第39卷第24期 数学的实践与认识 Vol. 39 No. 24
2009年12月 MATHEMATICS IN PRACTICE AND THEORY Decem., 2009

消费者不确定性的测度——基于异方差的视角

孙 凤[1]， 杨凯棣[2]

(1. 清华大学人文学院 统计学，北京 100084)
(2. 清华大学 自动化系， 北京 100084)

摘要： 预防性储蓄理论的核心是对消费者不确定的测度. 认为异方差是不确定性的很好代理变量. 本文以极大似然法估计异方差，并运用消费者家计抽样调查数据，讨论收入和支出的异方差对消费者行为的影响.

关键词： 异方差；极大似然估计；不确定性；预防性储蓄

1 问题的提出

近年来中国消费者行为研究的理论框架，主要是建立在预防性储蓄理论上(袁志刚、宋铮1999；孙凤2001，2002；万广华、张茵、牛建高2001)[1]. 预防性储蓄理论的核心是对不确定性是否存在及其影响程度进行测度和检验. 国外研究者对不确定性的测度，大致可以归纳为以下三类：① 使用间接变量替代不确定性，如用户主的职业、收入以及消费的方差进行替代(Skinner，1988)[2]；② 在调查问卷中设计出对未来不确定性认识的问题，如消费者信心调查中对未来经济形势、购买环境、就业预期等的看法(Luigi Guiso 等，1994)[3]；③ 将不确定性看作是潜变量，通过分析工具将其分离出来(Caballero，1991).

我国学者在不确定性的计算上，基本上是采纳第一种和第三种方法，有研究者已对其进行归纳，见表1.

表1 国内研究模型中有关不确定性衡量指标简述

作 者	指 标	特 点
宋铮(1999)	居民收入的标准差	主要是使用统计指标的离散程度衡量收入偏离均值的程度
孙凤(2001，2002)	居民收入、支出的方差	
万广华 张茵 牛建高(2001)	收入增长的预测误差的平方	对预测模型的精度要求高
袁志刚 宋铮(1999)	基尼系数	使用收入分配的不平等状况对居民自身感受进行衡量，刻画个人风险

引自田岗(2005)《不确定性、融资约束、与我国农村搞储蓄现象的实证分析》，《经济科学》第1期.

从表1看，国内学者多数是采用时间序列，从宏观角度展开讨论的. 他们将历年按居民可支配收入分组或消费支出分组的组间方差作为解释变量，引入时间序列模型，用以刻画不

收稿日期： 2007-05-28

基金项目： 教育部人文社科基金《跨时选择与消费者行为研究》(批准号 05JA90044)，和清华大学亚洲研究中心课题《金融危机下的中、日、韩消费者行为比较研究》的资助

万方数据

确定性. 事实上,这种做法已经受到学者的批判(Luasardi, 1998)[4],他指出以组间方差替代不确定性是相当不准确和不科学的,因为其中包含了大量的确定性因素. 同样,以收入增长的预测误差平方作为不确定性,这种刻画方法完全依赖于预测模型的设定,所带来的弊端是显而易见的,即不确定性的测度不是唯一的.

比较来看,由于不确定性是个体感受,从微观角度展开讨论应该更加深入和细致. 但由研究者自己设计问卷,借助抽样调查获取不确定信息需要支付高昂的调查费用,一般研究者是很难为之的. 本文讨论研究者如何从现有的家庭收支调查数据中获取不确定信息.

本文以为异方差是衡量个体之间不确定性差异程度的一个很好指标. 尽管传统计量经济学理论集中在异方差的检验和如何剔除异方差上,但笔者认为异方差的存在是客观必然,如果能够将异方差的影响分离出来,将比简单剔除更符合数据的实际状况[5]. 基于此,本文主要探讨异方差的极大似然估计方法,并以消费者不确定性的测度为例,提出异方差极大似然回归模型在研究不确定性对消费支出影响方面所具有的重要作用.

2 异方差的极大似然估计

经济计量建模研究中一个不可回避的问题就是"异方差性"(Heteroscedasticity)问题. 传统计量经济理论认为异方差的存在是有危险的,如果存在异方差还用最小二乘法去估计参数,会产生以下严重后果:① 参数估计值不再具有最小方差的特性;② 解释变量的显著性检验失效;③ 模型预测精度降低.

针对异方差的危险性,涌现出大量的检验方法,常见的有:图示检验法、等级相关检验法、Glejser 检验,Battlett 检验、Breusch-Pagan 检验、Goldfeld-Quandt 检验、Wald 检验、拉格郎日乘数检验、似然比检验和 White 大样本检验. 这些检验的共同思想是设法通过误差的估计量来检验误差方差与解释变量间是否存在相关性. 若存在明显的相关,则原模型存在异方差性;否则,认为原模型满足方差齐性条件.

由上可见,传统计量经济学理论对异方差的认识集中在对残差项的认识上,透过残差项的异方差性来反映因变量 y_i 的方差特征. 这带给我们的问题是,既然因变量 y_i 是已知数据,由参数 β 和 σ_i^2 共同决定着 y_i 的实际分布状况,为什么不直接从因变量 y_i 的分布了解齐方差或异方差性呢?

事实上,如果我们尝试从因变量的角度去思考异方差问题,发现异方差的存在其实是客观事实. 在微观计量分析中,数据多为个体数据,由于个体的差异性,异方差的存在应该是一种共性,而齐方差性则是特例. 如果我们仅仅为了研究问题的便利,而要求所有因变量 y_i 的方差都是相同的,是违背客观事实的.

因此,本文认为异方差性的存在不具危险性,同时认为判别异方差性是否存在固然重要,但如何估计异方差,揭示出异方差性的存在对客观事物的影响程度则更加必要. 那么,如何估计异方差呢?

传统计量经济模型中,对于存在异方差的模型,是通过一些方法剔除掉异方差性,如加权最小二乘法,广义最小二乘法等,目前使用最为广泛的是广义最小二乘法(GLS):

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{1}$$

(1)式中,Ω 是残差项的方差一协方差矩阵.

万方数据

尽管GLS法能够纠正由于异方差的存在而使最小二乘法存在偏差的估计量，但该方法并不能测算出有关异方差性影响程度大小的结果，这限制了我们对个体差异性影响程度的研究．本文以因变量 y_i 服从正态分布展开对问题的讨论．

当方差为齐性时：

$$y_i \sim \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{(y_i - x_i\beta)^2}{\sigma^2}\right]} \tag{2}$$

将方差加入下标，可以得到异方差的表达式：

$$y_i \sim \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\left[\frac{(y_i - x_i\beta)^2}{\sigma_i^2}\right]} \tag{3}$$

(3)式与(2)式相比，每一个观察值都有一个方差，由(3)式产生两个问题：

第一个问题是模型的识别问题．因为在该模型中，待估参数为 $k+N$，其中 k 为 β 的个数，而 N 为 σ_i^2 的个数．当待估参数值多于观察值时，模型是不可识别的，参数亦估计不出来．为了克服该问题，我们建立 σ_i^2 与一个可观察的外生变量和相关参数的关系模型，此外生变量的维度从 N 降到 s，当 $k+s$ 小于或等于 N，则该模型即可识别，参数亦可识别．

第二个问题是 σ_i^2 的约束条件应为非负．尽管许多函数形式可以满足该约束条件，但简单与可操作性则是我们首先需要的．为了达到该目标，假定 z_i 是一个 $1\times s$ 的外生变量，影响回归模型的方差，假定 γ 是 $s\times 1$ 的系数向量，则

$$\sigma_i^2 = e^{z_i\gamma} \tag{4}$$

(4)式的技术处理可以保证 σ_i^2 的非负性．

我们通过用参数替代 σ_i^2，建立极大似然函数模型：

$$y_i \sim \frac{1}{\sqrt{2\pi e^{z_i\gamma}}} e^{-\frac{1}{2}\left[\frac{(y_i - x_i\beta)^2}{e^{z_i\gamma}}\right]}$$

$$\begin{aligned}\ln L(y/x, z, \beta, \gamma) &= \sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi e^{z_i\gamma}}} e^{-\frac{1}{2}\left[\frac{(y_i - x_i\beta)^2}{e^{z_i\gamma}}\right]}\right) \\ &= -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{N}\frac{1}{2}\ln e^{z_i\gamma} - \frac{1}{2}\sum_{i=1}^{N}\left[\frac{(y_i - x_i\beta)^2}{e^{z_i\gamma}}\right] \\ &= -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{N}\frac{1}{2}z_i\gamma - \frac{1}{2}\sum_{i=1}^{N}\left[\frac{(y_i - x_i\beta)^2}{e^{z_i\gamma}}\right]\end{aligned}$$

通过求偏导、求极大值，我们就可以得到参数的估计值．本例由于存在非线性关系 $z_i\gamma$，需要借助计算机迭代完成．一旦得到参数的估计值，方差就可以通过公式(4)简单计算求得．

3　消费者不确定性的异方差估计

为了分析消费者预防性储蓄的强度，需要对消费者不确定感受进行估计，笔者认为变量的异方差是反映不确定性的一个很好的代理指标，因为如果我们假定个体的方差是相同的，这意味着所有被调查者对家庭人均可支配收入和人均消费的风险判别是相同的．本文认为

万方数据

由于被调查者所处的社会经济环境不同，其风险意识是不同的，这种风险意识的差异即为异方差. 个体之间内在风险意识的差异性，会对消费支出产生影响. 因此，就本文来说，估计出异方差，意味着估计出人们内在的风险意识，这为我们深入研究居民的消费行为提供了重要信息.

本文的分析数据来自天津市统计局2003年城市住户预算调查，调查样本为1433户. 本文选取的变量有家庭人均可支配收入、家庭人均消费支出、家庭生命周期虚拟变量(包括单身阶段、新婚阶段、满巢阶段和空巢阶段). 选用家庭生命周期虚拟变量是由于家庭在不同阶段有不同风险特征，家庭生命周期理论考虑了人口年龄和家庭规模等因素，隐含了较多的影响消费的非经济因素，从而在解释问题的时候具有更大的力度. 本文的理论模型为：

$$y_i = \alpha + \beta x_i + life_i + \sigma_i + \varepsilon_i \tag{5}$$

式中 x_i 为人均可支配收入的对数，y_i 为人均消费的对数，$life_i$ 为家庭生命周期特征虚拟变量，σ_i 为异方差.

本文分别以支出的异方差、收入的异方差，代表支出的不确定性和收入的不确定性，并分别建立四个模型. 模型1为不考虑不确定性的普通最小二乘法回归模型；模型2和模型3分别为将支出的异方差和收入的异方差作为不确定性建立异方差极大似然回归模型；模型4为将支出的异方差和收入的异方差作为不确定性引入同一模型，建立的异方差极大似然回归模型.

本文的分析软件是R1.9.1. R软件是当前国际上社会科学研究中新近发展起来的、运用非常广泛的统计研究软件，该软件在进行极大似然估计和贝叶斯估计上具有独特的功能.

我们首先通过建立异方差的极大似然估计模型估计异方差，然后将包含异方差的模型与普通最小二乘法回归模型的进行比较分析.

R软件需要写程序，在下列程序中，TJ.RData为数据库名，x 为可支配收入、y 为人均消费，LIFE1为独身阶段、LIFE2为新婚阶段、LIFE3为满巢阶段、LIFE4为空巢阶段；$absx$ 和 $absy$ 是对变量 x 和 y 重命名，以计算该变量的异方差，MLhetreg是建立异方差的极大似然估计模型，lm是建立普通最小二乘法回归模型. 计算结果见表2.

```
> load("E:/TJ.RData")
> attach(TJ.RData)
> x<-log(INCOME)
> y<-log(SPEND)
> absx<-abs(x)
> absy<-abs(y)
> hr1<-MLhetreg(y, cbind(x,LIFE1,LIFE2,LIFE3,LIFE4), cbind(absy))
> hr2<-MLhetreg(y, cbind(x,LIFE1,LIFE2,LIFE3,LIFE4), cbind(absx))
> hr3<-MLhetreg(y, cbind(x,LIFE1,LIFE2,LIFE3,LIFE4), cbind(absy,absx))
> summary(lm(y~x+LIFE1+LIFE2+LIFE3+LIFE4))
> summary(hr1)
> summary(hr2)
> summary(hr3)
```

万方数据

表 2　各个模型的参数估计结果与检验值

	Model 1 最小二乘估计量	Model 2 消费的异方差	Model 3 收入的异方差	Model 4 收入、消费的异方差
常数	6.44299***	6.1588137***	6.44222186***	6.2195916***
收入	0.24984***	0.2683983***	0.24978063***	0.2590964***
独身阶段	0.23523	0.3178102	0.23779937	0.3431486***
新婚阶段	0.11948***	0.1089089***	0.12167143***	0.1185151***
满巢阶段	0.40947***	0.4242328***	0.40904020***	0.4281051***
空巢阶段	0.48715***	0.5388218***	0.48835915***	0.5560611***
Z 常数		**−6.7679391*** **	**−0.86715636**	**−5.1716572*** **
消费的异方差		**0.5825748*** **		**0.6706724*** **
收入的异方差			**−0.07256879**	**−0.2620108*** **
Log likelihood	R-Squared：0.1483	428.4258	379.1564	437.1332
Wald statistic	F-statistic：49.68	92.43472	1.567874	103.5509

注：表 2 中的黑体字，代表的是公式(4)中的 z(常数)和参数 λ(消费或收入的异方差).

计算结果显示：

1）Model 1 是以最小二乘法建立回归模型，从表 2 中可以看出，除了家庭处于“独身阶段”阶段的影响不显著外，其他变量均有显著影响. 但该模型的总体相关系数极低，仅为 0.1483.

2）Model 2 是以消费的异方差作为不确定性建立异方差的最大似然回归模型，家庭处于“单身阶段”的影响不显著，但消费支出异方差有显著的影响，异方差影响的常数项与截矩项的乘积为负数，说明支出的不确定性对消费支出有显著的负影响. 比较模型2 和模型1，我们发现似然比估计量有显著提高，亦通过 Wald 显著性检验.

3）Model 3 是以收入的异方差作为不确定性建立异方差的最大似然回归模型，家庭处于“单身阶段”的影响不显著，收入异方差的影响亦不显著，并且未通过Wald 显著性检验，说明收入的不确定性对支出的影响不显著. 比较模型 1 和模型 3，二模型各变量的参数几乎相同.

4）共同引入收入的异方差和支出的异方差建立异方差的最大似然回归模型，各个变量均通过显著性检验，特别是二个异方差变量指标均对消费支出有不同程度的影响，其中消费支出异方差的影响更大，该值与异方差系数值相乘为负数，这意味着支出不确定性的影响将会降低支出水平. 比较似然比估计量，我们发现该模型明显高于其他两个模型，亦通过Wald 显著性检验.

4　结　　论

1）本文以为以异方差作为消费者不确定性的代理变量，能够较好地解释消费者的预防性储蓄行为. 与传统的以分组数据的收入、支出方差以及基尼系数作为不确定性的代理变量相比，更多地考虑了个体风险意识的影响，在信息利用上较前者更丰富.

2）本文计算得出的消费者支出的不确定性大于收入的不确定性，是其不确定性的主要

万方数据

来源，这与中国的现实状况相符，医疗、教育、养老、住房改革，使消费者支出的不确定性不断上涨，已严重影响了消费者的日常预算安排.

3）以极大似然法估计异方差的意义不仅仅在于风险、不确定性的计算上，实际上在社会科学研究中，特别是在对个体行为差异的研究上，异方差对于刻画个体行为差异具有独特的功能.

参考文献：

[1] 田岗. 不确定性、融资约束、与我国农村搞储蓄现象的实证分析[J]. 经济科学，2005，(1)：31-37.

[2] Hans-Georg Muller, Ulrich Stadtmuller. Estimation of heteroscedasticity in regression analysis[J]. The Annals of Statistic, 1987, 15(2): 610-625.

[3] Gragg J G. More efficient estimation in the presence of heteroscedasticity of unknown form[J]. Econometrica, 1997, 51(3): 751-764.

[4] 孙凤. 消费者行为数量研究[M]. 上海人民、上海三联出版社，2002：109-192.

[5] 孙凤. 异方差的估计与实证[J]. 中山大学学报(自然科学版)，2005，44：s2. 103-106.

Estimation of Consumer's Uncertainty ——Visualangle from Heteroskedastiy

SUN Feng[1], YANG Kai-di[2]

(1. School of Humanities and Social Sciences, Tsinghua University, Beijing 100084, China)

(2. Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: The key of precautionary saving theory is how to estimate uncertainty. The paper suggests that heteroskedastiy is a good index of uncertainty. In this paper, I use the ML to estimate heteroskedastiy; and discuss some coefficients of heteroskedastic income and heteroskedastic spending based on a survey of consumer budgets.

Keywords: heteroskedastiy; maximum likelihood estimate; uncertainty precautionary saving

万方数据