

How to Get Started on Research in Social Science? An Essential Guide to Planning Your Project

Weiwei Zheng (郑维伟)

Antai College of Economics and Management
Shanghai Jiao Tong University

2024 年 12 月 4 日

Highlights

- 1、为什么做研究？
- 2、确定怎样的选题？
- 3、如何进行高效文献检索？
- 4、怎么安排理论分析与假说提出？
- 5、检验假说的主要方法
- 6、论文写作与发表

社会科学研究的意义

研究 = 研发？

- **研究**：发现新知识、发明新技术、揭示新理论
- **研发**：基于已有的知识和技术进行研制、开发

科学研究扩展人类的知识

没有科学研究就没有技术进步

引申：**社会科学是不是就没有意义？ NO**

- 与自然科学差异性：研究对象、研究目的、研究领域、研究方法
- 现代社会具有合理化、分权，结构分化、独立自主的组织和个人复杂特征
- 社会科学作为专业知识和专家系统的发展成为社会必要

把你的工作告诉同行

- 经过同行评审 (Peer Review)，成为科学文献

通过毕业论文答辩是获得学位的重要前提

持续发表是学者（事业）稳步发展的路径

确定怎样的选题？想法 (idea)

遍地是黄金

- “水喝多了，总会有料” —— 季羨林
- 研究领域不存在“好”“坏”之分，只要做得**足够**深入都能做出好的工作
- 特定时期某些领域可能更活跃，相对出成果更多

可遇不可求 (开创性)

- 从生活出发，而不是从文献出发！—— **我发现了一个有意思的现象**
- 从 Topic 开始，先要进入一个具体的分支领域，并获得必要的了解
- **ESG** (谁来评级[明晟、富时罗素…自身]、对谁评级 [集团、母公司] ?)
- 《**基金 ESG**投资承诺效应——来自公募基金签署 PRI 的准自然实验》

选题需要具有一般性，考虑国际读者兴趣

- **北京非首都职能** v.s. **区域市场一体化**
- 《From agglomeration to dispersion: How does China's noncapital functions' relief affect regional development?》
- 选题时就需要考虑好大概可以投到哪些期刊
- 学科交叉时，要考虑侧重哪个领域的期刊

文献检索

中文文献

- **CNKI旧版高级检索**、万方、维普、国家哲学社会科学文献中心[free]…
- 关注**特定期刊** [主页、过刊浏览]：如学院奖励期刊List
- 擅用**高级检索**筛选期刊等级 CSCI
- **优秀学位论文参考**

外文文献

- Google Scholar [**镜像**]、Web of Science、Elsevier、Springer…
- 关注**特定期刊** [Journal Home、History Issues]
- 作者**个人主页**、Scopus
- **Working Paper** 追热点：NBER、SSRN、arXiv…
- Bing搜索引擎 [v.s. Baidu] 包括三方结果 ResearchGate、EconPapers…
- 擅用 **Advanced Search**

ETASeminar Minimalistic Search Engine

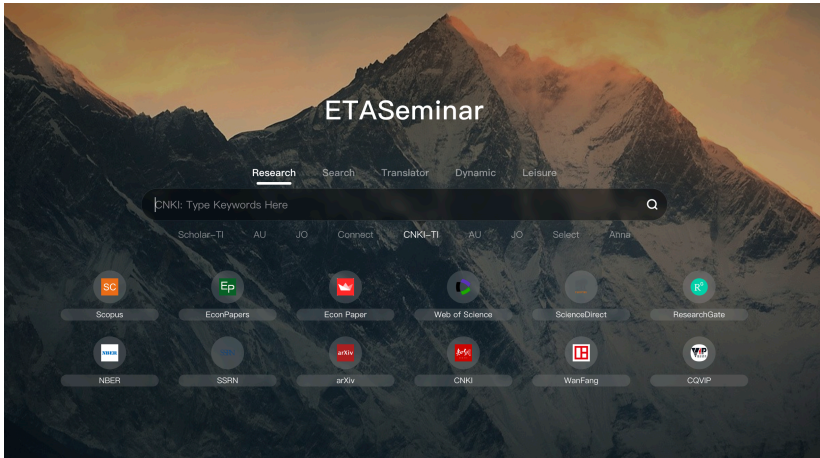


图: 1 ETASeminar Mini Search: Click <https://vividzheng.eu.org/mini/>

ETASeminar Search Engine

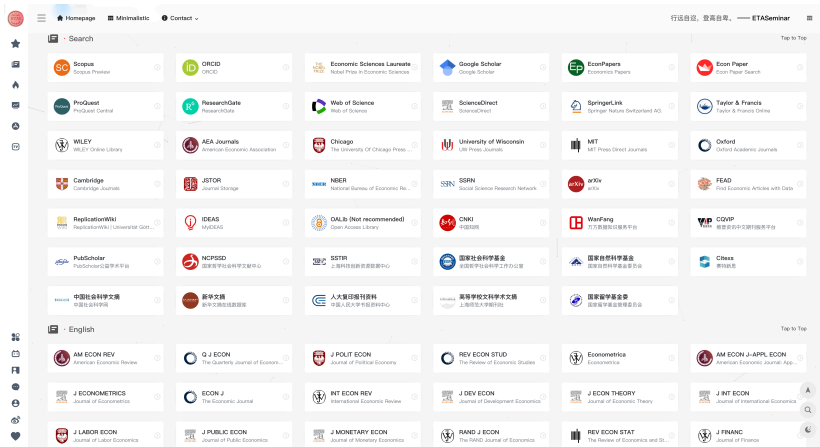


图 2 ETASeminar Search: Click <https://vividzheng.eu.org/>

文献综述的写作方式

擅用并列、递进、转折逻辑连接词

- 第一，第二，…，最后；首先，其次，再次，紧接着，最后
- First, Second, Third, …, Last
- 一方面，另一方面；On the one hand, on the other hand
- 虽然…，但是；On the contrary

多用间接引用，少用直接引用：尽量避免完全一致的描述

- 少用文献罗列式：Lee (2022) 发现；Zhao (2023) 认为…
- 多用观点式（间接）引用：信用债券大量违约将加大信用利差震幅，并通过投资、产出、消费等渠道快速传导至宏观经济，放大经济波动（Gomes & Schmid, 2021；Xing et al., 2023）

既要分门别类汇总已有文献，又要对引用的文献评述优缺点

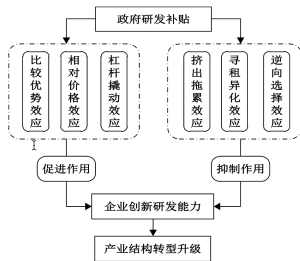
- 在 A1 方面，Lee (2022) 发现；Zhao (2023) 进一步认为；He (2024) 的观点与之相反
- 尽管已有部分文献探讨了 A1，然而，当前对 A2 的研究却十分有限，而 A2 在…背景下非常重要
- This paper fills the gap by incorporating …

理论假说 (Mechanism)

通常一篇论文**集中回答一个问题**

- **太多假说？** Hypothesis 1a; Hypothesis 1b; ...; Hypothesis N **NO!**
- 管理、金融、会计领域期刊一般需要提炼一个以上 Testable Hypothesis
- 经济学期刊通常不需要提炼 Hypotheses，直接说明要研究的问题 [近年来较大变化]
- 查阅拟投期刊过往论文风格，参考是否需要提炼 Hypothesis

图示理论框架：一图胜千言！ YYDS



图：3. 政府研发补贴对产业结构转型升级的作用机制图

研究方法选择

问卷调查法

- 问卷设计 [信度、效度]、确定调查对象、发放问卷、回收问卷、数据分析…

实验设计

- 随机对照试验 (Randomized Controlled Trial, RCT) 思想
- 两 (多) 阶段实验内容设计

简约式估计

- 线性回归、非线性回归、断点设计、受限因变量
- 标准化回归系数计算和解释
- Tips: **Probit 不能 FE!** (伴生参数问题)

博弈模型 & Simulation

- 博弈主体、策略、信息公开情况…

结构式估计

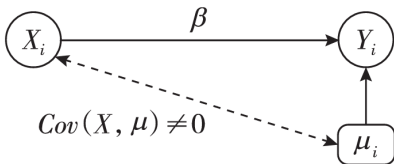
- 效用函数、生产函数、劳动力出清、资本出清、产品出清、一般均衡

…

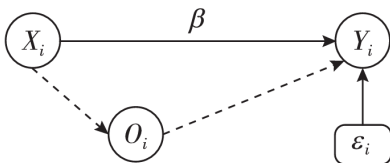
缓解内生性担忧

- 《追踪研究中的内生性问题：来源与应对》
- 《管理学研究中的内生性问题及修正方法》
- 《管理研究中的内生性》
- 回顾内生性的起源、一般化定义及基本原理，分门别类地梳理和总结管理研究中常见的内生性问题主要来源，并结合管理学经典研究案例，针对每种来源分别给出有效的解决方案
- 总结了**遗漏变量** (Omitted Variable)、**双向因果** (Simultaneity)、**测量误差** (Measurement Error)、**选择偏差** (Selection Bias) 等常见内生性来源，针对每种内生性问题归纳了相应解决方案
- Oster (2019) **系数稳健性检验思想**

内生性问题示意图



图：4 内生性问题



图：5 内生性问题一：遗漏变量

内生性问题示意图

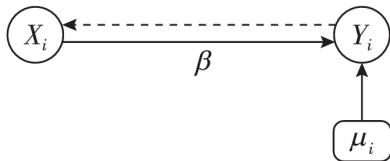


图: 6 内生性问题二：双向因果

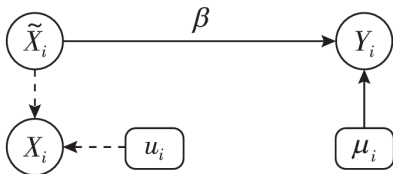
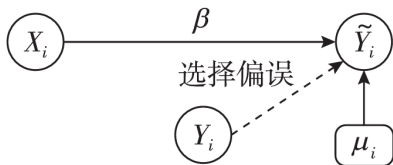
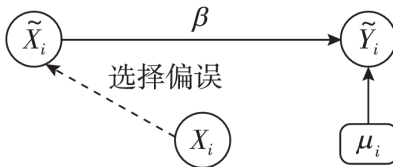


图: 7 内生性问题三：测量误差

内生性问题示意图



图：8 内生性问题四：样本选择偏误



图：9 内生性问题五：自选择偏误

机制检验

调节效应 (Moderation Effect)

- **回归中 M 、 X 以及 $X \times M$ 需要同时控制!**

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 (X \times M) + \epsilon$$

分组回归 (Group Regression)

对于 $G = 0$:

$$Y = \alpha_{00} + \alpha_{10} X + \epsilon_0$$

对于 $G = 1$:

$$Y = \alpha_{01} + \alpha_{11} X + \epsilon_1$$

模型异同

- 都能进行分组别的异质性效应检验
- 调节效应假定两组数据除**分组变量对 Y 的影响可能不同**外，其余所有变量对 Y 的影响不存在组间差异
- 分组回归放松了调节效应的假定，允许两个组别**所有变量对 Y 的影响均存在组间差异**
- 分组回归理论上需要做“**组间系数差异检验**” [Hausman、suest、Chow]

思考: **哪个模型更 General?**

调节效应与分组回归结果

	Base	Base	Base	M=Comp	M=Comp	M=Restr	M=Restr
Cfin	-0.001** (-1.96)			-0.002*** (-3.03)		0.007*** (4.95)	
Gfin		-0.004** (-1.96)			-0.010*** (-3.13)		-0.018*** (-2.93)
Sfin			-1.17e-04 (-0.50)				
Cfin*M				-0.019*** (-2.65)		-0.002*** (-5.95)	
Gfin*M					0.011** (2.16)		0.007*** (4.11)
M				0.009*** (7.57)	0.008*** (12.80)	0.010*** (13.91)	0.010*** (10.73)
Constant	0.010* (1.91)	0.011** (2.10)	0.012** (2.11)	0.004 (0.71)	0.005 (0.89)	0.196*** (12.19)	0.208*** (12.96)
控制变量	控制	控制	控制	控制	控制	控制	控制
Observations	28,822	28,822	28,822	27,852	27,852	27,852	27,852
R-squared	0.377	0.377	0.377	0.379	0.379	0.381	0.380

图: 10 调节效应回归结果

调节效应与分组回归结果

表 8 宗族文化与个体“传宗接代”观念

	(1)	(2)	(3)	(4)	(5)
	全样本	男性	女性	城市	农村
宗族文化	0.0959 (0.0861)	0.2085* (0.1192)	-0.0687 (0.1278)	-0.0496 (0.1977)	0.1354 (0.0951)
个体特征	是	是	是	是	是
家庭特征	是	是	是	是	是
村/居及省份特征	是	是	是	是	是
地区固定效应	是	是	是	是	是
观测值	644	365	279	252	392
R^2	0.1156	0.0810	0.2360	0.1673	0.1356

图: 11 分组回归结果

调节效应与分组回归图示

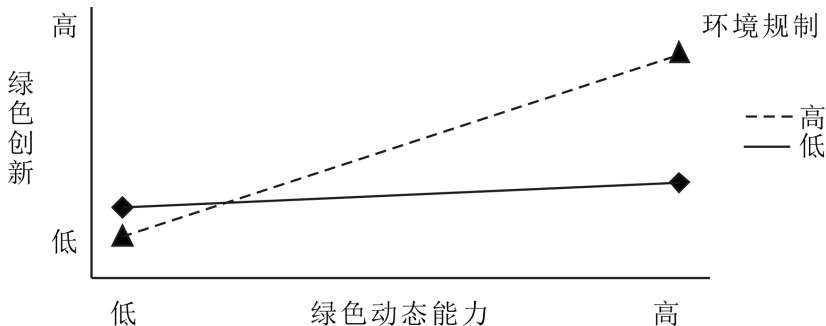


图: 12 调节效应图示

调节效应与分组回归图示

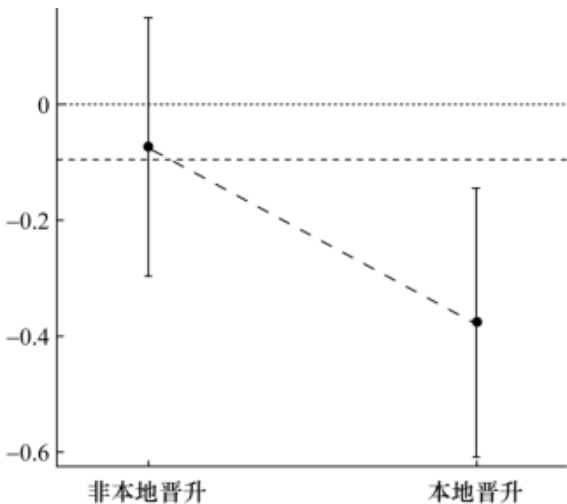


图: 13 分组回归图示

机制检验

慎用中介效应！

从自变量到中介变量的路径 (a 路径)

$$M = \alpha + aX + e_1$$

从自变量直接到因变量的路径 (c' 路径, 直接效应)

$$Y = \beta + c'X + e_2$$

从自变量和中介变量到因变量的路径 (b 路径和 c' 路径, 总效应为 $c = c' + ab$)

$$Y = \gamma + c'X + bM + e_3$$

问题出在哪里？

- 内生性问题！——**遗漏变量**

中介效应回归图示

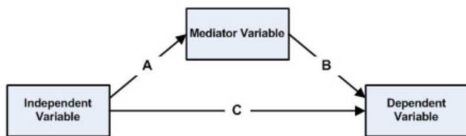


图: 14 中介效应图示

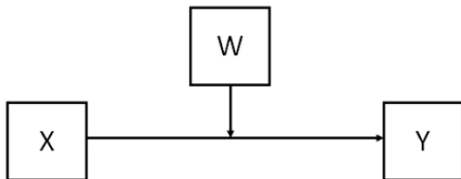


图: 15 中介效应图示

Roll up! Involution (内卷的一代)

以速度取胜

- 首发、抢发现象层出不穷
- 冰山一角之《对学术界失望至极，本来不想说出来的，但是觉得还是要说》

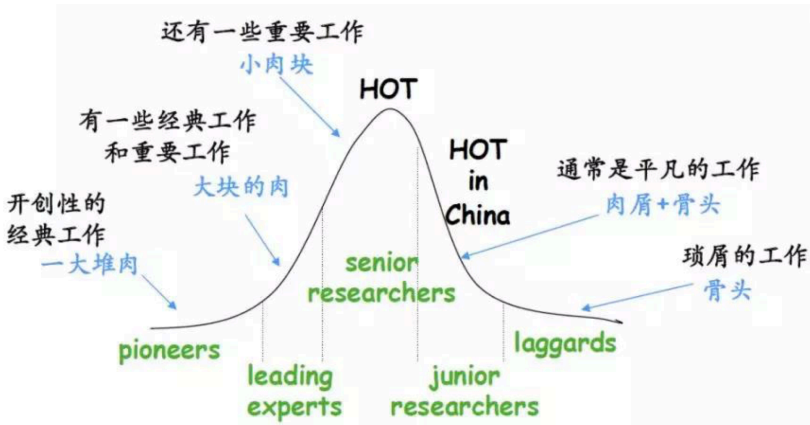


图: 16 Involution

好论文是“改”出来的

论文是“做”出来的，不是“写”出来的

- “这个月我要写篇英文论文” No
- 需要先有一个好的研究，然后才考虑写论文
- “写”的时间其实最多只占 10%

好论文是“改”出来的

- 我的第一篇论文经历了近两年时间、13 次修改后被接收

研究过程: **Phenomenon** -> Topic -> Problem -> **Idea** -> Concrete -> **Writing**
-> Submit -> (Desktop Rejection) -> Revision -> (Rejection) -> Accept -> Publish

最好的状态: 2 篇在外审、2 篇在写作、2 篇在构思

但行好事 (前路), 莫问前程; 心之所向, 无问西东!

Thanks

Questions?