

第 32 章 如何做规范的实证研究

我们已经学了不少计量知识。到底应如何应用？

“实证研究” (empirical research)一般有哪些步骤与注意事项？

有人把计量经济学理解为概率统计的一门分支，将注意力集中于定理的推导与证明。

事实上，计量知识只是一种手段，而最终目的是通过计量知识的指导来做好实证研究。

比如，计量知识告诉我们，在何种情况下，应该使用何种计量方法，其假设与局限性又是什么，如何检验，等等。

32.1 计量理论与现实数据

计量经济学的理论总是建立于一些理想化的假定基础之上，而现实的经济数据常常或多或少地不符合这个或那个假定。

尽管计量理论是可以严格证明的一门科学，但实证研究的实践本身在一定程度上却是一门艺术。

Kennedy(2003, p.390)认为，“理论计量经济学家”(econometric theorists)与“计量实践者”(applied econometricians)常常缺乏交流，而“不使用真实数据的计量理论”在某种意义上更容易些。

其结果是，计量实践者在处理现实数据时有时不知所措，甚至盲目进行。

事实上，无论多么高深的计量方法，如果原始数据的质量有问题，也只能是“垃圾进去，垃圾出来”(garbage in, garbage out)。

Griliches(1994)在给美国经济学会做的主席演讲(presidential address)中指出，由于经济学家不够注意数据的来源及产生过程，经常错误地解释数据，导致研究的进展缓慢；如果不提高数据质量，计量理论方面的重大进展将无用武之地。

对于大多数从事应用研究的学者而言，主要是使用别人(比如，统计局)提供的数据。

即便如此，也应该对数据的质量有一个清醒的判断，并使用相应的计量方法(至少在做出实证研究的结论时，应考虑到数据质量的影响)。

另一方面，对于计量的初学者来说，与真实数据打交道是加深对计量经济学理解的重要途径。

只有弄脏你的手，才能学会实证研究 (So let's get our hands dirty)。

32.2 实证研究的主要步骤

只有站在实证研究的框架上才能更好地回答“如何使用计量知识”这一问题。

一般来说，规范的实证分析包括以下几个步骤。

1. 提出研究问题

对于刚开始进行研究的人来说，常常不知该如何选择研究课题。研究者通常知道自己想要研究的领域(比如，经济增长)，但这还不是一个具体的研究问题(research question)。

对实证分析而言，研究问题通常是有关“ x 对 y 有何作用”之类的问题。

如果想研究“金融发展对经济增长的作用”，就更具体了，此处 x 是“金融发展”(可以包含多个变量)，而 y 是“经济增长”。

研究问题可以来源于理论(比如，检验资产定价模型 CAPM 是否成立)，也可以来自对经济现象的观察(比如，媒体的热点问题报道，社会调研)，可以研究某个政策的效应(比如，新劳动法对失业率的影响)，也可以对文献中已有论文进行改进。

提出好的研究问题，也许你的研究就成功了一半。什么是好的研究问题呢？总的来说，研究问题越具体、越有趣、越新颖、越有可行性，则越好！

(1) **具体**：简单来说，在以上“ x 对 y 有何作用”这样的句型中，应能明确 x 与 y 具体是什么。

(2) **有趣**：你研究的问题别人会感兴趣吗？为什么我们要在乎你的问题(Why should we care)？知道你的问题的答案后，能影响人们对世界的某个方面的看法吗？你的研究问题为什么重要？

(3) **新颖**：论文的核心价值在于它的创新性，即做出了文献中所没有的边际贡献。这种边际贡献可以是研究了新的现象、使用了新的(更好的)计量方法、或使用了新的数据集。做研究的过程是创造新知识的过程，在本质上不同于学习已有(旧)知识的过程。

(4) **可行**：即使你的研究问题很具体、很有趣、很新颖，如果找不到相应的数据，则不可行。

有了潜在的研究问题后，就可以进行初步的探索性研究 (exploratory study)，看看它是否满足以上标准。

比如，通过查文献，看别人是否已经做过类似的研究，或大致了解数据是否可得。

备选的研究问题或想法越多越好，因为能“存活”下来的选题通常不多。

2. 文献回顾

论文贵在有新意。

假设你找到了一个具体、有趣而可行的研究问题，那么它究竟有多少新颖性，这就不可避免地需要查阅文献，看看文献中是否已有类似的研究。

对于中文论文，可在 CNKI 中搜索；

对于英文论文，可在 JSTOR 或 EconLit 中搜索。

某些工作论文则可通过百度或 Google 搜索。

对于刚开始进行研究的研究生们而言，似乎自己能够想得到的研究题目，都已经被别人做过了。

要想做出新的边际贡献，当然可以在前人的基础上，继续拓展与改进(改进计量方法，增加变量，使用新数据等)。

也可以完全撇开前人，去研究全新的现象。比如，20 世纪 70 年代末中国农村实行了家庭联产承包责任制改革，这是史无前例的。到了 1990 年左右，就涌现出一批研究农村改革对中国农业产出影响的论文。

2008 年中国开始实施“新劳动法”。假以时日(有了足够的数据后)，就可以研究新劳动法对劳动力市场的影响。

即使是前人已经研究过的现象，也常常可以用新的眼光、新的视角去观察。

重要的是，要有敏锐的观察力，并“像经济学家那样去思考”(Think like an economist)。

对于刚起步的研究者，应尽量避免已经被研究得很烂、或过于有挑战性的题目。

什么时候开始看文献，即看文献的时机，也很重要。

如果从一开始就大量地看文献，则可能被文献所淹没，望洋兴叹，自觉渺小。

更好的方法是，只有当自己已经有了一定的想法后，再去系统地看文献。

这样才会知道自己究竟看什么，也更能带着批判性的眼光去看。

阅读文献的态度十分重要。

虚心地从经典论文中汲取营养，才能“站在巨人的肩膀上”(Standing on the shoulders of giants)，但仍应带着批判性的眼光。

由于经济现象的复杂性，任何论文一定都有缺点(比如，忽略了某些可能重要的因素)，也都有可以改进的空间(甚至可能推翻作者的结论)，不必太迷信“权威”。

既然经济学还不是科学，获得诺贝尔奖的经济学家可能持有相反的观点，那么又哪来的权威呢？

重要的是，使用逻辑与实证的方法对不同观点进行甄别与质疑。

如果认为前人所做的研究已经十全十美，你都赞同，那么，你怎么可能做出新的边际贡献呢？

3. 建立理论模型

虽然实证研究可以没有理论模型，但如果有好的理论模型作为基础，则更有说服力。

具体来说，回归分析一般只能说明变量之间的相关性，要对变量之间的因果关系做出判断，常常需要依赖于经济理论。

即使无法提供完整的理论模型，也应该进行一定的理论分析。最理想的情形是，从理论模型中推导出计量模型(econometric model)，即待估计的回归方程。

4. 收集数据

在正式开始研究之前，应大致知道你所需的数据是存在的，而且可以得到。

数据要么是别人提供的(比如，统计局)，要么是自己收集的(比如，问卷调查)。

如果使用别人提供的数据，要注意其数据是如何定义、计算的(统计口径)，是否是与你的理论模型中的变量相对应。

如果确实不知道该从哪里找数据，可以看文献中同类研究的数据来源，并追溯其本源。

近年来，《美国经济评论》(*American Economic Review*)与《应用计量经济学杂志》(*Journal of Applied Econometrics*)等期刊都在其网站上公开了发表论文中使用的数据集。

对于非电子版的数据，则需要自己耐心地手工输入数据，并注意检查，防止出错。

5. 使用适当的计量方法

有了数据之后，就可以根据数据的类型与特点，来决定使用合适的计量方法。

比如，被解释变量为二元变量，则可以使用 probit 或 logit;

如果是非负整数，则可以使用泊松回归;

如果是面板数据，则可以考虑固定效应、随机效应、时间效应、动态面板等;

如果是时间序列，则应先判断是否含单位根，再决定使用相应的计量方法。

对于一般的数据，通常先作 OLS，看看结果，作为一个参照系。

作完 OLS 后，可以画残差图(residual plot)，大致看看扰动项是否符合经典假定，然后进行严格的检验。

如果有所违背(比如, 存在异方差、自相关), 则做相应的处理。

对于时间序列, 还可以检验是否存在结构变动(邹检验)。

另外, 应该对数据的质量进行检验, 判断是否存在多重共线性、极端值、弱工具变量等, 并做相应的调整。

对于回归函数的具体形式, 可以考虑线性、对数(如果解释变量只取正数且有指数增长趋势)、双对数、非线性(边际效应不是常数)等。

由于受数据可获得性的限制, 遗漏变量几乎不可避免。

不外乎两种情况。第一, 存在遗漏变量, 但与解释变量不相关(需要说明为什么不相关), 故可以不做处理。

第二，存在遗漏变量，且与解释变量相关，则必须进行处理，处理方法包括增加控制变量、寻找代理变量、使用工具变量、使用面板数据等。

另一常见问题是内生解释变量。此时，一般需要找到有效的工具变量才能得到一致的估计。

由于面板数据可以在一定程度上克服遗漏变量问题，故比横截面数据或时间序列更有说服力。

如果可以获得面板数据，则应尽力争取。

大多数的实证论文都希望说明 x 对 y 的因果作用。而从回归分析的相关关系到因果关系，是一个很大的飞跃。

32.3 实证论文的结构

实证论文的一般结构如下。

- (1)引言(Introduction)
- (2)文献回顾(Literature review)
- (3)理论框架(Theoretical framework)
- (4)计量模型(Econometric model)
- (5)数据说明(Data description)
- (6)计量方法(Estimation method)
- (7)回归结果(Regression results)
- (8)结论(Conclusion)

对于初学者，要注意观察经典论文的文章结构与风格(包括如何用表格来汇报回归结果)，以便模仿。

比如，中文论文可以模仿《经济研究》或《经济学季刊》，而英文论文则可参照《美国经济评论》等。

32.4 计量实践的十诫

Kennedy (2003, p.390)提出了以下“计量实践的十诫”(The Ten Commandments of Applied Econometrics)，作为指导实证研究的十个准则。

1. 运用常识与理论(Use common sense and economic theory)

在进行模型设定时，应尽量使用常识与经济理论。

比如，将“人均变量”(如人均消费)与“人均变量”(如人均 GDP)相匹配；使用实际汇率来解释实际进出口；避免使用带时间趋势的变量来解释不带时间趋势的变量。

2. 避免第 III 类错误(Avoid Type III Error)

在作统计检验时，要尽量减少第 I 类错误与第 II 类错误。而第 III 类错误则指的是，对错误的问题给出了正确的答案(the right answer to a wrong question)。

比如，考虑 FDI 对经济增长的作用。由于 FDI 起作用需要时间，如果仅把当年的增长率对当年的 FDI 进行回归，则可能没有太大意义。比较适当的做法是，考虑期初的 FDI 对随后 5 年(或若干年)经济增长率的作用。

3. 熟悉经济现象(Know the context)

实证研究者应该知道所研究现象的历史、制度与文化背景，以及数据的来源与产生过程。换言之，你应该知道得比计算机更多。

比如，如果你研究中国的失业问题，那么你应该知道，中国的失业率指的是“城镇登记失业率”，其统计口径与标准教科书中以及西方国家的失业率概念很不相同。

4. 察看数据(Inspect the data)

研究者的一个常见误区是，只知道按电脑键盘，而不去熟悉原始数据(raw data)。

察看数据的方法包括，计算数据的主要统计量(summary statistics)，比如，均值、最大最小值、标准差、相关系数等；画图；以及增加对数据的感觉(get a feel for the data)。

在察看数据的过程中，还要注意观察数据中是否存在不一致的(inconsistent)地方，比如，出现了不可能、不现实或可疑的取值。

如果发现，则要进行处理(比如，可能是数据输入错误)，这被称为“数据清理”(data cleaning)。

5. 模型应尽量简洁(Keep It Sensibly Simple, KISS)

尽量不要把模型搞得太复杂。“从小到大”(specific-to-general)的建模方法简单易行，但可能偏差较大(因为存在遗漏变量)。

“从大到小”(general-to-specific)的建模方法偏差小，但却不易执行。

常的折中方案是，选择简单而有解释力的模型。

6. 理解计量结果(Use the interocular trauma test)

计算机输出结果常常很多页，包含密密麻麻的表格与数字。

如果你尝试了各种不同的计量方法与控制变量(alternative specifications)，则结果就会更复杂。

如何看这些结果？只能用双眼一个字一个字地看，直到看明白为止。

比如，系数的符号对吗，大小合理吗，主要变量在统计上显著吗，与经济理论矛盾吗？如果把结果告诉外行，对方会发笑吗？

7. 理解数据挖掘的成本与收益(Understand the costs and benefits of data mining)

不好的“数据挖掘”(data mining)指的是，根据数据的特征来设定计量模型。

这种数据挖掘之所以不好，是因为模型设定是由数据产生的，如果再用这个数据去检验由它产生的模型，那就不是客观的检验。

比如，进行回归之后，发现主要变量的系数不显著，有些研究者就想使用“何种方法”，能使得它变得显著。

“系数不显著”本身也是一种结果，也可以在论文中汇报。

如果数据本身没有问题，则它或许预示着，实证数据与理论相矛盾，故可能需要对理论进行修正。

好的“数据挖掘”指的是，对数据进行各种“实验”，以期发现数据中的某种规律性，发现模型设定的错误，以此改进理论。

一定程度的数据挖掘是不可避免的，而这两种形式的数据挖掘的界限并不清晰，正如 Heckman(2000)指出，“尽管使用数据来检验受到该数据启发的理论存在严重的问题，但如果拒绝从数据中学习并修改理论，则会导致更严重的问题”。

8. 权衡理论与现实(Be prepared to compromise)

在一般情况下，计量实践中的问题与理论计量经济学中最接近的情形会有一定的距离。

比如，使用的代理变量，是否满足教科书中对代理变量的要求？内生解释变量问题是否严重？单位根检验的结果是否可信？

通常需要在现实与理论之间找到一个适当的妥协。

9. 区分统计显著性与经济显著性(Do not confuse statistical significance with meaningful magnitude)

通常，样本容量越大，则估计值的标准误越小，该系数的“统计显著性”(statistical significance)就越高。

但如果该系数估计值为 0.001，则可能不具有“经济显著性”(economic significance)，即这个解释变量的变化对被解释变量的影响不大，尽管这种影响的幅度被估计得很精确。

10. 进行敏感度分析(Report a sensitivity analysis)

为得到论文的结果，研究者通常需要做一系列的假定。

论文的主要结果是否对这些假定很敏感？

有必要放松论文的假定，看结果是否稳健(基本不变)，这被称为“敏感度分析”。

比如，改变样本区间、函数形式、控制变量等。只有稳健的结果才是有说服力的。

32.5 结 束 语

在很多情况下，经济理论并不能给我们确切的答案。

比如，有关“中国的内需不足问题”，学界与媒体已经提出不下十种解释。究竟哪个(些)因素是导致内需不足的主要原因呢？

要真正回答这个问题，则必须仔细地收集数据，进行深入的实证分析。或许可以说：“理论是灰色的，而实证之树常青”。

有关如何做研究、写论文以及经济学方法论方面的讨论，可进一步参考 Friedman (1953), Stigler and Becker (1977), 林毅夫(1995, 2013), Varian (2001, 2009), Cochrane (2005), 李子奈(2011), 傅十和、佩雷拉(2012)等。