则会造成预测结果灰度过大。因此对原始数据进行灰色关联度分析,能够有效提高模型的预测精度。

设收集的原始数据矩阵为
$$S^{(0)}$$
,则母序列为: $S_1^{(0)} = \left(S_1^{(0)}(1), S_1^{(0)}(2), \cdots, S_1^{(0)}(k)\right)$
子序列为: $S_j^{(0)} = \left(S_j^{(0)}(1), S_j^{(0)}(2), \cdots, S_j^{(0)}(k)\right)$ ($j = 2, 3, \cdots, n$) ($k = 1, 2, \cdots, m$)
母序列、子序列无量纲化处理得: $S_j^{(0)}(K) = S_1^{(0)}(k) / S_1^{(0)}(1)$
 $S_j^{(0)}(k) = S_j^{(0)}(k) / S_j^{(0)}(1)$ 从而可得到序列 $S_1^{(0)} = S_j^{(0)}(k) / S_j^{(0)}(k)$

其中 $\Delta j(k) = \left|S_1^{\prime(0)}(k) - S_j^{\prime(0)}(k)\right|$ 。 ρ 为分辨系数,当 ρ < 0.55 时,模型灰色关联度分析的分辨率较好。于是可计算母序列 $S_1^{(o)}$ 与第 j 个子序列 $S_j^{(0)}$ 的关联度为:

 $\gamma_j = \frac{1}{n} \sum_{i=1}^{m} \ell_j(k)$ 如果 $\gamma_j \ge 0.5$,则认为母序列与子序列有关联。

序列 ρ 取 0.3,通过逐步回归得到的变量有食品和烟酒、居住、交通和通信。

分别将三组变量带入数据得到 CPI 与食品和烟酒的关联度 γ_1 为 0.8199,CPI 与居住的关联度 γ_2 为 0.5895,CPI 与交通和通信的关联度 γ_3 为 0.239,小于 0.5,所以接下来的分析去掉变量交通和通信。

4.2. 数据光滑性处理

考虑到 CPI 受食品和烟酒、居住等因素影响,采集的原始数据必然存在一定的波动性,而 GM(1,N) 模型在处理具有波动性的数据序列时存在一定的局限性,原因在于原始数据不能满足预测模型对光滑性的要求。因此为了更准确地预测 CPI,首先对筛选后的原始数据进行预处理,使处理后的数据满足预测模型对光滑性的要求。

设 $Q^{(0)}$ 为 $S^{(0)}$ 经过灰色关联度分析筛选后的原始数据矩阵, $X^{(0)}$ 为 $Q^{(0)}$ 经过对数处理后的建模数据矩阵。 为提高原始数据的光滑性,下面对初始数据参数作对数处理,即: $X_j^{(0)}(k) = \ln q_j^{(0)}(k)$ 其中 $Q_j^{(0)} = \left(q_j^{(0)}(1), q_j^{(0)}(2), \cdots, q_j^{(0)}(k)\right)$ 将处理得到的数据作为模型预测的原始建模数据,预测完成后再对预测结果进行指数还原,可得出模拟值和预测值。

4.3. 模型背景值优化

背景值的构造形式是影响模型预测精度的主要因素之一,对背景值的改进能够有效降低模型预测误差,增强模型适用性。本文在研究背景值构造的基础上,提出 GM(1,N)模型背景值的优化方法,从而提高预测 CPI 的精度。

引入加权因子 γ 对 $Z^{(1)}(k)$ 进行重构,即: $Z^{(1)}(k) = \gamma X^{(1)}(k) + (1-\gamma) X^{(1)}(k-1)$

将 γ 称为加权系数(0 < γ < 1)GM(1,N)模型中的 $Z^{(1)}(k)$ 是 γ = 0.5 的情况。当原始值与模型预测值之差的平均相对误差达到最小,即使公式 3 最小:

此时满足上述条件的 γ 值计算出来的预测结果误差最小,即为最优解。对原始建模数据进行一次累加生成累加数列,利用重组的 $Z^{(1)}(k)$ 构建背景值函数,建立改进后的GM(1,N)模型,求得满足 $\overline{\Delta}$ 最小时的加权因子 γ 。通过 python 测试不同的 γ ,得到使 $\overline{\Delta}$ 为最小的 γ 为 0.09。利用优化后的背景值求得 B 的值,根据最小二乘法求出 a,b_1,b_2 ,的值为(223.47, 16.34, 212.025)得出 $\hat{X}^{(1)}(k+1)$ 。

4.4. 实例分析

为了突出本文改进的 GM(1,2)模型进行 CPI 预测的优越性,以 2017、2018 年 CPI 为例,将预测结果分别与逐步回归模型、传统 GM(1,3)模型进行对比分析,预测结果和相对误差如表 5 所示。

Table 5. Comparison of the analysis results of the rural CPI model 表 5. 农村 CPI 模型分析结果对比

年份	实测值 -	MLR 模型		GM(1,3)		改进的 GM(1,2)	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
2017年	101.3	102.45	1.352%	101.10	0.197%	101.104	0.193%
2018年	102.1	103.17	1.048%	103.09	0.970%	102.729	0.678%

由 MLR 模型在确定食品烟酒、居住、交通和通信这三个因素为农村总 CPI 的主要影响因素时,为进一步优化其相应参量的贡献系数,即构建 GM(1,3)模型,由表 5 可知灰色系统模型拟合效果好。而相比之下改进的灰度模型效果最好,预测方法误差最小,说明在小样本情况下,进行 CPI 预测时,本文的改进模型要更优。

GM(1,3)模型结果为:

$$X_1^{(0)}(k) + 2.09Z_1^{(1)}(k) = 0.33X_2^{(1)}(k) + 0.94X_3^{(1)}(k) + 0.93X_4^{(1)}(k)$$

通过此模型相关变量的系数大小可以确定各分类价格指数对农村总 CPI 的波动变化情况。

 X_1 , X_2 , X_3 , X_4 分别代表农村总 CPI、食品烟酒、居住、交通和通信的消费价格指数。由上式可知食品烟酒对农村总 CPI 有影响,食品烟酒献系数较其他系数都小,说明我国农村的生活水平偏低,食品方面需要相应的改善;这三个相关性变量中交通和通信及居住的贡献系数很高,可见我国农村居民在居住及交通和通信等方面较其他因素的经济负担严峻许多,说明我国在相应制度上需进一步完善和健全。

改进的 GM(1,2)模型结果为:

$$X_1^{(0)}(k) + 228.47Z_1^{(1)}(k) = 16.34X_2^{(1)}(k) + 212.025X_3^{(1)}(k)$$

此模型剔除了交通通信的影响,保留食品及居住的影响。由于对数据进行了光滑性处理,系数的大小发生了相应的变化,但能看出居住的贡献系数依然远高于食品及烟酒。

5. 结论与建议

5.1. 结论

改进 GM(1,N)模型通过灰色关联度分析、数据光滑性理和背景值优化,提高了模型与实际数据的拟合效果,预测结果更加符合实际情况,充分发挥灰色系统预测所需数据少、预测精度高的优势,且对原始数据的要求不高,无须具有典型的分布规律,相比之下更具有实用性。对农村 CPI 预测的准确性来看,本文提出的改进的 GM(1,N)模型要明显优于线性回归模型和 GM(1,N)模型,在对农村 CPI 预测中具有较强的实用性。

通过对 MLR、GM(1,N)模型及改进的 GM(1,N)模型分析能得出我国农村存在的两大问题:第一,我国农村对食品的重视程度较低,说明我国农村的生活水平较低;第二,居住等方面的支出对我国农村总 CPI 走高具有较大的推动作用。由模型变量的相关系数可以确定各分类价格指数对总 CPI 的波动变化情况,较好地反映各因素对我国农村居民生活水平的推动作用,也给政府决策层提供较好的理论参考价值及可行的政策取向。为改善农村发展状况,政府有责任建立健全社会保障制度体系,同时加大对农村居民居住环境方面的扶持力度,为实现农村更好地发展作努力。

5.2. 建议

为缩小城乡 CPI 差异,使城乡都能得到更好的发展,我们建议应当从以下几方面着手:

1) 提供优质服务

随着农村居民生活水平的提高、消费结构的升级,我们应当鼓励商家推出更多符合农村居民消费习惯的商品,提供更加优质精准的服务。

2) 改善城乡居民居住条件

十九大报告提出,加快建立多主体供给、多渠道保障、租购并举的住房制度,让全体人民住有所居。 执行目标需要具体有效的措施执行,要强化房地产长效机制建设,全面落实房地产调控政策措施,大力 实施保障性安居工程,着力满足居民居住需求。同时要大力发展公租房市场,建立筹集分配长效机制, 实行市场租房补贴政策,将符合一定条件的非京籍家庭纳入保障范围,满足新市民住房需求。

参考文献

- [1] 沈赟. 中国城市和农村 CPI 差异研究——基于 2001~2014 年的数据分析[J]. 价格理论与实践, 2015(4): 42-46.
- [2] 邓海云,周莹,刘辉,程碧辉,张同全.基于 GM(1,N)模型的农村居民消费价格指数研究[J]. 统计与决策, 2013(17): 4-7.
- [3] 朱颜杰, 樊顺厚, 雷怀英. 基于 SARIMA 模型的居民消费价格指数的预测[J]. 统计与策, 2014(4): 37-39.
- [4] 陈梦根, 任桃萍. 新常态经济的 CPI 预测模型——构建与实证比较[J]. 调研世界, 2020(2): 3-8.
- [5] 黄娜. 基于 BP 神经网络改进的 GM(1,N)模型在经济预测中的应用——以河南省经济为例[J]. 南阳理工学院学报, 2009, 1(6): 76-79.
- [6] 张长聪, 方其庆, 刘庆华, 夏亮, 施端阳. 基于改进 GM(1,N)模型的对空情报雷达费用预测[J]. 兵器装备工程学报, 2019, 40(11): 144-148.