

高级计量经济学

Lecture 1: Introduction

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室 粤海校区汇文楼1510
E-mail huangjp@szu.edu.cn
Website <https://huangjp.com>

课程网站

- 直达链接：

<https://huangjp.com/teaching/econometrics.html>

- 首页链接及路径（个人网站）：

<https://huangjp.com/>

首页 > Teaching > Advanced Econometrics (graduate)

第四届全国国际商务专业学位硕士（2018年）优秀学位论文评选

全国国际商务专业学位研究生教育指导委员会 http://chinamib.org/html/2019/tongzhigonggao_1029/130.html

获奖等级	论文题目	是否使用计量方法
一等奖	研发投入、跨组织网络嵌入与创新绩效研究：以中国汽车企业为例	是
一等奖	“丝绸之路经济带”沿线国家商品贸易网络研究	是
二等奖	创新产品视角下百威啤酒品牌个性塑造研究	否
二等奖	“一带一路”倡议对我国在沿线家跨国并购的影响——基于双重差分的实证研究	是
二等奖	均胜电子连续海外并购绩效研究	否
三等奖	支付宝第三方跨境支付在东南亚的发展策略研究	否
三等奖	“一带一路”倡议下中国FTA合作伙伴选择与经济效应研究	否
...
合计：一等奖2名，二等奖3名，三等奖5名，提名奖6名		7/16

- 论文中使用的其他定量研究方法：最优化 (1)、主成分分析 (2)、构筑指数 (2)、问卷调查 (4)
- MIB 学位论文的形式：理论与政策研究论文、案例分析、市场调研报告、商业计划书、项目可行性报告等 http://www.chinamib.org/html/2016/jibenyaqiu_1202/75.html

课程简介

课程目标

- 课程目标：在本科计量经济学课程的基础上，加深对**核心理论**的理解
 - 必要的背景知识储备：
微积分、线性代数、概率与统计
尤其是**矩阵运算**和**条件概率**是必须了解的
 - 需要了解的初级计量经济学内容：
线性回归模型、最小二乘估计、假设检验、非线性回归函数、
虚拟变量、内生性等内容
- 需要自主学习的内容：
 - 数据的处理、计量经济软件的使用（未接触过计量的同学可以用 gretl，有基础的同学建议使用 R，可参考 <https://huangjp.com/teaching/> 上提供的信息)
 - 文献阅读

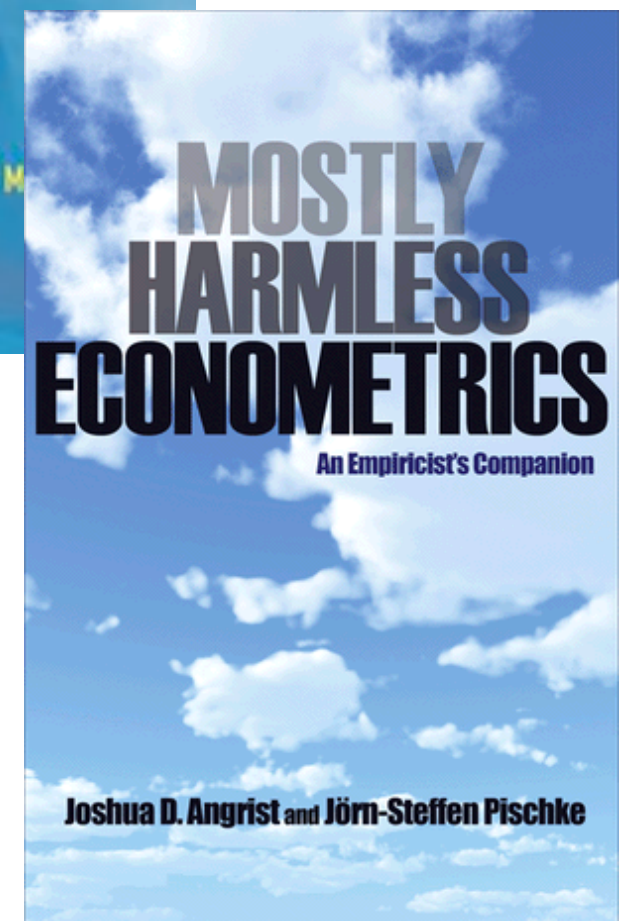
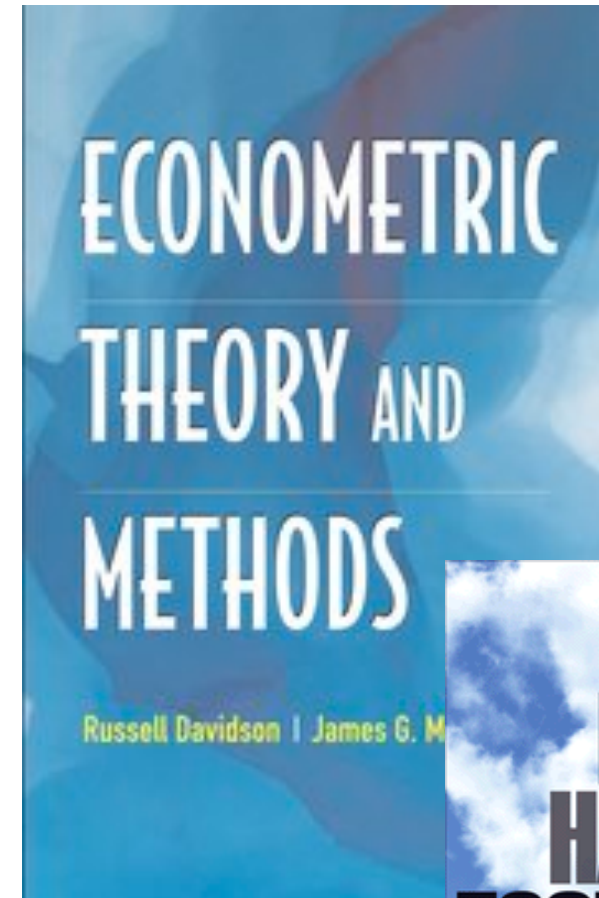
课程内容

- 第一部分：线性回归模型和 OLS 估计量
 - 线性模型的矩阵表达、OLS 估计量的几何特征、Frisch-Waugh-Lovell 定理、Gauss-Markov 定理、假设检验、等
- 第二部分：其他模型和方法
 - 非线性回归、工具变量估计、面板数据模型、最大似然估计*、核密度估计*、等
- 第三部分：因果推断
 - Neyman-Rubin 因果模型、匹配、双重差分法、合成控制法、断点回归设计、工具变量法和 LATE、等

课程教材

主要参考以下书籍

- Davidson & MacKinnon (2021), *Econometric Theory and Methods*, Oxford University Press.
可以从作者网站下载
<http://qed.econ.queensu.ca/ETM/>
- Angrist & Pischke (2008), *Mostly Harmless Econometrics*, Princeton University Press.



成绩评价标准

- 出席率和课堂参与度 20%
- 作业（或小组汇报） 30%
 - 计划布置一至三次作业
- 期末考试 50%
 - 拟采用随堂开卷考试形式

对学习方法的建议

- 课堂 1 小时、课后 x 小时
- 建议做笔记
- 建议参考不同的参考书，并通过文献学习理论的进展
- 读优质论文，并尝试提出问题
 - 如何培养对文献的鉴别能力？批判性思维 + 阅读量 + 有效的讨论
- 不依赖二手信息
 - 公众号的推送 \neq 论文原文的内容

部分优秀英文专业期刊

- **American Economic Review**
- **Econometrica**
- **Quarterly Journal of Economics**
- **Journal of Political Economy**
- **Review of Economic Studies**
- Review of Economics and Statistics
- Economic Journal
- American Economic Journal: Applied Economics
- RAND Journal of Economics
- Journal of Applied Econometrics
- Journal of Business and Economic Statistics
- International Economic Review Journal of Econometrics
- Scandinavian Journal of Economics
- Journal of Labor Economics
- Labor Economics
- Journal of Public Economics
- Journal of Economic Growth
- Journal of Health Economics
- Health Economics
- Journal of the European Economic Association
- European Economic Review
- Oxford Bulletin of Economics and Statistics

初级计量经济学内容复习

计量经济学 = 计量方法 × 经济学问题

- 计量方法：运用**数学模型**和**数据**探索经济变量间的关系
 - 数学模型：对经济现象的抽象总结
 - 数据：随机变量的样本，因此需要探究隐藏在数据面纱下的真实关系
- 计量经济学的两个侧面：解释、预测
 - 试图**解释**变量间的因果关系（causal inference）：自变量的变化如何引起（cause）因变量的变化
 - 在给定自变量的取值时**预测**因变量的值

从经济学模型到计量经济模型

以Gary Becker的犯罪模型为例

- 犯罪的经济学模型

- 犯罪行为作为个人选择的结果，其中包含理性决策和非理性行为。传统经济学的核心是对理性选择的研究，因此可以被用来分析犯罪行为。
- 诺奖得主 Gary Becker 将针对犯罪行为对社会带来的损失进行度量，并探讨了使损失最小化应该付出的社会成本，将经济学方法运用到犯罪的研究中。
- 我们可以把犯罪活动的强度（次数、时间等）总结为包含一系列变量的函数

$$y = f(x_1, x_2, \dots, x_n)$$

此式中 在经济学中，我们通常不去假设 f 的具体形式，而是假设它的某些特征，如增减、凹凸等

y = 犯罪活动的强度

x_1 = 犯罪活动带来的回报（如非法所得的钱财、内心需求的满足等）

x_2 = 犯罪活动的机会成本（如同样时间内合法工作的回报等）

...

其他变量包括逮捕率、量刑等犯罪成本，教育程度、收入等个人特征

从经济学模型到计量经济模型

以Gary Becker的犯罪模型为例

- 犯罪的计量经济模型

$$\text{crime} = \beta_0 + \beta_1 \text{wage}_m + \beta_2 \text{othinc} + \beta_3 \text{freqarr} + \beta_4 \text{freqconv} \\ + \beta_5 \text{avgsen} + \beta_6 \text{age} + u$$

此式中

计量经济学需要假设具体的函数形式，这里以线性函数为例

crime = 参与犯罪活动的频率

wage_m = 从事合法工作带来的报酬

othinc = 其他收入（投资、财产继承等）

freqarr = 被逮捕的频率（作为被逮捕概率的近似）

freqconv = 被判刑的频率

avgsen = 平均服刑年数

回归分析

- 核心问题：从样本数据推测总体中变量间的关系
 - 在线性回归模型中，我们通过推测因变量的系数探讨变量间的关系
- 参数模型的拟合：根据样本观测值计算最优的参数组合
 - 什么是最优？使目标函数最小，例如最小二乘法就是使残差平方和最小
- 参数的估计：根据样本观测值猜测参数的真实值（不可观测）
 - Gauss-Markov 定理指出：线性回归模型的最小二乘估计量是 BLUE 的
 - 估计的准确性（是否非偏）和精确性（误差的大小）往往不可兼得
 - 根据研究目的确定估计量的选择标准：目的为解释时更注重非偏性、目的为预测时更注重精确性。

最小二乘估计量仅在线性非偏估计量中是最优的

线性回归模型

Linear regression model

我们假设观测数据服从回归模型，因此此式也被称作
data generating process

- 多变量线性回归模型

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$$

若假设 $E[u_i | x_{i1}, \dots, x_{ik}] = 0$ ，则

$$E[y_i | x_{i1}, \dots, x_{ik}] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

- 线性指的是上述回归函数关于系数 $\beta_0, \beta_1, \dots, \beta_k$ 是线性的。
- 回归分析的主要目的：
 - 正确选择回归函数（基于经济学理论或对问题的理解）
 - 正确估计系数 $\beta_0, \beta_1, \dots, \beta_k$ 的值（基于样本数据）
 - 对总体中系数的值进行推断（基于模型假设和统计学理论）

最小二乘估计

Ordinary Least Square Estimation

- 假设单变量回归模型 $y_i = \beta_0 + \beta_1 x_i + u_i$
- 定义残差平方和

$$SSR = \sum_i (y_i - b_0 - b_1 x_i)^2$$

- 最小二乘估计量是使 SSR 取值最小的 (b_0, b_1) 的值
- 解法：求一阶条件，然后解联立方程

$$b_1 = \sum_i (x_i - \bar{x})(y_i - \bar{y}) / \sum_i (x_i - \bar{x})^2$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Gauss-Markov 定理

- 最小二乘估计量 (OLS) 是众多估计量中的一种，因为它满足很多好的性质，因此是最常用的估计方法。
- 估计量是观测值的函数，因此也是随机变量。若 $\hat{\beta}$ 是 β 的估计量，而 β 的真实值是 β_0 ，我们说
 - 当 $E[\hat{\beta}] = \beta_0$ 时， $\hat{\beta}$ 是非偏的 (unbiased)。
- OLS估计量是非偏的。
- OLS估计量是线性的 (它是 y_i 的线性函数)。
- 在适当的假设下，OLS估计量在所有线性非偏估计量中是方差最小的，因此被称作 the best linear unbiased estimator (BLUE)。

内生性

Endogeneity

- 内生性和外生性（exogeneity）源自联立方程模型。
 - 当一个变量可以被系统内其他变量解释时，该变量被称作内生变量，一般出现在方程左侧。当该变量被用来解释其他变量时，也可以出现在方程右侧。
 - 当一个变量无法被系统内其他变量解释时，该变量被称作外生变量，一般只出现在方程右侧。
- 若变量是内生的，则有可能和误差项相关，因此现在对内生性的定义，主要指解释变量和误差项之间存在相关性。
- OLS的主要假设是**不存在内生性**，表达为 $E[u | x] = 0$ 。

异方差、自相关、误差项的分布

- 异方差性 (heteroskedasticity) 是指不同观测值的误差项的方差可能不同。OLS一般假设同方差，即

$$\text{Var}[u_i | x] = \sigma^2$$

- 自相关 (autocorrelation) 是指不同观测值的误差项之间存在相关。OLS一般假设不存在自相关，即

$$\text{Cov}[u_i, u_j | x] = 0$$

- 有时我们也假设误差项服从多变量正态分布。

理论在实证研究中的重要性

估计边际生产率

- 让我们考虑一下如何估计生产要素的边际生产率的问题
- 以劳动 (labor) 为例, 令总产出为 Y , 劳动投入为 L , 则劳动的边际生产率定义为 $\frac{\partial Y}{\partial L}$ 。

- 如果假设 Cobb-Douglas 生产函数 $Y = AL^\alpha K^\beta$, 则

$$\frac{\partial Y}{\partial L} = \alpha AL^{\alpha-1} K^\beta = \alpha \frac{Y}{L}$$

这里假设了经济学模型

- 因此, 在实证研究中通常通过估计 α 和 β 来估计边际生产率, 即考虑下面的回归模型

$$\log Y_{it} = \log A + \alpha \log L_{it} + \beta \log K_{it} + u_{it}$$

这里将经济学模型转化为计量模型

此处, i 可以代表地区、企业等生产单位。

理论在实证研究中的重要性

Wang, H. & Lall, S. (2002). Valuing water for Chinese industries: a marginal productivity analysis. *Applied Economics*, 34:759-765.

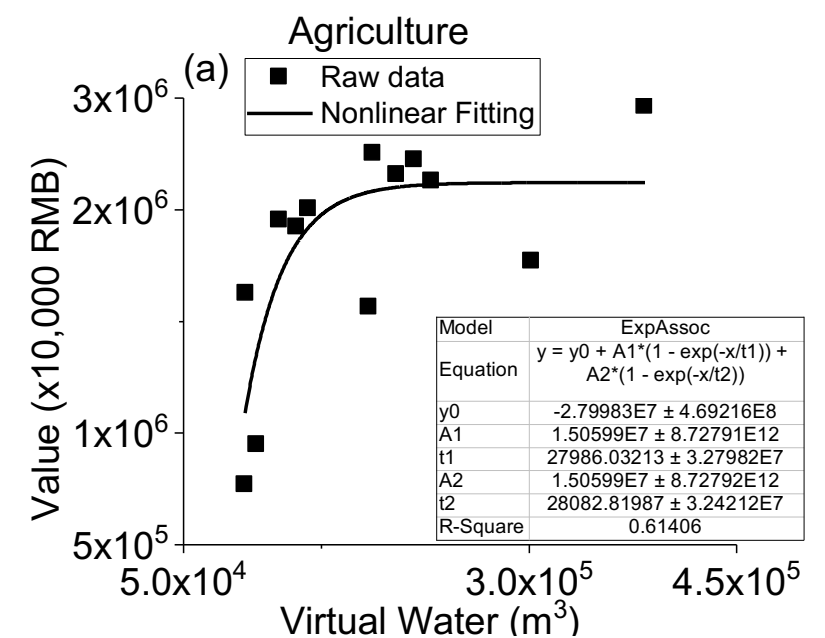
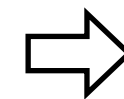
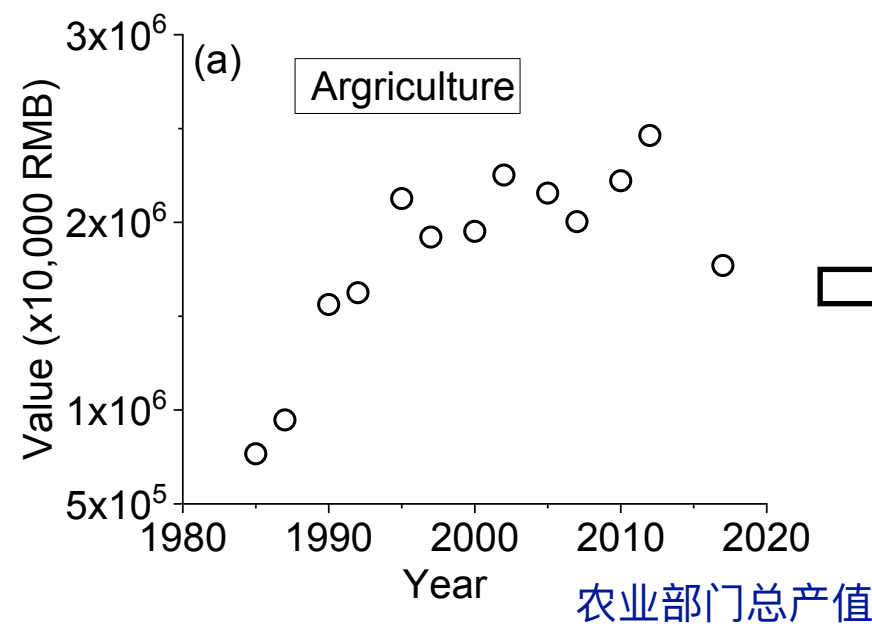
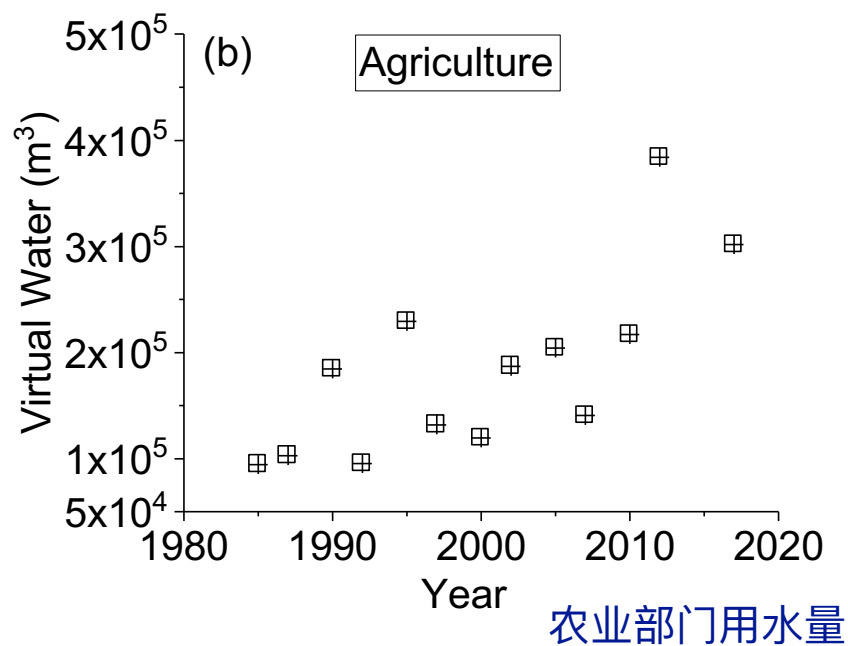
- 在实证中估计生产函数并不容易，最大的障碍往往是无法获得有效数据！
- Wang & Lall (2002) 用企业级数据估计了中国各产业部门的水资源边际生产率。
 - 模型：translog production function
$$\log Y_i = \beta_0 + \beta_1 \log K_i + \beta_2 \log L_i + \beta_3 \log W_i + \beta_4 \log E_i + \beta_5 \log M_i + \text{二次项}_i + u_i$$

水 能源 原材料
 - 数据：作者在世界银行工作，通过和中国环境保护部（现生态环境部前身）的合作研究项目，获得了1993年国有大中型企业中约2000家工厂的环境数据，变量包括 Y, K, L, W 以及其他特征变量
 - 文章内容比较简单，但是当你想要参考它做类似研究时，你会发现很难获得数据

理论在实证研究中的重要性

一个失败的硕士论文

- 在关于水与能源的边际生产率及相关研究中，企业级数据极难获得。公开可获得的数据多为地区级投入产出表（基于普查和估算，时间间隔2、3或5年不等）。
- 文章作者为了估计北京地区不同生产部门的水和能源的边际生产率，利用1985-2017年间13个年份的投入产出表计算了水资源投入量、能源投入量、以及总产值。以水为例，作者对用水量 and 总产值间的关系进行了非线性拟合，继而通过对拟合出的函数进行求导，得出“**边际生产率**”的函数：



- 那么这位作者错在哪儿呢？

课后学习

- 阅读 D&M (2021) Chapter 1，特别注意关于矩阵和条件概率的内容。
- 复习本科阶段线性代数、概率与统计、微积分的内容，特别是
 - 矩阵演算、线性方程
 - 概率分布、假设检验
 - 多变量函数的最优化
- 可参考 Online Appendices of Greene:
<https://pages.stern.nyu.edu/~wgreene/Text/Greene-EA-7&8ed-Appendices.pdf>