
中国经济学应用研究的可信性分析

张成思 陈曦*

内容提要 本文根据科学研究可信性的评判标准,以国内经济学领域五大领军期刊的作者群作为国内学者的代表,对稳健性分析意识、国内发表论文的竞争程度和国际 SSCI 期刊的发表记录情况等进行了统计和分析。结果表明,近年来中国经济学应用研究的可信性进步明显。本文以动态时序模型为例,阐释了应用研究中模型选择与设定的具体改进方案,以期提升相关应用研究的可信性。

关键词 经济学 应用研究 可信性 时序分析

一 引言

可信性(credibility)是经济学应用研究的关键问题。为了综合评述经济学应用研究在可信性方面的发展状况,《经济学视角期刊》(*Journal of Economic Perspectives*,也译做《经济学展望期刊》)在 2010 年发表了多篇专题文章(第 24 卷第 2 期)。其中,Angrist 和 Pischke(2010)以《经济学经验研究的可信性革命:如何提高研究设计》为题,探讨了西方经济学界过去 30 年间在应用研究可信性方面取得的革命性进步。根据作者的分析,这些进步主要源于研究设计的大幅改进(通过直接实验或者准实验方法),而且这些进步在劳动经济学、财政学和发展经济学领域表现得特别明显。作者最后指出(但没有论证),类似的革命性进步风潮在宏观经济学等领域也是存在的。

Stock(2010)认为 Angrist 和 Pischke(2010)所提出的通过实验推动的可信性革命

* 张成思、陈曦:中国人民大学财政金融学院 中国财政金融政策研究中心 100872 电子信箱;chenxi361@gmail.com;陈曦:中国人民大学汉青经济与金融高级研究院。

本研究受到教育部人文社会科学重点研究基地重大项目(12JJD790039)的资助。

只是经济学应用研究领域可信性发展的一个方面,而模型识别假设、识别条件以及敏感性分析等问题在过去30年间在可信性方面也取得了长足进步(Sims,1980;Hendry,1980;Leamer,1983;McAleer等,1985)。事实上,关于经济学应用研究的讨论可以回溯到上世纪三四十年代关于经济学分析中所使用的计量方法是否“科学”的争论。尽管以上作者在讨论应用研究可信性进步过程中的侧重点不同,但至今西方学界已经达成一个基本共识,即经济学应用研究(至少是公开发表的成果)的可信性程度已经显著提升,应用研究结果的可参考价值也相应大幅提高。

基于国际学界讨论的“可信性革命问题”,国内学者对应用研究可信性革命三次大讨论的内容进行了系统、深入地阐释,从模型的随机性设定、变量间因果关系识别和模型“统计适切性”^①评价等三个方面归纳了应用研究的可信性来源,并指出当前中国经济学者亟需提高应用研究的可信性(王美今和林建浩,2012)。这一研究对可信性问题的理论探讨系统、深入,对相关前沿理论发展脉络的评述逻辑清楚、严谨专业,充分反映了作者扎实的计量理论功底和对前沿问题的敏锐把握。

但是我们注意到,国内已有文献在对应用研究可信性问题的探讨过程中,主要从理论层面进行阐述,缺少实践层面的具体改进建议。例如,王美今和林建浩(2012)是在“模型的统计适切性评价”内容中,指出动态时序分析中模型设定检验比模型选择对研究结果的可信性更重要。但遗憾的是,对于应用研究中如何处理模型设定检验与模型选择之间的关系才能提升研究可信性,文章并没有给出具体的解决方案。本文认为,模型设定检验和模型选择对研究的可信性同等重要,而且在实践中需要利用二者的相互验证来确定实证模型形式。为了说明问题,本文在已有文献理论探讨的基础上,详细阐释了动态时序分析的可信性问题与应用改进方案(本文第三部分)。

另外,我们还注意到,已有文献对国内学者应用研究可信性的现状评价较低,特别是王美今和林建浩(2012)指出:“一部分研究者由于不了解计量模型方法具体的应用背景和适用条件,陷入一种滥用和错用的误区…。随意性和错误随处可见”。本文认为,对国内学者应用研究可信性的这种评价(如随意性和错误随处可见)有失偏颇,而且缺乏严谨的数据分析基础。事实上,应用研究中滥用和误用现象是指研究人员对基本方法使用不了解甚至不懂原理(只会操作软件)而导致应用研究结果不可信的情

^① “统计适切性”是“statistical adequacy”的直译,并非一个新的专业术语,即“统计完备性”之意。“适切性”的翻译略显晦涩,容易引起歧义(误认为是一个新生概念或术语)。专业术语的精准对译是确保外文资料在对应中文过程中尽量减少信息损失的关键。不准确的对译还有更典型的例子,如将稳健性分析对应的英文“robustness”直接音译成“鲁棒性”等。

况。而国际学界所提出的“可信性”问题是指在诸如数据质量、稳健性估计方法以及更科学明晰的研究设计(vigorous and clearly articulated research design)等方面不够完美而影响研究结果的可信性(Angrist 和 Pischke, 2010)。显然,这两种情况并不在同一个层次上(“滥用和误用”是更低层次的问题),后者是应用研究的科学可信性(scientific credibility),既包含诚信性(trustworthiness)又包括专业性(expertise),而且主要是指应用研究在多大程度上遵循了科学方法的基本原理;而前者至多是对应于诚信性层面的问题,而不是科学可信性的主要问题。因此,不应将二者并行或者混淆起来讨论。而对于科学可信性程度的评判,最常用的方法是考察研究结果是否能够通过同行评审并公开发表(Stephen, 2004),关于这一点,我们在下文还将进一步说明。

从中国经济学应用研究的实际情况来看,尽管可能与国际学界的最高水平还存在一定差距,但是应用研究的科学可信性程度仍然有明显进步和提高;同时,在可信性有所提升的基础上,应用研究还需要重视模型设定检验和模型选择的相互验证,从而进一步增强应用研究的可信性。

本文结构安排如下:第二部分根据(科学)可信性的评判标准,分析国内学者在稳健性分析意识和国内发文竞争性方面的变化情况,并统计《经济研究》等国内经济学领域五大领队期刊作者群在国际 SSCI 期刊发表论文的相关数据,阐释近年来中国经济学应用研究的可信性进步情况;第三部分在前述分析的基础上,提出对科学可信性进一步提升的具体改进方案,即以动态时序模型的选择与设定问题为例,提出此类模型在应用研究中进一步提升可信性的具体途径;第四部分总结全文。

二 国内经济学应用研究可信性进步的证据

要从定量角度说明国内经济学者应用研究的(科学)可信性进步情况并不容易,这涉及对经济学者涵盖范围(人数)的界定以及对不同时期经济学者研究成果可信性程度的鉴定。^①对于经济学学者的涵盖范围,我们以国内大经济学类中公认的五大领队期刊的作者群作为研究主体,这五大期刊包括《经济研究》、《管理世界》、《世界经济》、《金融研究》和《统计研究》。^②对于可信性程度的评价指标,我们分别根据 Leam-

^① 因为本文所讨论和统计的研究成果一般都含有应用分析内容,因此下文中所述“研究”即指应用研究。

^② 我们考察了中国“985”高校经济学科认定的核心期刊情况,综合结果表明以上五大期刊是最具有代表性的经济类领队期刊,其中《统计研究》虽然是统计学领域期刊,但是其历年发表的文章主要是与经济(计量)等相关的研究成果。以《统计研究》2011年刊登的文章情况为例,当年共刊文202篇,其中164篇属于经济领域(超过总数80%)。

er(1983)提出的“稳健性检验”、国内作者在国内领队期刊的发表难度和 Stephen (2004)提出的“能够通过同行评审并公开发表”作为评判基础,对以下三项内容进行统计:一是国内五大期刊所刊登的文章中含有稳健性检验文章数量的历年数据;二是五大期刊在 2006~2011 年发文数量和作者群变化情况;三是五大期刊作者群同期在国际高水平期刊发文数量的时序数据。其中,国际期刊的选取以社会科学引文索引数据库(SSCI)中包含的期刊为标准(均为实行匿名评审制度的期刊)。通过第一个指标的变化情况,可以直观地考察国内经济学应用研究成果在“稳健性检验”意识层面的可信性进步情况;通过第二个指标的变化情况,可以间接判断国内学术研究的竞争性程度(竞争性越高暗示可信性越高),从而作为国内经济学研究可信性进步的度量指标;另外,通过第三个指标,不仅可以反映国内学者研究成果的可信性进步情况,而且还可以用来考察国内经济学者(即国内五大期刊的作者群)在国际学术平台上的被认可程度,间接反映国内经济学者总体研究水平的进步情况。

(一) 稳健性检验意识

对于五大期刊所刊登的文章中含有稳健性检验的文章数量的统计,我们以 2000 年 1 月至 2011 年 12 月为考察区间,逐月计数五大期刊每期刊登文章中含有稳健性检验的文章篇数。图 1 描绘了五大期刊各个月份刊发文章中含有稳健性分析的文章占比情况(占全部刊发文章总数的比例,除去广告、会议、征文启事和纪念性文章),其中稳健性分析的标准依据明确出现稳健性检验、放松假设结论依然成立以及更改样本范围依然成立等情况进行统计。从图 1 可以看到,所有五个期刊发表的含有稳健性分析的文章数量占比随着时间推移都存在明显变化,特别是在 2006 年以后,大部分期刊的这一指标有较大幅度上升。从各个期刊论文稳健性分析数量占比时序来看,近年来《世界经济》峰值最高,接近 80%,《经济研究》和《金融研究》的峰值接近 70%,《管理世界》和《统计研究》的占比峰值也接近 50%。尽管各个期刊稳健性分析的占比在不同时期存在起伏涨落的变化,但从总体趋势来看,稳健性检验指标的时序变化表现为上升趋势,从而说明国内经济学者的稳健性检验意识日益增强。

虽然稳健性检验不是可信性进步的唯一度量指标,但图 1 所显示的总体上升趋势至少说明国内经济学者对研究结果的稳健程度越来越关注,这从一个侧面反映了研究结果可信性的日益提升。在实践中,我们还逐篇阅读了各期文章具体的稳健性分析内容,发现国内相关研究的稳健性检验设计在内容和层次上都日益丰富和完备。例如,早期刊登的文章对样本变化可能对研究结果的影响进行稳健性分析,多是进行简单变换回归样本测试结果的敏感性,而近期刊登的文章中已经发展到基于严格统计检验

(如未知断点检验)结果为基础的样本分割稳健性测试。另外,从五大期刊稳健性分析的具体统计数据来看(为节省篇幅未做报告),各期刊稳健性分析的绝对数量也呈明显上升趋势,这进一步说明国内经济学研究在稳健性分析层面有明显提升。

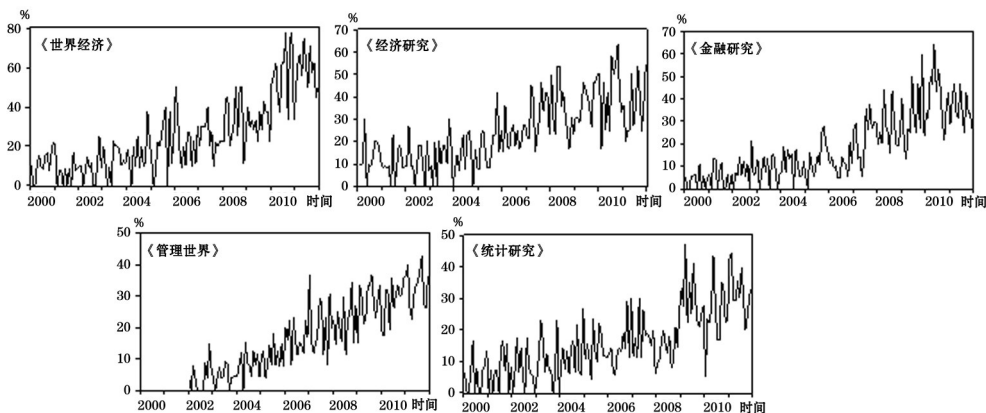


图1 “五大”期刊中含稳健性分析的文章占比(2000年1月~2011年12月)

说明:《管理世界》从2002年开始由双月刊改为月刊,故其数据从2002年开始统计。

(二)国内学者在五大期刊发文的竞争性

虽然高水平学术论文发表的数量可以反映学术研究可信性的进步情况(Stephen, 2004),但是如何界定学术论文的水平高低却是一个比较困难的问题。不过,为了考察国内经济学应用研究与国外研究相比是否水平较低、可信性较差,我们可以通过国内学者在国际学术期刊的发表数量来判断这一问题。在此之前,我们仍然需要对所研究的国内经济学学者群体加以界定才能进一步开展相应的统计分析,而对国内学者群体界定的同时,还可以考察这些学者在学术研究可信性层面的提升情况。这是因为,在其他条件基本不变的情况下,学术研究成果越难发表,研究人员就越需要提升研究成果的可信性,这样才有可能通过严格的双向匿名评审环节。

在这样的思路指引下,我们统计了2006~2011年国内五大经济学领队期刊历年的作者人数和各期刊刊登文章的篇数(“读者来信”、“开会通知”、“论文征集”和各种广告类内容不在统计之列),统计结果报告在表1中。首先,从每个期刊历年刊登文章的篇数来看,没有出现明显的趋势性上涨现象,反而大部分期刊所刊登文章的数量在逐年减少。而且,每年在五大期刊上发表的总体文章数量也呈现下降态势(从2006年的1036篇下降到2011年的921篇)。其次,相对于国内从事经济学研究的学者数量而言,五大期刊每期平均刊文数量都非常有限,其中《世界经济》每期平均只刊登9

篇文章(即 650/72),《经济研究》和《金融研究》每期分别刊登文章 13 篇和 16 篇左右,《统计研究》和《管理世界》则分别控制在每期 19 篇和 22 篇。

表 1 2006 ~ 2011 年中国大经济学类五大期刊发论文情况统计

	2006	2007	2008	2009	2010	2011	合计
《世界经济》篇数	123	118	108	101	105	95	650
《经济研究》篇数	140	162	165	163	169	186	984
《金融研究》篇数	235	192	194	186	173	180	1160
《管理世界》篇数	311	302	316	325	288	259	1801
《统计研究》篇数	227	243	256	213	210	202	1351
“五大”刊文总数	1036	1017	1039	988	945	921	5946
“五大”发文人数	1674	1664	1699	1735	1708	1586	6826

说明:根据各期刊历年内容计算;“五大”即指五大期刊;“五大”发文人数合计”等于历年发文人数加总再减去重复人数。

样本期内共有 6826 名学者在国内五大期刊上发表过总计 5946 篇学术文章,平均每人发表学术论文不足 1 篇。图 2 还刻画了这些文章在五大期刊上的分布情况,其中《管理世界》刊登的文章数量最多(占 30%),而《世界经济》刊登的文章数量最少(只有 11%)。

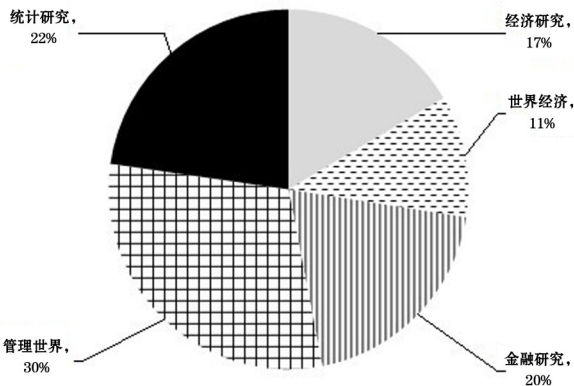


图 2 五大期刊发表文章数量的分布情况 (2006 ~ 2011 年)

为了进一步说明问题,我们还统计了 2006 ~ 2011 年每年五大期刊的人均发文数量。图 3 描绘了相应的统计结果。从中可以看到,五大期刊人均发表数量在 2006 年为 0.63 左右,此后逐渐下降,虽然 2011 年略有回升,但总体上看五大期刊的人均发表数量呈逐年递减的态势。在实践中,我们还统计了

各个期刊每年人均发文数量情况,对应结果的时序走势也多数呈现逐年下降趋势。

以上数据说明,经济学学者在国内五大领军期刊上发表文章的难度在逐年增加。而在国内经济学科不断向前发展、经济学研究人员数量不断增加的背景下,上述统计分析暗示出,国内学者在国内经济学类领军期刊发表文章过程中的竞争程度日益增

强,只有提高经济学应用研究的可信性程度,才有可能通过领队期刊的筛选流程(包括匿名评审),最后在这些期刊发表。

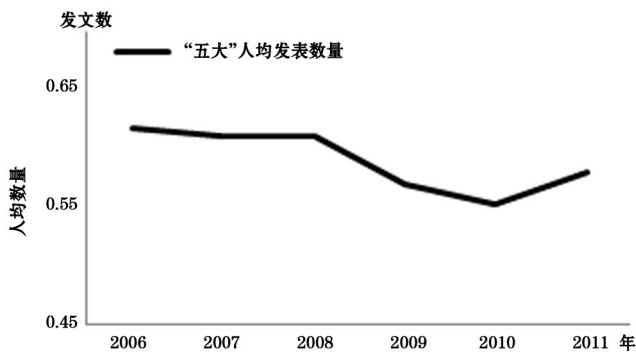


图3 “五大”期刊人均发文篇数年际变化(2006~2011年)

(三) 国际学术界认可度

Stephen(2004)提出,评判学术研究可信性程度(特别是科学可信性)的重要指标是考察研究结果能否通过同行评审并最后在学术期刊上公开发表。因此,我们可以通过计算历年来中国经济学学者的应用研究成果在 SSCI 学术期刊(均实行匿名评审)刊登的数量,

来考察中国经济学应用研究的可信性进步情况。

在具体计算和统计五大期刊作者群的国际发表情况过程中,我们首先逐篇统计五大期刊自 2006 年 1 月至 2011 年 12 月期间的所有作者名录(同一篇文章的多个作者均包含在内),然后根据原文署名信息逐一确定作者单位是否属于国内高校,在此名单基础上排除重复信息,最终确定国内作者群的准确名录。接下来,我们在 Thomson Reuters 旗下的 Web of Knowledge 数据库中,按确定下来的作者名录拼音来检索该作者(姓氏拼音)所发表的 SSCI 期刊文章,然后通过逐篇核对该文章的作者信息(姓氏拼音)确认此作者(姓名拼音)是否为本文作者群中对应的作者。^① 在我们统计的发表记录中,包括 *American Economic Review*、*Econometrica* 等国际一流经济类期刊,因此这些统计数据可以客观反映国内经济学者应用研究的可信性进步情况。

按照以上说明,表 2 归纳了 2006~2011 年国内五大期刊作者群在 SSCI 期刊上发文情况的统计结果。通过表 2 中的历年数据变化情况可以看出,国内经济学者在 SSCI 期刊上的发文数量呈逐年上升趋势:五大期刊作者群在 SSCI 期刊的发文篇数从 2006 年的 292 篇上升到 2011 年的 900 篇,年平均增长率超过 20%,而且 2011 年五大

① 检索工作比我们预想的要庞杂得多。在检索过程中,样本区间内五大期刊作者群共包含 6826 位作者,我们对每一位作者的名字进行检索,并且逐一核对该文章的作者信息,从而判断该作者是否来自五大期刊作者群。当遇到重名作者或者拼音(读音)相同作者时,无法简单依靠数据库检索结果进行筛选,此时需要将所有相关信息下载后再筛查甄别。此外,还有部分作者有多个工作单位或者工作单位在样本区间内发生了变动,此时我们根据作者的工作变动路径进行二次审核,从而确保检索结果准确可靠。我们保存了所有作者名录及发文信息的检索结果,如有需要可向作者索取。

期刊作者群在 SSCI 的发文数量与其在五大期刊上的发表数量相当。这些数字说明,国际学界(同行)对国内经济学学者的研究水平的认可度日益提高。

表 2 2006 ~ 2011 年五大期刊作者群在 SSCI 期刊发表情况统计

年份	2006	2007	2008	2009	2010	2011	合计
“五大”作者发表 SSCI 篇数	292	471	737	706	762	900	3868
“五大”作者发表 SSCI 人数	187	293	426	451	475	510	1217 ^a

说明:根据各期刊历年内容计算;^a表示这一数字等于历年 SSCI 发文人数加总再减去重复人数。

从五大期刊作者群中能够在 SSCI 期刊发文的人数来看,也存在明显的逐年递增趋势。2006 年在五大期刊上发表过学术论文的 1674 位作者(数字见表 1)当中,仅有 187 人有学术成果发表于 SSCI 期刊(比例不到 12%)。而到了 2011 年,五大期刊的 1586 位作者中有 510 人有成果发表于 SSCI 期刊(占比超过 32%)。综合 2006 ~ 2011 年的总体情况来看,共有 1217 位五大期刊作者的总计 3868 篇文章发表于国际 SSCI 期刊,平均每人在 SSCI 上发表论文 3.2 篇文章。

综上所述,近年来国内经济学者应用研究水平有了明显提高,这种提高不仅体现在稳健性检验意识的增强和相关学术成果在国内学术平台上发表的竞争程度加剧等方面,而且还反映在中国经济学者的应用研究水平得到国际同行认可度的日益增强上。如果国内学者缺乏学术研究的科学可信性,那么研究成果是很难得到国际学术界认可。因此,已有文献中关于“国内经济学应用研究水平低下和缺乏可信性”的判断并不客观,也缺乏科学依据。当然,尽管过去几年间国内经济学应用研究可信性具有明显进步,但是在应用研究过程中模型设立和模型选择等方面仍有改进空间。本文在下面部分以动态时序模型为例,在已有文献研究的基础上,阐述应用研究中如何科学设定和选择动态时序模型,以期对提升相关应用研究的科学可信性有所帮助。

三 应用研究可信性的进一步改进方案:以动态时序分析为例

动态时序模型在经济学应用研究中的使用频率很高,特别是在宏观经济金融问题的应用研究过程中更为常见。应用研究人员经常要面临动态时序模型的设定和选择问题,而如何科学平衡模型设定与选择的先后顺序,对应用研究的可信性程度至关重要。针对这一问题,王美今和林建浩(2012)在其“模型的统计适切性评价”部分以时

间序列分析为例,指出时序模型(如向量自回归(VAR)模型)的设定检验比模型选择对研究结果的可信性更重要。该文还特别指出,如果时序分析中的模型选择只是在平衡拟合和简洁性之间的选择,那么并不能保证残差必然满足独立同分布假定。从本质上看,此处提到了一个关于动态时序分析模型选择和模型设立的非常重要的问题,但遗憾的是已有文献并没有指出应用分析过程中如何处理此类问题才能提升研究可信性。

为此,我们针对动态时序分析问题,详细讨论提升此类研究可信性的具体解决方案。不过需要说明的是,与已有文献的观点不同,我们认为模型设定检验和模型选择对研究的可信性同等重要,而且在实践中需要利用二者的相互验证来确定动态时序分析中的实证模型形式。为了说明问题,我们以宏观时序分析中经常使用的 VAR 模型为例进行阐释。假定研究人员已经基于相关经济学理论判定向量 X 中的 n 个变量 (x_1, x_2, \dots, x_n) 存在动态互动关系,并适用 VAR 模型分析,即:

$$X_t = \Phi(L)X_{t-1} + e_t, e_t \sim i. i. d. (0, \Omega_e) \quad (1)$$

其中, $\Phi(L)$ 表示向量滞后算子多项式, e_t 表示服从独立同分布(*i. i. d.*)的扰动项向量。基于以上 VAR 模型进行回归分析和格兰杰因果关系检验等,从技术角度讲都相对简单,但科学的模型设定并不容易。对于此类动态模型,研究人员需要将模型选择(如最优滞后阶数)和模型设定检验结合起来进行相互验证才可能获得科学可信的结果。这一细节问题在已有研究中(包括国内和国际)并未引起足够重视,特别是对模型选择(滞后阶数)所依据的信息准则与模型设定中残差是否具有序列相关性的协同判断问题缺少明确说明(张成思,2012a),而这恰恰是可能降低研究结果可信性的关键问题之一。因为对于 VAR 这样的动态时序模型而言,如果残差存在序列相关性,那么模型估计和格兰杰因果关系检验等统计推断将既不具有无偏性又不具有一致性。此时,即使运用王-林文所提到的稳健推断方法(如 Newey-West 异方差和序列相关性修正标准差)也无法解决问题,因为稳健性标准差估计只是对非球面分布扰动项标准差估计的修正,而无法解决扰动项与自变量之间的非正交问题。问题的关键在于,当 e_t 存在序列相关性时(以 1 阶序列相关为例,高阶类似),即:

$$e_t = \rho e_{t-1} + \varepsilon_t, \varepsilon_t \sim i. i. d.(0, \sigma_\varepsilon^2) \quad (2)$$

$$\text{结合模型(1)和模型(2),则有: } e_t = \rho[X_{t-1} - \Phi(L)X_{t-2}] + \varepsilon_t \quad (3)$$

不难看出,扰动项 e_t 与 VAR 模型右手侧变量(即 X_t 的滞后项)非正交,因此基于这一 VAR 模型的传统估计(如最小二乘估计)结果的任何相关统计推断将既不具有统计无偏性也不具有一致性。

那么在实践中应该如何进行 VAR 模型设定?事实上,对于类似 VAR 这样的动

态时序模型,要获得准确有效的统计推断,最重要的是模型无序列相关性与简约性的综合平衡,既要满足模型的扰动项没有显著的序列相关性,又要尽可能选择简约的模型。为此,模型选择和设定检验的相互验证过程可以采取以下步骤:首先,在给定最大滞后阶数的条件下(如季度数据给定8期),利用AIC准则确定最优滞后阶数;然后,在AIC选定的阶数下,应用VAR模型序列相关性检验(VAR-LM检验)确定模型残差是否具有序列相关性。如果检验结果显示无显著序列相关性,则此滞后阶数即为最优;反之,则需要重新选择模型的最优滞后阶数。此时,可以遵循从一般到特殊的模型设立原则,先从最高阶(如季度模型的8阶)向下依次检验对应模型的序列相关性,在没有序列相关性的组内选择AIC最小值对应的滞后阶数即为最优。这样,通过模型选择和设定检验来相互验证共同判定最优滞后阶数,既能确保模型残差没有显著序列相关性,又能尽量满足计量建模的简约性要求。

以上讨论还可以拓展到更加复杂的动态时序模型的设立及估计上来。我们以近年来国际学术界广泛争议的新凯恩斯菲利普斯曲线(NKPC)模型的设立与估计问题为例进一步说明。Gali和Gertler(1999)的经典文献(Google Scholar的数据显示至今被引2347次)所提出的基于微观基础的NKPC模型为:

$$\pi_t = c + \gamma_e E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \delta y_t + \eta_t \quad (4)$$

其中, π_t 表示通货膨胀率, c 是常数项, $E_t \pi_{t+1}$ 表示基于时刻 t 及以前的信息集对 $t+1$ 期通胀率的预测序列,通胀率滞后项 π_{t-1} 用以捕捉通货膨胀的惯性特征, y_t 表示国内真实产出缺口, η_t 表示随机扰动项,其余各系数在模型中具有较直观的解释。

虽然模型(4)的设立形式比较简单,但是要对该模型(4)进行估计并获得科学可信的结果却非常困难。问题的关键在于Gali和Gertler(1999)在原始模型推导过程中对后顾型(backward-looking)企业定价模式的假设过于僵化(其假设后顾型企业在定价过程中仅考虑过去1期的通胀信息),最终导致NKPC模型(4)的设立形式过于程式化(stylized),制约了模型中动态机制的丰富程度(因变量仅有1期滞后出现在模型右手侧),这样在应用季度数据进行模型估计时就很可能无法消除扰动项的序列相关性。而一旦模型(4)存在序列相关性,那么即使运用学界普遍采用的广义矩(GMM)估计方法(用于克服自变量内生性),也无法获得精确可信的检验结果。

说得再具体一点,当 η_t 序列相关时,通过与(1)~(3)式的类似说明,可以看到 η_t 与NKPC模型(4)右侧的 π_{t-1} 非正交,而由于相关文献所使用的GMM估计工具变量中一般都包含通胀率(以及与通胀率相关的其他变量,如产出缺口、利率、广义货币增长率等)的滞后项,所以即使运用GMM估计,获得的结果也是无效的。事实上,如果

η_i 存在序列相关性,那么即使从 GMM 估计的工具变量中剔除通胀率的所有滞后项也无法解决问题,因为其他工具变量都与通胀率相关,所以如果 η_i 与通胀率的滞后项非正交,则 η_i 与其他工具变量也非正交 (Zhang 等,2009;张成思,2012b)。

因此,当序列相关性出现在使用工具变量估计的动态时序模型中时,问题就会变得相当复杂和困难。此时应该如何处理相关问题?首先,对于模型(4)的分析来说,提高结果可信性的重要环节之一是进行正确的序列相关性检验。注意,由于模型(4)右侧含有当期以及预期变量,所以不难判断模型扰动项与这些变量很可能存在相关性(非正交),这也是学界通常采用工具变量估计的基本原因。但是在工具变量估计的背景下,传统的序列相关性检验(如 DW 检验、Breuch-Godfrey LM 检验)就不再适用。关于这一点,Godfrey 等(1988)、Cumby 和 Huizinga(1992)、Davidson 和 MacKinnon(1993)以及 Godfrey(1994)的系列研究已经给出了详尽的理论阐释,只是在宏观动态时序的应用分析领域没有引起足够的重视。

为此,对类似模型(4)这样的情况进行序列相关性检验,需要根据 Godfrey(1994)提出的工具变量估计下的序列相关性检验方法进行检验,此检验与传统的 Breuch-Godfrey LM 检验过程完全不同,我们此处对该检验的核心内容进行阐释(张成思,2012b;Zhang 等,2008)。其基本思想是基于如下辅助回归方程:

$$\tilde{\eta}_i = Xb + \rho(L)\tilde{\eta}_{i-1} + \varepsilon_i \quad (5)$$

其中, $\rho(L)$ 表示滞后算子多项式, $\tilde{\eta}$ 是原始模型使用工具变量估计(2SLS)后获得的残差序列, X 表示原始模型中的自变量矩阵, ε 是辅助方程的扰动项(允许存在异方差,但不能序列相关)。为便于说明,我们需要将模型(5)重新写成以下向量形式:

$$\tilde{\eta} = X_1b + X_2\rho + \varepsilon \quad (6)$$

其中, X_1 代表原始模型的自变量矩阵, X_2 表示模型(5)中 $\tilde{\eta}$ 的滞后项。现在假定最初的工具变量集合为 Z ,那么用于估计辅助模型(5)或者(6)的工具变量需要在 Z 的基础上再增加 $\tilde{\eta}$ 的滞后项,增广后的工具变量矩阵用 W 表示。另外,定义投影矩阵 $P_w = W(W'W)^{-1}W'$ (其他投影矩阵据此定义),并令 $\tilde{\varepsilon}_r$ 表示 $\rho = 0$ 时模型(6)对应的残差序列, $\tilde{X}_i = P_w X_i$ ($i=1,2$), $M_1 = I - P_{\tilde{X}_1}$ (I 表示单位阵), $S_r^2 = \tilde{\varepsilon}_r' \tilde{\varepsilon}_r / (T - r)$ (T 表示样本大小, r 表示矩阵 X_2 的维度), $P_{M_1 \tilde{X}_2} = (M_1 \tilde{X}_2) (\tilde{X}_2' M_1 \tilde{X}_2)^{-1} \tilde{X}_2' M_1$ 。

这样,就可以通过计算下面的 LM 统计量对 $\tilde{\eta}$ 是否具有序列相关性进行检验(原

假设是 $\rho=0$):

$$LM = \tilde{\eta}' P_{M_1 \tilde{x}_2} \tilde{\eta} / r S_r^2 \quad (7)$$

其中, r 就是约束条件个数, 检验统计量服从渐近 $F_{(r, T-r)}$ 分布。我们在分析中计算并报告该统计量对应的 p 值。

在实际应用(编程)过程中, 序列相关性检验统计量 LM (以及对应的 p 值)的计算还可以通过另一种途径获得。首先, 使用增广后的工具变量 W 对模型(5)进行工具变量估计, 分别获得有约束条件下的残差序列 $\tilde{\varepsilon}_r$ 和无约束条件下的残差序列 $\tilde{\varepsilon}_\eta$ 。然后, 分别通过 $\tilde{\varepsilon}_r$ 和 $\tilde{\varepsilon}_\eta$ 对增广工具变量 W 进行最小二乘回归估计, 从而分别获得以下被解释残差平方和(explained sum of squares), 即:

$$ESS(\tilde{\varepsilon}_r) = \tilde{\varepsilon}_r' P_W \tilde{\varepsilon}_r \quad (8)$$

$$ESS(\tilde{\varepsilon}_\eta) = \tilde{\varepsilon}_\eta' P_W \tilde{\varepsilon}_\eta \quad (9)$$

这样, 序列相关性检验统计量 LM 就可以计算为:

$$LM = \frac{\{ESS(\tilde{\varepsilon}_r) - ESS(\tilde{\varepsilon}_\eta)\} / r}{\tilde{\varepsilon}_r' \tilde{\varepsilon}_r / (T - k)} \quad (10)$$

可以证明, 公式(10)中的 LM 统计量与公式(7)中的统计量完全一致。注意, 根据 Kiviet(1986)的研究结论, 在实际编写计量程序过程中, 滞后操作带来的缺失值需要用 0 进行插补, 以使检验结果更加精确。如果上述检验显示模型(4)确实存在显著序列相关性, 则需要重新考虑 NKPC 模型的设立问题, 例如考虑通过拓展模型右手侧的动态机制或者预期机制等来减弱或消除可能存在的序列相关性问题。

当然, 通过与以上类似的步骤还可以进一步检验 $\tilde{\eta}_i$ 和 $\tilde{\eta}_{i-n}$ (如 $n=2, 3, 4, 5$) 之间是否存在序列相关性, 此时需要注意的是, 要将 $\tilde{\eta}_{i-1}$ 从式(5)中删除而将 $\tilde{\eta}_{i-5}$ 添入到方程(5)的右手侧。这一检验可以用来甄别扰动项 $\tilde{\eta}_i$ 服从一阶移动平均过程(即 MA(1))还是一阶自回归过程(即 AR(1))。利用这一方法甄别 MA(1) 和 AR(1) 的基本思想是: 如果 $\tilde{\eta}_i$ 服从 MA(1), 那么联合显著性检验 $\tilde{\eta}_{i-n}$ ($n=2, 3, 4, 5$) 的系数同时等于 0 应该不具有显著性。当然, 在这一检验过程中, 对应辅助方程所使用的增广工具变量需要相应修改。

另外需要特别指出的是, 由于应用研究人员(包括国际学术界)在传统的最小二乘估计背景下经常习惯于使用 DW 检验进行序列相关性诊断, 所以在对动态时序模型进行序列相关性检验时必须注意, DW 检验(以及 Ljung-Box Q 检验)都是无效的(in-

valid),只有 Breuch-Godfrey LM 检验才有效,其本质原因就是动态模型中含有因变量的滞后项。关于动态模型中 DW 检验等无效的理论证明和推导,经典计量经济学教材(Davidson 和 MacKinnon,1993)已经做了详尽地阐释,此处不再赘述。

在进行应用研究分析的具体实践中,研究人员面对的问题可能远比以上分析内容复杂。例如,当应用 VAR 模型进行分析时,可能样本只有 10 年甚至 5 年的季度数据(即 40 或 20 个观测值),在利用 AIC 或者 SIC 作为模型滞后阶数选择标准时,最大可选滞后阶数如何给定? 尽管此时可以考虑使用 SIC(或者 AIC 与 SIC 同时使用)并以最大 4 期为基础,但此时研究人员仍然需要在序列相关性和模型简约性之间做出相对科学的研判。另外,如果在给定的最大阶数内不同信息准则给出的滞后阶数不同又该如何与序列相关性检验进行相互验证? 这些非常现实的应用问题显然没有一个简单普适的答案(张成思,2012c)。一个基本准则就是:此类动态时序模型的选择与设定要首先确保模型无序列相关性。在此条件下,才可以再进一步在模型拟合与简约性之间进行权衡取舍。

四 结论

近年来国内相关文献对国际学术界关于经济学应用研究可信性革命的讨论进行了归纳和梳理,总结了应用研究的可靠性来源,为国内学者了解应用研究可信性问题的国际前沿进展提供了有益资料。但是,部分文献对国内学者应用研究的可信性评价不够全面,特别指出了国内学者在应用研究中存在滥用和错用现象,随意性和错误随处可见,中国经济学应用研究面临提高可信性的重要问题,需要全面吸收和借鉴国际学界对于可信性问题的成果,改进研究模式和教学模式。

本文担心这些研究对国内经济学者应用研究水平低下的论断和对中国经济学应用研究可信性的负面评价,可能会给国内经济学者的学术声誉和研究成果的学术意义及应用价值带来负面影响。为此,本文统计并分析了《经济研究》等国内经济学领域五大领军期刊的作者群在国内和国际期刊的发表记录等相关数据,从国内学者的稳健性分析意识、在国内发文的竞争度和受国际学界的认可度等不同层面,阐明国内经济学者在应用研究可信性方面的进步情况。尽管相关统计工作庞杂,需要付出大量时间和精力,但是我们相信这项工作所提供的数字依据对客观准确评价中国经济学应用研究的可信性水平意义不凡。另外,本文还针对动态时序分析中模型选择与设定问题进行了应用研究层面的阐述,围绕动态模型最优滞后阶数与序列相关性问题的给出了提高应用分析可信性的具体解决方案。

本文的分析结果表明,中国经济学应用研究的可信性在多个层面都取得了明显进步,这些进步至少在中国经济学领队期刊的作者群中是有事实依据的,因此不宜用“随意性和错误随处可见”来评价国内经济学者的应用研究水平。尽管经济学应用研究仍然存在很多待解的问题(如动态模型的科学选择和设立等),可信性也有很大的提升空间,但这些是国内外经济学者共同面临的问题,并非只针对国内学者。另外,对于计量经济学的教学而言,理论讲解辅以软件示范可能仍然是有效的教学模式,对这种教学方式的批判需要特别谨慎。而在评价国内应用研究的可信性问题时,不能将教学过程中遇到的学生犯的一些基础性错误与科学研究层面的可信性问题混为一谈,也不宜将“中国经济学者”的涵盖范围无限扩大,否则相关立论就会失去针对性,对应的结论就失去了应有的意义。

参考文献:

- 王美今、林建浩(2012):《计量经济学应用研究的可信性革命》,《经济研究》第2期。
- 张成思(2012a):《通货膨胀、经济增长与货币供应:回归货币主义?》,《世界经济》第8期。
- 张成思(2012b):《全球化与中国通货膨胀动态机制模型》,《经济研究》第6期。
- 张成思(2012c):《金融计量学——时间序列分析时间》,中国人民大学出版社。
- Angrist, J. and Pischke, J. “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics.” *Journal of Economic Perspectives*, 2010, 24, pp.3-30.
- Cumby, R. and Huizinga, J. “Testing the Autocorrelation Structure of Disturbances in Ordinary Least Squares and Instrumental Variables Regressions.” *Econometrica*, 1992, 60, pp.185-195.
- Davidson, R. and MacKinnon, J. *Estimation and Inference in Econometrics*. Oxford: Oxford University Press, 1993.
- Gali, J. and Gertler, M. “Inflation Dynamics: A Structural Econometric Analysis.” *Journal of Monetary Economics*, 1999, 44(2), pp.195-222.
- Godfrey, L. “Testing for Serial Correlation by Variable Addition in Dynamic Models Estimated by Instrumental Variables.” *The Review of Economics and Statistics*, 1994, 76, pp.550-559.
- Godfrey, L.; McAleer, M. and McKenzie, C. “Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models.” *Review of Economics and Statistics*, 1988, 70, pp.492-503.
- Hendry, D. “Econometrics—Alchemy or Science?” *Economica*, 1980, 47, pp.387-406.
- Kiviet, J. “On the Rigour of Some Misspecification Tests for Modelling Dynamic Relations.” *The Review of Economic Studies*, 1986, 53, pp.241-261.
- Leamer, E. “Let’s Take the Con Out of Econometrics.” *The American Economic Review*, 1983, 73, pp.31-43.
- McAleer, M.; Pagan, A. and Volker, P. “What Will Take the Con Out of Econometrics?” *The American Economic Review*, 1985, 75, pp.293-307.
- Sims, C. “Macroeconomics and Reality.” *Econometrica*, 1980, 48, pp.1-48.
- Stephen, B. *Nature’s Experts: Science, Politics, and the Environment*. New Brunswick, NJ: Rutgers University Press, 2004.
- Stock, J. “The Other Transformation in Econometric Practice: Robust Tools for Inference.” *Journal of Economic Perspectives*, 2010, 24, pp.83-94.
- Zhang, C.; Osborn, D. and Kim, D. “The New Keynesian Phillips Curve: From Sticky Inflation to Sticky Prices.” *Journal of Money, Credit and Banking*, 2008, 40, pp.667-699.
- Zhang, C.; Osborn, D. and Kim, D. “Observed Inflation Forecasts and the New Keynesian Phillips Curve.” *Oxford Bulletin of Economics and Statistics*, 2009, 71, pp.375-398.

(截稿:2013年8月 责任编辑:宋志刚)