

The Statistical Regression Model of Consumer Price Index

Yi Zhang

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: zgws9999@163.com

Received: Feb. 10th, 2015; accepted: Feb. 21st, 2015; published: Feb. 28th, 2015

Copyright © 2015 by author and Hans Publishers Inc.

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

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



Open Access

Abstract

We obtain the best regression model by using stepwise regression method and analyzing the consumer price index and its impact factors; then we get some most important impact factors of it. The feasibility of the regression model is proved in the paper. The results show that the accuracy of model reaches to 99.8982%, so it have feasibility. Some suggestions based on the results are given at last.

Keywords

Consumer Price Index, Stepwise Regression Method, Model Accuracy

居民消费价格指数的统计回归模型

张 艺

云南财经大学统计与数学学院, 云南 昆明
Email: zgws9999@163.com

收稿日期: 2015年2月10日; 录用日期: 2015年2月21日; 发布日期: 2015年2月28日

摘 要

对居民消费价格指数及各影响因子进行分析, 运用逐步回归法得到最优回归模型, 得出居民消费价格指

数最重要的几个影响因子；对回归模型的可行性进行了验证，结果显示所得模型精度达到99.8982%，具有可行性，并对结果进行分析，提出一些建议。

关键词

居民消费价格指数，逐步回归法，模型精度

1. 引言

居民消费价格指数(CPI)，是一个反映居民家庭一般所购买的消费商品和服务价格水平变动情况的宏观经济指标。它是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相对数，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。

居民消费价格统计调查的是社会产品和服务项目的最终价格，一方面同人民群众的生活密切相关，同时在整个国民经济价格体系中也具有重要的地位。它是进行经济分析和决策、价格总水平监测和调控及国民经济核算的重要指标。其变动率在一定程度上反映了通货膨胀或紧缩的程度。它主要包括食品类、烟酒及用品类、衣着类、家庭设备用品及维修服务类、医疗保健和个人用品类、交通和通信类、娱乐教育文化用品及服务类、居住类。了解CPI的变化规律和价格波动对居民消费的影响程度，为采取相应的措施提供依据，具有重要的意义[1]。

关于居民消费价格指数的研究也比较多，例如王晖[2]对不同地区居民消费价格水平做了比较分析；何维，炜田皓[3]对居民消费价格指数影响因素进行了分析；石柱鲜，吴泰岳，邓创[4]分析了我国居民消费价格变动的主要影响因素；杨凌云，王凡彬，潘瑞，梁杰[5]研究了CPI指数预测的统计回归模型；张愿章，王淑敏[6]基于多元回归分析研究了河南省居民消费价格指数的数学模型。

本文主要运用SPSS软件对居民消费价格指数进行多元回归分析[7]，首先对自变量进行共线性诊断，然后运用逐步回归法得到最优回归模型，得出居民消费价格指数最重要的几个影响因子；并对回归模型的可行性进行了验证，结果显示所得模型精度达到99.8982%，说明所得结果具有可行性，可以通过对结论中影响最大的几个影响因子进行合理的调控，从而更好进行经济分析和决策、国民经济核算以及监测价格的总水平。

2. 数据来源及变量

为了研究居民消费价格指数，本论文收集了2002年至2012年之间居民消费价格指数及食品类居民消费价格指数等9个变量的数据，此数据来源于中华人民共和国国家统计局网站。

其中 y 是指居民消费价格指数； x_1 是指食品类居民消费价格指数； x_2 是指烟酒及用品类居民消费价格指数； x_3 是指衣着类居民消费价格指数； x_4 是指家庭设备用品及维修服务类居民消费价格指数； x_5 是指医疗保健和个人用品类居民消费价格指数； x_6 是指交通和通信类居民消费价格指数； x_7 是指娱乐教育文化用品与服务类居民消费价格指数； x_8 是指居住类居民消费价格指数。

3. 共线性诊断

变量间的关系可由所谓的方差膨胀因子VIF来判断。记 R_j^2 为预测变量 x_j 关于所有其他预测变量作回归得到的复相关系数的平方。因此 x_j 的方差膨胀因子为：

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, \dots, p$$

其中, p 为预测变量的个数。若 x_j 与其他预测变量有较强的线性关系, 则 R_j^2 将接近 1, 从而 VIF_j 会很大。当方差膨胀因子的值大于 10 时, 通常被看做数据有共线性问题。

另外, 对预测变量做主成分分析得出的特征向量的特征值 λ 。若某个 λ 比其他小很多(接近于 0), 那就存在多重共线性。接近 0 的 λ 的个数就等于数据中存在的不同多重共线性的组数。此外, 主成分分析得出的相关矩阵的条件数也是各变量总的多重共线性的一种度量, 较大的表明强共线性。当条件数超过 15 时, 数据中的共线性的负面效应变得很强。

因此, 为了判断哪些自变量是造成多重共线性的, 通过 SPSS [8] 做主成分, 计算自变量的方差膨胀因子 VIF、特征向量的特征值及相关矩阵的条件数。结果见表 1、表 2。

从表 2 可以看出各变量的方差膨胀因子 VIF。可以看出 x_1 、 x_2 、 x_4 、 x_6 、 x_8 的方差膨胀因子较大, 均大于 10。因此更加确定这 8 个变量间存在共线性。

从表 1 可以看出各变量的特征值和条件数。其中除了常数项外, 其余变量的特征值都接近 0。相关矩阵的条件数也是除了常数项外都大于 15。因此该数据存在多组多重共线性, 并且具有很强的负效应。为此, 必须要对数据的多重共线性进行处理。

4. 逐步回归法的应用

对于数据的多重共线性, 可以采用逐步回归法进行处理, 通过 SPSS 对变量进行线性回归分析, 筛选和剔除不合适的变量, 得到以下 6 组模型, 即表 3。

从表 3 可以看出, 6 组模型中 F 检验的概率值均小于 0.05, 且 6 组模型的决定系数 R^2 均接近 1。下面继续判定 6 组模型中那一组成立。

由表 4 至表 9 中 6 种模型的回归系数, 这 6 种模型可以分别表示为:

$$\text{模型 1: } y = 56.474 + 0.434x_1$$

$$\text{模型 2: } y = 7.805 + 0.391x_1 + 0.538x_6$$

$$\text{模型 3: } y = 14.841 + 0.361x_1 + 0.379x_6 + 0.115x_8$$

$$\text{模型 4: } y = 7.951 + 0.302x_1 + 0.283x_2 + 0.12x_6 + 0.214x_8$$

$$\text{模型 5: } y = 8.431 + 0.283x_1 + 0.375x_2 + 0.253x_8$$

$$\text{模型 6: } y = 16.033 + 0.309x_1 + 0.213x_2 + 0.092x_3 + 0.224x_8$$

对于以上模型, 系数估计值显著的模型需要符合以下条件: 回归系数的估计值在 T 检验中的概率均小于显著性水平 0.05, 且其变量的方差膨胀因子小于 10。由表 4 至表 9 可知, 符合此条件的模型为模型 1 和模型 6。

由以上的分析可知, 只有模型 1 和模型 6 的拟合效果最好。比较模型 1 和模型 6, 模型 1 的 R^2 等于 0.947; 模型 6 的 R^2 等于 0.999。由于模型 6 的拟合优度比模型 1 的拟合优度更大, 所以优先选取模型 6, 从表 9 中也可以看到模型 6 的回归系数估计值。因此, 利用逐步回归法建立的回归模型为:

$$y = 16.033 + 0.309x_1 + 0.213x_2 + 0.092x_3 + 0.224x_8$$

5. 回归模型的可行性检验

利用建立的回归模型, 对模型精度进行检验, 所得结果如表 10。

由表 10 可知, 该模型的平均相对误差为 0.001018, 模拟精度达到 99.8982%, 模拟精度较高, 具有可行性。

6. 结论

通过以上的讨论, 首先得出了回归方程的模型, 通过对模型的分析可知, 食品类居民消费价格指数

Table 1. Collinearity diagnosis table

表 1. 共线性诊断表

模型	(常量)	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
特征值	8.997	0.002	0.000	0.000	0.000	5.409E-05	1.472E-05	5.883E-06	1.982E-06
条件数	1.000	72.987	136.980	209.617	286.565	407.835	781.864	1236.660	2130.772

Table 2. Inspection table of VIF

表 2. 方差膨胀因子检验表

模型	(常量)	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
VIF		21.105	18.876	8.241	12.386	3.715	22.447	6.631	30.470

Table 3. F -test of each regression model and inspection table of coefficient of determination表 3. 各回归模型的 F 检验及决定系数检验表

模型	1	2	3	4	5	6
F 值	162.035	291.631	440.227	581.420	781.115	1114.056
Sig	0.000	0.000	0.000	0.000	0.000	0.000
R 方	0.947	0.986	0.995	0.997	0.997	0.999
	因变量	y				
模型 1	预测变量	(常量)	x_1			
模型 2	预测变量	(常量)	x_1	x_6		
模型 3	预测变量	(常量)	x_1	x_6	x_8	
模型 4	预测变量	(常量)	x_1	x_6	x_8	x_2
模型 5	预测变量	(常量)	x_1	x_8	x_2	
模型 6	预测变量	(常量)	x_1	x_8	x_2	x_3

Table 4. T -test of model 1 and inspection table of VIF表 4. 模型 1 的 T 检验及方差膨胀因子检验表

预测变量	(常量)	x_1
回归系数		56.474
Sig		0.000
VIF		1.000

Table 5. T -test of model 2 and inspection table of VIF表 5. 模型 2 的 T 检验及方差膨胀因子检验表

预测变量	(常量)	x_1	x_6
回归系数	7.805	0.391	0.538
Sig	0.471	0.000	0.001
VIF		1.240	1.240

(x_1)、烟酒及用品类居民消费价格指数(x_2)、衣着类居民消费价格指数(x_3)、居住类居民消费价格指数(x_8)对居民消费价格指数有显著性的影响,且有正向的推动作用,它们之间呈一种线性关系。

Table 6. *T*-test of model 3 and inspection table of VIF
表 6. 模型 3 的 *T* 检验及方差膨胀因子检验表

预测变量	(常量)	x_1	x_6	x_8
回归系数	14.841	0.361	0.379	0.115
Sig	0.078	0.000	0.004	0.013
VIF		1.797	1.749	2.535

Table 7. *T*-test of model 4 and inspection table of VIF
表 7. 模型 4 的 *T* 检验及方差膨胀因子检验表

预测变量	(常量)	x_1	x_6	x_8	x_2
回归系数	7.951	0.302	0.120	0.214	0.283
Sig	0.239	0.000	0.369	0.004	0.046
VIF		8.299	5.933	8.249	7.470

Table 8. *T*-test of model 5 and inspection table of VIF
表 8. 模型 5 的 *T* 检验及方差膨胀因子检验表

预测变量	(常量)	x_1	x_8	x_2
回归系数	8.431	0.283	0.253	0.375
Sig	0.206	0.000	0.000	0.000
VIF		3.735	2.129	2.202

Table 9. *T*-test of model 6 and inspection table of VIF
表 9. 模型 6 的 *T* 检验及方差膨胀因子检验表

预测变量	(常量)	x_1	x_8	x_2	x_3
回归系数	16.033	0.309	0.224	0.213	0.092
Sig	0.022	0.000	0.000	0.029	0.036
VIF		5.846	2.972	6.272	3.139

Table 10. The result of simulate data
表 10. 模拟数据检验结果

年份	原值	模拟值	误差	相对误差
2002	99.2	99.3831	-0.1831	0.00185
2003	101.2	101.109	0.091	0.000899
2004	103.9	104.1073	-0.2073	0.002
2005	101.8	101.8675	-0.0675	0.00066
2006	101.5	101.6467	-0.1467	0.00145
2007	104.8	104.9486	-0.1486	0.00142
2008	105.9	105.9634	-0.0634	0.0006
2009	99.3	99.3784	-0.0784	0.00079
2010	103.3	103.3146	-0.0146	0.00014
2011	105.4	105.456	-0.056	0.00053
2012	102.6	102.6895	-0.0895	0.00087

食品价格的上涨对 CPI 的上涨有很强的推动作用。因此，应该促进第一产业的发展，提高农业生产效率，保证农产品的供求平衡。加大对农业科技创新的投资，鼓励对农业科技的研究，推广节约成本又成效高的农业技术。政府可制定相关的措施进行主要农产品的市场调控，稳定食品价格；也可指派专业人员深入农户，对种植和养殖技术及产品的储存进行指导，对运送农产品的物流公司给予相关福利政策，从而使农业成本和浪费率降低，产量增高，进而达到稳定农产品价格的效果。

烟酒及用品类价格的上涨对 CPI 的上涨的推动作用也很大。一些香烟和酒类等通过炒作等手段使价格增长的较高，甚至出现天价，政府也可制定相关措施和政策对其价格进行调控，例如给各个档次的烟酒制定一个价格最高的上限，且不同档次的区间长度选取要适中，不宜过大，从而避免出现价格失调的现象。

衣着类价格的上涨也是 CPI 上涨的一个主要推动因素。对此，我们可以通过鼓励棉花等原料的种植以及绵羊等动物的养殖，通过增大衣着原料的产量来降低工厂的原料采集成本，进而达到降低衣着类价格的效果。

CPI 持续上涨的主要驱动因素中居住类的上涨趋势也非常明显。近年来，房地产价格上涨幅度较大，对整体物价的上涨具有很大的刺激作用。对此应继续加强和改善房地产调控，在引导商品房产业健康发展的同时，结合旧城区改造，大力建设保障性住房，增加住房的有效供应，采取长效措施遏制房价的过快上涨[9]。

致 谢

审稿人及编辑提出的修改意见使本文的内容更加完善，在此对审稿人及编辑表示衷心的感谢！

参考文献 (References)

- [1] 李慧 (2014) 我国居民消费价格指数存在的问题及改进措施. *赤峰学院学报(自然科学版)*, **11**, 139-141.
- [2] 王晖 (2005) 不同地区居民消费价格水平比较分析. *经济师*, **8**, 30-31.
- [3] 何维, 炜田皓 (2007) 居民消费价格指数影响因素分析. *统计与决策*, **23**, 66-68.
- [4] 石柱鲜, 吴泰岳, 邓创 (2009) 我国居民消费价格变动的主要影响因素分析, *延边大学学报*, **2**, 99-103.
- [5] 杨凌云, 王凡彬, 潘瑞, 梁杰 (2010) CPI 指数预测的统计回归模型. *重庆文理学院学报*, **1**, 38-41.
- [6] 张愿章, 王淑敏 (2010) 基于多元回归分析的河南省居民消费价格指数的数学模型. *华北水利水电学院*, **1**, 103-105.
- [7] 于秀林, 任雪松 (2006) 多元统计分析. 中国统计出版社, 北京.
- [8] 杨小平, 刘刚, 赵晋 (2008) 统计分析方法与 SPSS 应用教程. 清华大学出版社, 北京.
- [9] 王沪怡, 尹长明 (2011) 广西 CPI 指数统计数据多元回归分析. *合作经济与科技*, **417**, 12-13.