

## 第 20 章 如何做实证研究

### 20.1 什么是论文

我们已经学了不少计量知识。到底应如何应用于实证研究 (empirical research)?

撰写实证论文一般有哪些步骤与注意事项?

经济学的研究大致可分为两类，即理论与实证研究。

理论研究通过建立数学模型，来解释现实世界的经济现象。

理论模型虽然可以告诉我们变量之间的可能关系，但通常无法提供这种关系的作用系数。

这种理论关系是否成立，也仍需通过现实数据来检验。

实证研究主要通过收集样本数据，使用统计推断的方法，估计并检验变量之间的因果关系，或进行预测。

理论与实证研究的根本区别在于，前者基本不用数据，仅仅依赖数理逻辑得到结论；后者则主要基于数据进行统计推断。

根据理论与实证研究的区别，经济学论文也可大体分为理论论文与实证论文两大类。

也有一些论文既提出新理论，又直接进行实证检验；但这种论文一般篇幅较长，比较少见。

理论论文也可以援引实证数据或经济现象作为佐证。

实证论文也可以包含简单的理论框架。本章仅关注实证论文。

作为本科生，你必须完成一篇毕业论文才能毕业。

由于理论论文需要有理论上的创见，更为困难；故本科毕业论文一般都是实证论文。选修的课程也可能需要提交“学期论文”(term paper)。

为了保送研究生或出国深造，一篇尚未发表的工作论文(working paper)也会使你增色不少。

究竟什么是论文？

简单地说，论文就是对新的研究成果的汇报。

为什么一位成绩优秀的本科生，在撰写毕业论文时可能觉得一筹莫展？

因为平时上课做题，主要学习已有知识，只需被动消化吸收即可，有固定模式可循；而做研究写论文，则需主动创造(哪怕是一点点)新知识。

刚起步研究的学生，面临着从学习知识(学生)到创造知识(研究者)的转型。

论文与一般的文章或散文不同，后者可以仅仅表达某种情感，或记录一些事情。

经济学论文必须用十分严谨的数理逻辑或统计推断，来一步一步地得到结论，保证每个环节都丝丝入扣、经得起推敲；而不能随便发表议论，或轻率地下结论。

论文贵在创新，其价值主要在于其原创性(originality)或新颖性(novelty)，即对于已有文献的边际贡献(marginal contribution)，参见图 20.1。

本科论文并不要求有太多创新，但至少应有一点点创新；而绝不能是“山寨版”或抄袭。

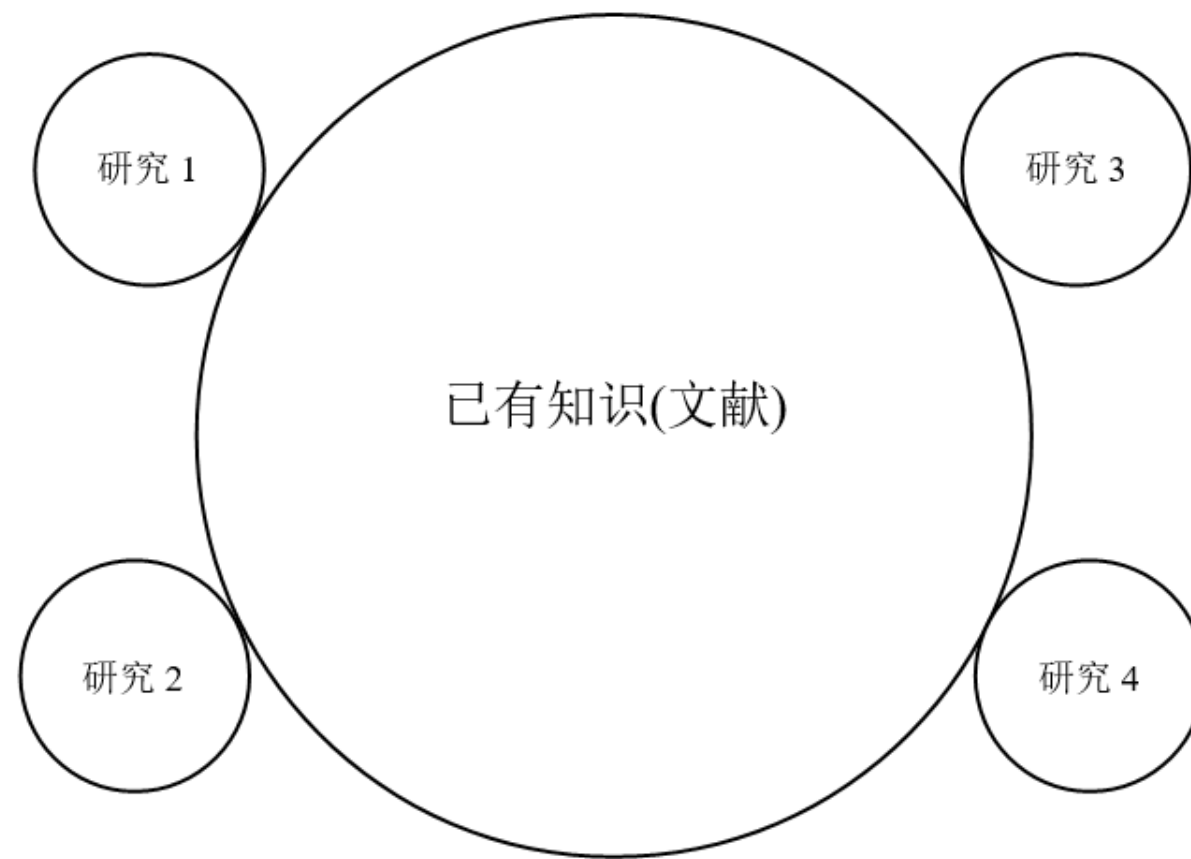


图 20.1 已有知识与新研究的关系

规范的实证研究包括以下几个步骤，即准备阶段、选题、探索性研究、收集数据、建立计量模型、选择计量方法、解释回归结果、论文写作、与同行交流、提交论文或投稿。

## **20.2 准备阶段**

要开始真正的研究，需要一系列的准备工作。

首先，必须掌握一定的经济理论，以获得观察经济现象的必要视角(perspective)、参照系(reference 或 benchmark)与分析工具(analytical tools)。否则，即使看到经济现象，也可能无从下手。

正如钱颖一(2002, p.2)所指出：

我在哈佛大学做博士生的时候，韦茨曼(Martin Weitzman)教授问我，受过现代经济学系统训练的经济学家和没有经过这种训练的经济学家究竟有什么区别？他研究比较经济制度，经常去苏联访问，问这个问题是从与苏联经济学家交往中有感而发的。

韦茨曼的回答是，受过现代经济学系统训练的经济学家的头脑中总有几个参照系，这样，分析经济问题时就有一致性，不会零敲碎打，就事论事。



这正是经济学界常说的“像经济学家那样思考”(Think like an economist)。爱因斯坦更早就说过类似的话，甚至更为深刻：

你能不能观察到眼前的现象取决于你运用什么样的理论，理论决定着你到底能观察到什么。

认为可以不需要任何理论指导而直接去“看真实世界”的想法，或许过于天真了。

经济理论的学习，主要体现在微观经济学、宏观经济学以及经济学的各专业课程上，比如金融学、财政学、发展经济学、产业经济学、劳动经济学等。

其次，为了进行实证研究，还必须掌握一定的计量方法与统计软件(比如 Stata)，这正是本课程的要旨所在。

即使你收集到相关的数据，但数据也不会“自己说话”，仍需要使用统计软件，运用适当的计量方法进行统计推断。

计量经济学对于实证研究不可或缺。

具备一定的理论功底与计量训练后，即可开始做实证研究。

## 20.3 选题

*Everything has been thought before, but the problem is to think of it again.* -- Johann Wolfgang von Goethe

实证研究的第一步就是选题，即选择研究的题目。

对于刚起步的研究者，常常不知如何选题。

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”之类的因果关系。

如果想研究“家庭联产承包责任制对农业经济增长的作用”， $X$  指“家庭联产承包责任制”， $Y$  指“农业经济增长”(Lin, 1992)。

实证研究也可以只有  $Y$  而没有  $X$ ，比如对于某个统计指标  $Y$  的测算；但纯粹描述性的研究已比较少见。

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)。

也可来自对经济现象的观察(比如媒体报道、社会调研)。

可以研究某政策的效应(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。

如果没有任何研究想法, 则建议先浏览一些经济学的顶级期刊。

比如, 经济学中文期刊的“四大金刚”, 即《经济研究》、《经济学(季刊)》、《世界经济》、《管理世界》。

经济学英文期刊的“Top 5”，即 *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*。

这些顶级期刊都是经济学的一般性期刊 (general interest journal)，涵盖经济学的各个领域。

如果确定研究经济学的某个领域，比如金融学，还可关注《金融研究》等专业期刊(field journal)。

浏览这些期刊中的论文，可大致知道当前的经济学者都在研究哪些问题，取得了哪些成果，还有哪些未解之谜；

进一步，可评估他(她)们的研究方法是否可靠，以及可能的改善空间。

如果能提出好的研究问题，也许你的研究就成功了一半。

什么是好的研究问题呢？

研究问题越具体、越有趣、越新颖、越有可行性，则越好！

(1)**具体**：在以上“ $X$ 对 $Y$ 有何作用”的句型中，应能明确 $X$ 对 $Y$ 具体是什么。

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

(3)**新颖**：论文的核心价值在于其创新性，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，或者使用了新的数据集。

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。

对于刚开始选题的学生而言，似乎自己能想到的题目，都已被别人做过了。其实未必。

要想做出新的边际贡献，当然可在前人的基础上，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(比如，Lin, 1992)。

即使是前人已经研究过的现象(太阳底下没有新的事物)，也可用新眼光、新视角去观察。

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)。



对于刚起步的新手，应尽量避免已经被研究得很烂、或过于富有挑战性的题目。

备选的研究问题越多越好，因为能“存活”下来的研究想法通常不多。

作为选题的示例(也为了启发思路)，下面以 2014 年 4 月《经济学(季刊)》发表的论文为例，分别列出每篇论文的  $X$  (关键解释变量)与  $Y$  (被解释变量)。

表 20.1 2014 年 4 月《经济学(季刊)》论文(按刊登顺序排列)

	论文题目	<i>X</i>	<i>Y</i>
1	中国资本回报率上升之谜	*	*
2	转轨时期我国行业层面资本积累的研究——资本存量 and 资本流量的测算	无	我国行业层面的资本存量与资本流量
3	监管资本、经济资本及监管套利——妥协与对抗中演进的巴塞尔协议	*	*
4	银行资产证券化的风险自留监管：作用机制和福利效果	*	*
5	健康对中国中老年人劳动供给的影响	健康(高血压)	中老年人劳动供给

6	中国农村人均家庭收入流动性研究：1986—2010 年	无	中国农村人均家庭收入流动性
7	养老保险企业缴费对员工工资、就业的影响分析	养老保险企业的缴费比例	企业人均工资、人均福利与雇佣人数
8	基于双胞胎数据的教育收益率估计	个体受教育年数	个体收入对数
9	中国大城市的工资高吗？——来自农村外出劳动力的收入证据	城市规模	农村外出劳动力的收入对数
10	平均相关系数与系统性风险：来自中国市场的证据	股票组合与债券组合的平均相关系数	可观测资产组合的超额收益率
11	中国投资者具有信息获取能力吗？——来自“业绩预告”效应	投资者信息获取行为(股票简称的百度超额搜索指数)	公布业绩预告前后的累积收益率变化

	应的证据		
12	限行政策是否改善了北京市的空气质量?	北京市的限行政策	空气质量
13	遗传距离对跨国收入差距的影响研究:理论和中国的实证分析	中国与其他国家的遗传距离(追溯到共同祖先所过去的时间)	中国与其他国家人均 GDP 对数的差距的绝对值
14	劳动力转移对中国农业生产的影响	农户劳动力外出(农户是否为壮年、男或女劳动力转移)	农户退出农业、农户土地净转出率、农业生产的各项指标
15	中国各省经济增长是否收敛?	期初各省人均 GDP (1952-1955 年)	期末各省人均 GDP (2008 年)
16	学区房的溢价究竟有多大:利用“租买不同权”和配对回归的实证估计	小区是否为重点小学学区房	小区二手房平均交易均价对数的差价

17	“哥伦布大交换”终结“气候—治乱循环”吗？——对玉米在中国引种和农民起义发生率的一项历史考察	各省的玉米种植时间(玉米种植强度的代理变量)	各省每年的农民起义的发生数(1470-1900 年)
18	创意产品出口、模仿威胁与知识产权保护	进口国的知识产权保护水平(问卷调查)	美国的创意产品出口额(2006-2010 年)

注：打“\*”号者为理论论文，较少实证分析。

在该期 18 篇论文中，纯理论的论文仅占 3 篇，而实证论文多达 15 篇(少数论文兼具理论与实证分析)。

在这 15 篇实证论文中，绝大多数的研究范式均为考察  $X$  对  $Y$  的因果效应。

从研究内容来看，既有不少传统的金融学、劳动经济学、发展经济学的题目(但均在研究方法上有所创新)，也有对于限行政策、遗传距离、学区房、哥伦布大交换等较为新颖的研究。

实证经济学的研究方法几乎可以应用于所有的社会经济现象。

## 20.4 探索性研究

*If I have seen further it is by standing on the shoulders of giants.*  
-- Isaac Newton

尽信书，则不如无书。 -- 《孟子·尽心章句下》

有了潜在的研究问题后，首先需要进行初步的“探索性研究”(exploratory study)，看看它是否具有新颖性与可行性。

比如，通过查找文献，考察别人是否已经做过类似研究，并大致了解数据是否可得。

### **(1) 通过文献回顾评估选题的新颖性**

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，但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研究。

如果别人已做过很相似的研究，则通常须更换题目；除非另辟蹊径，找到很不相同的方法或数据。

对于中文论文，可在 CNKI(China National Knowledge Infrastructure)中搜索。

对于英文论文，可在 JSTOR(Journal Storage)或 EconLit with Full Text (美国经济学会)输入关键字进行搜索；二者均全文收录了许多经济类英文期刊，但前者有几年滞后。

对于二者未覆盖的经济类期刊，可通过一些主要出版社(集团)搜索，比如 Elsevier Science Direct, Springer Link, Taylor & Francis, Wiley 等。某些工作论文则可通过百度或谷歌搜索。

以山东大学图书馆为例，其电子资源的第一页提供了如下资源(参见图 15.2)，翻页即可获得更多资源：





山东大学图书馆

Shandong University Library

IFC易瑞授权访问系统 V6.1  
IFC IRAS Authorized Access System

电子资源 Electronic Resources

常见问题 Question

公告 Announcement

留言 Message

退出 | 修改密码

今天是 2014年11月10日 17时11分

search:  搜索

您收藏的电子资源有:

资源名称 Name	简介 Introduction	访问统计 Statistics	收藏 Favorites	全部类型 Type	语言 Language
ACM Digital Library(清华镜像)	简介	14348次	收藏	期刊	外文
ACS	简介	52633次	收藏	期刊	外文
AIP	简介	1020次	收藏	期刊	外文
AIP(scitation)	简介	325次	收藏	期刊	外文
Apab教参类电子书	简介	6009次	收藏	图书	中文
Apab电子书	简介	19984次	收藏	图书	中文
APS	简介	4984次	收藏	期刊	外文
CALIS西文期刊目次数据库(CCC) 新版	简介	22601次	收藏	期刊	外文
Cell Press	简介	3825次	收藏	期刊	外文
CNKI中国学术文献总库(远程包库)	简介	1299128次	收藏	期刊 学位论文	中文
CNKI年鉴、工具书数据库	简介	23613次	收藏	其他	中文
EBSCO (ASC&BSC)	简介	20403次	收藏	期刊	外文
EconLit with Full Text	简介	4647次	收藏	期刊	外文
Elsevier ScienceDirect	简介	162129次	收藏	期刊	外文
Emerald全文期刊库	简介	3221次	收藏	期刊	中文
Emerald回溯内容全国在线	简介	1133次	收藏	期刊	外文

首页 上一页 1/6 下一页 尾页 共89项 去往  页 go

图 20.2 山东大学图书馆电子资源首页

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阅读文献，则可能被文献所淹没，望洋兴叹，自觉渺小。

更好的方法是，当自己有了一定的想法之后，再去系统地看文献。这样才知道自己究竟要看什么，也更能带着批判的眼光去看。

阅读文献的态度也十分重要。

虚心地从经典论文中汲取营养，才能站在巨人的肩膀上，但仍应带着某种批判性的眼光。

由于经济现象的复杂性，任何论文都有一定缺点，也都有可以改进的空间，故不必太迷信权威。

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的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，你怎么可能做出新的边际贡献呢？

## **(2) 确定所需数据是否可得**

在正式开始研究之前，还应大致知道所需要的数据不仅存在，而且可以得到。

数据从何而来？

一般来说，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)。

寻找数据可以从网络搜索开始(比如谷歌或百度)，也可以询问专家或同行。

如果确实不知道该从哪里找数据，还可关注文献中同类研究的数据来源，然后溯本及源。

阅读一定文献之后，就应该基本了解该研究领域的常见数据来源了。

近年来，一些国内外期刊已在其网站公开了发表论文中所用的数据集与估计程序。本章附录也列出了一些常用的数据来源。

## 20.5 收集与整理数据

从数据的来源格式来看，数据可分为电子版与非电子版两大类。

对于非电子版的数据，需耐心输入数据(通常先输入 Excel 表，再导入 Stata 中)，并注意检查，防止出错。

即便下载电子版数据，也应检查可能存在的错误。

实证研究的关键材料乃是数据。

如果数据质量不高，则“巧妇难为无米之炊”。

无论多么高深的计量方法，如果原始数据质量有问题，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)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。

将数据导入 Stata 后，需仔细察看数据(inspect the data)。

一个常见误区是，研究者只知进行回归，却不去熟悉原始数据(raw data)，或增加对数据的感觉(get a feel for the data)。

察看数据的常见方法为，计算变量的主要统计特征(summary statistics)，包括均值、最大值、最小值、标准差、相关系数等，并根据经济常识判断它们是否合理。

比如，虚拟变量的最小值与最大值必然为 0 与 1；否则，此变量有误。

如果数据有时间维度(比如时间序列或面板数据)，还可画时间趋势图。

如果发现在某个时点上的变量取值异常波动，则应考察此数据是否有误；即使数据无误，也应考虑异常波动的原因。

在察看数据的过程中，主要观察数据中是否存在不一致(inconsistent)的地方；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)。

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、世界银行)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，应考虑数据质量的影响)。

在使用别人提供的数据时，还应注意其定义及统计口径，是否是**与理论模型中的变量相对应**。

比如，中国的失业率一般指的是“**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有很大不同。



对于计量的初学者来说，与真实数据打交道也是加深对计量经济学理解的重要途径。

只有弄脏你的手(get your hands dirty)，才能真正学会做实证研究。

## **20.6 建立计量模型**

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。

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。

一篇好的实证论文，需要讲一个好的“故事”(story)，然后用数据来证实或检验此故事。

对于回归函数的具体形式，可以考虑线性、对数(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。

在进行模型设定时，应尽量使用常识(common sense)与经济理论(economic theory)。

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。

又比如，考虑 FDI 对经济增长的作用。

由于 FDI 起作用需要时间，如果把当年的增长率对当年的 FDI 进行回归，可能没有太大意义。

比较适当的做法是，考虑期初的 FDI 对随后五年(或若干年)经济增长的作用(这样做也可缓解双向因果关系)。

如果不确定该如何设定计量模型，可借鉴文献中同类研究的模型设定。

模型既不能过于简单(解释变量过少),也不宜过于复杂,而应当保持适当的简洁(keep it sensibly simple)。

在选择解释变量时,“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);而“从大到小”(general-to-specific)的建模方法偏差小,但却不易执行。

实践中,常采用折衷方案,即选择简单而有解释力的模型。

## **20.7 选择计量方法**

有了计量模型与数据之后,即可根据数据类型与特点,选择合适的计量方法。

比如，被解释变量为虚拟变量，则可使用 Probit 或 Logit。

如果是面板数据，应考虑固定效应、随机效应、时间效应等。

如果是时间序列，则须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，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)。

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。

受数据可得性(availability)的限制，遗漏变量几乎不可避免。

很有必要在实证论文中对此进行讨论。不外乎以下两种情况。

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，例如增加控制变量、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。

一般需找到有效的工具变量，或使用其他因果推断方法，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。

比如，对于中国的宏观变量，如果使用全国的时间序列，则一般样本容量较小。

可考虑收集省际面板(provincial panel)的相应数据。

大多数的实证论文都希望说明  $X$  对  $Y$  的因果作用。

从回归分析的相关关系升华到因果关系，是很大的飞跃，需要使用适当的计量方法来识别这种因果关系。

在这部分应该说明，为什么所用的计量方法是最恰当的。

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些假定。

尽管计量理论是可以严格证明的一门科学，但实证研究在一定程度上却是一门艺术，通常需要在理论与现实之间找到适当的妥协(*be prepared to compromise*)。



## 20.8 解释回归结果

*There are two things you are better off not watching in the making: sausages and econometric estimates.* -- Edward Leamer

使用计量方法估计模型后，计算机软件(比如 Stata)将输出相应的计量结果。

此结果可能较长，包含密密麻麻的表格与数字。

如何看这些结果？只能一个字一个字地看，直到看明白为止。

当然，也有诀窍。计量结果可能很复杂，但真正重要的信息通常不多，比如回归系数(含符号)、 $p$ 值，以及样本容量、拟合优度等统计量。

以一元回归为例(工资对数对教育年限回归), 回归结果参见图 15.3。

Source	SS	df	MS	Number of obs = 758		
Model	35.2039946	1	35.2039946	F( 1, 756) = 255.70		
Residual	104.082155	756	.137674809	Prob > F = 0.0000		
Total	139.28615	757	.183997556	R-squared = 0.2527		
				Adj R-squared = 0.2518		
				Root MSE = .37105		

lnw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
s	.0966245	.0060425	15.99	0.000	.0847624	.1084866
_cons	4.391486	.0821136	53.48	0.000	4.230288	4.552684

图 20.3 工资对数对教育年限的回归结果

变量  $s$ (教育年限)的回归系数符号为正(与经济理论相符)，系数估计值为 0.0966245， $p$  值为 0.000(在 1%水平上具有统计显著性)，样本容量为 758，而拟合优度  $R^2$  为 0.2527(教育年限可解释工资对数约四分之一的变动)。

在图 20.3 中，左上角的残差平方和、右下角的置信区间，乃至常数项等信息，基本可以不关心(除非有需要)。

在解释回归系数时，还应注意区分统计显著性与经济显著性。

**统计显著性**(statistical significance)主要通过  $p$  值来考察，如果  $p$  值小于或等于 0.05，则意味着该系数在统计上显著地不等于零；反之，则在统计上不显著，在统计上可将此系数视为零(不存在)。

**经济显著性**(economic significance)主要通过系数的绝对值来考察，须特别注意变量的取值单位。

在上例中，解释变量教育年限  $s$  的单位为年，而被解释变量工资对数  $\ln w$  可解释为工资的百分比变化，故  $s$  的回归系数为 0.0966245 意味着，每增加一年教育，未来工资收入将提高 9.66%，具有很高的经济显著性(可能过高了)。

反之，假如  $s$  的回归系数为 0.01 或 0.001，则意味着每增加一年教育，未来工资收入只会上升 1%或 0.1%，在经济意义上不显著。

统计上显著而经济上不显著，则意味着解释变量对被解释变量的影响很小(经济上不显著)，尽管这种影响的幅度被估计得很精确(统计上显著)。

在进行计量检验时(比如, 豪斯曼检验), Stata 可能会输出很多结果, 但最需要关注的只是原假设以及  $p$  值; 因为知道二者就可以进行检验了, 而其余信息都是细节。

研究者通常花费较长时间收集与整理数据。

将数据导入 Stata, 然后输入相应的回归命令, 则是“见证奇迹的时候”(moment of truth)。

如果关键解释变量兼具统计与经济显著性, 符号也与理论预期一致, 而其他控制变量的符号与显著性也大体与预期相符, 则会感到十分欣慰, 过去收集整理数据的辛劳也都值了。

但有时, 所得计量结果未必尽如人意, 比如关键解释变量不显著, 甚至符号与预期相反。此时应怎么办呢?

出现这种情况，可能有如下三种原因。

(1) 使用计量方法不当。

比如，在上述一元回归中，显然遗漏了许多变量，可能存在遗漏变量偏差，导致 OLS 估计不一致。

如果存在内生性而未加以处理，将导致不一致的估计，使得本应显著的变量变得不显著。

(2) 数据质量有问题。

如果数据存在较大的度量误差，所用代理变量与真实变量相差较远(由于真实变量不可观测)，或者数据输入中的人为错误，都有可能影响估计的一致性。

### (3) 经济理论有问题。

在排除了以上两种可能性之后，最后一种可能性是，经济理论不正确。

经济理论所预期的某种效应可能不存在；或者同时存在其他作用机制，使得净效应的符号相反。

实证研究的目的之一就是检验经济理论。

如果发现已有理论与经验证据不符，则说明此理论尚有改进空间，甚至需要放弃。

正如林毅夫(2001, p. 75)所指出：

如果发现理论推论和我国经验事实不一致,要坚持的不是现有的理论,而是进一步去了解我国的经验现象,然后,根据经验现象构建一个可以解释这个现象的理论。所以,当发现这种不一致时,不要死抱理论,成为现有理论的俘虏,也不要巨人的面前而感到自己渺小。其实,这正是对理论发展做出贡献的绝好机会。

在计量实践中,研究者经常根据计量结果而调整模型,以期得到更为理想的结果,并且只在论文中汇报最佳的结果,而将寻找此结果的过程隐去。



这实际上是数据挖掘(data mining)的一种形式。

数据挖掘既有成本(缺点)，也有收益(优点)。

数据挖掘的优点是，可对数据进行各种“实验”，以期揭示数据中的某种规律性，发现模型设定的错误，以此改进理论或计量模型。

数据挖掘的缺点则是，由于它根据数据特征来设定计量模型，故模型设定由数据产生，如果再用此数据去检验由它产生的模型，就不是客观的检验，由此导致偏差。

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”<sup>①</sup>。

解决数据挖掘所带来的偏差的方法之一是进行稳健性检验,而不是仅汇报最佳的结果。

## 20.9 诊断性检验

任何计量方法都有其适用的前提条件;如果前提不成立,则无法使用此计量方法(可能导致不一致的估计)。

---

<sup>①</sup> 英文原文为 “While there are serious problems in using the data that suggest a theory to test that theory, even more important problems arise from refusing to learn from the data in revising economic models.”

在估计完模型后，应对计量方法的前提条件进行**诊断性检验**(diagnostic checking)或作出定性说明。

比如，使用工具变量法进行 2SLS 估计后，应进行弱工具变量检验、过度识别检验(假设存在过度识别)、解释变量内生性检验；并从定性的角度说明“排他性约束”(exclusion restriction)为什么成立。

又比如，使用时间序列估计自回归(AR)或向量自回归模型(VAR)，则应检验残差是否为白噪声(无自相关)。

即使进行 OLS 回归，也应说明解释变量为什么外生，或者遗漏变量偏差为什么不重要。

## 20.10 稳健性检验

*We have to learn...that the prime virtue of any econometric procedure is robustness.* -- Robert Solow

为了使用特定的计量方法，研究者通常需要做一系列的假定。

问题是，论文的主要结果是否对这些假定很敏感？

有必要放松论文的某些假定，看结果是否稳健或基本不变，这称为**稳健性检验** (robustness check) 或**敏感性分析** (sensitivity analysis)。

比如，通过改变样本区间(或去掉极端值)、函数形式、计量方法、控制变量、变量定义、数据来源等，来考察计量结果的稳定性。

在计量实践中，研究者通常会通过数据挖掘，找到“最佳”的计量模型。

如果仅汇报此最佳模型，则会导致偏差。

有必要适当地改变模型的设定，比较其主要结果的变化。

只有稳健的结果才有说服力，故稳健性检验已成为高质量实证论文不可或缺的一部分。

## 20.11 论文写作

*Learn to write but also write to learn.* -- William Thomson (2001, p.1)

得到较为理想的实证结果之后，即可开始写论文。

论文就是对研究成果的汇报。

为了便于读者更快地从论文中获取信息，经济学论文通常有一定的结构，而论文写作本身也是一门精益求精的艺术。

经济学家曼昆曾问过加尔布雷斯(John K. Galbraith)写作成功的秘密；加尔布雷斯回答说，他写的所有东西都会修改很多次，通常直到第 5 稿时才会基本满意(Mankiw, 1996)。

## 1. 标题、关键字、摘要

论文的首页通常包括标题、作者、摘要、关键字等信息。

**标题(title)**是论文的标签，正如商品的商标或名称。

一般应选择简洁而有吸引力的标题，并能让读者知道该文主要做什么。

在论文写作乃至成文之后，都有可能修改论文题目，使之更为贴切有趣。

在题目之下一一般为作者姓名，而将具体的作者单位、联系方式、感谢语(包括基金资助)以及“文责自负”等声明放在脚注里。

如果有多位作者，一般需选择其中一位作者作为“通讯作者”(corresponding author)，负责投稿并与编辑部保持联系。

在题目与作者之下，一般为**摘要**(abstract)，通常在 100 字左右。

摘要需突出论文的重要意义、研究方法与主要结论。

一般读者会先看摘要，再决定是否看全文。

论文摘要应字斟句酌，凸显本文的主要贡献，并激起读者进一步阅读的兴趣。

摘要通常在论文主体完成后才撰写，因为此时作者对于论文的主要内容会有更清晰的概念。

在摘要的下面，通常还需提供几个**关键字**(key words)，以便读者能很快地根据关键字搜索到此文。关键字常常来自论文的题目。

在关键字之下，还可能提供 JEL 分类号，这是美国经济学会主办的 *Journal of Economic Literature* 杂志所用的经济学各领域的分类编号。



## 2. 正文部分

经济学实证论文的正文一般依次包括以下部分：引言、文献回顾(可归入引言)、理论框架或背景介绍(可省略)、数据说明、计量模型与估计方法、回归结果、稳健性检验 (可归入回归结果)、结论。

### (1) 引言(Introduction)

引言虽是全文的第一部分，却经常最后写。

引言集中了全文的卖点(selling points)，最难撰写，须反复修改。

引言概括了全文的内容，只有在全文大体完工后，才能准确地总结与提炼。

引言通常包括以下内容：

本文研究了什么问题，此问题为什么重要(研究意义)；本文使用了什么数据(最好在数据来源上有所创新或挖掘)，实证研究的计量方法是什么，得到了哪些主要结论；此研究与已有文献的关系，本文的主要创新与边际贡献等。

引言将论文的精华部分以非技术性的方式呈现给读者，可视为扩展版的摘要，是“销售”此文的重要手段。

许多读者在浏览论文时，常常先看引言与结论，然后再决定是否细读正文；可见引言的重要性。

引言的写作大致有两个套路。

传统的套路是，在提出研究问题之后，首先回顾已有文献的相关研究以及不足之处，然后顺势引出本文的研究方法与主要贡献(比如，填补了文献的空白)。

传统套路的优点是，比较有逻辑性，能自然地呈现学术发展的脉络；其缺点在于读者需要有一定耐心，先回顾主要文献，然后才知道本文的主要工作。

现代的套路是，提出问题之后，马上直奔主题，介绍本文的研究方法与主要结论，然后再回头介绍本研究与现有文献的关系。

这两种套路各有优缺点，适合不同的论文，但直奔主题的现代套路似乎日益流行。

另外，引言的最后一段通常提供全文的路标(roadmap)，告诉读者本文的其余部分在结构上如何安排，以便于读者阅读。

## (2) 文献回顾(Literature Review)

文献回顾如果较短，可以归入引言部分；反之，如果文献回顾较长，则可单独作为论文的一个部分。

对于文献的回顾一般按文献出现的时间先后进行，着重介绍重要的文献，而其他文献可以简略介绍、放入脚注，甚至略去。

文献回顾的写作切忌只是堆砌罗列一些文献，而未进行深入分析。

文献回顾的根本目的是为了厘清本文的研究与已有文献的关系，以凸显本文的边际贡献及其在文献中的地位。

在肯定现有文献的原创贡献外，难免会指出其不足之处(或者被忽略的方面)。

应注意语气委婉，因为这些文献的作者有可能正是未来的审稿人或编辑。

另一方面，你又希望突出本文的独特贡献(当然必须实事求是)。

在指出现有文献的不足与突出本文的贡献之间，需要找到措辞与语调上的平衡。

### (3) 背景介绍(Background Information)或理论框架(Theoretical Framework)

实证论文并非仅仅是找一堆数据，然后汇报回归结果。

只有告诉读者有关经济现象的背景，完整地述说一个经济故事，才能使得计量结果更有说服力。

比如，Nunn and Qian (2011)研究引入“新世界”(New World)作物土豆对“旧世界”(Old World)人口增长与城市化的影响，在其第二节背景部分，即以大量篇幅介绍土豆的优点(virtues of the potato)、土豆如何从新世界传播到旧世界，以及其他新世界作物。

实证研究者的工作并不仅仅是下载数据进行回归，还需要熟悉所研究现象的历史、制度与文化背景，乃至数据的来源与产生过程。

如果可能，在此部分可引入一个简单的理论模型(theoretical model)或思想框架(conceptual framework)，为后续的实证研究提供理论基础。

对于实证论文而言，其理论部分不宜太过复杂，以致喧宾夺主。

如果经济现象过于复杂，没有现成的理论，也可根据常识(common sense)直接写下计量模型或回归方程。

#### (4) 数据说明(Data description)

实证论文的结论是否可靠，首先取决于数据的质量。

在数据说明部分，应详细说明数据的具体来源，并评估其可靠性。

介绍数据来源的详细程度，应使读者能按图索骥得到同样的数据，以保证科学结果的可重复性。

如果对原始数据进行了一些处理或加工，也应一一说明。

如果学术界对于数据的质量有质疑，则应说明这些潜在的数据质量问题，对于你的研究有何影响。



如果数据来自问卷调查，则应说明随机抽样如何进行，问卷如何发放与执行等，并在附录中附上具体的问卷。

介绍数据来源之后，通常以表格形式给出主要变量的统计特征(summary of statistics)，比如样本容量、均值、标准差、最小值、最大值等。

有时，还会提供关键变量的相关系数矩阵(matrix of correlation)，作为对变量之间关系的初步证据。

#### (5) 计量模型与估计方法(Econometric model and estimation)

在此部分，需要结合所研究的问题以及已有数据，给出具体的计量模型，即回归方程。

通常会有一个基准(baseline 或 benchmark)的计量模型，然后在此基础上对模型设定(model specification)有所变化，比如增加或替换变量。

此部分着重需说明论文的估计策略(estimation strategy)，即究竟应使用什么计量方法来识别主要变量之间的因果关系。

初学者易犯的错误是，在论文中直接使用某计量方法，而未说明为什么这是最合适的计量方法。

任何计量方法都有适用的前提条件，需要研究者仔细甄别与判断。

如果有两个计量方法，各有优缺点，则可二者都用，然后作为稳健性检验，比较二者的结果。

## (6) 回归结果(Regression results)

介绍计量方法之后，即可汇报回归结果，通常以表格形式来呈现，主要包括以下信息：

被解释变量与解释变量的名称、回归系数估计值、标准误(或 $t$ 统计量)，以星号表示统计显著性，以及相关的统计量(样本容量、 $R^2$ 等)。

在正文中，需要对回归结果进行解读，包括回归系数的统计显著性与经济显著性，符号是否与理论预期相符等。

## (7) 稳健性检验 (Robustness checks)

在实证论文中仅仅汇报一个回归结果显然是不够的，因为变量的显著性可能在不同的模型设定下变化。

只有在不同的模型设定下，都能得到类似的结果，才是稳健与可信的。

对于稳健性检验的结果汇报，如果篇幅比较短，可归入上一部分的“回归结果”。

反之，如果做了较多的稳健性检验，则可单独作为论文的一个部分。

## (8) 结论(Conclusion)

结论是论文的最后部分，对全文所作工作进行总结，并给读者留下最后的印象。

结论部分通常概要地回顾本文的研究问题、计量方法与主要结论，也可重申本文的独特贡献。

由于任何论文都有局限性，故也可指出未来的改进空间与研究方向。

许多读者会先看引言与结论，再决定是否看正文，故结论部分也十分重要。

### 3. 参考文献(References)

几乎所有研究都建立在前人成果之上，故必然会在文中引用他人的论文或著作。

这些论著的详细出处，则收集于文末的参考文献。

文中所有引用的论著，都应包括在参考文献中；反之，所有参考文献中的论著，都应在正文中被引用。

参考文献的顺序一般按照作者姓氏的字母(拼音)进行排列，对于同一作者的作品则按发表年代排序。

不同期刊对于参考文献的具体格式也有不同要求；在投稿前一般需按所投期刊的要求进行修改。

## 4. 附录(Appendix)

有些论文还有附录，主要收集不影响正文阅读，但篇幅较长的细节。

比如，对于理论文章，可能把繁琐的证明放在附录。

对于实证论文，有时会把过长的数据说明放在附录。

如果数据来自问卷调查，则通常把具体的问卷放在附录。

## 5. 写作风格

经济学论文属于科学类的论文，并不需要过于华丽的词藻，而应首先注意行文的简洁与逻辑性。

优美流畅的文笔对于提高论文可读性、吸引读者注意力十分重要。

对于初次写论文者，首先要注意“书面语”与“口语”的区别，避免过分口语化。

不能嘴上怎么说，笔下就怎么写，而应使用更为洗练到位的书面语言。

在下笔之前，可先在脑海里构思文章的结构与写作风格。

写作的过程也是使思路更加清晰的过程。



对于论文中的方程式,可使用 Word 文档中的“insert”→“object”→“Microsoft Equations”进行编辑,使得方程更为美观。

论文中所有单独成行的方程式,都应按顺序编号,以(1)、(2)、(3)等表示,以便于检索。

对于论文中的表格与图片,也应注意其格式。

表格的标题应在表的上方;而图片的标题则在图的下方。在表格或图片的下方,还可以有注释,说明数据来源、变量定义等相关信息。

对于初学者,建议仔细观察经典论文的文章结构与风格,并注意模仿。

比如，中文论文可以模仿《经济研究》或《经济学(季刊)》

英文论文则可参照 *American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics* 等。

正如古语所云，“熟读唐诗三百首，不会作诗也会吟”。

有关经济学论文的写作风格，可进一步参考傅十和、佩雷拉(2012)。

## **20.12 与同行交流**

论文初稿完成后，通常不宜直接投稿，或作为毕业论文提交。

在研究与写作过程中，难免受到个人先入为主的主观限制，出现这样或那样的偏差或疏忽。

因此，恳请导师、同行或朋友阅读你的论文，并提出批评与修改意见，是十分必要的。

更正式的渠道包括将论文提交至学术会议，或应邀到相关学术机构作报告；以便收集有益反馈，然后进一步修改论文。

当代论文的复杂程度越来越高，需要考虑的问题也越来越多，如果一味闭门造车则难免挂一漏万，难以保证论文的高质量。

## 20.13 提交论文或投稿

经过与同行交流并将论文修改完善后，可考虑提交毕业论文，或将论文投稿到合适的期刊。

在选择期刊时，首先要评估论文的重要性与质量，即该文是否研究了一个重要或有趣的问题，以及所用方法是否严格、结论是否可信；然后再将论文投给相应档次的期刊。

这里所说的“重要”，并不一定非要是影响国计民生的重大课题，也可以只有学术上的意义。

即使只是一个有趣的小问题，如果使用了严格的研究方法，也可能很有价值。

反之，如果研究方法有漏洞，即便研究的是大问题，也可能大而无当。

如果不清楚论文该投给哪个杂志，可请教导师或有投稿经验的前辈。

在投稿时，切忌“一稿多投”，即将一篇稿件同时投给多个杂志，造成编辑部的审稿资源浪费。

这是投稿的基本规则；如果违背，可能导致严重的后果。

只有在被拒稿或主动撤稿之后，才能将稿件投给另外一个期刊。

期刊编辑部在收到稿件后(通常为电子投稿)，一般由主编(editor)或共同主编(co-editor)先行浏览，并决定是否送外审。

如果不送外审，则会直接在案头拒稿(desk reject)。

对于送外审的论文，主编通常选择 2-3 位匿名审稿人(anonymous referee)进行审稿，并要求在规定的时间内(比如一个月或更长时间)提交审稿人报告(referee report)。

有时匿名审稿人也不知道论文作者的身份，这称为“双向匿名审稿”(double blind review)。

在互联网时代，审稿人通常不难查到作者身份(很多作者会把工作论文挂在网上)，双向盲审也就失去了意义。

*American Economic Review* 从 2011 年开始改用“单向匿名审稿”(single blind review)，即作者不知道审稿人身份，而审稿人知道作者身份。

主编在收到全部审稿人报告后，通常根据这些审稿意见，对论文采取以下决定：(1) 直接接受；(2) 直接拒稿；(3) 修改再投 (Revise and Resubmit, 简记 R&R)。

第一种情况(直接接受)非常少见，而第三种情况则说明此稿件有希望发表，应根据审稿人的建议进行认真修改。

有时，“修改再投”可能会发生 2-3 轮，而且越是顶尖的期刊，修改再投后被拒稿的可能性越大。

从论文投稿到期刊发表，即使在最顺利的情况下，也通常需要一年时间(除非编辑部对重要稿件加急处理)；而英文期刊的发表周期则可能更加漫长。

在此期间，还可能需要几经修改，甚至转投多个期刊。

这时你会发现，写论文并不难，难的是发表论文。

唯有切实提高论文质量，才是发表论文的根本保证。

## **20.14 写作伦理**

在论文写作过程中，应特别注意引用的规范性，并杜绝抄袭。

究竟“引用”(citation)与“抄袭”(plagiarism)有何区别？

二者最本质的区别在于，引用给出了信息的出处；而抄袭未提供出处，让读者误以为是作者的原创。



抄袭可以定义为“将已经存在的思想或产品‘偷来’作为自己的思想或产品”(The Modern Language Association of America, 2009, p. 52)。

抄袭是一种严重违背学术规范与职业道德的行为，可能导致无可挽回的严重后果。

首先，它将别人的思想占为己有，等于“偷窃”了别人的知识产品。

其次，将别人的知识产品作为自己的成果发表以获得好处，这等同于“欺诈”(fraud)。

初学者在使用别人的研究成果时，一定要注意通过正确的引用来注明出处。

如果直接引言别人的原话，应加上双引号，并注明文献来源。

如果大段地复制已有文献而未标明出处，则为赤裸裸的抄袭，应坚决杜绝。

即使是间接引用，比如用简洁的语言概述前人的思想，或将已有模型作了小的改动，也应及时注明其出处。

## **20.15 结束语**

如何才能做出高水平的实证研究？如何才能写出高质量的经济学论文？

更进一步，如何才能成为好的经济学家或经济工作者？

要达到这些目的，绝非单一学科就能胜任，而需要全方位的学识与素养。引用凯恩斯的一段话作为结束语，与大家共勉：

经济学研究似乎并不需要任何极高的特殊天赋。与更高深的哲学或纯科学相比，经济学不是……一门极其容易的学科吗？一门容易的学科，但这个学科中很少有人能出类拔萃！这个悖论的解释也许在于杰出的经济学家应该具有各种天赋的罕见的结合。在某种程度上，他应该是数学家、历史学家、政治家和哲学家。他必须了解符号并用文字表达出来。他必须根据一般性来深入思考特殊性，并在思绪奔放的同时触及抽象与具体。

他必须根据过去、为着未来而研究现在。他必须考虑到人性或人的制度的每一部分。他必须同时保持果断而客观的情绪，像艺术家一样冷漠而不流俗，但有时又要像政治家一样脚踏实地。

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, 2001, 2013), Varian (2001, 2009), Cochrane (2005), 田国强(2005), 洪永淼(2007), 李子奈(2011), 傅十和、佩雷拉(2012), 加里·金等(2014)等。