

逐步回归分析法及其应用

游士兵, 严 研

(武汉大学 经济与管理学院, 武汉 430072)

摘 要:逐步回归分析是多元回归分析中的一种方法,在经济研究建模中发挥着重要的作用。文章系统介绍了逐步回归分析,并分析了逐步回归分析在经济研究(建模与预测)中的应用步骤与需要注意的问题。分别应用向前引入法、向后剔除法、逐步回归法分别对中国农村居民收入影响因素模型进行实证分析。

关键词:逐步回归分析;农村居民;收入影响因素

中图分类号:O212

文献标识码:A

文章编号:1002-6487(2017)14-0031-05

0 引言

逐步回归分析是多元回归分析中的一种方法。回归分析是用于研究多个变量之间相互依赖的关系,而逐步回归分析往往用于建立最优或合适的回归模型,从而更加深入地研究变量之间的依赖关系。目前,逐步回归分析被广泛应用于各个学科领域,如医学、气象学、人文科学、经济学等。

经济现象纷繁复杂并多变,经济问题往往需要研究一个(或多个)变量(如一国的产出)与其他变量(如资本、劳动力、人口、技术等)相互间的关系,从而揭示经济现象背后的经济规律,用于解决现实中的经济问题,制定经济政策等。因此,多元回归分析在现代经济学,尤其是其分支学科计量经济学中常常得到应用。随着统计学与计量经济学软件的开发与应用,经济研究者在经济学的实证分析过程中越来越多的采用逐步回归分析的方法来建立多元回归模型。因此,研究逐步回归分析及其在经济学研究中的应用具有理论与实践意义。

目前,国内外对于逐步回归分析的研究主要停留在理论叙述,不少学者提出了修正逐步回归法的新方法,如类逐步回归法、基于单元步的逐步回归法、“宜取回归方程”的逐步回归法等。赵希男(1994)提出并在理论上证明可将逐步回归算法从静态系统拓广到其他系统(动态系统、分布参数等系统)。传统的逐步回归法通常用于静态系统,即自变量为非随机变量时。当自变量为分布型变量时,基于传统的逐步回归法,由于统计量的分布非常复杂,基于统计量的算法难以实现。该文运用泛函分析,按照传统逐步回归法一致思想和相似的步骤,加以拓广,以有效性(剩余相关系数)为标准剔除或是引入变量,从而使得逐步回归法应用范围更广。张华嘉、舒元(1998)提出单元步的概念来修正传统逐步回归法,并以实例证明这种方法的

优势。逐步回归法的自变量个数可能偏小,而向后回归法可能纳入更多的变量,同时带来多重共线性。为了避免多重共线性并包含更多的变量,他们分析向后回归法较逐步回归法而多余的变量,以一个或几个变量为一单元步(当去掉该单元步中所有变量时其余变量显著,而去掉其中某一部分变量,单元步中剩余的变量总存在不显著变量)。确定单元步变量之后,以单元步为整体,确定是否剔除单元步。陈全润、杨翠红(2008)提出了基于残差项绝对值加权和最小准则下的回归估计,以有效性为指标来选择剔除或是删除变量的类逐步回归法。

而经济学家多直接采用这一方法研究各种经济问题,而忽略这种方法可能存在的问题与局限性。而且在具体应用逐步回归分析法时,不同的研究者常采用不同的算法。

逐步回归法作为建立最优线性回归模型的一种方法,在经济研究中也得到广泛的应用,尤其是在经济建模与预测中。因为逐步回归法简单易行,所得的回归方程的变量较少,并保留了影响最显著的重要变量,而且在实践中这种方法也被证明较为有效,预测精确度较高;同时经济变量之间往往存在相互关系,即经济变量可能存在多重共线性,而逐步回归在一定程度上可以修正多重共线性。

本文将全面、系统地阐释逐步回归分析,从经济建模与预测角度进行探讨,并说明逐步回归分析在经济研究中的具体应用、局限性以及可能存在的问题,将中国农村居民收入影响因素模型作为一个经济实例,给予实证说明。

1 逐步回归分析的基本思想

逐步回归分析建立于线性回归模型的基础上,本文将经典的多元线性回归模型为例进行阐释。在建立模型过程中,需要选择合适的变量。如果变量过多,可能会导致预测精度下降,或者有时收集某些变量的数据成本很高,从而不得不放弃某些变量。当然如果影响显著的变量

作者简介:游士兵(1964—),男,湖北天门人,教授,博士生导师,研究方向:国民经济学、统计学、数量经济学。

严 研(1982—),男,湖北仙桃人,博士研究生,研究方向:统计学。

没有包括在模型内,也会影响预测精度。因此,需要进一步解决的问题是建立“最优”的回归方程,这里“最优”的回归方程,通常是指模型中包含所有对因变量 Y 影响显著的自变量,而不包含对因变量 Y 影响不显著的自变量。建立“最优”的回归方程主要有三种方法。一种是计算量很大的全子集法,以某一确定准则来确定最优回归子集;另一种是“最优”子集的变量筛选法,该方法基于偏回归平方和检验逐步引入变量或是逐步剔除变量或是两者结合,包括向前引入法、向后剔除法与逐步回归法(即逐步回归分析);第三种是计算量适中的选择法。

逐步回归法可以认为是向前引入法与向后剔除法的综合。逐步回归法克服了向前引入法与向后剔除法的缺点,吸收两种方法的优点。逐步回归法是以向前引入为主,变量可进可出的变量选取方法。它的基本思想是,当被选入的变量在新变量引入后变得不重要时,可以将其剔除,而被剔除的变量当它在新变量引入后变得重要时,又可以重新选入方程。

2 多重共线性和逐步回归

在经济研究的建模中,由于各个因素之前常常有着各种联系,因此模型常会受到多重共线性的影响。而逐步回归法则可以很好地解决这一问题。

不妨设经典的计量经济学线性回归模型为:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_m X_{im} + \varepsilon_i$$

其中 $E(\varepsilon_i) = \mu$, $\text{Var}(\varepsilon_i) = \sigma^2$, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, i, j = 1, 2, \dots, n$ 。

为了保障最小二乘法下的参数估计量具有良好的性质,除了上述对随机干扰项的假设外,还须对解释变量(自变量)做一些基本假设:解释变量是确定型变量,不是随机变量(解释变量之间不相关)。随机干扰项与解释变量之间不相关。当违背了上述一个或几个基本假设时,则运用最小二乘法估计模型参数将会产生一些问题,即最小二乘法“失效”。

而本文将要讨论的正是多重共线性,它违背了解释变量不是随机变量且互不相关这一假设。因此,必须采取措施来解决多重共线性。而逐步回归法是其中一种方法。

2.1 多重共线性

设经典的线性回归模型为:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_m X_{im} + \varepsilon_i, \text{ 其中 } i = 1, 2, \dots, n_0$$

有一个或多个解释变量相关,则称解释变量存在多重共线性。

即若存在关系式:

$$C_0 + C_1 X_{i1} + C_2 X_{i2} + \dots + C_m X_{im} = 0, \text{ 其中 } i = 1, 2, \dots, n$$

则称 X_1, X_2, \dots, X_m 存在完全(或精确)共线性。

如果存在关系式:

$$C_0 + C_1 X_{i1} + C_2 X_{i2} + \dots + C_m X_{im} + v_i = 0, \text{ 其中 } i = 1, 2, \dots,$$

n, v_i 为随机干扰项,则称 X_1, X_2, \dots, X_m 存在近似共线性(或称交互相关)。

实际经济问题中以近似共线性更为常见。一些基本的经济变量具有相关的共同趋势,如时间序列样本、截面数据等,引入滞后变量以及样本资料限制等都可能导致(近似)多重共线性。

多重共线性将导致很多严重的问题。完全多重共线性将导致最小二乘法不存在参数估计量(由于此时 $(X'X)^{-1}$ 不存在,参数估计量 $(X'X)^{-1}$ 不存在)。而近似共线性将导致参数估计量的方差变大($\text{COV}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, 由于 $|X'X|$ 接近 0, 则 $(X'X)^{-1}$ 的主对角元素很大)。这将进一步导致参数估计量经济含义不合理,以及变量的显著性检验与预测检验失效。

2.2 应用逐步回归法修正多重共线性

逐步回归法用于多重共线性的检验。逐步回归法的基本思想是逐个引入新的变量。考虑是否引入新的变量时,若偏回归平方和变化显著,则可以引入,否则不引入。此时,若偏回归平方和经检验显著,则表明可以认为新变量是独立的解释变量,而不可以由其他解释变量(近似)线性表示,否则说明新变量不独立。

逐步回归法也可以用于修正多重共线性。逐步回归法有两大基本步骤。以偏回归平方和来考虑是否引入新变量,若在一定的显著性水平(α_m)下偏回归平方和较大(或者说变量的显著性检验下 $p < \alpha_m$),则表明可以引入,此时新变量与已选变量不大可能存在相关性。若在一定的显著性水平(α_{out})下偏回归平方和较大(或者说变量的显著性检验下 $p < \alpha_{out}$),则不应剔除该变量,此时该变量应该不是引起多重共线性的变量,否则偏回归平方和应该不显著。换言之,逐步回归法是通过偏回归平方和对变量进行显著性检验考虑是否引入或是剔除变量,而偏回归平方和的检验大体上可以与变量间的多重共线性(即该变量可否用其他变量来近似线性表示)检验相统一。这就是逐步回归法用于检验与修正多重共线性的原因,也是其在经济模型中广泛应用的最主要原因。

下面对上述观点(偏回归平方和的检验大体上可以与变量间的多重共线性,即“该变量可否用其他变量来近似线性表示”检验相统一)进行证明。

在用最小二乘法估计参数前提下,本文假设线性模型为:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (1)$$

若 X_1 和 X_2 近似共线性,即 $X_2 = aX_1 + b + v$, 则代入模型(1),得

$$Y = \beta_0 + b\beta_2 + (\beta_1 + a\beta_2)X_1 + \varepsilon + \beta_2 v \quad (2)$$

令假设线性模型:

$$Y = \alpha_0 + \alpha_1 X_1 + \mu \quad (3)$$

此时模型(3)的估计为:

$$\hat{\alpha}_0 = \hat{\beta}_0 + b\hat{\beta}_2$$

$$\hat{\alpha}_1 = \hat{\beta}_1 + a\hat{\beta}_2$$

此时模型(3)的残差平方和为:

$$Q_3 = \sum_{i=1}^n [y_i - (\hat{\alpha}_0 + \hat{\alpha}_1 x_{i1})]^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + b\hat{\beta}_2 + (\hat{\beta}_1 + a\hat{\beta}_2)x_{i1})]^2$$

而模型(1)的残差平方和为:

$$Q_1 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})]^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + b\hat{\beta}_2 + (\hat{\beta}_1 + a\hat{\beta}_2)x_{i1} + v_i\hat{\beta}_2)]^2$$

比较 Q_1 与 Q_3 , $Q_1 - Q_3 \approx 0$ 。故模型(3)中引入(或是模型(1)中剔除) X_2 后偏回归平方和变化很小,即残差平方和或是回归平方和没有显著变化,因此借助偏回归平方和的系数检验($H_0: \beta_2 = 0$)的结果为接受原假设,即剔除 X_2 , 显然它正是引起多重共线性的变量。这就说明了引起多重共线性的变量往往可以通过偏回归平方和检验而甄别出,即逐步回归法可以应用于检验与修正多重共线性。

2.3 逐步回归分析应用中应注意的问题

建立经济模型是分析研究经济问题的一个有效而常用的途径。将逐步回归法应用于经济建模与预测必须将其置于经济大背景下,把研究对象作为经济问题,而非单纯的统计学问题来考察。应用逐步回归法建立经济模型时,应该基本遵循以下步骤,并设定理论模型为多元线性回归模型:第一,选取合适的变量;第二,样本数据的收集与整理;第三,模型参数估计;第四,应用逐步回归法并结合其他的模型检验,确定最优的线性回归方程;第五,根据上述模型,分析实际问题,得出结论或是提出政策建议等。

逐步回归的基本思想是逐个引入新变量,每引入一个新变量时考虑是否剔除已选变量,直至不再引入新变量。这种方法既保障了方程能保留影响显著的变量,又能够剔除非显著的变量。它以前回归为主,结合向后剔除法,通常可以获得较好地拟合效果,因此,被广泛应用于经济建模与预测。不过,这并不代表逐步回归法就是毫无缺陷的方法。理论上,在变量个数较少时,全子集法是能够获得在某一准则下的最优回归方程,而且计算量也在可以接受的范围,相对于逐步回归法可能更优。而逐步回归法则依赖于人为确定的显著性水平,不同的显著性水平可能得到的结果不一样,而且也很难给出理论上证明其回归方程的最优性。换言之,逐步回归法只是一种实践上较为有效的方法,理论上它没有被证明所得到的回归方程即为最优的回归方程。

考察运用逐步回归法建立最优回归模型的经济论文,能够发现一些问题。归纳总结这些问题,才能更好地运用这一方法解决经济问题。

(1)在经济建模与预测的应用中,对逐步回归的具体操作有很大不同。有些学者使用向前引入法,逐个将变量引入方程;而有些学者使用的本质上是向后剔除法,逐个剔除变量;还有部分学者,做了多次逐步回归,然后以一些标准综合比较结果以选择最优的。这些情况都充分说明了经济研究中本质上采纳的是逐步回归法的基本思想,而不一定与多元统计分析理论当中的逐步回归分析完全一

致,也体现了逐步回归分析在经济建模运用中的灵活性。但是,对于同一经济问题建立模型,向前引入法与向后剔除法、严格的逐步回归法得到的结果是否一致,结果是否最优都是值得思考的问题。

(2)在经济建模中,逐步回归法选取或剔除变量往往同时结合各种检验,包括经济意义检验、统计学检验、计量经济学检验、模型预测检验等。逐步回归法虽然一定程度上避免了多重共线性,但是其他问题(如随机干扰项的异方差性、序列相关性、随机解释变量问题)却无法全部得到解决,因此逐步回归法需要经过经济意义检验、统计学检验、计量经济学检验等多重检验。经济模型一个重要的应用在于经济预测,模型的预测精确度是考察模型优劣性的一个重要指标,所以逐步回归法需要通过模型预测检验。

(3)在经济建模的整个过程中,一定要考虑解释变量间、解释变量与被解释变量间的经济关系,或者考察经济环境,不能忽视经济理论与经济行为规律,毕竟研究目的不仅是单纯地处理分析样本数据,而是想要通过揭示变量的客观的数据关系,从而进一步揭示变量之间的规律。

3 运用实例

根据理论和经验分析,农村居民收入通常由农业规模、农产品价格、劳动力分配、技术、政府投入等因素决定。依据经济理论和客观事实,考察数据的可得性,从而选取影响农村居民收入的变量,而且符号的正负将是经济意义检验的依据,符号正负性的判断则依赖于经济理论。因此可以认为影响中国农村居民收入(Y)的主要因素有农业总产值(X_1)、农业机械总动力(X_2)、从事农业生产的劳动力比例(X_3)、农产品生产价格总指数(X_4)、支农支出占国家财政支出的比重(X_5)。

建立中国农村居民收入函数影响因素模型为:

$$Y = AX_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} X_4^{\beta_4} X_5^{\beta_5}$$

可用如下线性回归模型进行估计:

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3 + \beta_4 \ln x_4 + \beta_5 \ln x_5$$

3.1 多重共线性的检验

根据《中国统计年鉴 2009,2013》、《中国农业统计年鉴 2005,2012》、《中国农村统计年鉴 1990,2008》获得 1983—2012 年的农村居民纯收入、农林牧副渔总产值、农业机械总动力、从事农业生产的劳动力人口比例、农产品生产价格总指数、支农支出占国家财政支出的比重的数据资料,制成表 1(见下页)。

根据样本数据(表 1)以及 Eviews 软件,检验自变量之间是否存在多重共线性。

用 OLS(最小二乘法)估计 $\ln Y$ 与 $\ln X_1$ 、 $\ln X_2$ 、 $\ln X_3$ 、 $\ln X_4$ 、 $\ln X_5$ 的回归方程参数。得到:

$$\ln y = 5.932869 + 0.413 \ln X_1 + 0.444 \ln X_2 - 1.721 \ln X_3 - 0.078 \ln X_4 - 0.078 \ln X_5$$

$$t \text{ 值: } 1.4426 \quad 6.959 \quad 2.553 \quad -2.879 \quad -0.4606 \quad -0.4405$$

表1 中国农村居民收入相关资料

年份	农村 居民 纯收入 (元/人)	农林牧 副渔 总产值 (亿元)	农业 机械 总动力 (万千瓦)	从事农业生产 劳动力人口占 总从业人口的 比重(%)	农产品 生产价格 总指数 (上年=100)	支农支出占 国家财政支出 的比重(%)
1983	309.8	2508.2	18022	67.08373	104.4	9.4
1984	355.3	2815.6	19497	64.04548	104	8.3
1985	397.6	2912.2	20913	62.41854	108.6	7.7
1986	423.8	3010.7	22950	60.94536	106.4	8.4
1987	462.6	3185.1	24836	59.98712	112	8.7
1988	544.9	3309.7	26575	59.35326	123	8.6
1989	601.5	3412.8	28067	60.04988	115	9.4
1990	686.3	7662.1	28708	60.09977	97.4	10
1991	708.6	8157	29389	59.69981	98	10.3
1992	784	9084.7	30308	58.50012	103.4	10.1
1993	921.6	10995.5	31817	56.40043	113.4	9.5
1994	1221	15750.5	33802	54.2999	139.9	9.2
1995	1577.7	20340.9	36118	52.2001	119.9	8.4
1996	1926.1	22353.7	38547	50.50036	104.2	8.8
1997	2090.1	23788.4	42016	49.89974	95.5	8.3
1998	2162	24541.86	45208	49.79968	92	10.7
1999	2210.3	24519.06	48996	50.09945	87.8	8.2
2000	2253.4	24915.8	52574	50.00069	96.4	7.8
2001	2366.4	26179.6	55172.1	50.00068	103.1	7.7
2002	2475.6	27390.75359	57929.9	50	99.7	7.2
2003	2622.2	29691.8	60386.5	49.09985	104.4	7.2
2004	2936.4	36238.98976	64027.9	46.90027	113.1	8.2
2005	3254.9	39450.88734	68397.8	44.80053	101.4	7.2
2006	3587	40810.82661	72522.1	42.61911	101.2	7.9
2007	4140.4	4889.3	76589.6	40.80004	118.5	6.8
2008	4760.6	58002.2	82190.4	39.59994	114.1	7.3
2009	5153.2	60361.0	87496.1	38.10004	97.6	8.8
2010	5919.0	69319.8	92780.5	36.69995	110.9	9.0
2011	6977.3	81303.9	97734.7	34.79979	116.5	9.1
2012	7916.6	89453.0	102559.0	36.60059	102.7	9.5

数据来源:《中国统计年鉴2009,2013》、《中国农业统计年鉴2005,2012》、《中国农村统计年鉴1990,2008》。

p 值:0.1663 0.0000 0.0200 0.0100 0.6506 0.6648

方程的 R^2 (调整的可决系数) 为 0.9928, F 值为 496.596 (对应的 p 值为 0), 这说明方程的总体线性关系显著, 但是观察各解释变量的 t 统计量值 (简称为 t 值) 及其对应的 p 值, 可以发现 C 、 $\ln X_4$ 、 $\ln X_5$ 经检验其系数不显著, 因此说明解释变量间应该存在多重共线性。

3.2 最优回归方程的建立

3.2.1 向前引入法

向前引入法是从常数项开始逐个引入变量的方法。利用 Eviews 软件, 使用该方法, 得到结果制成表 2。

表2 向前引入法 Eviews 结果

	$\ln Y$	$\ln X_1$	$\ln X_2$	$\ln X_3$	$\ln X_4$	$\ln X_5$	调整的 R^2
第一步 引入变量 $\ln X_1$	参数估计值 t 值 p 值	0.755 242.33 0.0000					0.967
第二步 引入变量 $\ln X_2$	参数估计值 t 值 p 值	0.777 16.77 0.0000	-0.021 -0.501 0.0000				0.966
第三步 引入变量 $\ln X_3$	参数估计值 t 值 p 值	0.420 8.176 0.0000	0.662 7.391 0.0000	-0.959 -7.851 0.0000			0.991

至此不再引入变量 (若进一步引入, 则引入变量 $\ln X_4$ 发现不合理, 向前引入法到此结束)。

由表 2 可知, 向前引入法的最终回归模型为:

$$\ln Y = 0.420 \ln X_1 + 0.662 \ln X_2 - 0.959 \ln X_3$$

3.2.2 向后剔除法

向后剔除法是从全模型 (即 $\ln Y$ 与 $\ln X_1$ 、 $\ln X_2$ 、 $\ln X_3$ 、 $\ln X_4$ 、 $\ln X_5$ 建立的回归模型) 开始逐个剔除变量的方法。利用 Eviews 软件, 使用该方法, 得到结果制成表 3。

表3 向后剔除法 Eviews 结果

	$\ln Y$	$\ln X_1$	$\ln X_2$	$\ln X_3$	$\ln X_4$	$\ln X_5$	调整的 R^2
第一步	参数估计值 t 值 p 值	0.429 7.152 0.0000	0.646 6.102 0.0000	-0.923 -3.953 0.0000	0.015 0.091 0.9282	-0.057 -0.315 0.7562	0.991
第二步 剔除变量 X_5	参数估计值 t 值 p 值	0.420 7.982 0.0000	0.659 6.910 0.0000	-0.972 -5.650 0.0000	0.017 0.110 0.9134		0.991
第三步 剔除变量 X_4	参数估计值 t 值 p 值	0.420 8.176 0.0000	0.662 7.391 0.0000	-0.959 -7.851 0.0000			0.991

其最终回归模型与向前引入法相同。在第三步剔除变 $\ln X_4$ 之后, 并没有进一步剔除变量。若进一步剔除变量 $\ln X_3$, 则模型的调整系数反而有所下降, 而且从经济理论上而言, 参与农业的劳动力人口比例与现代化的农业规模及农业机械化水平线性关系并不明显, 应加以保留。

由表 3 可知, 向后剔除法的最终回归模型为:

$$\ln Y = 0.420 \ln X_1 + 0.662 \ln X_2 - 0.959 \ln X_3$$

3.2.3 逐步回归法

逐步回归法逐个引入新变量, 每引入一个新变量时考虑已进入模型的变量是否可以被剔除, 直至不再引入新的变量。

逐步回归过程如表 4 所示。

表4 逐步回归法的 Eviews 结果

步骤	结果
第一步	引入变量 $\ln X_1$
第二步	引入变量 $\ln X_2$
第三步	考虑可否剔除变量 $\ln X_1$ (否)
第四步	引入变量 $\ln X_3$
第五步	考虑可否剔除变量 $\ln X_5$ (否), 再考虑可否剔除变量 $\ln X_4$ (否)
第六步	不再引入新的变量

最终可以得到回归方程:

$$\ln Y = 0.420 \ln X_1 + 0.662 \ln X_2 - 0.959 \ln X_3$$

t 值: 8.176 7.391 -7.851

p 值: 0.0000 0.0000 0.0000

3.3 分析

在中国农村居民收入影响因素模型函数的这个实例中, 向前引入法与向后剔除法、逐步回归法得到的是同一个回归模型。该模型的 R^2 (调整的可决系数) 为 0.991, 说明方程总体的线性显著。而且各解释变量的参数都拒绝了原假设, 即各解释变量影响显著, 模型的模拟效果较好。如前所述, 逐步回归法所建的模型必须通过各种检验, 如异方差检验、序列相关性检验、预测检验等。

该实例分析进一步说明, 在经济建模当中, 逐步回归法、向前引入法、向后剔除法简易可行, 克服了多重共线性, 得到的模型也比较优良。虽然这个实例中三种方法得

到的结果完全一致,但这可能是由于该实例中变量比较少,而且变量间的相关性相对简单导致的,并不是所有的案例都会出现这种情况。不过,这也验证了在经济建模当中,这三种方法都可以使用,关键是看哪种方法得到的模型最优、最具可操作性;而通常而言,逐步回归法从模型的优良性的角度考虑,应该优于向前引入法和向后剔除法。

从实例中可以看出,中国农村居民收入主要受农业规模、农业机械化水平、从事农业生产的劳动力人口比例这三个因素影响较大,而受农产品价格水平及政府的转移支付影响较少。现代化农业的生产已经不再依赖于单纯劳动力的增加,进行科学的大规模经营、提高农业投入的机械化水平对于提高农业收入的作用更加明显。而且在我国,农民的收入增加越来越多的来自于非农收入,如外出打工的工资性收入、农业资源租赁的租金收入等,因此农村居民收入与从事农业生产的劳动力人口比例呈现负相关。这与经济理论和我国的实际情况相一致,也进一步说明在提高农村居民收入的实践中应该加大对农业的技术和资本投入,改进农业生产方式和方法,改良农业管理方式,促进农村劳动力合理转移。分析结果也说明从事农业的劳动力人口的增加不再是提高农村居民的主要途径,只要补偿足够技术、资本投入,即使劳动力减少也不一定会造成农业收入的减少。而且农业收入已不再是我国农村居民的主要收入来源,无论是政策制定还是经济研究都必须了解考虑这一事实,才能有效地提高我国农村居民的收入水平。

4 结束语

经济问题的分析不同于单纯的统计学研究,在经济研究建模中,不仅要考虑统计学检验,更要把实际经济意义等因素纳入考虑和检验范畴。本文在阐述逐步回归分析的统计理论基础,总结其在经济建模分析中的使用方法及步骤,研究其在经济问题分析中的应用方法。

逐步回归分析是一种被广泛应用于各个领域并极具价值的多元线性回归中的自变量选择方法。它的基本思想在于逐个引入显著性变量,同时剔除不显著性变量,以期获得最优的回归模型。在经济建模与预测中,它可以避

免多重共线性,获得变量较少、相对简单的模型,在模型中通常仅保留住少数的、影响显著的变量,因此估计的模型较优,具有较好的模拟预测效果。虽然在实践中获得较好效果,但是,逐步回归分析本身存在不足,其回归结果在理论上的最优性证明仍有待于进一步的探讨。

参考文献:

- [1]高惠璇.应用多元统计分析[M].北京:北京大学出版社,2005.
- [2]李子奈,潘文卿.计量经济学[M].北京:高等教育出版社,2005.
- [3][美]S.韦斯伯格.应用线性回归[M].北京:中国统计出版社,1998.
- [4]王松桂.线性统计模型[M].北京:高等教育出版社,1999.
- [5][美]约翰逊,威克恩.实用多元统计分析[M].北京:清华大学出版社,2008.
- [6]李松臣,张世英.基于逐步回归法的人口出生率影响因素分析[J].统计与决策,2008,(4).
- [7]佟玉权.我国出境旅游市场影响因素的逐步回归分析[J].市场论坛,2008,(11).
- [8]王红,童恒庆,魏平.进出口贸易额预测的逐步回归建模研究[J].统计与决策,2006,(1).
- [9]赵希男.逐步回归方法的拓广研究[J].应用数学,1994,(4).
- [10]陈全润,杨翠红.“类逐步回归”变量筛选法及其在农民收入预测中的运用[J].中山大学学报,2008,(11).
- [11]郝勇.对影响上海人均寿命与综合环境因素的逐步回归分析研究[J].经济师,2003,(11).
- [12]赵云河.逐步回归法在税收预测中的应用研究[J].云南财贸学院学报,2007,(1).
- [13]周丹.中国各地区产业发展影响因素的逐步回归分析[J].商场现代化,2009,(1).
- [14]李晓虹,杨有.一种基于逐步回归的国家财政收入模型[J].统计与决策,2006,(12).
- [15]赵芳林,牛雅莉.应用逐步回归分析的方法建立教学评估模型[J].科技信息,2008,(11).
- [16]杨巍,张莉莉.逐步回归分析在经济林产品需求预测中的应用[J].林业经济,2009,(8).
- [17]周文芳,李氏.逐步回归分析法的一点不足之处[J].西北水电,2004,(4).
- [18]杨有,李晓虹.多重共线性的逐步回归检验分析[J].重庆三峡学院学报,2003,(3).
- [19]张华嘉,舒元.逐步回归分析的拓展[J].中山大学学报:自然科学版,1998,(5).

(责任编辑/亦 民)

Stepwise Regression Analysis and Its Application

You Shibing, Yan Yan

(School of Economics and Management, Wuhan University, Wuhan 430072, China)

Abstract: Stepwise regression is a method of multivariate regression analysis, and it plays an indispensable role in economic research and modeling. This paper systematically introduces the stepwise regression method, and analyzes its application procedure and problems needing our attention in economic research (modeling and forecasting). And then the paper respectively uses forward selection, backward selection and stepwise regression method to empirically analyze the Chinese rural income model.

Key words: stepwise regression; rural residents; income determinants