

基于我国财政收入统计模型的应用回归分析案例教学研究

张宣昊

(上海第二工业大学理学院, 上海 201209)

摘要: 收集了我国宏观经济指标数据, 利用多元线性回归分析方法建立了我国财政收入的统计模型。揭示了影响我国财政收入的主要经济因素, 为政府制定经济政策提供依据。同时也为统计专业主干课程“应用回归分析”提供一个综合性教学案例, 帮助学生更好地理解并掌握回归分析的统计方法。

关键词: 回归分析; 逐步回归; 财政收入; 教学案例

中图分类号: O212.4

文献标志码: A

0 引言

“应用回归分析”是统计专业的主干课程。回归分析是确定两个或两个以上变量间相互依赖定量关系的一种常用统计分析方法, 运用十分广泛。综合性案例分析是“应用回归分析”教学的一个重要手段, 通过实际统计案例能更好地帮助学生快速掌握回归分析这一实用的统计方法。

本文收集了我国宏观经济指标数据(1955~2014)^[1], 采用逐步回归方法筛选变量并做异方差的处理, 建立了我国财政收入的统计回归模型, 揭示影响我国财政收入的主要经济因素。具体指标及模型中对应的变量如下: y 为财政收入(亿元); x_1 为交通运输仓储邮电通信业(亿元); x_2 为批发零售贸易及餐饮业(亿元); x_3 为黄金储备(万盎司); x_4 为第三产业(亿元); x_5 为建筑业(亿元); x_6 为工业(亿元); x_7 为第一产业(亿元); x_8 为第二产业(亿元); x_9 为人均国内生产总值(元); x_{10} 为国内生产总值(亿元); x_{11} 为年末总人口(万人); x_{12} 为出口总额(亿元); x_{13} 为进口总额(亿元); x_{14} 为外汇储备(亿美元); x_{15} 为进出口总额(亿元)。

通过我国财政收入回归分析的实际案例, 可使學生充分理解回归分析方法的统计意义及其应用价值, 能帮助教师提高“应用回归分析”课程的教学效果。

1 建立回归模型

多元线性回归是指有多个解释变量的线性回归, 用于揭示被解释变量与多个解释变量间的线性关系, 多元线性回归的数学模型为:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

其中: y 为响应变量; x_i 为自变量 ($i = 1, 2, \cdots, p$); $\varepsilon \sim N(0, \sigma^2)$ 为随机误差项^[2-3]。

合理有效的回归模型要求自变量“少而精”^[4-5], 学生在初步学习回归分析时容易将所有收集到的指标变量加入模型而违背了“少而精”的原则, 使得模型效果不理想。通过 SPSS 统计软件, 以财政收入为响应变量, 对 15 个常用经济指标筛选变量^[6-7], 使学生通过案例更直观地了解筛选变量的过程和方法。

1.1 自变量的选择

利用逐步回归法, 经过 7 步变量筛选得到表 1 所示的逐步回归系数。

根据表 1 第 7 步回归系数得回归模型:

$$y = -2\,347.009 - 0.828x_3 + 0.446x_4 - 4.964x_9 + 0.298x_{10} + 0.051x_{11} + 0.983x_{14} + 0.031x_{15}$$

注: 回归分析中自变量的选择最常用的有效方法是逐步回归法——从 15 个经济指标中逐个筛选并检验后加入回归模型中, 直到所选变量均通过显著性检验。

收稿日期: 2016-08-24

通信作者: 张宣昊(1981-), 男, 上海人, 讲师, 硕士, 主要研究方向为概率统计、最优试验设计。E-mail: xhzhang@sspu.edu.cn。

基金项目: 上海第二工业大学校重点学科 (XXKPY1604), 上海第二工业大学校重点课程建设 (A01GY16F022) 资助

1.2 多重共线性的检验与处理

一般经济指标之间本身有较强的相关性,即存在多重共线性现象,而表 1 所列逐步回归第 7 步的

多个指标存在 VIF 统计量远大于 10 的情况,模型拟合效果不佳。依次去除 VIF 值最大的变量,重新建立模型,如表 2 所示。

表 1 逐步回归系数

Tab. 1 Stepwise regression coefficients

指标变量	非标准化系数		标准系数	t	Sig.	共线性统计量	
	B	标准误差	β			容差	VIF
常量	-2 347.009	358.839		-6.541	0.000		
第三产业 x_4 /亿元	0.446	0.016	0.951	28.211	0.000	0.001	921.090
外汇储备 x_{14} /亿美元	0.938	0.039	0.283	24.095	0.000	0.009	111.704
人均国内生产总值 x_9 /元	-4.964	0.236	-1.699	-21.072	0.000	0.000	5 268.928
国内生产总值 x_{10} /亿元	0.298	0.022	1.392	13.559	0.000	0.000	8 545.047
年末总人口 x_{11} /万人	0.051	0.006	0.037	8.624	0.000	0.068	14.798
进出口总额 x_{15} /亿元	0.031	0.004	0.068	7.101	0.000	0.013	74.466
黄金储备 x_3 /万盎司	-0.828	0.206	-0.020	-4.011	0.000	0.048	20.961

表 2 回归系数

Tab. 2 Regression coefficients

指标变量	非标准化系数		标准系数	t	Sig.	共线性统计量	
	B	标准误差	β			容差	VIF
常量	14 630.934	4 307.600		3.397	0.001		
黄金储备 x_3 /万盎司	13.084	2.807	0.322	4.660	0.000	0.113	8.822
年末总人口 x_{11} /万人	-0.296	0.058	-0.214	-5.063	0.000	0.303	3.299
进出口总额 x_{15} /亿元	0.373	0.024	0.833	15.761	0.000	0.194	5.149

可见最终 3 个自变量的 VIF 统计量均小于 10,即消除多重共线性后的回归模型为

$$y = 14\,630.934 + 13.084x_3 - 0.296x_{11} + 0.373x_{15}$$

1.3 异方差的检验与处理

异方差是常见的影响回归模型效果的问题,例如建立居民收入与消费水平回归模型时,由于低收入家庭购买的差异性比较小,高收入家庭购买行为的差异很大,导致模型的随机项具有不同的方差,即产生了异方差。当存在异方差时,普通最小二乘估

计将不再是最小方差线性无偏估计且参数的显著性检验失效。学生在做回归分析时最容易疏忽的问题之一就是直接对回归参数做最小二乘估计而不做异方差的检验。以下对上述筛选变量后的财政收入模型加以说明与处理。

随着我国经济的高速发展,我国财政收入不断增长^[8],确实有必要对模型做异方差的检验与处理。由图 1 可见,残差分布极其不均匀,模型可能存在异方差。继续用等级相关系数法进一步检验。由表 3 得残差绝对值与 x_1 的等级相关系数为 0.709,残差

绝对值与 x_3 的等级相关系数为 0.632, 残差绝对值与 x_{14} 的等级相关系数为 0.807, 可以认为残差绝对值与各自变量显著相关, 即模型确实存在异方差。

消除异方差一般可使用加权最小二乘法重新估计参数, 选取等级相关系数最大的自变量 x_3 构造权函数, 得到输出结果如表 4 所示, 即最终回归模型及标准化模型为:

$$y = 11\,702.247 + 0.080x_3 - 0.015x_{11} + 0.368x_{15}$$

$$y^* = 0.066x_3^* - 0.443x_{11}^* + 0.726x_{15}^*$$

且模型判定系数 $R^2 = 0.926$, 模型有很好的拟合效果。

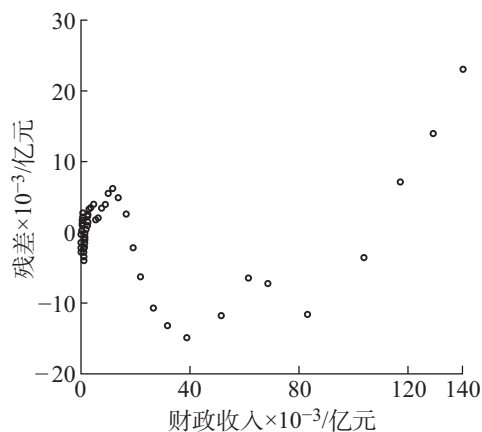


图 1 残差图

Fig. 1 Residual plots

表 3 等级相关系数

Tab. 3 Rank correlation coefficients

相关系数检验	ABSE	黄金储备 x_3 /万盎司	年末总人口 x_{11} /万人	进出口总额 x_{15} /亿元
Pearson 相关系数	1	0.709**	0.632**	0.807**
显著性检验 P 值		0.000	0.000	0.000
样本量 N	60	60	60	60

表 4 回归系数

Tab. 4 Regression coefficients

指标变量	未标准化系数		标准化系数		t	Sig.
	B	标准误差	β	标准误差		
常数	11 702.247	162.723			-4.316	0.000
黄金储备 x_3 /万盎司	0.080	0.104	0.066	0.086	2.768	0.046
年末总人口 x_{11} /万人	-0.015	0.003	-0.443	0.086	5.138	0.000
进出口总额 x_{15} /亿元	0.368	0.021	0.726	0.042	17.447	0.000

2 结 论

影响我国财政收入规模的经济因素有很多^[9], 其中最直接、最能反映财政收入变化规律的是进出口总额、年末总人口数、黄金储备等几个指标。由财政收入的多元线性回归方程得: 在其他变量不变的情况下, 黄金储备每增加 1 万盎司, 将使得财政收入平均增加 0.080 亿元; 同样在其他变量不变的情况下, 年末总人口每增加 1 万人, 财政收入平均减少 0.015 亿元, 进出口总额每增加 1 亿元, 财政收入平均增加 0.368 亿元。另由标准化的回归方程得: 进出口总额对财政收入的影响最大, 黄金储备对财政支

出的影响相对较小。根据影响财政收入的主要经济因素制定宏观经济政策能使财政收支平衡, 是社会总供求平衡和国民经济健康发展的有力保证^[10]。

财政收入回归分析过程较充分地运用了多元回归分析过程中的逐步回归法和异方差检验处理法, 在教学中可作为一个综合性的案例帮助学生了解筛选变量和做异方差检验的必要性, 从而更好地掌握回归分析这一应用广泛的统计方法。

参考文献:

- [1] 国家统计局. 中国统计年鉴 (2015 光盘版)[M]. 北京: 中国科学出版社, 2015.

- [2] 何晓群, 刘文卿. 应用回归分析 [M]. 2 版. 北京: 中国人民大学出版社, 2011.
- [3] 王松桂, 陈敏, 陈立萍. 线性统计模型——线性回归分析与方差分析 [M]. 北京: 高等教育出版社, 1999.
- [4] 王学民. 应用多元分析 [M]. 2 版. 上海: 上海财经大学出版社, 2004.
- [5] 何晓群. 多元统计分析 [M]. 4 版. 上海: 中国人民大学出版社, 2015.
- [6] 薛薇.SPSS 统计分析方法及应用 [M]. 北京: 电子工业出版社, 2004.
- [7] 时立文.SPSS 19.0 统计分析从入门到精通 [M]. 北京: 清华大学出版社, 2012.
- [8] 刘宏杰. 中国财政收入与经济增长之间的定量研究: 1978~2006——基于 VAR 模型的经济计量分析 [J]. 河北经贸大学学报, 2008, 29(6): 5-9.
- [9] 曾五一, 刘小二. 中国财政收入与经济增长关系的实证分析 [J]. 统计与信息论坛, 2009, 24(7): 34-39.
- [10] 李涌波. 平衡财政收支保证国民经济长期稳定发展 [J]. 中国财政, 1987(9): 28-29.

Case of Teaching Study of Applied Regression Analysis Based on Statistical Model of China's Government Revenue

ZHANG Xuanhao

(School of Sciences, Shanghai Polytechnic University, Shanghai 201209, China)

Abstract: The data of macroeconomic indicators of China was collected and a statistical model of government revenue using of multiple linear regression analysis was established. It revealed the major economic factors affecting our government revenue and provided the basis of government economic policy. It provided a teaching case of the main statistics courses "Applied Regression Analysis" and helped students to understand and master the regression analysis better.

Keywords: regression analysis; stepwise regression; government revenue; teaching case