

# 关于计量经济学模型随机扰动项的讨论<sup>\*</sup>

李子奈 李鲲鹏

**内容提要:** 论文指出了计量经济学模型中源生的随机扰动项和衍生的随机误差项之间的区别;讨论或证明了,如果模型存在总体设定误差和变量观测误差,在很多情况下将导致随机误差项对 Gauss 假设以及正态性假设的违背。

**关键词:** 计量经济学模型;随机扰动项;模型设定误差;变量观测误差

**中图分类号:** 0212 **文献标识码:** A **文章编号:** 1002-4565(2009)02-0062-06

## Discussion about the Stochastic Disturbance Term of Econometric Models

Li Zina & Li Kunpeng

**Abstract:** The paper highlights the distinguish between the original stochastic disturbance term and the derived stochastic error term, suggests that if the relationship error of model or the measurement error of variables exist in an econometric model, in the most of cases the stochastic error term will not follow the normal distribution assumption and some other Gauss Assumptions.

**Key words:** econometric model; stochastic disturbance term; relationship error of model; measurement error of variables

### 一、随机扰动项的源生性

随机扰动项在计量经济学模型中占据特别重要的地位,也是计量经济学模型区别于其它经济数学模型的主要特征。李子奈(2008)将计量经济学应用研究的总体模型设定归纳为:将影响被解释变量的因素集进行有效分解,按照与被解释变量关联关系的恒常性和显著性两个维度,分解为显著的恒常性因素集、显著的偶然性因素集和无数单独影响可以忽略的非显著因素集;所有显著的恒常性因素作为解释变量;显著的偶然性因素对被解释变量的影响,则通过对数据进行奇异点诊断后采用技术手段予以消除;而无数非显著因素对被解释变量的影响则用一个随机扰动项(stochastic disturbance term)表示,并引入模型。W. H. Greene(2000)指出,没有什么模型可以期望处理经济现实的无数偶然因素,因此在经验模型中纳入随机因素是必须的,被解释变量的观察值不仅要归于已经清楚了解的变量,也要考虑来自人们并不清楚了解的偶然性和无数微弱因素的影响。于是,对于单方程计量经济学模型,总体回归模型的一般形式为:

$$y_i = f(X_i, ) + \mu_i \quad i = 1, 2, \dots, n$$

其中, $y$ 称为被解释变量; $X$ 称为被解释变量,可能包括多个变量; $\beta$ 为反映解释变量和被解释变量之间关系的参数,一般称为结构参数; $n$ 表示随机抽取的样本数量; $\mu$ 即为随机扰动项。

显然,这样界定的随机扰动项具有源生性。从经济学意义上, $X$ 包含了所有对于 $y$ 具有显著影响的因素, $f(X_i, )$ 表达了这些因素与 $y_i$ 之间的动力学关系,生成了 $y_i$ 的条件期望值。但是,无数不显著因素的影响对于生成 $y_i$ 的观测值是不可忽略的,“不显著”不是“没有影响”。从统计学意义上,在 $X_i$ 的条件下重复抽样,无数不显著因素对 $y_i$ 的均值没有影响,但是在一次抽样中,它们对 $y_i$ 的个值的影响是不可忽略的。

在基于随机抽样的截面数据的经典计量经济学模型中,这个源生的随机扰动项 $\mu$ 由大数定理保证其满足 Gauss 假设,由中心极限定理可以证明其服从正态分布。于是,建立在 Gauss 假设和正态分布假设基础上的统计推断具有可靠性。

<sup>\*</sup> 本文获国家社会科学基金重点项目“计量经济学模型方法论基础研究”(08AJY001)的资助。

## 二、随机误差项的衍生性

正如 W. H. Greene (2000) 指出的,在确定性模型中引入随机扰动,并不是为了掩盖确定性模型的不足之处。因此,如果所谓的未被解释的随机扰动并不是真正的不能被解释的因素,模型就是不适当的。牢记这一点对计量经济学是非常重要的。统计推断的理论不像确定性理论那样,会被仅仅一个不符实际的观察否定。引入随机要素后,对预期结果的描述从确切的表述转化为可能性的描述,除非有占优证据(占优本身则是很难清楚界定的),很难否定随机模型。当然,如果未被解释的随机扰动并不是真正的不能被解释的因素,即使这样的模型难以被否定,也是建模者自欺欺人。Greene 的担忧在很多情况下成了现实:在很多计量分析中,随机误差项成了确定性模型不足之处的遮羞布。在大部分计量经济学教科书中,例如 D. N. Gujarati (2003),在第一次引入随机扰动项的概念时,都将它定义为“被解释变量观测值与它的期望值之间的离差”,即

$$\mu_i = y_i - E(y | X_i)$$

用一个平衡式代替定义式,并且将随机扰动项(stochastic disturbance term)与随机误差项(stochastic error term)等同。一个“原生”的随机扰动项变成了一个“衍生”的误差。而且在解释它的具体内容时,一般都在“无数非显著因素对被解释变量的影响”之外,加上诸如“变量观测值的观测误差的影响”、“模型关系的设定误差的影响”等。国内出版的计量经济学教科书也是这样。

将“原生”的随机扰动变成“衍生”的误差,有许多理由可以为此辩解。如果不对数据生成过程的理论结构做出假定,即进行总体模型设定,就无从开始模型研究。但是,相对于物理学,经济学家对经济现实所知较少,总体模型被研究者有限的知识所确定,因此误差在所难免,只能将总体模型方程的误差项设定为衍生性的。

问题在于,关于随机扰动项的 Gauss 假设以及正态性假设,都是基于“原生”的随机扰动而成立的。如果存在模型设定误差、变量观测误差等确定性误差,并将它们归入“随机误差项”,那么它是否满足这些基本假设?如果不满足,进而进行的统计推断就缺少了基础。对于这个问题,一般的计量经济学教科书没有进行讨论,有的只是进行简单的说明,例

如,在 W. H. Greene (2000) 的教科书中就有以下的说明:鉴于我们对随机误差来源的描述,中心极限定理的条件一般都成立,至少近似成立,因此,正态性假定在多数情况下也都是合理的。

## 三、包含模型设定误差的随机误差项

### (一) 存在模型关系误差的情形

所谓“关系误差”,指模型的总体设定不能准确反映所研究的经济系统中的动力学关系。这样动力学设定之间的偏差将不可避免的进入随机扰动项,影响其正态性。

对于一个回归的模型,假定真实的数据生成过程是:

$$y_i = f(X_i, \beta) + \mu_i \quad i = 1, 2, \dots, n$$

这里的随机扰动项  $\mu_i$  服从经典假设,即独立同分布、高斯正态。假定模型被错误的设定为:

$$y_i = g(X_i, \gamma) + v_i \quad i = 1, 2, \dots, n$$

简单的数学变换后可得:

$$v_i = \mu_i + (f(X_i, \beta) - g(X_i, \gamma))$$

显然这里的分布与  $(f(X_i, \beta) - g(X_i, \gamma))$  有密切的关系,这里我们分两种情况讨论:

第一种情况:  $X_i$  是非随机的。这时关键是如何看待  $\gamma$ , 由于  $\gamma$  是在模型错误设定下的参数,因此没有很好的定义。不过对每一个给定  $\gamma$ ,  $(f(X_i, \beta) - g(X_i, \gamma))$  是确定性变量的函数之差,因此错误模型中的误差  $v_i$  是一个正态随机数  $\mu_i$  与非随机数  $(f(X_i, \beta) - g(X_i, \gamma))$  之和,因此仍然是正态的。考虑到  $\gamma$  的任意性,因此  $v_i$  将是正态的。

第二种情况:  $X_i$  是随机的。这种情况下,  $(f(X_i, \beta) - g(X_i, \gamma))$  将必然是一个随机数,而且这个随机数受到了三个因素的影响: (a) 模型的正确动力学关系  $f$ ; (b) 模型被误设的动力学关系  $g$ ; (c) 随机回归元  $X_i$  的分布。注意到:

$$v_i - \mu_i = (f(X_i, \beta) - g(X_i, \gamma))$$

因此误差  $v_i$  是一个正态随机数的充要条件是  $(f(X_i, \beta) - g(X_i, \gamma))$  是正态的。而在上面提到的三个因素的作用下,  $(f(X_i, \beta) - g(X_i, \gamma))$  的正态性即使在大量样本下,也不能为任何数学定理所保证。

### (二) 存在遗漏显著变量的情形

可以通过一个例子来说明。假如模型的正确设定应该为:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \mu_i \quad \mu_i \sim i. i. d. N(0, \sigma^2) \quad i = 1, 2, \dots, n$$

即被解释变量由两个解释变量来解释。如果将模型错误设定为:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \mu_i \quad i = 1, 2, \dots, n$$

那么,该模型的随机误差项为

$$\mu_i = \mu_i + \beta_2 x_{2i}$$

其中  $\mu_i \sim i. i. d. N(0, \sigma^2)$ 。显然,该随机误差项的分布不仅受到  $\mu_i$  还受到  $x_{2i}$  的影响,如果  $x_{2i}$  是正态零均值,独立同分布的,则  $\mu_i$  仍然满足零均值、正态独立同分布的性质;如果  $x_{2i}$  是正态非零均值,独立同分布的,则  $\mu_i$  将不再具有非零均值,但具有正态独立同分布性质;如果  $x_{2i}$  是正态非零均值,同分布但不独立,则  $\mu_i$  将是正态非零均值,同分布但不独立;如果  $x_{2i}$  是非正态的,则  $\mu_i$  也将不再具有正态性。

### (三) 存在函数变换的情形

通过函数变换将一些简单的非线性模型转换为线性模型,是计量经济学模型估计中常用的,但是,人们对于变换后模型随机扰动项的性质却缺少讨论。需要强调的是,这种变换与前面讨论的模型误设有所不同,这里以常用的 Cobb - Douglas 生产函数模型为例来予以说明,假定用于估计的模型形式为:

$$\ln y_i = \beta_0 + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} + \beta_3 \ln x_{3i} + \mu_i \quad i = 1, 2, \dots, n$$

虽然  $x_1, x_2, x_3$  已经包括了所有对产出量  $y$  有显著影响的投入要素,我们仍然不能对模型随机扰动项

的分布给出明确的界定,因为它不是源生的,而是由下列模型

$$y_i = Ax_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} \mu_i \quad i = 1, 2, \dots, n$$

经过对数变换后“衍生”得到的,它的分布不仅取决于原模型随机扰动项的分布,而且取决于用于变换的函数。如果原模型  $\mu_i \sim i. i. d. N(0, \sigma^2)$  中,那么显然  $\ln \mu_i$  将不再服从正态分布。如果原模型是如下形式:

$$y_i = Ax_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} + \mu_i \quad i = 1, 2, \dots, n$$

其中  $\mu_i \sim i. i. d. N(0, \sigma^2)$ 。该模型虽然具有更合理的经济学解释,但是却使得经过对数变换得到的线性模型的随机扰动项更加复杂。

以上分别讨论了模型设定误差的几种形式对模型随机扰动项的影响,表明这些误差的存在,可能导

致随机扰动项对 Gauss 假设以及正态性假设的违背,随之进行的模型估计及统计推断的基础将受到损害。

## 四、包含观测误差的随机误差项

现在讨论观测值的观测误差。所谓“观测误差”,指变量的样本观测值不能准确反映变量的实际状态。观测误差是普遍存在的,正如 W. H. Greene (2000) 指出的,在理论上确定变量之间的关系并不难,但要得到这些变量的准确度量则完全是另外一回事。例如,合理度量利润、利率、资本存量或资本存量提供的服务流量的困难,是经验文献中反复出现的主题;在极端情况下,甚至找不到理论变量的可观测的对应物。所以,观测误差构成了“衍生”的随机误差的主要内容。

为了方便,下面分别讨论被解释变量的观测误差和解释变量的观测误差存在的情况下,随机误差项是否满足关于扰动项的 Gauss 假设和正态性假设。

### (一) 存在被解释变量观测误差的情形

假定没有观测误差时,第  $i$  个被抽样的值是  $y_i^n$ , 有观测误差时,第  $i$  个被抽样的值是  $y_i^e$ , 根据以上假设,应该有  $y_i^e = y_i^n + e_i$ , 这里  $e_i \sim i. i. d(0, \sigma_e^2)$  是观测误差。进一步,没有观测误差的回归方程是:

$$y_i^n = X_i \beta \quad \mu_i \sim i. i. d. N(0, \sigma^2) \quad i = 1, 2, \dots, n$$

将  $y_i^e = y_i^n + e_i$  代入得到:

$$y_i^e = X_i \beta + (\mu_i + e_i)$$

如果  $e_i$  是正态的,则  $(\mu_i + e_i)$  也应该是正态的,经典假设不被破坏,基于正态分布的统计推断仍然成立。如果  $e_i$  是非正态的,显然  $(\mu_i + e_i)$  就不可能再是正态分布,此时基于中心极限定理给出的所有的统计推断都是近似成立。

这里需要说明:为什么要将  $\mu_i$  和  $e_i$  区别看待? 这是因为  $\mu_i$  和  $e_i$  的本质是不一样的,  $\mu_i$  是无数的无法辨别地非显著的影响的总和,我们无法对  $\mu_i$  的来源做一个清晰的界定。但是  $e_i$  则不同,它的来源清晰明确,就是抽样时的测量误差,它的影响也是显著的。如果从林德贝格—费勒中心极限定理的角度来看待  $(\mu_i + e_i)$ , 则上面的分析等同于这样的论断,林德贝格条件要求每一个误差因子要“一致地”小,因而其对总和的极限分布不产生影响,但是  $e_i$  这个

因子不具有这一特征,破坏了林德贝格条件,正态性不再成立。

众所周知,如果扰动项不服从正态分布,统计量的分布在渐近意义下仍然是正态的。基于此,我们很自然地要问,上面讨论的问题在实际中多大程度上是重要的。下面我们通过计算机模拟来回答这个问题。

我们模拟三种测量误差分布对检验结果的影响,这三种分布是贝努利分布 (bernoullian distribution)、均匀分布 (uniform distribution)、正态分布。为了使三种分布对结果的影响具有可比性,我们使计算机上生成的三种误差具有相同的均值(都是 0)和方差。引入标准差比参数  $\sigma = \sigma_e / \sigma_{\varepsilon}$ 。具体来说,我们设定的贝努利分布是:  $P\{e_i = 1/3\} = 0.9, P\{e_i = -2/3\} = 0.1$ ;均匀分布是:  $e_i = 2\sqrt{3}(\varepsilon_i - 0.5)$ , 这里  $\varepsilon_i$  是一个服从  $[0, 1]$  均匀分布的随机数;正态分布是:  $e_i = \sigma \varepsilon_i$ , 其中  $\varepsilon_i$  是一个标准正态随机数。

在模拟试验中,真实的数据生成过程 (data generating process) 是:

$$y_i^n = 1 + 1.5x_i + \mu_i$$

这里  $\mu_i$  是标准正态随机数,即  $\mu_i \sim N(0, 1)$ 。回归元  $x_i$  可能是随机的也可能是非随机的,两种情形我们都予以考察,如果  $x_i$  是非随机的,我们设定  $x_i = i$ ;如果  $x_i$  是随机的,设定  $x_i = i + \varepsilon_i$ , 这里  $\varepsilon_i$  也是一个标准正态随机数,且与  $\mu$  和  $\varepsilon$  各期都独立。通过生成  $x_i$  和  $\mu_i$  计算出  $y_i^n$ ,再通过前面的误差  $e_i$  生成程序,计算出  $y_i^e (y_i^e = y_i^n + e_i)$ 。据此进行  $y_i^e$  对 1 和  $x_i$  的回归,计算出  $x_i$  的  $t$  统计量。将上述方法重复 10000 次,找出 5% 的两侧分位点。

上述模拟中,标准差比  $\sigma$  和样本数量  $n$  是允许变动的两个参数,我们让  $\sigma$  的取值是  $\sigma = 0.1, \sigma = 1, \sigma = 20$ ,这样可以考察不同的测量误差大小对正态性的破坏程度。当然样本的数量大小是一个需要关注的参数,随着样本的增大,统计量的质量将会得到改进,因此需要予以控制。

计算机模拟结果报告如下:

从上面的模拟中我们发现如下的结果:

1. 就伯努利分布而言,如果标准差比值  $\sigma$  偏低 ( $\sigma = 0.1$ ),即使在小样本下 ( $n = 10$ ),实际的  $t$  值分布与通常的  $t$  值分布差距也不是很大(实际的  $t$  值分布的 5% 双侧分位点是 2.317,通常的  $t$  值分布的

标准差比 样本数 $n$	回归元是确定的			回归元是随机的			真实 $t$ 值
	伯努利	均匀	正态	伯努利	均匀	正态	
$n = 10 \quad \sigma = 0.1$	2.317	2.319	2.301	2.358	2.340	2.347	2.306
$n = 10 \quad \sigma = 1$	2.261	2.301	2.296	2.259	2.343	2.305	2.306
$n = 10 \quad \sigma = 20$	2.048	2.319	2.311	2.186	2.316	2.313	2.306
$n = 20 \quad \sigma = 0.1$	2.095	2.111	2.102	2.102	2.115	2.088	2.101
$n = 20 \quad \sigma = 1$	2.084	2.128	2.094	2.121	2.124	2.098	2.101
$n = 20 \quad \sigma = 20$	1.956	2.127	2.116	1.944	2.121	2.115	2.101
$n = 30 \quad \sigma = 0.1$	2.023	2.054	2.050	2.037	2.038	2.037	2.048
$n = 30 \quad \sigma = 1$	2.052	2.105	2.069	2.031	2.024	2.037	2.048
$n = 30 \quad \sigma = 20$	1.943	2.037	2.058	1.975	2.014	2.059	2.048

5% 双侧分位点 2.306),使用通常的  $t$  值分布以及 5% 的双侧分位点,不会对结果造成严重的影响;但是如果标准差比值  $\sigma$  较高 ( $\sigma = 20$ ),在小样本下,实际的  $t$  值分布与通常的  $t$  值分布差距显著,使用通常的  $t$  值分布以及 5% 的双侧分位点,会使检验的显著性水平产生较大的扭曲(实际的  $t$  值分布的 5% 双侧分位点是 2.048,通常的  $t$  值分布的 5% 双侧分位点 2.306)。这一扭曲程度随着样本的增大得到明显的改善,在一个偏小样本下 ( $n = 30$ ),这一扭曲程度已经不到 10% 了。

2. 就均匀分布而言,表现比伯努利分布要好一些,综合各种情形的结果来看,  $t$  值的偏误都在 1.5% 以内。

这里有一点需要指明,就是误差是均匀分布的结果与误差是正态分布的结果在我们的报告中优劣难以分辨。这是因为在计算机上生成的正态随机数也不是标准的正态随机数,而是伪正态随机数,导致了正态情形下模拟的  $t$  值与正确的  $t$  值有一定的偏差而且有时偏差甚至很大的结果。同时这也可以解释上面的结果:既然均匀分布不是正态的,而伪正态分布也不是正态的,那就很难比较它们之间孰好孰坏了。

总结上面模拟试验的结果。一般而言,如果因为测量误差的原因导致误差项正态性假设受到了一定程度的破坏,即使在中等偏小的样本情况下,假定正态性仍然是无害的。但是如果测量误差相对于回归扰动项的标准差特别的大、样本长度又特别的小、误差的形式相对较为“病态”,这时仍然假定误差项是正态的假设会导致检验显著性水平产生一定程度的扭曲。

需要说明,在我们设定的回归中,  $x$  前系数的收敛速度不再是  $\sqrt{n}$  而是  $n\sqrt{n}$ 。

## (二) 存在解释变量观测误差的情形

我们将解释变量向量第  $i$  次抽样分为没有观测误差  $X_{1i}$  和有观测误差  $X_{2i}$  两类,假定不存在观测误差的抽样值是  $X_i^n = (X_{1i}, X_{2i}^n)$ ,由于解释变量  $X_{2i}$  存在观测误差,于是引入观测误差向量  $\varepsilon_i$ ,则实际存在观测误差的抽样值  $X_i^e = (X_{1i}, X_{2i}^e)$ ,据此应该有  $X_{2i}^e = X_{2i}^n + \varepsilon_i$ ,这里  $\varepsilon_i \sim i. i. d(0, \sigma^2)$  是观测误差,与  $\mu$  的各期独立。

这里进一步假定回归方程形式  $f(X_i^n)$  已知,且  $f(X_i^n)$  充分光滑,性质良好。在没有误差的情况下,回归方程是:

$$y_i = f(X_i^n) + \mu_i = f(X_{1i}, X_{2i}^n) + \mu_i$$

将  $X_{2i}^e = X_{2i}^n + \varepsilon_i$  代入并将其泰勒展开:

$$y_i = f(X_{1i}, X_{2i}^e - \varepsilon_i) + \mu_i = \mu_i + f(X_{1i}, X_{2i}^e) + \frac{\partial f(X_{1i}, X_{2i}^e)}{\partial X_{2i}^e} (-\varepsilon_i) + \frac{1}{2} \frac{\partial^2 f(X_{1i}, X_{2i}^e)}{\partial X_{2i}^e \partial X_{2i}^e} (-\varepsilon_i)^2 + \dots$$

这样回归方程的误差项是:

$$\mu_i + \frac{\partial f(X_{1i}, X_{2i}^e)}{\partial X_{2i}^e} (-\varepsilon_i) + \frac{1}{2} \frac{\partial^2 f(X_{1i}, X_{2i}^e)}{\partial X_{2i}^e \partial X_{2i}^e} (-\varepsilon_i)^2 + \dots$$

上面误差项的分布形态依赖于  $f(x)$  和  $\varepsilon_i$ ,当  $\varepsilon_i$  是非正态分布的情况下,线性的  $f(x)$  就足以导致误差项的非正态分布;而即使  $\varepsilon_i$  是正态的情况下,一个二次型的  $f(x)$  也会使得误差项是由一个正态分布和一个类卡方分布所组成,正态性的假设同样受到破坏。正态性的假设要能够成立只有在极其严格的条件下,也就是  $f(x)$  是线性的,且误差项是正态分布的情况下才会有出现,而这时估计的参数又是非一致的,下面继续说明这一问题。

依上,设没有误差项的回归方程是:

$$y_i = X_i^n + \mu_i = X_{1i} + X_{2i}^n + \mu_i$$

$$\mu_i \sim i. i. d. N(0, \sigma^2) \quad i = 1, 2, \dots, n$$

将  $X_{2i}^e = X_{2i}^n + \varepsilon_i$  代入得到:

$$y_i = X_i^e + (\mu_i - \varepsilon_i) = X_{1i} + X_{2i}^e + (\mu_i - \varepsilon_i)$$

据此可以得到回归的参数:

$$\hat{\beta} = \left( \sum_{i=1}^n X_i^e X_i^e \right)^{-1} \left( \sum_{i=1}^n X_i^e y_i \right) = \left( \sum_{i=1}^n X_i^e X_i^e \right)^{-1} \left( \sum_{i=1}^n X_i^e (\mu_i - \varepsilon_i) \right)$$

进一步将  $X_i^e = (X_{1i}, X_{2i}^e)$  代入得到:

$$\hat{\beta} = \left( \sum_{i=1}^n X_i^e X_i^e \right)^{-1} \left( \sum_{i=1}^n \begin{pmatrix} X_{1i} \\ X_{2i}^e \end{pmatrix} (\mu_i - \varepsilon_i) \right)$$

则:

$$\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} = \left( \frac{1}{n} \sum_{i=1}^n X_i^e X_i^e \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} X_{1i} \\ X_{2i}^e \end{pmatrix} (\mu_i - \varepsilon_i) \right)$$

依据大数定理:

$$\frac{1}{n} \sum_{i=1}^n X_{1i} (\mu_i - \varepsilon_i) \xrightarrow{P} 0$$

$$\frac{1}{n} \sum_{i=1}^n X_{2i}^e (\mu_i - \varepsilon_i) \xrightarrow{P} -\beta_2$$

由此我们可知  $\hat{\beta}_1$  是有偏估计,  $\hat{\beta}_2$  是一个有偏估计量。

综上,当解释变量存在观测误差时,正态性的假设依赖于回归函数的形式和误差的分布形态。一般情况下,正态性的假设都是不能成立的,只有在回归函数是线性的,且误差分布是正态的特殊情形下,正态性才成立,而在此情况下,参数的估计是非一致的。

## 五、结论与启示

在计量经济学模型中,用“衍生”的随机误差代替“原生”的随机扰动,特别是将模型的设定误差和变量的观测误差对被解释变量的影响引入随机误差项,在很多情况下将导致随误差项对 Gauss 假设以及正态性假设的违背,随之进行的模型估计及统计推断的基础将受到损害。这个问题应该引起计量经济学理论研究的关注,也应该引起应用研究的高度重视。

就十分重要的随机误差项正态性假设而言,具体结论如下:当存在模型关系误差时,如果解释变量是随机的,随机误差项的正态性将得不到保证。当模型遗漏了显著的变量,如果遗漏的变量是非正态的随机变量,随机误差项将不具有正态性。如果待

说明,导出下面的条件是需要额外条件的。例如在  $(x_i, \mu_i, \varepsilon_i)$  是独立同分布的条件下,需要  $x_i, \mu_i, \varepsilon_i$  的一阶绝对矩存在,这是 kolmogorov 大数定理的条件;如果去掉同分布的条件,则一般需要一阶更高绝对矩存在的条件,这是 markov 大数定理的条件,当然还可以继续放宽独立不同分布的条件,代之以更高的矩条件。关于这点 White (1984) 对此有详细的论述,这里不多赘言,而默认这些条件是满足的。

估计的模型是原模型经过函数变换得到的,随机误差项将不再服从正态分布。当模型存在被解释变量的观测误差,如果观测误差相对于随机误差项的标准差特别大、样本长度又特别小,随机误差项的正态性假设会导致显著性水平产生一定程度的扭曲。当模型存在解释变量观测误差时,一般情况下,随机误差项的正态性假设都是不能成立的;只有在回归函数是线性的,且观测误差分布是正态的特殊情形下,随机误差项的正态性才成立。

克服这个问题的根本途径是设定正确的总体模型,取得准确的变量观测值。计量经济学应用模型的总体设定,即经济系统的主体动力学关系分析,不是理论经济学的任务,而是计量经济学的任务。变量观测值的获取,不是统计学的任务,也是计量经济学的任务。一项计量经济学应用研究课题,或者一篇计量经济学应用研究论文,必须将大部分工作或者大部分篇幅放在模型的总体设定方面,放在数据的采集和处理方面。否则研究课题是不可能成功的,研究论文也是没有人愿意阅读的。另外,对模型的随机误差项进行正态性检验,应该成为应用模型研究的一个不可忽略的步骤。

参考文献

[ 1 ]李子奈. 计量经济学模型方法论的若干问题[J]. 经济学动态, 2007(10) :22 - 30.

[ 2 ]李子奈. 计量经济学应用研究的总体回归模型设定[J]. 经济研究,2008(8) :136 - 143.

[ 3 ]Milliam H. Greene , Econometric Analysis (4th edition) [ M]. Prentice Hall Inc. ,2000.

[ 4 ]Damodar N. Gujarati , Basic Econometrics (4th edition) [ M] , McGraw-Hill Companies, 2003.

[ 5 ]H. White , Asymptotic Theory for Econometricians [ M]. Academic Press. 1984.

作者简介

李子奈,男,1946 年生,江苏阜宁县人。1970 年毕业于清华大学工程物理系,1971 年获清华大学工学硕士学位。现为清华大学经济管理学院教授,博士生导师。研究方向为计量经济学理论、方法与应用。

李鲲鹏,男,1980 年生,湖北荆州市人。2007 年毕业于华中师范大学经济学院,获经济学硕士学位。现为清华大学经济管理学院数量经济学专业博士研究生,研究方向为计量经济学理论、方法与应用。

(责任编辑:李峻浩)

《统计研究》主要栏目中图分类号简明对照表

主栏目	分栏目	分类号
统计工作的改革与发展	法律法规	C829. 2
	统计方法制度	C829. 21
	统计管理体制	C829. 22
	统计资料管理,统计信息化建设,统计数据库	C829. 23
		C816
国外统计工作		C829. 1
经济统计学		F222
国民经济核算		F222. 33
统计方法的应用与创新		C81
	统计调查、抽样与抽样分布:	C811
	概率论	0211
	数理统计方法(如非参数统计、参数估计、假设检验、时间数列、贝叶斯统计、相关分析与回归分析)	0212
	统计指数	C813
统计实证分析		C812
	统计模型的应用	F222. 3
统计史	C829. 29	
统计教育		C829. 29