

# 计量经济学的方法论回顾

---

赵博

南开大学金融学院

2025 年 6 月 4 日

- 本学期所学计量经济学回顾
- 还有什么计量经济学可以学
  - 如何继续学习计量经济学
- 如何看待（利用）计量经济学

# 本学期所学计量经济学回顾

---

# 最小二乘法 Ordinary Least Squares vs. 线性回归 Linear Regression

- 最小二乘法 (Ordinary Least Squares, OLS) 是一种**算法**
- 线性回归 (Linear Regression, LR) 是一种**模型** (注意: 所有的那些假设, SLR, MLR 1-6, 等等, 都是对 LR 的假设, 和 OLS 无关)
- 区别:
  - **模型**描述的是对现实的理解 (由其他知识获得, 比如哲学观、经济理论、等等)
  - **算法**是估计这个模型的方法: OLS 利用的是最小化残差平方和
- 人们常常说的词 “做回归”, 实际上说的就是 “用最小二乘法估计了线性回归式子”
  - 是一种方便的说法, 可以理解
  - 但严格来说, 混淆了这二者
  - 请注意这个区别, 虽然以后也许不会有人强调这一点

- 技术上的区别：

- LR:  $E(\mathbf{x}u) = \mathbf{0}$ , 是假设, 不可从统计上验证的假设, 来源于其他知识的假设 (同理: IV 时,  $E(\mathbf{z}u) = \mathbf{0}$  也是不可从统计上验证的假设; 但注意什么是 overidentification test)
- OLS: 自动给出  $\frac{1}{N} \sum_{i=1}^N \mathbf{x}_i u_i = \mathbf{0}$
- 而大数定律说:  $\frac{1}{N} \sum_{i=1}^N \mathbf{x}_i u_i \xrightarrow{P} E(\mathbf{x}u)$
- 换句话说: 如果用了 OLS 来估计 LR, 那么隐含的意思就是, 你认为模型不存在导致  $E(\mathbf{x}u) \neq \mathbf{0}$  的问题 (遗漏变量 omitted variables、测量误差 measurement error、反向因果 simultaneity problem (考试时因为翻译差异的原因, 保险起见, 可以把英文写上))
  - 通常没有人会这么绝对, 因此, 稳健性检验才是重头戏

## 其他内容概括

- 数学
  - 矩阵代数是基础（矩阵运算更容易，考试时写矩阵代数给分）
  - 最优化（求 MSE 的最小值）
- 统计
  - 无偏（期望等于真值）、有效（方差计算、比较大小）、一致（大数定律）、中心极限定理
  - t 检验，F 检验，渐进正态分布
  - $R^2$

- 线性、对数、量纲
- 虚拟变量
- 异方差（自相关为补充内容）
- 工具变量法
  - 遗漏变量、测量误差、反向因果
  - 两个条件：无关性、相关性
  - 弱工具变量问题
  - 弱工具变量检验（不可识别检验）、工具变量外生性检验（overidentification test），内生变量检验（Hausman test）。
  - 实践中，弱工具变量检验要报告、外生性检验要报告。不可识别检验可以不报告、内生变量检验可以不报告
  - 注意：从根本上来说，底层的那个核心 (IV 外生) 无法检验。因此，外生性检验、Hausman 检验，都只是一种外围的、希望让人更放心的佐证而已，触不到核心内容

还有什么计量经济学可以学

---



- 广泛地分类：描述性 vs. 因果
  - 描述性：统计建模思路，用一个模型来描述现象，不一定是因果
  - 因果：强调  $x$  导致了  $y$ ，而不是反向，或者有其他干扰
- 时间序列（课程名称通常为“金融计量经济学”）例如：
$$y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t$$
  - 对现实现象的描述（不强调因果）
  - 统计学其实就是这样的思路
- 既有横截面数据、又有时间数据：面板
  - 短面板（最常见）： $N$  大， $t$  小。仍然是横截面性质为主。利用时间可以更好地观察因果（强调因果）
  - 长面板： $N$  小， $t$  大。时间序列性质为主
  - $N$  大、 $t$  也大：取决于研究的问题，看是否用“因果”讨论
- “因果”实际上是统计、或者说计量中的一套独特的语言
  - 不是唯一的语言
  - 因此，不要说计量经济学研究因果，机器学习不研究因果
  - 但强调“因果”的确是计量经济学独特于统计学的地方

# 大致分类

- 微观计量经济学：因果；宏观计量经济学：现象。但两者会有交叉的地方
  - 如果想讨论的问题类似于：企业发行绿色债券（微观的）会影响业绩（微观的）吗？用因果模型
  - 如果想讨论的问题类似于：货币供应（宏观的）如何影响就业率（宏观的）或者相互影响？用时间序列模型
- 结构方程模型：基于理论进行建模
- 以上是基于建模方式的分类。
- 另外的分类方式：
  - 统计理论的不同：频率 (frequentist) vs. 贝叶斯 (Bayesian)
    - 频率：存在客观不变的数；通过抽样估计。贝叶斯：概率是主观的；数据可以更新这个主观的判断
  - 方法论的不同：估计 (estimation) vs. 校准 (calibration)

- 起源：联立方程模型的失败  $\Rightarrow$  反思（Lucas 批判、Sims 的 Vector Autoregression, VAR）
  - 时间序列很长、横截面不多，自然的选择就是 VAR，或者长面板
  - 注意：如果用 VAR 做时间序列，注意采用的语言。“因果”此时要谨慎使用。如果说因果，考虑用 Structural VAR
- Kydland & Prescott (1982) 实际经济周期模型（Real Business Cycle, RBC）：
  1. 建立家庭效用模型  $\Rightarrow$  生产函数限定了预算约束  $\Rightarrow$  最大化效用
  2. 确定底层的一些参数：偏好、技术、技术冲击等：利用微观研究确定偏好；利用宏观的一些稳态变量匹配（例如消费收入比、资本收入比）；自由参数，人为选定
  3. 用模型来产生模拟的时间序列，弄出一大堆样本
  4. 计算出来的数据，与真实数据对比（这种对比没有统计理论来支撑到底是否“足够近”）
- Dynamic Stochastic General Equilibrium (DSGE) 基于此发展出来

# 校准 vs. 估计

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：Kydlan & 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：Kydlan & Prescott(1982)。

- 计量：模型可以变，根据哲学观、经济理论等，改变模型，以便更符合数据
- 校准：模型就这样，对模型有十足的信心。数据是否贴合模型的结果，不 care。模型当然可以还不够好，但那是理论层面的讨论
- 校准的方法论实际上和 Austrian 的底层方法论很像，只是 Austrian 不用数学，也不用统计学

# 如何继续学习计量经济学

- 做计量研究：数学多学点
- 不做计量研究：没必要紧跟计量发展。根据具体的问题，选择合适的计量模型。搞清楚模型的大致含义，有程序直接用程序；没程序看一下关键的几个估计量的矩阵代数，编程套用

# 如何看待（利用）计量经济学

---

# 计量经济学的本质

- 计量经济学方法论本质上仍然是“非预测”的——这由社会科学性质决定
  - 并不是不预测，所有的理论都隐含着“预测”的意思。只是社科类预测的归因极难说清楚
- 计量经济学是“归纳法”，而不是“演绎法”
  - 看一亿只白天鹅，也不能总结出“天鹅一定是白的”
  - 从统计上看，有可能推出“公鸡叫醒了太阳”
    - 公鸡通常在天没亮就叫，时间上在太阳升起之前（时间先后）
    - 全世界的公鸡大概都这样（样本很大）
    - 偶尔会有不叫的时候（有误差）
  - 只有明白了太阳和地球的关系，才能理解太阳升起这件事的本质。这跳出了搜集“公鸡——太阳”数据进行归纳的方法论范畴
- 真正的“因果关系”，只能是从理论（演绎法）出发得到。计量经济学是一种“佐证”，而不是“证明”
  - 因此，所谓的计量经济学中的“因果关系”，不是真正意义上的因果关系（很多人反对这个论点，有一些人支持）

- 这意味着：客观评价标准相对缺失（并不是完全缺失）
  - 比如，做计量模型开发：

描述已有模型不足 → 提出统计量 → 推出大数定律、中心极限定理 → 做数据模拟 → 实际数据测试

- 做应用研究：

描述问题 → 用什么计量模型合理 → 这种模型可能存在的问题（稳健性检验） → 作用机制的探讨 → 模型的意义

- 这些是标准。但这些做法本身的意义不像自然科学那样不言自明
- 因此，评论者们（学术圈：编辑、审稿人；业界：买方）的态度很大意义上决定了产出的最终去向



- 但，请一定要做稳健性检验、机制探讨。虽然不能“证明”，但多少让人宽慰
- 这也是为什么经济学、金融学论文的核心论点通常很简单，但篇幅那么长的原因（自然科学论文最多 10 几页）
- 不要随便找了个  $x$  和  $y$ ，弄出来几颗星星，开始大谈特谈“政策建议”

- 计量经济学的核心：很多观测值
- 现实可能只有几个观测值，但人类也可通过其他方式理解道理。
  - 历史只经历了一次。即便有些类似的历史经历了多次，但显然未必能满足大数定律的条件。但人类通过对于人性的洞察，仍能规避一些错误
- 不要把这二者混杂：
  - 计量上的不显著，不妨碍你获得对于问题本质的真知
  - 但：有多少证据，说多少话

欢迎入坑，少走弯路