

# 如何正确的运用计量经济模型进行实证分析？<sup>1</sup>

## —————实证计量经济学中的数据、模型 与参数

德国哥廷根大学教授 于晓华

### 摘要：

本文从实证计量经济学的的数据种类，模型结构，以及参数估计的稳健性三个角度出发，具体指出如何正确使用计量经济模型来分析实际经济问题，得出稳定，合理，可靠的参数估计值，为政策分析提供重要的参考。

**主题词：** 计量经济模型，模型结构，数据种类，参数估计，外生性，政策分析

## 1. 引言

随着中国经济市场化改革的深化，经济学教育也发生了翻天覆地的变化。在不到 20 年的时间内，伴随着一大批从西方留学归来经济学者的努力推广，西方经济学已经被完整地介绍到了中国，占领了中国高等教育机构的讲台；现代西方经济学的理论框架和分析工具，也已经成为了政府经济政策的主要参考依据。中国学者对经济学的研究，也逐步与世界接轨（于晓华，2008；Yu and Gao, 2010）。

跟随世界主流经济学的研究范式，数量化研究已经成为了中国经济研究的主流。经济学，作为社会科学中数量化程度非常高的一门学科；其本身还是脱离不了社会科学本身的限制。经济本身是一个复杂系统，各种变量——可观察的以及不可观察的变量，错综复杂，这限制了经济学科学化，使得经济学对社会经济发展的预测几乎不可能。经济学的主要使命是帮助我们认识复杂的经济世界，更多时候是对经济现象提出合理解释。

一篇严谨的经济学论文，一般需要三个基本的要素：视点(Perspective)，参照系(Benchmark)，以及分析方法 (Analytical Tool) (钱颖一，2002)。视点为论文所要论证的观点；参照系为大家理解经济现实提供了一些基本的比较标尺；而只有分析工具才能够真正帮助我们深入分析纷繁复杂的经济世界。分析工具也通常被称之为“经济学模型”。

模型是对现实世界的一种抽象。由于经济现象本身的复杂性，在实际分析中，我们需要剥离一些对关注的现象无足轻重的变量，抽象出关键变量，根据一些基本或者显而易见的假设，分析这些变量之间的关系，得出一些通常出人意料，但又合理，并且对我们了解现实具有帮助的结论。很多经济学的重要结论并不是显而易见；如果显而易见，那也不需要经济学这门学科。比如，贸易理论中非常重要的比较优势模型，对一些训练有素的经济学家来说是非常简单的结论；但对普通大众来说，却有可能显得异常高深而难以理解 (Pfleiderer, 2014)。

---

<sup>1</sup> 于晓华为德国哥廷根大学 Courant 贫困、平等与增长研究中心，以及农业经济与农村发展系教授。感谢哈佛大学经济系黄炜博士生，上海财经大学经济学院周德威博士，以及澳大利亚阿德莱德大学何晓波博士等为本文的初稿提出了宝贵的修改意见，但文责由本人负责。

支撑现代经济学分析框架就是模型。经济模型在一篇经济论文中，起着非常核心的作用，模型的结构和正确应用，对文章的观点稳健与否，正确与否起着决定性支撑作用。经济学模型可以是描述性模型，也可以是数量化的模型。前者虽非主流，但也不能否定其存在。一个典型代表即为科斯定理。而后者则为现代经济学的主流，在主流经济学期刊和教科书大行其道。非数量化的模型的学术论文，已经很难在主流经济学期刊占有一席之地。因为经济学数量化的优势在于能够容易检验结论的稳健性，和逻辑结构。如果一个经济学结论不合理，数量化的模型可以帮助我们比较容易检视其出错原因：是假设的问题，还是论证过程的问题。

经济学数量化模型大体可以分为数理模型（Theoretical Model）和计量经济模型（Econometric Model）。数理模型从一些简单并且显而易见的现实假设出发，通过理性经济人假设等机制，得出一些有助于我们理解经济现实，通常出人意料的结论，这属于现代理论经济学的主流研究范畴。计量经济模型是结合了经济学理论和统计学方法的对经济现象进行定量分析的方法。这是现代经济学实证分析的主要手法，可以具体的得出一些现实变量之间的实证关系。

Pfleiderer（2014）指出，如果不注重经济学理论模型的假设条件，很有可能导致理论模型结果不稳健，产生“变色龙”一样的结果。这样的现象也同样存在于实证计量经济模型分析中，由于计量经济模型的概率基础，“变色龙”的现象在实证模型中可能尤甚。无论是数理模型，还是计量模型，他们都属于模型的范畴。模型有一些共同的结构特性。不遵守模型的规则，滥用模型，会导致分析结果缺少稳健性，合理性和可靠性。Pfleiderer（2014）已经对现有理论经济学模型中的“变色龙”问题作了一个非常详细总结，所以本文主要对应用计量模型在实证分析中存在的问题作一个总结。

实证分析是计量模型和数据的一个有机结合。理论计量经济学主要关注于开发出适用于现实经济分析的计量经济学模型，是一门严谨的科学。而应用计量经济学（Applied Econometrics）是应用计量经济模型和数据结合来分析实际中的经济问题。一个好的值得得信赖的实证分析，需要注意的操作问题非常多而且琐碎；甚至比理论模型存在的问题更为麻烦。从某种意义上讲，应用计量经济模型进行实证分析是一门艺术。

本文试图提供一个应用计量模型进行实证分析框架。这个框架试图包含实证分析的主要问题。接下来首先论述一般模型的结构特征，然后再从计量经济分析的数据特征，计量模型结构，以及参数稳健性等角度，来分析如何才能得出一个合理（Reasonable）且值得信赖（Reliable）的实证结果。

## 2. 经济模型的结构

凡模型（不仅是经济模型），其结构必存在三部分：环境(Environent)，机制(Mechanism)，以及求解过程(Solution Concepts)。

### 2.1 环境

所谓环境，即为模型的假设条件。任何模型都有一定的适用环境。比如要在长江建一座大桥，首先需要建一个模型。长江上游和下游的环境显然不一样：水流，运输，以及地质条件存在很大的差异。如果不考虑这些条件，直接把上游大桥的模型移植到下游，其结果必然造成很大损失。

同样，任何一个经济模型，也有其适用环境。即每个经济模型都存在一些严格的假设条件。这些假设条件可能是显在的，也可能是隐含的。经济学家的一个主要任务就是要发现模型中的假设条件，放宽或者改变这些假设条件，然后分析其可能的后果。经济学理论的每一次突破，基本上都是改变这些假设条件造成的。不注重经济模型假设的后果是非常严重的；有可能导致论文的结论象“变色龙”一样，从而使文章失去严谨性（Pfleiderer 2014）。

应用计量模型进行实证分析，其假设条件更加复杂苛刻。一方面，实证分析所参照的经济理论存在一些假设条件；另一方面，计量模型本身也存在一些假设。忽视前者的假设条件，通常会导致实证结果无法解释，或者解散存在偏差。忽视后者的假设条件，使得计量模型的参数估计存在偏差。这两方面的例子在实证分析中不胜枚举。

经济学研究者的初学者，通常会把国际权威学术杂志上分析一国（通常为美国）的经济模型应用到另一国（通常为中国）中去。国家之间经济，政治以及文化背景可能存在很大差异。如果不了解这些差异，会导致模型出现适用偏差，或者实证结果的解释出现很大错误。一个例子，现有的实证研究发现在美国欧洲等发达国家，肥胖和教育存在显著的负相关。这样的结果很好理解：良好教育的人很注重自己的健康与形象。但是，一些学者对印尼和越南等发展中亚洲国家的研究却发现肥胖和教育存在显著的正相关。这样的结果使得某些西方学者的解释出现了错误。按照西方的逻辑，为了减少肥胖，难道要降低教育水准？这样的结果差异其实是出于文化的差异。因为在印尼，越南等亚洲发展中国家，肥胖通常是显示一种社会地位。按照中国的俗话，这是一种“富态”。不了解这样的隐含假设，文章结果的解释可能会出现一些低级错误。

计量经济模型本身也存在一些固有的假设，这些假设条件必须在实证分析时小心检视。比如，我们最常用的最小二乘法（OLS），一般存在5个基本假设（Greene, 2012）。这些假设包括：（1）模型是线形的；（2）不存在多重共线性；（3）自变量和误差项之间不相关；（4）误差项同方差，并且误差项之间不相关；以及（5）误差项服从正态分布<sup>2</sup>。对于这5个假设条件，每一个条件都至关重要，保证了最小二乘法结果的存在性，一致性，有效性以及可检验性。一个严谨的应用最小二乘法的实证分析，必须要对这5个条件逐一检视或者检验。如果检验拒绝了这些假设条件，必须要校正模型。比如，如果变量关系不是线性的，我们就要采用非线性的模型。如果自变量和误差项之间不相关（自变量外生性）的假设被拒绝，即模型存在所谓的内生性问题，工具变量方法就成为了一个可行的解决方案。

## 2.2 机制

一旦有了假设环境，我们要识别出模型中的关键变量，找出这些变量之间的关系。这就是模型的机制问题。比如，最小二乘法模型中的机制就是：估计系数使得误差项的平方和最小。

在实证模型中，有些人喜欢非常复杂的模型，喜欢包含尽量多的变量在计量模型中。包含无关的冗余变量在模型中，在样本有限的情况下，可能导致模型的效率下降，使估计结果显得不稳定；在解释结果时也会造成不必要的困难。在另外一些情况下，比如模型为非线性，包含太多的变量，或者使用过分复杂的机制，可能会导致模型无解。相反，如果丢失一个关键变量，有可能造成模型的内生性问题，使得模型的估计结果不能满足一致性要求。权衡计量模型的变量需要很高的技巧，这是一门艺术。

---

<sup>2</sup> 根据渐进近似理论，在大样本的情况，根据大中心极限定理，误差项会趋向于正态分布。

一个好的经济模型不是以复杂程度来衡量；只要能够说明问题的模型都是好模型。通常情况下，“简单最好（simple is the best）”原则可能模型选择。

### 2.3 求解

最后，在确定模型假设条件和机制确定后，求解模型就成了水到渠成的一个问题。在实证分析中，模型的解就是所谓的模型参数估计。

作为一个训练有素的经济学者，应该对计量模型的参数估计会有有一个合理范围的猜测。模型的参数估计在这个范围内，才显得有道理（Make Sense）。如果一个计量模型的估计结果不合理，我们必须回头检视数据，或者模型的假设以及机制，来确认为什么会出现违背经济学常识的现象。

## 3. 经济数据的特征

实证分析是经济计量模型和数据的一个结合。一个好的实证论文首先需要好的数据支撑。对于实证所需数据，我们要注意三个问题：数据的真实性，数据来源以及数据的种类。数据的真实性和数据来源紧密相关；数据的真实性决定分析结果的稳健性；而数据的种类通常会决定具体的模型选择。

### 3.1 数据的真实性

数据的真实性与数据的来源密切相关。用来进行经济分析的数据，数据的真实性出现扭曲是很正常的现象。数据本身不会说谎，会说谎的是人。造成这些扭曲的原因非常复杂，可能是数据收集者或者提供者无意造成的测量误差（比如，在家庭调查中，被调查者对自己的年收入通常没有很准确的记录，只是通过记忆给出一个大体的数字）；也可能是他们故意的扭曲（比如，在家庭调研中，有些富裕的被调查者不愿意告诉别人真实的收入，害怕惹麻烦）。如果数据本身出现了问题，分析结果出现偏差，那就毫不意外。

数据的扭曲，大体上可分为：测量误差（Measurement Error）和样本选择（Sample Selectivity）。

- 测量误差

测量误差对实证结果造成的扭曲在理论计量经济学中已经得到了充分研究。Deaton（1997，99-101页）指出，在最小二乘法中，如果只有一个自变量存在非系统性测量误差，那它的回归系数的绝对值会比真实值小，这称之为“测量误差的铁律（Iron Law of Measurement Error）”；但是如果多个变量存在非系统性测量误差，那回归系数的扭曲方向没有一定的规律。但是，如果测量误差是系统性的，那会造成模型的内生性问题。何谓系统性误差？在这里指误差项和自变量存在相关关系。这需要用工具变量方法来解决。

- 样本选择

所谓样本选择？就是某些样本由于某些外在的系统性原因，导致观察值丢失。一个典型的例子就是：在调查居民工资收入时，某些居民可能由于市场工资低于自己的保留工资水准而选择不工作的，导致这部分居民的工资收入无法观察。显然，他们有自己的市场工资水平，这个工资不会等于零。如果以零作为他们的工资，这显然是不对的；如果去除这部分样本，也会

导致回归模型的系数产生和真实值不一致的问题。Heckman（1979）解决了这个问题，并因此在2000年获得诺贝尔经济学奖。

### 3.2 数据来源

经济数据一般有三个来源，具体为调研数据（survey data），政府等机构的统计数据（Statistical Data），以及实验数据（Experimental Data）等。数据的来源和数据的真实性密切相关。

- 调研数据

在研究居民微观经济问题时（比如，消费，投资，以及储蓄等行为），研究者或者政府等机构通常会搜集微观调研数据。Deaton（1997）对调研数据的实证分析的方法作了一个很好总结。

在调研数据中，由于被调查对象的认知有限（比如，被受访者由于记忆的限制，无法准确提供一年内家庭收入，消费，以及投资数量），或者调查者在填写数据时发生笔误（比如，调查者在填写调查表时，不小心把男性填成女性，或者把年龄15岁听成75岁等），会造成一些非系统性误差。在数据处理时，我们可以运用一些统计工具把一些异常值（Outlier）去除。如果不加以处理，在样本较小时，这样的误差可能给回归系数带来很大的偏差。

比较棘手的是调研数据中的系统性误差。一个很好的例子就是国家统计局的农村住户调查统计和城镇居民住户调查统计中，由于富人拒访率过高，使得（1）富人的样本代表性过低，（2）富人的真实收入报告值低于真实值（Wang & Woo, 2011）。富人的样本过低属于上述的“样本选择”问题；富人的真实收入等报告值过低，这属于上述的“系统性测量误差”。这些低报的灰色收入，可能导致计算的储蓄率，以及GINI系数等一些列经济指标偏低。这些误差给计量模型的回归结果带来的误差通常也是非常严重的。

在信息时代，网络给我们提供了一种便捷的调研手法。但是很显然，网络调研也存在很严重的样本选择问题。在现阶段，不是所有的人都可以接触互联网络，那些没有网络接触的人就被过滤掉了。

我们已经进入了一个大数据的年代。由于现代计算机技术的发展，可以存储海量的数据。“大数据”可以帮助研究者减少因为样本选择造成的误差，有助于更准确的研究消费者行为。但是，这些大数据很多时候与一些商业目的相联，这就激励了大数据的造假。比如，某些手机企业为了增加市场影响力，在网络销售中可能会发生销售给自己代理人的行为，从而达到影响消费者的目的<sup>3</sup>。

有些系统性误差可以通过一些统计或者计量方法（比如用工具变量，加权回归，或者Heckman样本选择处理方法）来解决；而另外一些系统性误差却可能因为造成误差的信息不明确而根本无法解决（比如，根本没有被选择去掉的样本信息），甚至研究者根本就无法意识到这些误差的存在。

---

<sup>3</sup>资料来源：杨理：《中国式大数据都是骗人的》

<http://www.21cbr.com/html/magzine/2014/144/hot/2014/0417/18166.html>

对于调研数据，一个值得关注的问题就是数据的“簇效应”（Cluster Effect）。一般而言，为了避免调研数据的样本选择误差，通常会采用分层抽样办法（比如在农户调研中，通常是每村或者每个社区都要有代表性样本）。但是，分层抽样也造成了所谓的“簇效应”（Deaton, 1997, 74-78 页）。在计量实证分析中，“簇效应”会导致计量模型的回归系数和方差混合了簇间效果和簇内效果，必须要用“簇效应”纠正方差，或者采用稳健性的标准误差。否则，这会对回归系数的显著性检验存在很大的影响。

### • 统计和行政管理数据

数据的另一个主要来源为政府等机构提供的统计和行政数据。很多研究者认为，政府的数据具有很强的权威性，所以在研究中受到广泛采用。在实证分析中，由于数据来源的便利性，广受研究者青睐的一种分析方法是用省级的面板统计数据。

很可惜，中国政府的统计数计一直以来广受质疑。其中最受关注，也最受广大研究者质疑的一个基础统计变量为 GDP。长期以来，GDP 与官员的政绩密切关联，所以 GDP 被人为扭曲是一个在所难免的现象。在中国统计数据报告中，通常发现省级的 GDP 总合远远大于国家层面总和的 GDP 值。

不仅是 GDP，中国的许多统计变量存在扭曲的现象。具体到农业统计数据，一个奇怪的现象是中国的耕地面积统计一直在增加；从 1997 年前的 15 亿亩，增加到 1997 年的 18 亿亩。而最近的第二次全国土地调查显示，中国有耕地 20.27 亿亩。这么一个关键的统计数据都不准确，这对经济分析和政策决定会造成重大伤害。

中国的很多农产品的生产量远远大于消费量。Yu & Abler (2014)对中国的猪肉统计作了一个系统性的分析，发现统计误差由很多原因造成的，包括生产的虚报，消费的低报，以及肉类加工中的损失等原因；其中最主要的原因还是生产的虚报，超过了误差的 50%。Yu & Abler (2014)同时也指出，中国的统计系统在改革开放初建立，已经很难适应现代的高度市场化的经济体系；必须要对现有的统计体系作重大的改革。

这种系统性误差，不言而喻，对实证结果的影响是巨大的。但是，现有的实证文献很少对这样误差造成的影响作分析，讨论以及纠正；基于这些实证基础上的政策建议的基础都不是稳健的。

基于以上分析，本文建议那些运用政府统计数据进行的实证分析，都应该重视这种系统性误差；运用工具变量等方法来检验和纠正计量结果的误差是一种通常的手法。如果无法纠正，对误差造成的可能后果作一个讨论是必要的。

### • 实验数据

由于传统的经济分析在某些实证分析中无法克服内生性问题，实验经济学便大行其道。实验经济学分为田野实验（Field Experiment），实验室实验（Lab Experiment），以及自然实验（Natural Experiment）。

实验经济学最重要的关注点是实验的设计（比如，样本是否随机），而数据的误差本身很少很受质疑。但是，发生删改和捏造数据，那是另当别论。

调研数据和统计数据一般是由随机产生机制产生；而实验数据本身是非随机的。这个区别本身不会影响计量模型估计的一致性和效率性。相对于调研数据和统计数据，由于成本的限制，实验数据一个重大缺陷是实验样本通常不会很大。所以在计量分析实验数据时，必须要注意小样本对计量结果造成的影响。在遇到小样本的时候，“自助法”（Bootstrapping）产生标准误差是一个通常的选择；此外，运用 Bell - McCaffrey (2002) 手法得出稳健的标准差也是一个可行的方法（Imbens，2012）。

### 3.3 数据类型

数据类型可以分为：截面数据，面板数据，以及时间序列数据<sup>4</sup>。每一类数据有自己的优势，也有自己的劣势。数据的类型，通常决定计量模型的选择。

- **截面数据**

截面数据由于收集的时间成本比较低，在经济分析中是一种常见的数据类型，它主要应用在微观经济行为的分析中，虽然在一些宏观政策分析中也偶有应用。

由于数据结构本身的限制，在计量实证分析中，隐含着一个假设条件是每一个经济主体的未观测到变量（即误差项）不存在系统性差异。但是，这样的假设条件有时过于强烈，而且在实证分析中也无法检验该假设。这有可能会产生一些实证分析产生不切实际的结果。

Stock 和 Watson（2007）在他们流行的计量经济学教科书的第八章中有一个经典例子：运用美国各州的截面数据发现啤酒税的高低和交通事故率之间存在奇怪的正相关关系。但是，运用固定效应模型的面板数据后，这个相关关系变成了显著的负相关。这是因为各个州对酒驾的态度不一样，截面数据模型无法控制这个变量，而导致了内生性问题；而面板数据能够通过控制州的效果而去掉这个变量的影响。这显示了面板数据的优势。

- **面板数据**

相对于截面数据，面板数据可以控制观测主体的未观测到的变量（Unobserved Heterogeneities），使得模型的估计变量满足一致性要求。此外，在实证分析中，很多研究者也偏好于使用滞后变量作为工具变量来克服模型的内生性问题。

在对中国的实证分析中，最广泛应用的面板数据为省级的统计数据。对以家庭住户为调查对象的面板数据收集，虽然需要很多的人力和财力，但是随着信息科技的发展以及政府等部门的大力支持，大规模需要长时间追踪的以家庭住户为对象的面板数据的收集在中国也变为实际。关于中国的微观调研数据，除了不开放利用的国家统计局的农村住户调查统计数据和城镇住户调查统计数据，在学术界数据广为使用的数据包括，农业部农村研究中心的“固定观察点数据”，美国北卡大学的中国健康与营养调查(CHNS)，以及北京大学的中国健康与养老追踪调查(CHARLS)。

---

<sup>4</sup> 此外还存在拟面板数据（Pseudo Panel Data），在这里就不作讨论。

对于面板数据的处理，除了不常见的差分法，比较常见的计量模型包括固定效应模型和随机效应模型。在本文开始就指出，每一个模型都有不同的假设条件。固定效应模型和随机效应模型的假设条件区别在于随机模型必须满足观测对象的固定效果变量和自变量之间不相关，而固定效果却没有这样的限制。他们之间的差异可以用 Hausman 检验来识别（Hausman 1978）。

从实际操作的角度来说，固定效应模型在任何时候都保证了一致性，所以固定效应模型虽然损失了一部分效率性，但更加稳健，所以在实证分析中占主导地位。除非是某些出现极端的情况：截面数目（ $N$ ）远远大于时间长短（ $T$ ）（即  $N \gg T$ ），使用随机效应才可以显著提高效率性。

在实证分析中，我们还要注意面板数据中的截面数目（ $N$ ）和时间长短（ $T$ ）之间的关系。在  $T > N$  的场合，面板数据更多展现出时间序列的特征。一些时间序列的分析方法有必要导入进来。

最后，在现实中平衡的面板数据比较少见，更多存在的是不平衡的面板数据，不平衡的面板数据在实际计量处理中和平衡的面板数据没有什么本质的不同。

## • 时间序列

随着现代市场交易频率不断增加，为我们实证研究提供了更多的大样本长时间序列数据。时间序列分析中最大的问题是数据的平稳性。如果数据是平稳的，最小二乘法模型基本特征可以移植到时间序列里面；通常的分析模型包括 ARMA, VAR, GARCH 等模型。

如果数据不是平稳的情况下，情况就变得非常复杂，回归系数的标准差有时不服从标准正态分布，从而造成了所谓的伪回归（Spurious Regression）问题，系数的显著性加大，原假设更容易被拒绝。

为了校正回归系数的分布问题，在用 VAR 模型作 Granger 因果分析的时候，即使采用差分的办法去除非平稳性，其回归系数的分布也可能不是标准正态。通常的方法是对于水平变量 VAR 在用 AIC 或者 BIC 基准选择最优滞后项的基础上，继续增加滞后项的方法（Surplus-Lag method）来得到一致的检验结果（Bauer 和 Maynard 2012），增加的滞后项的数目为非平稳变量的最高整合阶数（Maximum Order of Intergation）。

但是，如果非平稳变量之间存在存在协整（Cointegration），回归系数是超级一致的（Super Consistency）。所以，我们必须先识别计量模型的变量之间是否存在协整。协整检验的方法包括原始的 Engel-Granger 方法（Engel 和 Granger, 1987）和 Johansen 秩检验和特征值检验 等方法（Greene 2012）。Engel-Granger 方法比较直观，但缺点是只能识别是否存在协整，不能识别在多变量情况下的多个协整；Johansen 检验能够识别多变量情况下的复数协整个数，所以在实证中广为使用，但缺点是检验通常不够稳健。

检验时间序列数据是否平稳的检验方法很多。比较著名的检验包括 Augmented Dickey-Fuller 检验，Phillips-Perron 检验以及 KPSS 检验等。前两者主要检验时间序列是否存在单位根，其原假设是存在单位根；而 KPSS 检验的原假设是时间序列是平稳的。在实证中，最好能够同时报告 Augmented Dickey-Fuller 检验（或者 Phillips-Perron 检验）和 KPSS 检验，从不同的两方面来验证检验结果的稳健性。

最后，时间序列的滞后项数对回归或者检验的结果影响很大。选择科学合理的滞后项数就显得至关重要。实证中，通常选择之后项的方法为 Akaike 信息准则（AIC）和贝叶斯信息准则（BIC）。这两个准则的原理是基于数据本身对模型的吻合性，加上一个对滞后变量数目的惩罚因子。AIC 或者 BIC 的最小者为最好之模型。

#### 4. 计量模型与统计检验

计量经济学是结合了经济学理论和统计学的一门学科。不可否认，计量经济学的基础是统计学，但是它和统计学还是有本质的区别。我认为计量经济学更偏重于模型的估计，实证的计量经济分析需要经济学理论模型作为基础；而统计学更偏重于发掘数据之间相互关系，而他们之间的理论基础不太重视。

正是计量经济学和统计学之间的差异，导致了实证计量分析在现实中遇到了很多难点，而有些问题直到现在还没有从理论上得到根本的解决。计量实证分析中的有三个问题尤其要值得重视。这三个问题为：外生性（Exogeneity），伪回归（Spurious Regression），以及经济学上的显著性（Significance）。

##### 4.1 外生性

我们在第二部分已经指出，计量经济模型作为一个模型必然存在假设条件。为了使得回归结果有效，我们必须要保证计量模型中的自变量为外生的假设条件。如果变量是内生的，这会导致估计结果不能满足参数一致性条件。

- 一致性和效率性

参数的估计值结果只是一个随机变量，不是参数的真实值。从理论计量经济学的角度出发，判断一个参数回归估计是否为一个“好”的估计的标准有两个：一致性和效率性。另外一个标准是无偏性，但是无偏性在实际操作中显得无关重要。一些参数估计可能是有偏的，但还是满足一致性的条件。

一致性是参数估计的最核心基准。一致性是指当样本趋向于无穷大时，估计值收敛于参数的真实值。只要估计满足一致性，我们就可以增加样本的数量，得到比较接近真实值的估计。效率性是指一个估计值的反差在所有可能的估计值中最小。

一致性和效率性有时会发生冲突的问题。比如在面板数据的回归中，固定效应模型一定是一致的估计，而随机效应模型却是更有效率的估计。另一个例子就是在有效工具变量回归中，工具变量回归满足一致性；而最小二乘法却是更有效率的估计。在这两者发生冲突的时候，我们倾向于优先满足一致性。这也是 Hausman 检验的理论基础（Hausman 1978）。

- 内生性的产生原因

经济学中最困难的问题就是变量外生性的检验。在经济学理论中，外生变量为独立于经济模型系统，在系统外发生变动的变量。而计量经济学中的外生性是指自变量和模型的误差项之间不存在相关性。外生性相对应的一个概念就是内生性：自变量和模型的误差项之间存在相

关性。内生性问题会造成模型参数估计不满足一致性的严重后果。内生性问题是计量经济学中的重点与难点，各主要教科都对此有着墨（Chen 和 Pearl, 2013）

Stock 和 Watson（2007，336-338 页）指出造成内生性原因包括：（1）相关变量的丢失；（2）方程形式错误（把非线性方程设定为线性方程）；（3）变量中存在系统性测量误差；（4）样本选择；（5）联立因果关系；以及（6）误差项之间存在相关性。为了保证回归结果稳健，合理以及可靠，这些问题在我们的实证分析中必须要认真检视。

外生性假设对经济模型的估计结果衍生出来的政策建议具有至关重要的作用。只有满足了外生性假设的估计结果，才能对现实政策分析有意义。否则，我们只能解释为变量之间存在相关关系，而不能给出政策建议。

如果对变量的外生性存在怀疑，我们可以用工具变量回归方法来检验和校正，但前提是工具变量是有效的。

### • 工具变量的选择

工具变量回归方法中最重要的问题是工具变量的选择。从广义的角度来讲，最小二乘法就是用变量自身做自己的工具变量。Imbens（2014）对工具变量的选择有一个精彩的综述。工具变量的选择标准为：（1）工具变量和内生变量相关，（2）但和误差项不相关。

在实证分析中，工具变量的选择需要非常小心。我们首先要运用经济学理论来寻找工具变量，这可以从自然，历史或者制度安排的角度出发，发现一些变量会影响内生变量，但是不会直接响应变量。除此之外，在实际操作中选择工具变量的手法有两个简单方法：1，如果数据存在“簇效应”，我们可以用“簇效应工具变量”（Ji, Yu 和 zhong, 2012）；2，如果数据是面板数据，可以用内生变量的滞后期变量来作为工具变量。值得指出的是，这两种方法存在很强的假设：簇效应和滞后变量必须和误差项不存在相关；这个假设很多时候可能不能满足。在选择工具变量时，对工具变量的有效性的辩护和检验就显得非常重要。

Stock 和 Watson（2007，第 12 章），与 Cameron 和 Trivedi（2005，第四章）对工具变量的有效性作了详细的论述。对于工具变量的具体选择，我们要注意三个问题。

（1）弱工具变量问题（Weak Instrument）：内生变量和工具变量之间相关性不强。这会使得工具变量模型产生没有意义的估计结果，甚至比不使用工具变量的结果还要差。如果只有一个内生变量<sup>5</sup>，通常采用的方法用内生变量作为应变量和所有的外生变量以及工具变量进行回归，然后对工具变量进行联合检验；强工具变量的一般检验标准为联合检验的 F 值大于 10。

（2）过度识别问题（Overidentification）：所有工具变量必须是外生的。首先，如果工具变量数小于内生变量数，工具变量模型是无法识别的；如果工具变量数等于内生变量数，工具变量回归模型是正好识别的；如果工具变量数超过内生变量数，我们才可以检验工具变量模型是否被过度识别？我们通常采用 J-统计值或者 Sargan 检验，来检验工具变量是否存在过度识别。只有在存在过度识别的时候，所有工具变量才是外生的。值得注意的是：过度识别检验的原假设为“模型过度识别”，如果拒绝了原假设，可能至少存在一个工具变量为内生性的。

---

<sup>5</sup> 如果存在多个内生变量的情况，R<sup>2</sup>测度（R<sup>2</sup> Measure）是通常的检验方法，请参照 Cameron 和 Trivedi（2005，104-105 页）。

(3) 工具变量的冗余问题 (Instrument Redundancy)：工具变量中可能存在多余的无关重要的变量。冗余的工具变量会让估计结果损失效率性。在弱工具变量检验中，我们观察是否存在一些工具变量的回归系数的 t-检验不显著。如果不显著，他们可能就是冗余的。

在找到有效的工具变量后，我们可以用工具变量法估计出参数。很显然，有效的工具变量法估计出的参数是满足一致性，而非工具变量法估计的参数却是更有效率的。我们可以再次运用 Hausman 检验 (1978) 来比较他们之间的差异。如果两者不存在显著性差异，我们偏好非工具变量法；如果两者之间存在显著差异，我们就要依赖于工具变量的估计结果。

#### • 实验经济学和干预效果的估计

在实证分析中，工具变量的上述两个标准通常难以满足，所以一些重要的研究课题，在现实中根本无法深入定量研究。一个最典型的例子就是 GDP 和环境污染之间的关系，我们无法找出一个工具变量与 GDP 相关而与环境污染无关；我们也无法找一个变量与环境污染相关，而与 GDP 无关。虽然环境 Kuznets 曲线为我们提供了 GDP 和环境污染之间的一个定量关系，但由于我们无法识别他们之间的因果关系，我们的结论至多就是他们之间存在显著相关关系，这样的结果并不能提供很好的政策依据。另外一个有意思的研究课题为信息技术的发展和经济发展之间的因果关系，我们也很难做计量分析。

这时，就需要我们应用实验经济学的方法来解决内生性的问题。实验经济学通过人为的外生的随机干预来使得某些变量成为外生。关于对干预效果 (Average Treatment Effect) 的估计，Wooldridge (2010, 第 21 章) 提供了一个很好的计量方法的综述。

我在这里强调的是有些干预看起来是外生的，其实不然。(1) 如果干预真是外生且随机的情况，最小二乘法的估计结果是有效的；(2) 如果干预只和某些可观测的自变量相关，这时需要运用倾向评分匹配方法 (Propensity Score Matching) 来得出满足一致性的干预估计结果；(3) 如果干预变量和某些不可观测变量相关，或者干预是内生的情况下，我们还需要使用工具变量的计量方法来估计干预的效果。

## 4.2 伪回归

在计量经济的实证研究中，经常会存在伪回归 (Spurious Regression) 的问题。所谓的伪回归问题，就是本来没有关系的两个变量，或者没有显著关联的两个变量，在计量经济模型中却显示出了显著的相关性。这有可能是计量模型设定的问题，或者是计量经济模型本身特征造成的问题。

#### • 经济学理论和伪回归

如上所述，计量经济学是结合了经济学理论和统计学的一门理论。如果计量经济学模型离开了经济学理论，把不相关的变量放到一起做计量分析，很有可能出现伪回归现象。

在一篇广受争议的研究中，Oster (2005) 发现中国女孩出生比例失调的现象可能主要是由于中国人口中高比例的乙肝病毒携带人口。这篇文章虽然结果新颖，在统计上也存在显著的相关性，但是显然没有很强的经济学理论基础，或者医学基础。最后，在学术界广泛质疑的情况下，Oster 被迫运用新的数据，重新得出不显著相关的结论。

在实证分析中，我们一定要清楚计量模型的理论基础是什么？这样的计量模型是否有道理（Make Sense）。没有理论基础的计量模型，最后可能纯粹成为一个无本之木的数字游戏。

- 时间序列中的伪回归

我们已经在第三部分指出，时间序列如果在不平稳的情况下，很有可能出现所谓的“伪回归”。两个本来不相关的非平稳时间序列，最后的 t-检验结果可能显著性相关。这时我们需要运用协整检验来判定变量之间是否存在伪回归（Engel 和 Granger 1987）。

- 空间数据中的伪回归

在中国经济的实证分析文章中，由于其收集成本较低，省级面板数据被广泛使用。我们必须注意的是省级行政区在空间上可能存在互相影响。也就是说，这有可能违背了最小二乘法中的误差项不相关的假设。

更为致命的是，由于空间上的相关，和时间序列类似，变量可能在空间上不是平稳的，这可能会导致伪回归的问题（Fingleton 1999）。类似于时间序列，我们可以运用空间上的协整（Spatial Cointegration）的方法来检验伪回归。不幸的是，这个空间伪回归的问题并没有在实证分析中得到应有的重视，一个重要的原因可能是这样的检验在计量技术上还不够成熟。从理论计量经济学的角度来讲，这个方向还有很大的发展空间；从应用计量经济学的角度来说，我们需要对用省级面板（或者地区面板以及县级面板等）等具有空间上相关的数据的估计结果，保持一个谨慎的态度。

### 4.3 经济学上的显著性和统计上的显著性

计量经济学的基础是统计学，所以计量经济学中的很多检验事实上是基于统计学中的检验。我们一般所讲的显著性，即是指统计上的显著性，和经济上的显著性有很大区别。统计上的显著性只关注于损失方程的概率本身；而经济上的显著性更多的关注损失方程的总量。这个区别早就被 McCloskey（1985）以及 McCloskey 和 Ziliak（1996）作了详细论述。可惜的是在现有的实证研究中，这两者的差异并没有得到足够重视，这使得实证研究的结果，甚至计量经济学的方法本身都受到质疑。

在统计上，约定俗成的显著性是指  $p=5\%$ 。在某些情况下，统计上显著拒绝原假设，却不等于经济上有显著区别。这里举两个例子：

第一个例子，假设我们在研究 Cobb-Douglas 生产函数的规模报酬一定与否，原假设为  $\alpha + \beta = 1$ 。我们经过计量分析后发现  $\alpha + \beta = 0.99$ ，而且它们标准误差为 0.001。从统计上讲，原假设  $\alpha + \beta = 1$  在 1% 的显著水准（ $t=10$ ）被拒绝。但是，从经济学的角度讲，1 和 0.99 并没有显著的差异。

第二个例子，假设我们在飞机上导入一项新的设备（比如新的电池），这可能会影响飞机的安全。如果研究发现导入新的设备后，飞机飞行 200 次，会发生一次坠机危险。从统计上讲，我们在 1% 的显著性水准拒绝了飞机不安全的假设。但是，经过这样统计检验的飞机，有谁敢乘坐？

经济上的显著性问题，从理论计量的角度来说，到现在还没有一个很好的客观的检验标准。但是，按照 McCloskey（1985）的建议，我们在作实证分析的时候，必须要意识到这个问题，从而有必要对经济上的显著性做一定的讨论。

## 5. 参数估计值的稳健性和政策分析

经济研究的主要目的是为了帮助人们认识复杂的经济现象，并为政策制定提供必要的参考依据。计量经济模型估计得到的参数，通常会被用来作为政策分析的基准。从理论上讲，计量经济模型的估计值为随机变量；所以任何估计值都有其合理性。但是，为了使得实际政策分析更加有效，稳定（Invariant），合理（Reasonable），以及值得信赖（Reliable）的模型估计就显得至关重要。稳定，合理以及值得信赖，就成为了实践中评价参数估计值的标准。

### 5.1 估计值的稳定性（Invariance）

Lucas（1976）批评计量经济模型时指出，计量模型的参数估计值通常是条件于政策变量；当政策发生变化时，估计的参数也会随之发生变化，所以计量经济学对现实经济政策的分析作用不大。

如果不能保证参数估计值的稳定性（Invariant），条件于数据的模型估计结果就成为了“变色龙”：数据变化，模型估计结果也随之变化。为了应对 Lucas 批判，计量经济学家们对变量的外生性作了深入的研究，以期计量经济学的估计结果能够为政策分析提供值得信赖的结果。

根据 Engle, Hendry & Richard(1983)的定义，变量外生性可以分为三类：弱外生(weak exogeneity)，强外生(Strong Exogeneity)，以及超级外生 (Super Exogeneity)。所谓弱外生性，就是变量数据生成过程中的信息不会给参数估计提供任何信息，也就是说变量的生成过程对估计结果没有影响。在时间序列数据中，如果保证了变量弱外生，内生变量对该外生变量也没有 Granger 因果影响，我们可以定义这些外生变量为强外生性。也就是说，条件于强外生外生变量，模型并不会降低预测的精度。最后，超级外生性是指：即使外生变量的数据产生机制(Data Generating Process)发生变化，估计变量还会保持稳定不变（Invariance）。

对于政策分析来说，最重要的工作就是识别超级外生变量，因为只有条件于这样的变量，政策分析才能有效；不会因为政策变化而使得参数估计发生“变色龙”效应。Favero 和 Hendry（1992）对变量的超级外生性以及其检验方法作了一个具体的总结。

### 5.2 估计值的合理性（Reasonability）

做为一个训练有素的经济学研究者，对照参考文献以及经济理论，应该会对计量模型的参数估计值有一个大体的判断:这个估计值是否合理（reasonable，或者 make sense）；从经济学理论角度或者对照现有的文献，计量结果是否可以得到合理的解释。

比如，在估计中国的食物需求弹性时，如果估计出大米消费的收入弹性为 1.3，这显然没有道理。按照现阶段的中国居民收入水准状况，大米作为一个生活必需品，甚至可能是一种劣等商品，它的收入弹性应该不会很大，有道理的范围应该在 $[-0.1, 0.5]$ 之间。另外一个例子，Tian 和 Yu (2012) 在一个荟萃分析 (meta analysis) 中指出中国经济增长中的 TFP 增长率在改革开放后平均为 2%。如果某个研究估计出一个 10% 或者 -10% 的 TFP 增长率，这样的结果显然会受到质疑。

如果估计的结果不合理，我们也无法解释。我们必须回头检查出错的原因：这是数据原因，还是模型出错，还是经济学理论假设错了。

### 5.3 估计值的可靠性 (Reliability)

由于计量模型的结构和数据存在某些问题，计量模型的估计值有时对模型的结构非常敏感。增加一个变量，减少一个变量，或者使用不同的工具变量都有可能对模型的估计结果产生重大影响；有时，这种差异可能是本质性的。

由于计量经济模型的估计值是一个随机变量，任何变量估计都有其合理性。所以，获得一个可靠的估计值对论文的质量显得更加至关重要。

“不能因为偶然一枪打中了一只鸟儿，就认为自己是神枪手”。同样，不能因为某种特殊的模型结构而得到一个较理想的估计结果，而在论文写作中只报告这一组结果。对主要模型的结果，最好在论文中报告 2-3 组相近模型的结果，并对他们的差异进行比较和解释这些差异，这样会使计量模型的估计值让人觉得可靠。

## 6. 结论

计量经济学作为实证分析的主要手法，已经被广大中国经济研究者所接受。但是，正确运用计量经济模型，得出一个稳定，合理以及可靠的参数估计值，还没有一个很好系统梳理。由于计量经济学的统计学基础，不正确使用计量经济模型，可能会使得估计结果不能稳健，从而产生“变色龙”一样的实证结果，导致实证结果的政策分析被广受质疑。

本文从数据，模型和参数等三个角度出发，分析应用计量经济学模型在实证分析中要注意的问题。

首先，数据是进行实证分析的基础。数据按照来源，可以划分为微观调研数据，机构统计数据，以及实验数据。在广为使用的调研数据和统计数据中，系统性误差，包括测量误差和样本选择，常常存在。如果无视这些误差，可能使得估计结果不能满足一致性。如果数据存在系统性测量误差，工具变量方法通常是主要的解决方案；如果数据存在系统性的样本选择问题，Heckman 方法是广为使用的校正方法。

其次，从模型的角度来说，任何模型都是由环境假设，机制以及求解概念三个组成部分。这三个部分中，环境假设对计量经济模型的正确使用尤为重要。在运用计量模型时，必须要清楚了解他们的假设条件，并对这些条件作必要的检查和检验。计量经济模型区别于统计模型最重要的假设为：变量的外生性；许多因素可以造成变量内生性问题。工具变量是对于内生性的常见的检验和校正方法。可是有些研究中，工具变量无从寻找，我们就必须要依靠实验经济学的方法。

伪回归在计量分析中也不鲜见。伪回归可能是由模型本身原因造成的，也可能是数据结构造成的。计量经济学是结合了经济学理论和统计学的定量分析方法，没有经济学理论基础的计量经济分析，很可能导致伪回归结果。某些特殊的数据结构，比如非平稳的时间序列，或非平稳的空间数据，都有可能产生伪回归结果。

再次，计量经济学的基础虽然是统计学，但是两者之间还是存在一些差异。由于技术上的限制，现有的计量经济模型的检验还是基于统计检验，所谓“显著性”都是统计上的显著性，这不同于我们关注的“经济上的显著性”。在实证分析中，我们在讨论估计参数在统计上显著性的时候，也必须讨论经济上的显著性；后者有时可能更为重要。

最后，计量经济学的估计结果通常会被运用到政策分析中去，但是 Lucas 批判（1976）认为参数的估计值可能会随着政策的变化而变化，使得计量经济学无法为政策分析服务。为了应对 Lucas 批判，计量经济学家提出了变量超级外生性的概念。条件于超级外生的变量，数据产生机制对估计参数结果没有影响，这时的政策分析才有意义。

## 参考文献:

- Abler, D. (2010), "Demand Growth in Developing Countries", *OECD Food, Agriculture and Fisheries Papers*, No. 29, OECD Publishing.  
<http://dx.doi.org/10.1787/5km91p2xcsd4-en>
- Bauer D. and A. Maynard (2012) Persistence-robust Surplus-lag Granger Causality Testing. *Journal of Econometrics*. Vol. 169(2): 293-300.
- Bell, R., and D. McCaffrey, (2002), "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples," *Survey Methodology* , Vol. 28(2), pp. 169-181.
- Chen B. And J. Pearl(2013) ,, Regression and causation: a critical examination of six econometrics textbooks." *Real-World Economics Review*,No.65.
- Cameron A C. and P. Trivedi(2005 ): *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Deaton A. (1997) *The Analysis of Household Surveys*, John Hopkins University Press. Maryland U.S.A.
- Engle, Robert F.; Granger, Clive W. J. (1987). "Co-integration and error correction: Representation, estimation and testing". *Econometrica* **55** (2): 251–276.
- Engle, R. F., Hendry, D. F. & Richard, J. F. (1983). Exogeneity, *Econometrica*.51: 277–304.
- Ericsson, N. R. (1994). Testing exogeneity: An introduction, in N. R. Ericsson (ed.),*Testing Exogeneity*, Oxford University Press, Oxford, pp. 3–38.
- Favero C. And F.D. Hendry (1992) "Testing the Lucas Critique: A Review", *Econometric Review*. Vol. 11(3): 265-306.
- Fingleton Bernard (1999) "Spurious Spatial Regression: Some Monte Carlo Results with A Spatial Unit Root and Spatial Cointegration." *Journal of Regional Science*. Vol. 39(1):1-19.
- Greene, W. (2012). *Econometric Analysis*. 7th ed. Pearson Education, New Jersey.
- Hausman, J. A. (1978). "Specification Tests in Econometrics". *Econometrica* **46** (6): 1251–1271.
- Ji Y., X. Yu, and F. Zhong (2012): Machinery Investment Decision and Off-Farm Employment in Rural China. *China Economic Review*. Vol.23(1):71-80.
- Lucas, Robert (1976). "Econometric Policy Evaluation: A Critique". In [Brunner, K.](#); Meltzer, A. *The Phillips Curve and Labor Markets*. Carnegie-Rochester Conference Series on Public Policy **1**. New York: American Elsevier. pp. 19–46
- Heckman, James J, (1979). "Sample Selection Bias as a Specification Error," *Econometrica*, *Econometric Society*, vol. 47(1):153-61.
- Imbens G.W. (2014 ) "Instrumental Variables: An Economic Perspective." NBER working paper 19983. Cambridge. MA. USA. <http://www.nber.org/papers/w19983>
- Imbens G. W. (2012) " Robust Standard Errors in Small Samples: Some Practical Advice." NBER Working Paper 18478. Cambridge. MA. USA.

<http://www.nber.org/papers/w18478>

McCloskey D. and Ziliak S. T. (1996) "The Standard Error of Regressions", *Journal of Economic Literature*, Vol. XXXIV:97-114.

McCloskey D. (1985) "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests", *American Economic Review*. Vol.75(2):201-205.

Oster, Emily (2005) "Hepatitis B and the Case of the Missing Women," *Journal of Political Economy*, December 2005, 113 (6), 1163-1216.

Pfleiderer Paul (2014): *The Misuse of Theoretical Models in Finance and Economics*, Working Paper, Stanford University, available at: <https://www.gsb.stanford.edu/sites/default/files/research/documents/Chameleons%20-The%20Misuse%20of%20Theoretical%20Models%20032614.pdf>

钱颖一 (2002): "理解现代经济学"("Understanding Modern Economics"), 《经济社会体制比较》, 北京, 2002年第2期。

Stock J. and M. Watson (2007): *Introduction to Econometrics* (2<sup>nd</sup> edition). Pearson Education.

Tian X. and X. Yu (2012): *The Enigmas of TFP in China: A Meta-Analysis*, *China Economic Review*. Vol. 23(2): 396-414.

Wang, X., & Woo, W. T. (2011). [The size and distribution of hidden household income in China. \*Asian Economic Papers\*, 10, 1–26.](#)

Wooldridge J. (2010): *Econometric Analysis of Cross Section and Panel Data*, Second edition. MIT press.

于晓华(2008): "中国研究机构发表经济学英文论文的一个统计研究 (1998-2007)", 《经济学季刊》, 2008年第7卷第4期: 1477-1490。

Yu X. and Z. Gao (2010): "An Updated Ranking of the Economic Research Institutions in China (2000-2009)", *China Economic Review*, Vol. 21(4):571-581.

Yu X. and D. Abler. (2014): *Where Have All the Pigs Gone? Inconsistencies in Pork Statistics in China*. Forthcoming in *China Economic Review*.

## **Data, Model and Parameters in Empirical Econometrics: A Synthesis Review**

### **Abstract:**

This paper discusses how to correctly apply econometric models in empirical research from the perspectives of model structure, data type, and parameter estimation, in order for obtaining invariant, reasonable and reliable estimates, which could be used for policy analysis.

**Key words:** Econometric Models, Model Structure, Data Types, Parameter Estimation, Exogeneity, Policy Analysis.