

# 宏观经济学的经验研究方法：校准与估计之争\*

张明志 铁 瑛 林 娟

**内容提要：**本文分析了校准方法兴起的背景，基于方法论的角度厘清了校准方法与估计方法之间的逻辑关系，探讨了校准方法与估计方法争论的本质及其融合发展的趋势。分析表明，校准方法和估计方法的出发点是完全背离的，二者对模型优劣的评判标准和对模型现实性的看法大不相同。校准和估计方法各有其优缺点，也面对共同的难题。校准和估计方法都在努力寻找支配经济现象背后的运行规律，但从方法论的角度看，正是对经济运行规律寻求方式的不同导致它们之间存在着本质的差异。从发展趋势上看，校准和估计方法将在动态随机一般均衡的分析框架下获得更为深入的融合发展。

**关键词：**宏观经济学 方法论 估计 校准 动态随机一般均衡

## 一、引言

理论与数据之间的关系是现代宏观经济学关注的核心问题之一。在过去的很长一段时间里，估计方法在宏观经济学的经验研究中占据了主导地位。但自从20世纪80年代实际经济周期模型(RBC)创立以来，越来越多的经济学家特别是新古典宏观经济学家开始转向校准方法，并在相当程度上推进和丰富了宏观经济学的理论和经验研究。也正是从那时起，校准与估计之争就没有停息过。

在20世纪80年代，支持校准方法的学者(简称为“校准派”)和支持估计方法的学者(简称为“估计派”)形成了两个立场迥然不同的阵营。在凯恩斯和新凯恩斯主义经济学家的阵营内，估计方法仍然是主流。但在新古典宏观经济学家的阵营之内，校准方法却备受推崇，代表性的人物是罗伯特·卢卡斯(Robert E. Lucas)、芬恩·基德兰德(Finn E. Kydland)和爱德华·普雷斯科特(Edward C. Prescott)。虽然支持校准方法的经济学家多为新古典宏观经济学家，但并非所有的新古典宏观经济学家都支持校准方法，比如托马斯·萨金特(Thomas J.

Sargent)就对校准方法进行了广泛的批评，并认为估计方法仍然是宏观经济学经验研究较好的方法，其代表性论文是 Hansen & Sargent(1988)。随着两种方法论战的不断深入，一些学者尝试将校准方法纳入传统计量经济学的范畴，如 Manuelli & Sargent(1988)和 Gregory & Smith(1990)等。但他们尝试的出发点大都偏向于传统的估计方法，并且缺乏对这两种方法本质差异的探讨。关于校准与估计之争，国际学界较具影响力的一篇文献是 Hoover(1995)。该文对校准方法进行了全面的剖析，并基于方法论的角度对估计和校准方法的本质差异进行了研究。然而美中不足的是，这篇文献并没有注意到两种方法融合发展的可能性。而在一些最新的涉及动态随机一般均衡模型(DSGE)的文献中，校准和估计方法的交叉应用及其融合发展已经初露端倪，最为典型的是 Smets & Wouters(2007)。

国内的相关研究起步较晚，且大多集中于对估计方法的探讨。一方面，估计方法较早被中国学界接受并迅速成为国内宏观经济学经验研究的主流方法；另一方面，国内学界在估计方法的应用方面存在着比较严重的错用、误用和滥用现象。在这一背景

\* 张明志，厦门大学经济学院国际经济与贸易系，邮政编码：361005，电子邮箱：mzzhang@xmu.edu.cn；铁瑛，林娟，厦门大学经济学院国际经济与贸易系。基金项目：教育部哲学社会科学研究重大课题攻关项目(项目批准号13JZD010)、福建省社会科学规划一般项目(项目批准号：2013B181)、中央高校基本科研业务费项目(项目号：T2013221017)、厦门大学国际经济与贸易系教育发展基金(项目号：201112106)。感谢匿名审稿人的宝贵意见，当然，文责自负。

下,一些学者开始基于方法论的角度对估计方法的使用予以反思,代表性的文献有洪永森(2007)、李子奈(2009)、李子奈、齐良书(2010)和王美金、林建浩(2012)等。与此同时,随着国内宏观经济学研究水平的不断提升,RBC理论和校准及其与估计方法的比较也逐渐进入国内学者的视野。龚刚(2004)在详细分析RBC理论的过程中介绍了校准方法,但没有涉及到与估计方法的比较问题;黄贇琳(2005)以Cooley & Prescott(1995)为例在详细地介绍校准方法的同时,涉及到了校准与估计方法的比较问题,但没有详细展开分析;李凌、权衡(2009)针对RBC模型、动态因素模型(DFM)和结构向量自回归模型(SVAR)进行了探讨,虽然隐含着校准和估计方法的比较问题,但没有触及到两种方法争论的本质。总而言之,到目前为止,国内鲜有文献是基于方法论的角度对校准和估计方法进行系统和深入比较研究的。

本文分析了校准方法兴起的背景,基于方法论的角度厘清了校准方法与估计方法之间的逻辑关系,探讨了校准与估计方法争论的本质及其融合发展的趋势,从而有助于我们更好地理解 and 认识宏观经济学的演变过程及其未来可能的发展方向。

## 二、“卢卡斯批判”与校准方法的兴起

### (一)“卢卡斯批判”

20世纪五六十年代,起因于1929—1933年经济大危机的凯恩斯理论在宏观经济分析中居于核心地位。素有“萧条经济学”之称的凯恩斯理论认为,引起短期经济波动的主要原因在于总需求,因此平滑经济波动或者说解决经济危机的根本出路在于政府采取适当的稳定性政策来稳定总需求。要确定一个正确的稳定性政策,需要根据历史数据对总需求变量和政策变量之间的数量关系进行经验估计。而要使特定的稳定性政策实施能够达到预期的目标,要求政策变量与总需求变量之间的关系是稳定的,不会因为政策的实施而改变它们之间的稳定关系。

利用宏观计量经济学的估计方法,20世纪五六十年代菲利普斯曲线在经验上获得了支持,即失业率与通货膨胀率之间存在一种相对稳定的负相关关系。但20世纪70年代的石油危机却使失业与通货膨胀之间所存在的稳定关系出现了经验性失效。从理论上讲,菲利普斯曲线出现经验性失效是因为短期经济波动可能来源于总供给而非总需求层面的冲击。但20世纪80年代初的美国并不存在明显的较

大的供给冲击,通货膨胀却高于20世纪60年代的任何时期,因此供给冲击并不能完全解释菲利普斯曲线的失效。其背后的原因是,美国20世纪70年代的高通货膨胀改变了价格和工资确定的方式。

由菲利普斯曲线的经验性失效所引发的理论思考是,尽管在现实中可能存在失业率与通货膨胀率之间这种统计性的负相关关系,但并不意味着政府可以利用这种替代关系。尽管宏观计量模型也能够准确地估计出二者之间的数量关系,但并不意味着政府当局可以利用这种估计出来的数量关系进行政策分析和评价。特别是,对于总量变量之间的许多关系而言,预期可能是很重要的。而政策的变化很可能影响那些预期,结果是政策的变化可能改变总量变量之间业已存在的关系。简言之,如果政策制定者企图利用这种统计关系,通过预期所产生的效应可能使这些关系失效。这就是著名的“卢卡斯批判”。

### (二)校准方法的兴起

卢卡斯批判的本质是政策无效性命题。它对于凯恩斯理论范式而言是个沉痛的打击。卢卡斯批判表明,合理的政策分析不能基于简约式(reduced form)的关系。包括消费者和厂商在内的行为个体不仅受当前正在执行政策的影响,而且也受政府未来政策的影响。随着政策的改变,形成总需求的个体选择变量与政策变量之间的关系并不是不变的。Lucas(1980)指出,如果某宏观经济模型刻画的是实行某项经济政策之前的经济体,那么该模型就无法用于预测该政策执行之后的效果。因此,如果用传统的宏观计量经济学方法来估计总需求变量与政策变量之间的关系,计量模型中的参数估计实际上混合了结构参数和政策参数,<sup>①</sup>而这些参数在不同的政策制度下是不同的,传统计量经济学的估计方法在确定政府最优的政策选择上表现出明显的局限性。

为了评估不同政策准则的影响,我们需要一种新的研究范式。在这种新的研究范式下,行为主体对政策的反应是基于一些固定的特性,例如偏好、技术和禀赋等,这些特性不会随着政策的变化而改变。与此同时,模型的构建必须以经济中各种代表性行为个体的最优化行为为基础。罗伯特·卢卡斯强烈要求宏观经济学家应该建立这样一种结构模型(Lucas,1980),他也正是这种新研究范式最为积极的倡导者。而对这种新研究范式的发展做出卓越贡献的代表性学者则是芬恩·基德兰德和爱德华·普

雷斯科特,他们于1982年发表的奠基性的论文不仅标志着实际经济周期模型的创立,而且意味着宏观经济学新研究范式的重大突破,其突出表现就是全新的宏观经济学经验研究方法即校准和模拟(Calibration and Simulation)的出现。实际经济周期模型的出现,不仅有效规避了卢卡斯批判所指出的问题,而且彻底改变了传统宏观经济学的研究范式,从而开创了现代宏观经济学研究的新纪元。从这以后,实际经济周期模型的大部分追随者在不同程度上放弃了传统宏观计量经济学的估计方法,转而投向具有坚实微观基础的动态随机一般均衡模型,同时在经验研究上把注意力集中于校准和模拟上,导致校准方法在宏观经济学研究领域的兴起。

### 三、估计方法与校准方法的比较分析

#### (一)估计方法与校准方法的基本内涵

1. 估计方法的基本内涵及其应用条件。宏观计量经济学的估计方法,是在完成宏观计量模型的总体设定之后,从样本数据即真实数据出发估计计量模型中的参数并对其进行适当的检验,最终要么对已有的宏观理论模型进行验证或修正并推动宏观理论模型往前发展,要么发现现实经济中宏观经济变量之间新的关系或者说新的经济规律。当然,对于基于特定样本而建立的,并以估计和检验好的宏观计量模型而言,还可能利用它对样本外的经济主体的状态进行预测和进行宏观经济政策评价。宏观计量模型中的估计方法因为拥有坚实的数学基础、统计学基础及其良好的应用性,自凯恩斯革命之后在宏观经济学的经验研究中得到了广泛的应用。

估计方法在宏观计量模型构建中能够得以有效应用需要满足一些条件。一方面,需要对所估计的函数形式进行具体的设定,在函数设定过程中需要政策不会变化这一条件,因为政策的变化会使得所估计的方程发生结构性突变。从现实来看,宏观经济时间序列的结构变化是一个普遍的现象。而所估计方程的结构性变化意味着模型内部环境遭到严重破坏,这样估计也就失去其原先的意义。换言之,被估计的模型可能会很好地拟合特定历史数据,却不能保证它同样能够刻画所研究问题的未来分布。另一方面,估计方法应用的一个基本前提是需要有观测到的实际信息资料。但事实上,要对刻画现实世界一些特征的重要宏观经济参数进行估计,所需的相应信息资料很可能要么取得的成本非常之高,要么根本不可能获得。此外,多数宏观计量研究是基

于时间序列数据,而时间序列分析的一个基本要求便是大样本和平稳性。但在实际的宏观经济经验研究中,这个基本要求并不是那么容易得到满足。

2. 校准方法的基本内涵。校准方法是在对理论模型进行大量随机模拟的基础上,总结出有关宏观经济变量波动的统计特征,并与真实数据相比较,从而判断模型的有效性。其基本内涵蕴含在实际经济周期理论的分析框架之中。从实际经济周期理论的建模及其数值求解过程来看,校准方法一般经历如下步骤:

第一步是构建用来刻画宏观经济的理论模型。以Kydland & Prescott(1982)为例,在他们的模型中,代表性家庭的期望效用函数为:  $\max E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, \alpha(L)l_t)$ , 其中  $\beta$  是主观贴现因子,  $c_t$  是  $t$  期的消费,  $l_t = 1 - n_t$  是  $t$  期的闲暇,  $L$  是滞后算子, 且  $\alpha(L) = \sum_{i=0}^{\infty} \alpha_i L^i$ ,  $\alpha_i = (1 - \eta)^{i-1} \alpha_1$ ,  $0 < \eta \leq 1$ 。即期效用函数采用通常的CRRA形式:  $u(c_t, \alpha(L)l_t) = (c_t^\gamma (\alpha(L)l_t)^{1-\gamma})^\gamma / \gamma$ 。

家庭的预算约束为:  $c_t + i_t \leq f(\lambda_t, k_t, n_t, y_t)$ , 其中生产函数的形式为  $f(\lambda_t, k_t, n_t, y_t) = \lambda_t n_t^\theta [(1 - \sigma)k_t^{-\nu} + \sigma y_t^{-\nu}]^{-(1-\theta)/\nu}$ 。这里,  $\lambda_t$  是生产率水平,  $n_t$  是就业水平,  $k_t$  是资本存量,  $y_t$  是存货, 且  $0 < \theta < 1$ ,  $0 < \sigma < 1$ ,  $\nu > 0$ ;  $k_{t+1} = (1 - \delta)k_t + s_{j,t}$ ,  $s_{j,t+1} = s_{j+1,t}$ ,  $j = 1, \dots, J - 1$ , 这里,  $\delta$  是资本的折旧率,  $s_{j,t}$  是  $t$  时期建成的项目将成为  $j$  期后的生产资本, 并在  $t + j$  期用于生产;  $i_t$  是  $t$  期的投资, 且  $i_t = \sum \varphi_j s_{j,t} + y_{t+1} - y_t$ 。最后, 假定生产率水平  $\lambda_t$  具有如下形式,  $\lambda_t = \lambda_{1t} + \lambda_{2t} + \bar{\lambda}$ , 其中,  $\lambda_{1t}$  和  $\lambda_{2t}$  分别代表生产率冲击的长期和短期因素, 且  $\lambda_{1,t+1} = \rho \lambda_{1t} + \xi_{1t}$ ,  $\lambda_{2,t+1} = \xi_{2t}$ , 随机变量  $\xi_{1t}$  为独立同分布。假定向量  $x_t = (\lambda_{1t}, \lambda_{2t})$  服从一个马尔可夫过程。

由上可见,家庭的最优行为是通过求解模型中代表性家庭终生效用的最大化来反映。由于Kydland & Prescott(1982)考虑的是一个社会计划者问题,模型中生产者的最优行为并非清晰可见,而是通过对生产函数  $f(\lambda_t, k_t, n_t, y_t) = \lambda_t n_t^\theta [(1 - \sigma)k_t^{-\nu} + \sigma y_t^{-\nu}]^{-(1-\theta)/\nu}$  的界定,并融入家庭的预算约束之后隐含在模型的求解过程之中。需要说明的是,社会计划者问题与竞争性均衡问题实际上是等价的。因此,以Kydland & Prescott(1982)为

代表的实际经济周期模型所构建的是以经济中各种代表性行为主体即家庭和企业的跨期最优化行为为基础的结构模型,这种结构模型是建立在行为主体不会随着政策的变化而改变的一些固定的特性如家庭的偏好和生产者技术的基础上的。更为重要的是,这种模型所考虑的最优化问题中已经纳入了微观行为主体对未来的预期。

第二步是对模型参数进行赋值。在 Kydland & Prescott(1982)的论文中,在所构建的模型中需要确定的参数有:偏好参数  $\alpha_0$ 、 $\eta$ 、 $\varphi$ 、 $\gamma$  和  $\beta$ ; 技术参数  $\sigma$ 、 $\nu$ 、 $\theta$ 、 $J$ 、 $\varphi_i$ 、 $\delta$  和  $\lambda$ ; 技术冲击参数  $\rho$ 、 $var(\xi_i)$ 。

对于模型中这些参数的赋值,其来源是不同的。第一个来源是通过研究微观经济行为所得到的系数来为参数赋值,如上述模型中的  $\beta$  和  $\varphi$ ; 另外,由粗略的经验研究或不相关联的计量研究来获得参数值,如  $\rho$  和  $var(\xi_i)$ 。第二个来源是通过模型的稳态行为与经济运行的长期特征相匹配的过程来确定

参数值,如前述模型中的参数  $\gamma$ 、 $\theta$ 、 $\delta$ ,它们是通过将其表达成关于消费收入比、资本收入比等现实经济中长期不变的比率的函数来获得的。这些现实经济中的比率主要来自于国民收入核算表的一般事实。第三个来源是为了使模型准确模拟历史数据的某些特征或者说“典型事实”而人为选定的,即自由参数的选择。在这里,自由参数值的选择不能仅考虑到与模型试图解释的数据相匹配,而是根据独立的观测值进行选择。特别是,参数值的选择使得模型确定性均衡中的变量能够匹配它们数据中对应部分的平均值或者说使得模型在稳态下的预测与长期观测数据的平均值相匹配。比如,Kydland & Prescott(1982)论文中的  $J$ 、 $\varphi_i$ 、 $\alpha_0$ 、 $\eta$ 、 $\sigma$ 、 $\nu$  就是自由参数。

第三步是运用校准后的模型来产生模拟的时间序列。其间要经历足够多次(比如 5000 次或是 20000 次)的重复模拟实验,最终把模拟出的数据的二阶矩与真实数据的二阶矩进行比较,并由此进行相关模型的对比,进而比较各个模型的得失及其优劣。

表1 产出的自相关关系

自相关阶数(k)	1	2	3	4	5	6
模拟数据	0.71 (0.07)	0.45 (0.12)	0.28 (0.13)	0.19 (0.12)	0.02 (0.11)	-0.13 (0.12)
真实数据	0.84	0.57	0.27	-0.01	-0.20	-0.30

注:括号中的数据为相应的标准差,下同。资料来源:Kydland & Prescott(1982)。

表2 模型经济与真实经济的特征值比较(美国经济:1950:1-1979:2)

	产出	消费	投资	存货	小时	生产率
标准差(%)						
模拟数据	1.80 (0.23)	0.63 (0.09)	6.45 (0.62)	2.00 (0.20)	1.05 (0.13)	0.90 (0.10)
真实数据	1.8	1.3	5.1	1.7	2.0	1.0
与产出的相关系数						
模拟数据		0.94 (0.01)	0.8 (0.04)	0.39 (0.06)	0.93 (0.01)	0.90 (0.02)
真实数据		0.74	0.71	0.51	0.85	0.10

资料来源:Kydland & Prescott(1982)。

仍以 Kydland & Prescott(1982)的论文为例。表1给出了模型经济和真实经济中产出的自相关关系,可以看出,二者基本接近。表2比较了模型经济与美国真实经济的特征值,不难看出,无论是模拟数

据还是真实数据,从相关关系来看,消费、投资和劳动时间等变量都与产出有正向的相关关系;从标准差来看,产出、消费、劳动时间和生产率的波动幅度较小,而投资的波动幅度较大。但作为高度抽象和

简化的模型经济,模拟出的数据也存在与真实数据不相一致的方面,比如,模型经济中生产率与产出之间存在高度的相关关系,而真实数据表明二者之间的相关关系非常小。后来的学者针对 Kydland & Prescott(1982)模型所存在的局限性做了一些改进。比如,Hansen(1985)提出了不可分劳动模型,其模拟的结果较之于 Kydland & Prescott(1982)的模型更加接近于美国经济的特征事实,从而提高了 RBC 模型的解释力。

## (二)估计方法与校准方法的分歧点

1. 估计方法与校准方法的出发点是完全背离的。校准方法的出发点是一个稳定的理论模型,一旦最优求解完毕,那么理论模型将是不变的,而后的参数校准和经验评估等工作全部是在这个不变的理论模型的基础上展开的。从这个意义上说,在校准方法应用过程中,模型是投入。

而估计方法乃至整个计量经济学的出发点是现实的数据。正如一些学者所分析的,计量经济学对客观经济世界的探索,蕴含着从“现实经济世界到概率空间的映射——概率空间到概率模型的映射——概率模型到计量经济模型的映射”这一过程(王美今等,2012)。虽然估计方法通常也会借助某特定的经济理论来建立可供经验研究之计量模型或者说经济理论在计量经济模型的设定中具有重要的导向作用,但经济理论模型存在的意义却不是一切工作的终点,而恰恰相反,它只是一个起点。与此同时,数据关系的导向作用很大程度上影响着计量模型的设定,因此随着计量经验研究的推进和深入,经济理论模型可能被证实或被修正甚至被证伪。从这个意义上说,在估计方法应用过程中,模型是产出。

2. 估计方法与校准方法对模型优劣的评判标准是不同的。就估计方法而言,对模型优劣的评判取决于模型在多大程度上拟合了数据。估计方法需要有一系列严格的假设检验,在这些假设检验的帮助下,不断对模型进行各方面修正,最后寻求一个最高拟合优度的计量模型。在这里,初始的理论模型只不过是一个模子,可以对它进行各种各样的添减加工,使其成为最符合当前数据特征的计量模型。正如 Leamer(1983)所指出的,回归分析中模型假定以及控制变量选择的随意性导致了估计结果的脆弱性。在估计者们看来,似乎没有普遍正确的模型或是理论,而只有最适合于现实数据的模型或者理论。

但新古典宏观经济学家特别是卢卡斯、基德兰德和普雷斯科特却反对把统计意义上的拟合度作为

评判理论或模型优劣的标准。就校准方法而言,对模型的评价在很大程度上是由模型模拟的结果与真实数据的方差和协方差相匹配的程度决定的。它只要求模型模拟出的各变量二阶矩能够与真实数据相匹配就可以了,它没有一个类似于估计中拟合优度的衡量指标,可以帮助说明模型确切的质量到底如何。这也是校准方法长期被经济学界所诟病的一点,它不像估计方法那样有严谨的检验过程。

而对于校准派来说,之前选定的模型已经是绝对正确的了,那么我们就没有必要再纠缠于模型到底是否是对的,我们所要知道的就是模型是否足够好,即通过对参数的校准能否较好地模拟现实。比如,在 Kydland & Prescott(1982)的论文中,即使在工作时长并非是一个有效变量的情况下,他们依然对模型模拟的结果能够基本符合真实数据的二阶矩而感到满意。即便是近些年发展起来的动态随机一般均衡模型,因为它具有较为简单的动态结构,所以在校准之后其拟合的程度通常也会低于被估计的情形。但由于我们对长期参数值的了解会更多些也更有把握一些,从这个意义上说,更多关注模型长期解的校准法可能是一种更好的评判模型优劣的方法。

从校准的角度来看,利用估计方法能够很好地拟合数据的模型未必就是要选用的模型,而无法很好地拟合数据的数量化模型并不能成为拒绝一个核心理论的理由。换言之,每一个基于核心理论的特定模型若不能很好地拟合数据并不能成为拒绝该核心理论的理由。核心理论就是一些普遍成立的一般性规律,相当于一个拉卡多斯内核,<sup>①</sup>本身不会也不应该受到质疑。比如,Altug(1989)就 Kydland & Prescott(1982)论文中同样的问题进行了最大似然估计,而结果是否定了基德兰德和普雷斯科特所认为的满意的结果。但从校准的角度看,这种估计是没有意义的。Prescott(1998)认为,从校准的角度看,模型好比温度计。用模型来刻画经济运行好比用温度计来测量温度,模型和温度计都是衡量现实世界某些特性的工具。直接去估计模型本身就像对一个通用的温度计上的刻度进行重新设定一样,是没有意义的。

Hoover(1995)将校准方法描述为“适应性战略”(Adaptive Strategy),而将估计方法描述为“竞争性战略”(Competitive Strategy)就非常精辟地阐明了这个差异。校准的目的是要让校准后的模型能够模拟出现实的典型特征,从而使现实能够适应于所选定的经济理论或是模型。估计方法则是有一系

列的模型或是理论可以作为备选,这些模型或理论要互相竞争,最终最符合当前数据的那个模型将被认为是“对”的。

3. 校准方法与估计方法对于模型的现实性有不同的看法。估计方法的思想很简单,尽量使计量模型贴近现实数据即可,因为它本来要寻找的就是最贴近现实的模型。而校准方法因为已经强调了已有模型是正确的模型,所以要把握的是这个模型所能告诉的规律。而“经济周期都是类似的”(Lucas, 1977),理论模型只要能够使模拟出来数据的二阶矩和真实数据的二阶矩相匹配即可,这似乎是一种更加高屋建瓴的做法。可以认为,估计方法更多的是针对某些具体的或局部的问题,而校准方法关注的是一些更具有普遍性的或是一般性的规律。

### (三)校准方法和估计方法各自的优缺点

1. 相对于经济计量方法来估计模型,校准方法具有以下三个优点:第一,基于校准模型所进行的预测将不会出现卢卡斯批判所指出的问题,因为校准的模型是基于真正的不变量如技术、偏好和禀赋来建构的,同时在纳入预期的基础上考虑了微观主体的动态最优化行为。这也是校准方法最大的优点。比如,在 Kydland & Prescott(1982)中,对模型参数进行校准时把参数  $\varphi$  确定为  $1/3$  是考虑到美国大部分家庭通常分配  $1/3$  的时间用于市场活动这一事实。显然,家庭的这种时间分配即所校准的参数不会因政策变动而发生改变。类似的参数校准加之微观主体的动态最优化使校准的模型可以避免卢卡斯批判所指出的问题。从这个意义上说,被校准的模型比被估计的模型将具有更强的预测能力。

第二,在模型校准过程中,一部分参数值的选择是以微观经济资料为基础的,除了通常采用的信息之外,大量的微观信息得到了采用。特别是,校准通过对关键参数的微观估计,它不会因为模型在加总过程中导致个体行为人的信息丢失。比如, Kydland & Prescott(1982)论文中主观贴现因子  $\beta$  是由经济处于稳态时的利率水平来确定的,由于美国的年均实际利率为  $4\%$  左右,因此其季度实际利率为  $1\%$  左右。而估计方法则直接采用总量数据,并研究总量之间的关系,既缺乏微观基础,又有可能丢失大量的信息。实际上,校准方法正是利用微观经济证据来约束或限制模型相关的参数和函数形式,所以相对于估计的模型来说,被校准的模型建立在一个更高的标准之上。

第三,从估计方法的角度看,即便模型在一个不

重要的方面与数据拟合得不好,而在其他所有方面都拟合得很好,这个模型仍然可能在统计上被彻底拒绝。同时,不能拒绝一个模型的原因仅仅在于,数据与多种可能性相一致。此外,对所估计的模型在统计上被拒绝或不能被拒绝,其经济含义通常难以解释。而校准方法则不存在类似问题。由于模型在校准过程中通过足够多次的模拟实验来计算模型的二阶矩,因此校准方法在一定程度上可以克服估计方法由于缺乏实验的可能,同时又囿于现实中数据的可获得性所带来的局限性。

2. 相对于估计方法来说,校准方法有其自身的软肋:第一,校准方法的可靠性似乎不如估计方法。在早期,校准方法因为对于模型和理论的绝对信任,从而不会或很少会认真地对其结果进行类似于估计那样的稳健性检验,因此其结论往往并不像估计方法那样具有说服力;并且校准方法缺少类似于估计方法中的拟合优度这样的评判指标和一系列关于参数的检验方法,因此我们很难对校准模型的优劣进行评判。特别是,如果校准的参数来自其他相关研究,而其他相关研究很可能是基于局部均衡而非一般均衡的分析,由此,我们不能不对参数的这种直接借用的合理性表示疑问。但估计方法却可以让人一目了然地看出依据这种方法研究所得到的结论到底在多大程度上值得信任。另外,校准方法实际上在很大程度上依赖于选定的模型,因此不禁会让人产生一个疑问,即复杂的经济事实是否可以武断地由这样一个模型来描述。而模型校准过程中,无论是“基准年”的选择,还是自由参数值的选取都具有一定的随意性。这多少有点给人不放心的感觉。

第二,校准方法缺乏像估计方法那样的方法论基础。从方法论基础来看,前已述及,估计方法拥有很好的数学基础(主要是概率论基础)和统计学基础,但校准的方法论基础则不甚明了。但也必须指出的是,估计方法对于统计学和数学的应用也是有条件的,它需要一系列严格的假定得以满足,而这点往往会被一些研究者有意或无意地忽视。比如,估计方法对数据有非常苛刻的要求,不仅仅体现在数据量上,更对数据的质量和可靠程度提出很高的要求。而校准方法虽然缺乏明确的方法论基础,但它在模型的指引下,对现实数据的要求相对低些,同时可以借助数值方法和计算机编程来解决校准过程中涉及到的一些难题。

### (四)估计方法与校准方法共同面对的难题

与估计方法类似,在模型校准过程中,给定模型

形式的多样性,模型与数据的匹配具有相当大的灵活性。所以,当模型与真实数据重要的“矩”匹配得很好时,也不能简单地就说模型就具有较好的解释能力。此外,由于没有对这些模型与其备择模型进行检验,我们也不知道是否存在其他甚至完全不同的模型,它们同样能对真实数据相关的矩匹配得很好。有关“矩”的匹配是否应该作为校准模型的一个合意的特征目前尚无定论。

相比之下,估计方法更加依赖于繁琐的检验,在经历过重重检验之后,从结果上看似乎是很可靠的。但是事实也可能并非如此,因为这样估计出来的模型可能遭遇卢卡斯批判,导致基于估计出来的模型所进行的任何预测没有意义甚至完全错误。借用 Simon(1969)的一个类比,即一个人工制品能够正常发挥作用,需要内部环境和外部环境的有效运作。在校准和估计方法下,内部环境就是指模型或是理论,外部环境则是指外部的冲击,例如技术冲击、外部需求冲击和政策变化等。对于校准方法来说,就是预先设定好内部环境,并且这个内部环境是绝对稳定的,与外部环境是严格隔绝的,这样的结果似乎会更可靠。但不能不看到的是,这个可靠性是建立在内部环境本身值得信赖的基础上的。但“基准年”的随意选定以及自由参数的存在,都会让这个内部环境的可靠性蒙上一层阴影,而且就是校准方法的信奉者们一直自矜的内外部环境的绝对独立本身也是存疑的,因为参数取值的来源途径不仅有政府的统计,还有估计所得到的数据,校准方法并不能从根本上脱离估计方法,因此它本身对估计方法的一系列指责都有可能反映在它自己身上。

而估计方法所带来的的是一个变动的内部环境,并且内部环境和外部环境并不独立,因此内部环境的稳定依赖于外部环境的稳定。一旦外部环境发生变化,那么原来的内部环境就会随之崩溃,还遵照原来的内部环境进行运作将会导致致命的错误。即使能够对内部环境进行相应的调整,但这也意味着内部环境的震荡,那么人们不禁会问,在这种震荡的内部环境下输出的结果又有多少可值得信赖。

#### 四、估计方法与校准方法之争 的本质及其融合发展

##### (一)校准方法与估计方法之争的本质

1. 校准方法与估计方法之间的根本性差异。估计和校准这两种方法都在不同程度上依赖并利用经

验数据,努力寻找支配经济现象背后的运行规律。从表面上看,校准方法是估计方法的一种逆向推演,即先有模型,然后对参数赋值,最后跟真实数据进行比对。但它们之间的差别绝非逻辑推演方向的不同这么简单。看上去殊途同归的校准方法和估计方法,在思想上却存在着根本性的差异,这个差异体现在对经济现象背后运行规律寻求方式的不同。而它们在其他方面的所有差异也不过是这个根本性差异的具体体现。

具体而言,校准方法从一个稳定的宏观经济模型出发,随着研究问题的侧重点不同,模型可能会彼此有所差别,但是这一系列模型的根源却几乎是相同的。就实际经济周期模型而言,基于前面所讨论的“核”的思想,后续的系列经济周期模型只不过是这个“核”的基础上,根据所研究问题的实际需要,进行扩展或者少许的修改。比如,以 King, Plosser & Rebelo(1988)的 RBC 模型为基础, Benhabib, Rogerson & Wright(1991)分析的家庭生产模型, Burnside, Eichenbaum & Rebelo(1993)研究的劳动窖藏模型,以及 Baxter & Crucini(1993)把模型拓展到开放经济情形,等等,都只不过是在 RBC 模型“核”的基础上所做的扩展和延伸。在这个过程中,扩展的部分或是变动的部分有可能是错误的,但是“核”一定是没有错误的,因此这个“核”就是前文中所描述到的稳定的、不容置疑的规律。校准方法一定是承认这个“核”的存在,并且以它为基础进行建模,然后进行数值模拟和比较等一系列的工作。

估计方法却不一样。估计方法最为强调的是经济计量模型与现实的最大程度的拟合,所以对于它来说只有最适合现实的模型或是理论,而不存在一个绝对正确的模型或是理论。而在宏观经济学中,对于同一个研究对象,各种迥然不同的理论同时存在是很正常的。比如,以居民消费为研究对象,分别存在着凯恩斯消费理论、莫迪利安尼的生命周期假说、弗里德曼的永久收入假说和霍尔的随机游走假说等。估计方法所要进行的工作就是要在众多的可能正确的模型或是理论中找到和实际最相符合的模型或是理论。显然,从校准的角度看,这种状况是不可能出现的。

2. 校准方法与估计方法之争的本质。校准方法与估计方法的根本性差异可以追溯到早期学界关于经济学理论及其假设的现实主义问题的讨论,代表性的论文是 Friedman(1953)。该文被誉为战后经济学方法论的中心著作,其中心论题是经济学家

不必为他们假设的“现实”性问题而烦恼。他指出,理论被看作是一个主要是假说的实体,对它应该用它所要“解释”的那一类现象的预言力来评判。“对某一假说的合理性的惟一有关的检验,是将其预测与实际情况所做的比较。如果该假说的预测(‘频繁’或比来自另一种假说的预测更为经常地)与实践相抵触,那么该假说就遭到了否定;如果该假说没有与实践相抵触,那么它则为人们所认可;如果该假说业已多次成功地避免了可能出现的抵触现象,那么它则具有极大的可依赖程度。实际证据永远不可能‘证实’某一假说的正确性,它只能通过无法将该假说驳倒来显示该假说的正确性。当我们说到某一假说已经在实践中得到了确认(并不十分准确),我们通常所指的就是这个意思”(弗里德曼,1991,第197页)。实际上,在弗里德曼看来,重要的不是理论及其假设的现实性问题,而是理论的解释能力和预测能力。

Lucas(1980)进一步认为,一个“好”的模型并不一定比不好的模型更“真实”,但是会模拟得更好,从而更具有解释力。由此看来,倡导校准方法的宏观经济学家显然继承并发扬了弗里德曼的这一思想,即构筑在一定假设上的模型或者理论一定有其正确的部分,进一步,如果模型能够很好地模拟出可以反映现实经济规律的特征事实,那么它便是好的。但校准方法也并非是在搭建空中楼阁。校准方法更像一种实验室的方法,它利用模型构建的是另一套虚拟的经济系统,在经过校准之后,这个虚拟的系统既可以用来模拟现实,也可以用来进行政策模拟试验。在虚拟经济系统模拟现实的过程中,需要有一个桥梁将二者进行连接,这个桥梁往往是外部冲击。由于冲击必须是符合现实特征的,因此在模型的系统中进行经济波动模拟就具有现实意义。在Kydland & Prescott(1982)的论文中,所要探讨的问题是技术冲击是否可以解释美国经济波动。论文从美国的索洛剩余中寻找技术冲击,在校准后的模型中进行模拟,由于模拟出的结果和真实经济的主要特征相符,因此经校准的模型被认为成功模拟了现实。

而估计派显然并不完全支持弗里德曼的观点,他们更加信赖现实的数据,更加崇尚现实的绝对正确性。由此我们认为,校准和估计这两种方法的核心差异在于方法论思想的根本差异,在于是否承认宏观经济现象背后存在着一个稳定的、不容置疑的并且普遍成立的规律。这正是校准与估计之争的本质所在。校准与估计方法之争实际上也反映了宏观

经济学家对社会经济运行基本认识之不同。

## (二)校准方法与估计方法的融合发展

校准方法和估计方法各有其优点,也各有其缺点,因此简单地讲哪一个更好是欠思考的而且也是不负责任的。可以说,校准方法与估计方法适用的对象是大不相同的。从应用对象上来说,估计方法似乎在解决一些具体的问题上更有优势,因为它更专注于现实;而校准方法则能获得更具有普遍适用性的结论,更能够解决一系列共性的问题,因为它更侧重于现实背后的“规律”,尤其是在宏观经济研究领域。实际上,从校准方法走入经济学家视野至今,两个方法的支持者们之间的论战就没有停止过。而如今两种方法都活跃在经济学界,因此两种方法更可能是不分优劣,选用哪种方法可能取决于现实情况和研究对象之不同。但从总体上看,就宏观经济学经验研究而言,校准与估计方法的相互促进及其融合发展已然成为一种趋势。

首先,必须指出的是,校准方法与估计方法之间并非是完全相互排斥的。事实上,在模型校准过程中并不意味着对估计方法的拒绝,相反,估计方法在宏观经济模型的校准过程中得到了大量的应用。比如,Kydland & Prescott(1982)通过估计索洛残差来构建生产率时间序列,然后利用构建好的生产率时间序列来估计生产率冲击所遵循的马尔可夫过程。因此,在宏观经济理论的实际经验评估过程中,校准与标准的计量经济学方法即估计是相容的。有趣的是,在基德兰德和普雷斯科特看来,计量经济学与估计属于不同层面的范畴,他们并不拒绝计量经济学,而且认为校准是计量经济学。

其次,校准方法的提出在客观上促进了估计方法的改进和发展。校准方法的提出最初是为了解决卢卡斯批判所提出的问题,而恰恰是这一点为计量经济学估计方法的改进和发展指明了方向。为了应对卢卡斯批判,计量经济学在传统估计方法的基础上发展出了向量自回归模型(VAR)。虽然向量自回归方法由于缺乏坚实的经济理论基础以及脉冲响应分析不一定可靠而遭人质疑,但它却是估计方法针对卢卡斯批判所做出的一个重要改进。其改进的核心思想就是在原有静态估计模型中引入动态因子,从而具有一定的试验性并尝试使估计方法具有结构性的特质。VAR方法把ARIMA(差分自回归移动平均)模型发展到多个时间序列向量,用模型中所有当期变量对所有变量的若干滞后变量进行回归。它刻画了数据的动态表现,并加入了脉冲响应

分析。而正是脉冲响应函数的加入使得 VAR 模型具有一定的试验性。后来为弥补 VAR 模型的缺陷,并让 VAR 模型能够更好地进行结构分析,在其基础上又演化出了结构向量自回归(SVAR)模型。这样,SVAR 模型与经校准的模型从形式上看具有惊人的相似之处。此外,我们还可以通过模型校准,来为估计提供一系列新的想法与思路,特别是在估计的变量选取上。比如,Hansen & Sargent(1988)使用校准方法大大减少了待估模型的参数数量,从而简便了模型的估计。

再次,估计方法也为校准方法的改进提供了强有力的手段。借助估计方法,校准方法所做的改进主要围绕参数值选择的科学性和模型校准结果的可靠性来进行。具体的做法是,放弃 Kydland & Prescott(1982)等早期在自由参数选择上的随意性,而在参数值的选取上越来越多地赋予了估计的色彩。对于校准派来说,他们开始更加关注校准模型的稳健性。以 Smets & Wouters(2007)一文为例,该文使用贝叶斯估计来得到相应的参数值,在传统校准方法基础上加入了更多的随机因子,并通过样本外估计、子样本估计等一系列方法来检验经校准过的模型的解释能力和预测能力。前已述及,Kydland & Prescott(1982)在对技术冲击的参数设定上,依据的是对索洛剩余的估计,而 Smets & Wouters(2007)对于冲击的参数均采用贝叶斯估计来获取,同时对于不可观测的参数,如风险厌恶系数也使用贝叶斯估计。

估计方法不仅可以为模型的参数校准提供帮助,而且可以对校准的模型予以评价,即近期经济学界所提出的一些所谓的“最优校准方法”。使用最优校准方法时,第一步先要运用传统的估计方法将数据与模型进行拟合,这个模型就叫辅助模型,它可以是 VAR 或 VARMA(向量自回归移动平均模型)。第二步是校准 RBC 模型,并且通过模型产生人工数据(通常在获得模拟值时,并不需要对模型进行线性化处理)。第三步则是使用模拟出来的数据估计辅助模型。对于 RBC 模型的不同校准值,重复进行第二步和第三步。所谓最优校准是指,基于模拟的数据所估计出来的辅助模型最接近于基于观测数据所估计出来的辅助模型。可以通过多种方法将两组数据集进行比较:比较辅助模型估计出来的系数或比较似然函数比、似然函数值(迈克尔·威肯斯,2011)。

最后,动态随机一般均衡模型为校准与估计的融合提供了很好的平台。从以上分析可以看出,估

计方法的完善思路是在原本静态的模型中加入动态性和结构性,而校准方法则是通过吸收一些类似于估计的思想来获得更为精确和可靠的参数值,更加强了随机性,并在模型校准过程中增加类似计量经济学模型中的稳健性检验,从而提高了校准模型的可靠性。换言之,就估计方法和校准方法而言,前者纳入动态性和结构性,后者则吸收了随机性。而动态随机一般均衡模型完美地融合了动态性、结构性和随机性。从这个意义上说,动态随机一般均衡模型的发展不仅是两大宏观经济学理论流派即古典、新古典宏观经济学与凯恩斯主义、新凯恩斯主义宏观经济学的融合发展的成果,而且是估计方法与校准方法相互借鉴并彼此吸收对方优点的产物。基于 DSGE 的分析框架,校准和估计在宏观经济学经验研究中的交叉组合及其应用,并用来分析特定的问题已经成为一种常态,其融合发展之势渐趋明显。

## 五、结束语

综上所述,我们认为,校准与估计之争反映了现代宏观经济学两大流派即凯恩斯主义、新凯恩斯主义与古典、新古典宏观经济学派之间研究范式的分歧。从目前来看,宏观经济学无论就其理论体系还是就其研究方法来说,都处于不断的发展、完善和融合之中,而动态随机一般均衡方法则为这一系列的融合提供了一个很好的平台。在 DSGE 框架下,既可以容纳两种研究范式迥异的宏观经济理论,也可以同时利用两种风格完全不同的研究方法。因此,无论是从两个流派的理论发展还是从两种经验研究方法的整合发展来看,随着两种方法的相互吸收和借鉴,同时也随着正规的计量经济学方法对 RBC 类型的随机动态优化模型进行估计和检验技术的突破,可以预见,估计方法与校准方法的融合发展将会更加深入。

### 注:

- ① 现在所谓的“结构参数”是指在经济行为主体的动态最优化过程中,刻画偏好、技术、禀赋以及制度等因素的深层参数(Heckman, 2000; Reis & Wolak, 2007)。
- ② 拉卡多斯内核是拉卡多斯(Lakatos)在其《科学研究纲领方法论》中详尽论述的一个概念。简单说来,就是对于任一科学理论体系,一般都由核心和外围组成,核心是最基本的理论,外围则包含了许多假定。核心(很少或)不会同现实发生直接的联系,它依赖于外围来和现实加以联系并对现实问题进行推断。如果推断与现实不符,那么被证伪的并不是内核本身,而是作为桥梁的外围。

### 参考文献:

龚刚,2004:《实际商业周期:理论、检验与争议》,《经济学(季

- 刊)》第4期。
- 洪永森,2007:《计量经济学的地位、作用和局限》,《经济研究》第5期。
- 黄贇琳,2005:《校准法的原理、应用与发展方向》,《数量经济技术经济研究》第1期。
- 李子奈,2009:《计量经济学模型对数据的依赖性》,《经济学动态》第8期。
- 李凌 权衡,2009:《宏观计量经济学研究方法的演进与比较》,《经济学动态》第5期。
- 李子奈 齐良书,2010:《关于计量经济学模型方法的思考》,《中国社会科学》第2期。
- 王美今 林建浩,2012:《计量经济学应用研究的可靠性革命》,《经济研究》第2期。
- 米尔顿·弗里德曼,1991:《弗里德曼文萃》,高榕等译,北京经济学院出版社。
- 迈克尔·威肯斯,2011:《宏观经济理论:动态一般均衡方法》,段鹏飞等译,东北财经大学出版社。
- 朱晓冬,2005:《现代宏观经济分析的一个可行框架:关于 Finn Kydland 和 Edward Prescott“建设投资的时滞和总体波动”一文的评述》,《经济学报》第1卷第1辑,清华大学出版社。
- Altug, S. (1989), "Time to build and aggregate fluctuations: Some new evidence", *International Economic Review* 30:889-920.
- Baxter, M. & M. J. Crucini(1993), "Explaining saving-investment correlations", *American Economic Review* 83: 416-436.
- Benhabib, J., R. Rogerson & R. Wright (1991), "Home-work in macroeconomics: Household production and aggregate fluctuations", *Journal of Political Economy* 99: 1166-1187.
- Burnside, C., M. Eichenbaum & S. Rebelo(1993), "Labor hoarding and the business cycle", *Journal of Political Economy* 101(2):245-73.
- Christiano, L. J. & M. Eichenbaum(1992), "Current real business cycles theories and aggregate labor market fluctuation", *American Economic Review* 82:430-450.
- Friedman, M. (1953), "The methodology of positive economics", in: *Essays in Positive Economics*, University of Chicago Press.
- Gregory, A. W. & G. W. Smith(1990), "Calibration as estimation", *Econometrics Reviews* 9:57-89.
- Hansen, G. D. (1985), "Indivisible labor and the business cycles", *Journal of Monetary Economics* 16:309-28.
- Hansen, G. D. & T. J. Sargent(1988), "Straight time and overtime in equilibrium", *Journal of Monetary Economics* 21:281-308.
- Hoover, K. D. (1995), "Facts and artifacts: Calibration and the empirical assessment of real-business-cycle models", *Oxford Economic Papers* 47:24-44.
- King, R., C. Plosser & S. T. Rebelo(1988), "Production, growth and business cycle", *Journal of Monetary Economics* 21:195-232.
- Kydland, F. & E. C. Prescott (1982), "Time to build and aggregate fluctuations", *Econometrica* 50:1345-1370.
- Kydland, F. E. & E. C. Prescott(1991), "The econometrics of the general equilibrium approach to business cycles", *Scandinavian Journal of Economics* 93:161-78.
- Leamer, E. E. (1983), "Let's take the con out of econometrics", *American Economic Review* 73: 31-43.
- Liedo, D. (2006), "Priors from real business cycle model in dynamic factor models", Working Paper, ECARES, Université Libre de Bruxelles.
- Lucas, R. E. (1977), "Understanding business cycles", in: *Studies in Business - Cycle Theory*, Blackwell, Oxford, 1981.
- Lucas, R. E. (1976), "Econometric policy evaluation: A critique", *Carnegie Rochester Conference Series on Public Policy* 1:19-46.
- Lucas, R. E. (1980), "Methods and problems in business cycle theory", *Journal of Money, Credit and Banking* 12:696-715.
- Manuelli, R. & T. J. Sargent(1988), "Models of business cycles: A review essay", *Journal of Monetary Economics* 22:523-32.
- McCandless, G. (2008), *The ABCs of RBCs: An Introduction to Dynamic Macroeconomic Models*, Harvard University Press.
- Prescott, E. C. (1998), "Business cycle research: Methods and problems", Federal Reserve Bank of Minneapolis Research Department, Working Paper 590.
- Romer, D. (2006), *Advanced Macroeconomics*, McGraw-Hill.
- Smets, F. & R. Wouters(2007), "Shocks and frictions in US business cycles: A Bayesian DSGE approach", *American Economic Review* 97:586-606.
- Simon, H. A. (1969), *The Sciences of the Artificial*, MIT Press.
- Wichens, M. (2008), *Macroeconomic Theory: A Dynamic General Equilibrium Approach*, Princeton University Press.

(责任编辑:陈建青)