中央财经大学
Central University of Finance and Economics

计量经济学

中国财政发展协同创新中心
陈怡心
cyx@cufe.edu.cn

课程目标

1.了解现代计量经济学的特征、经济计量分析在经济学科的发展和实际经济工作中的作用。

2.掌握基本的经典计量经济学理论与方法，并对计量经济学理论与方法的扩展和新发展有概念性了解。

3.能够建立并应用简单的计量经济学模型，对现实经济现象中的数量关系进行实际分析，尤其应用于财政学领域的研究问题。

4.具有进一步学习与应用计量经济学理论、方法与模型的基础和能力。

5.初步掌握Stata应用，能够进行数据处理和分析，初步完成实证研究。

考核方式

期末总成绩=考试成绩（70%）+平时成绩（30%）

期末考试：试卷统一命题、统一考试

平时成绩：课程作业+随机签到+课堂表现

参考资料

- 陈强，《计量经济学及Stata应用（第二版）》，高等教育出版社，2023
- 陈强，《计量经济学及Stata应用》，高等教育出版社，2015
- 杰弗里·M·伍德里奇，《计量经济学导论：现代观点》，中国人民大学出版社，2018
- 陈强，《高级计量经济学及Stata应用》，高教出版社，2014
- 赵西亮，《基本有用的计量经济学》，北京大学出版社，2017

STATA应用

- 自行安装stata
- 网上资源：经管之家（原人大经济论坛）、知乎、微信推文（公众号：连享会、计量经济学及Stata应用、Stata and Python数据分析....等等）...等

Stata简介

- Stata（读作“stay-ta”，通常也可以写作 STATA）是 StataCorp 开发的一款通用统计软件包，用于数据处理、可视化、统计和自动报告。包括生物医学、经济学、流行病学和社会学等许多领域在内的研究人员都在使用它。
- Stata 最初由加利福尼亚州的计算资源中心（Computing Resource Center）开发，并于 1985 年发布了第一个版本 。1993 年，该公司迁至德克萨斯州，并更名为 Stata Corporation，即现在的 StataCorp。该司 2003 年发布了一个包括新的图形系统和所有命令的对话框的重要版本 。此后，每两年发布一次新版本 。当前版本为Stata 18，并于 2023 年 4 月发布 。

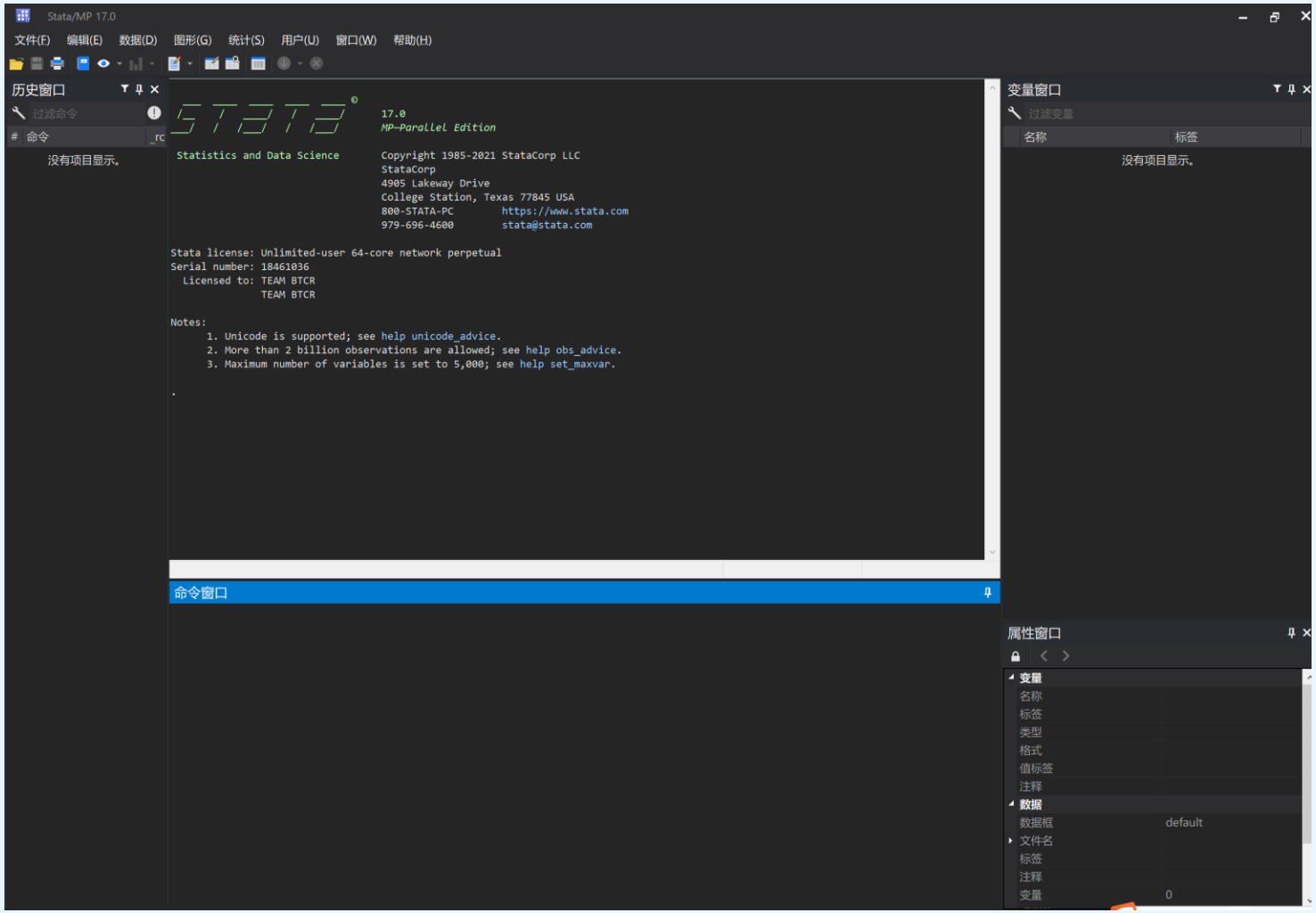
Stata简介

- Stata 有四个版本：**Stata/MP**、**Stata/SE**、Stata/BE 和 Numerics by Stata。Stata/MP 允许对某些命令进行内置并行处理，而 Stata/SE 和 Stata/BE 则存在瓶颈，只能使用一个内核。与 SE 或 BE 版本相比，Stata/MP 在四个 CPU内核上运行并行处理时，某些命令的运行速度要快 2.4 倍，约为理论最高效率的 60%。
- SE 和 BE 版本在数据集可使用的内存量上有所不同。Stata/MP 可以存储 100 到 200 亿个观测值和多达 12 万个变量，而 Stata/SE 和 Stata/BE 则分别可以存储 21.4 亿个观测值，处理 32767 个变量和 2048 个变量。在 Stata/MP中，一个模型中自变量的最大数量为 65,532 个，在 Stata/SE 中为 10,998 个，在 Stata/BE 中为 798 个。

Stata简介

StataMP 17 (64-bit)

中央财经大学
Central University of Finance and Economics



第1章 计量经济学基础

陈怡心
中央财经大学中国财政发展协同创新中心

什么是计量经济学

- 缩小班级规模对学生成绩有什么影响?
- 贸易是否能促进经济增长 ?
- 住房贷款市场存在种族歧视吗?
- 明年的通货膨胀率会是多少?
- 财政学领域的例子?

什么是计量经济学

实证研究的一般方法

- 从实际数据中发现有趣的经济现象或规律
- 经济理论/经济模型的推演
- 经济模型的估计与检验
- 应用：将经过验证的规律用于实践并提出政策建议。

什么是计量经济学

- “计量经济学”(econometrics)，是运用概率统计方法对经济变量之间的(因果)关系进行定量分析的科学。

- Wooldridge：计量经济学是一门基于统计方法的发展来估计经济关系、检验经济理论、评价和实施政府和商业政策的一门学科。

- Stock & Watson：计量经济学是使用经济理论和统计技术来分析经济数据的科学和艺术。

什么是计量经济学

- Frisch (1933)：计量经济学和经济统计决不相同。它与我们所说的一般经济理论也不相同，尽管该理论的相当一部分具有明确的定量特征。计量经济学也不应被视为数学应用于经济学的同义词。

- 经验表明，统计、经济理论和数学这三种角度都是真正理解现代经济生活中数量关系的必要条件，但其中任何一种都不是充分条件。这三者的统一才是强大的正是这种统一构成了计量经济学。

为什么要学习计量经济学

- **普遍性**：使用实验数据开展研究的经济学，在经济学中是少数。
- **适用性**：需要使用非实验性、观察性的数据做出统计推断。
- **重要性**：在将经济理论解释和应用于现实数据时显得非常重要。
- **明确性**：对于那些解释政策变化效应显得含糊不清的经济理论，如果使用计量经济学方法进行政策评估，会降低这种模糊性。
- **定量分析**：使用数据来检验经济理论或估计经济关系的经验分析。

为什么要学习计量经济学

- Angrist：我们(经济学家)工作的魅力在于利用这个世界作为实验室来探索经济变量之间的因果关系。然而，我们用数据而不仅仅使用热情来进行这种探索。

为什么要学习计量经济学

- 我们在检验经济变量关系时考虑的三种形式：
  - 关系的存在性：被解释变量是否和众多的解释变量之间存在被认为决定了被解释变量的经济关系。
  - 关系的方向性：被解释变量或者结果变量与某个解释变量之间的关系方向，以及其假设可以通过观察获得的关系方向。
  - 关系的大小性：被解释变量与解释变量之间关系的大小程度，其中包括统计显著性和经济显著性。

- 思考1：如何评价一款药品的疗效?
- 如果有人发明了一种新的抗癌药，为了证明这种抗癌药的效果，发明人请来了100位癌症患者，让他们按要求服用此药。几个疗程之后，发现这100位癌症患者中，有50%的患者都康复了。那么，是不是说明这种抗癌药的治疗效果有50%呢?

- 思考2：相对于非重点大学，“双一流”在多大程度上能提高人们的收入?
- 假如我们统计某双一流大学和某普通大学2024届大学生的毕业工资，并且发现双一流大学2024届毕业生的平均工资，要显著高于某普通大学毕业生工资约2000元。并且不管你用什么计量方法，控制什么个体特征，结果都很显著。计量结果既有干预组，又有控制组，潜在结果可观测。是否能够说明某双一流大学的教育作用显著高于某普通大学呢?

局限性

- 经济理论或模型只能抓住主要或最重要的因素
- 计量经济学是对大量现实中的“平均行为”的分析，然而某一次具体实验的结果有可能和"平均行为”相去甚远
- 经济是一个不可逆转或不可重复的系统，经济关系常常随着时间而变化
- 数据质量不一定是良好的

例1

- 我们希望从经验上决定家庭的食物支出，尤其是家庭的食物支出与家庭经济收入之间的关系。

- 样本数据由从所有家庭总体的38个家庭随机抽样组成。
- 随机抽样中的每个家庭，我们具有三个观察变量：

- $foodexp$:家庭的食物年度支出，千美元/年
- $income$:家庭的年度收入，千美元/年
- $hhsize$:家庭规模，使用家庭中人口数度量

- 问题1：这些样本数据产生了什么样的关系？
  - 相当于：这些观察样本对应的**数据生成过程**如何？

- 回答：我们试图模拟$foodexp$的每一个总体值，使用下面形式的经济变量关系

$$foodexp_i = \underbrace{f(income_i, hhsize_i)}_{the\ population\ regression\ equation} + u_i, i = 1,2,\dots,38$$

其中

- $\text{food}exp_i$:我们试图解释的每个家庭i的年度食物支出
- $income_i$: 我们认为可能解释变量$foodexp_i$的一个解释变量，家庭年收入
- $hhsize_i$:我们认为可能解释变量$foodexp_i$的一个解释变量，家庭人口数

- $f(income_i, hhsize_i)$:表示使用变量$income_i$和$hhsize_i$系统决定$foodexp_i$的总体回归函数。

- u_i:一个不可观测的误差项，这个误差项表示所有决定$foodexp_i$中部分成分的未知和不可测变量。

- 问题2：总体回归函数$f(income_i, hhsize_i)$该采取什么形式?

- 回答：我们一般假设**总体回归函数**是一个线性函数：

$$f(income_i, hhsize_i) = \beta_0 + \beta_1 income_i + \beta_2 hhsize_i$$

- 启示：**总体回归方程**就是：

$$\begin{aligned} foodexp_i &= f(income_i, hhsize_i) + u_i \\ &= \beta_0 + \beta_1 income_i + \beta_2 hhsize_i + u_i \end{aligned}$$

- 可观测变量：
  - $\mathrm{food}exp_i$:家庭i的年度食物支出
  - $income_i$: 家庭i的年收入
  - $hhsize_i$: 家庭i的人口数

- 不可观测变量：
  - u_i:总体中家庭i年度食物支出的随机误差项的值。

- 未知参数：回归系数$\beta_0, \beta_1, \beta_2$的总体值(真实值)是未知的。
  - β_0:截距项系数
  - β_1:变量$income_i$的斜率系数
  - β_2:变量$hhsize_i$的斜率系数

例2

- 我们希望从经验上研究工人工资率的决定因素。特别地，我们希望研究具有相同特征的男性工人和女性工人在工资率上是否存在显著差异。

- 样本数据由1976年美国人口调查中在劳动力市场526个受雇佣工人获得的工资数据组成。对于随机样本中的每个工人，我们具有6个可观测变量。
  - $wage$:每个工人平均小时工资，美元/小时
  - ed:每个工人完成的受教育年限，年度
  - exp:每个工人的工作经验，年度
  - ten:每个工人在当前岗位的工作年限，年度
  - $female$:如果是女性工人，取值为1；否则，取值为0
  - $married$:如果已婚，取值为1；否则取值为0

计量经济学中的两类基本变量

- 连续型变量，例如工资，受教育年限，工作经验，当前岗位工作年限数

- 离散型变量或者分类变量，例如是否为女性，是否已婚，这些取值为二值的变量通常被称为示性变量，或者虚拟变量。

- 问题1：这些样本数据产生的数量关系是怎么样的?
  - 数据生成过程是怎样的?

- 回答：我们试图模拟每个$wage$变量的总体值$wage_i$为以下的**总体回归方程**形式：

$$wage_i = f(ed_i, exp_i, ten_i, female_i, married_i) + u_i$$

- 其中:

- $wage_i$ =希望研究的被解释变量
  =每个受雇佣工人i的每小时工资数
- ed_i =我们希望用于解释被解释变量$wage_i$的解释变量或自变量
  =每个受雇佣工人i的受教育年限
- exp_i =我们希望用于解释被解释变量$wage_i$的解释变量或自变量
  =每个受雇佣工人i的工作经验
- ten_i =我们希望用于解释被解释变量$wage_i$的解释变量或自变量
  =每个受雇佣工人i在当前岗位的工作年限数
- $female_i$ =我们希望用于解释被解释变量$wage_i$的解释变量或自变量
  =每个受雇佣工人i的性别，如果是女性，取值为1，否则取值为0.
- $married_i$ =我们希望用于解释被解释变量$wage_i$的解释变量或自变量
  =每个受雇佣工人的婚姻状况，如果已婚，取值为1，否则取值为0.

- $f(ed_i, \exp_i, ten_i, female_i, married_i)$

  =系统表述被解释变量$wage_i$和解释变量

$ed_i, \exp_i, ten_i, female_i, married_i$关系的总体回归函数。

- u_i =表述所有决定个人工资$wage_i$的所有未知和不可测变量的随机误差项。

- 问题2：
- 总体回归函数$f(ed_i, \exp_i, ten_i, female_i, married_i)$应该采用什么样的数学形式？

- 回答：我们假设总体回归函数采用以下的线性函数形式：

$$f(ed_i, \dots, married_i) = \beta_0 + \beta_1 ed_i + \beta_2 exp_i + \beta_3 ten_i + \beta_4 female_i + \beta_5 married_i$$

- 启示：总体回归方程就是：
- $wage_i = f(ed_i, \dots, married_i) + u_i$

$$= \beta_0 + \beta_1 ed_i + \beta_2 exp_i + \beta_3 ten_i + \beta_4 female_i + \beta_5 married_i + u_i$$

- 可观测变量：
  - $wage_i$:被解释变量，第i个受雇佣工人的每小时工资
  - $ed_i, \exp_i, ten_i, female_i, married_i$：第$i$个受雇佣工人一系列的解释变量。

- 不可观测变量：
  - u_i:表示总体中第i个受雇佣工人在每小时工资总体中的随机误差项。

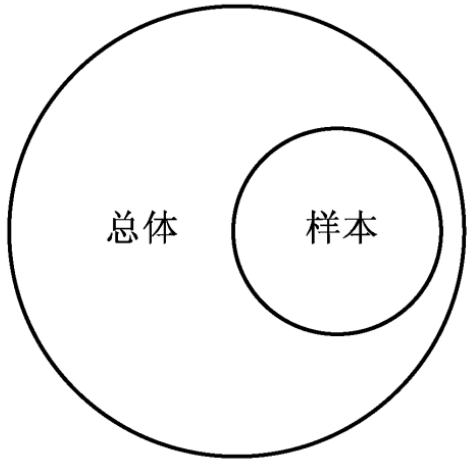
- 未知参数：回归系数$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$是未知的。
  - β_0 =截距项的系数
  - β_1 =变量ed_i的斜率系数
  - β_2 =变量exp_i的斜率系数
  - β_3 =变量ten_i的斜率系数
  - β_4 =变量$female_i$的斜率系数
  - β_5 =变量$married_i$的斜率系数

计量经济学的任务

- 学会从样本数据计算获得回归系数$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$数值的可靠估计。

- 学会从回归系数的样本估计量中对回归系数真实值的有关原假设进行检验。

统计推断的思想



总体与样本

- 计量经济学的主要方法是数理统计的**统计推断**(statistical inference)
- **总体**(population)：感兴趣的研究对象全体：$Y = \{y_1, y_2, y_3, y_4 \dots\}$
- **个体**(individual)：其中的每个研究对象
- **参数**(parameter/estimand)：描述总体某种特征或关系的参数, 被估计量

20岁男性的平均身高:μ

- 由于总体的难以获得性，我们希望通过样本来了解总体的特征
- **样本**(sample)：通过抽样而获得的总体的一部分:$\{y_1, \dots, y_N\}$
- **随机样本**（random sample）：总体中的每位个体都有相同的概率被抽中，且抽中的概率相互独立（独立同分布）
- **估计量**(estimator)：统计工具/方法，基于数据的函数
- **估计值**(estimates)：将估计量应用于数据后的函数数值

中国20岁男性的平均身高:

$\hat{\mu} = \bar{\mu} = 176cm$

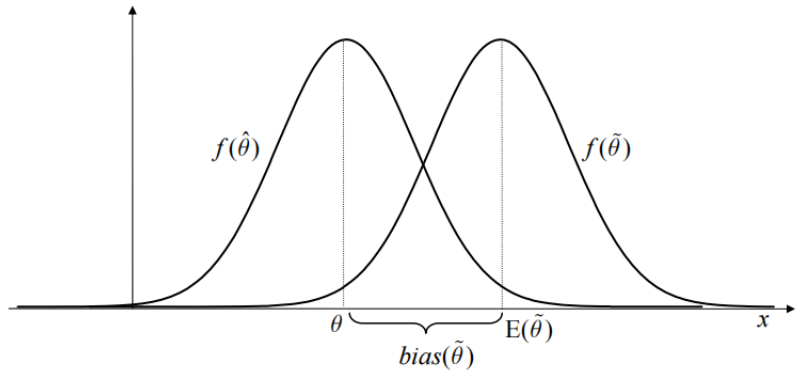
统计推断的思想

- 由于总体难以获得，我们希望通过样本来了解总体的特征

中国20岁男性的平均身高→20岁男性的平均身高：$\hat{\mu} \to \mu$

- 若想推论正确，我们希望我们的统计量能满足一些特性。
- 无偏性(unbiasedness)：$E(\hat{\mu}) = \mu$
- 有效性(efficiency)：$\mathrm{Var}(\hat{\mu})$最小化。
- 一致性(consistency)：$N \to \infty$，$\mathrm{Var}(\hat{\mu}) \to 0$，$E(\hat{\mu}) \to \mu$
- 充分性(sufficiency)

- 因此我们需要知道估测量的分布:
- 抽样
- 大数定理
- 中心极值定理



无偏估计量$\hat{\theta}$与有偏估计量$\tilde{\theta}$的概率分布

计量经济学的四元素

- 1.数据
  - 对非实验性的可观测数据进行搜集和编码，获得计量经济学的原始材料。
- 2.设定
  - 对设想中的样本数据生成过程（DGP）设定计量经济学模型。
  - 经济模型（基于经济关系）和统计模型（基于随机变量统计性质）。
- 3.估计
  - 使用模型中的可观测变量的样本数据计算所有未知参数估计量的数值。
- 4.推断
  - 使用样本数据计算的参数估计量检验用于描述总体行为的未知总体参数。

计量经济学研究的一般步骤

1. 明确自己的研究问题
2. 写出要估计的回归方程式（最重要的一步）
3. 选择适当的数据集
4. 使用统计软件用数据估计回归方程式
5. 分析并解释结果

相关关系与因果关系

- 相关关系（相关性）：
  - ✓ 是两种或两种以上事物之间的相互关系。通常，这是一种统计关系，两个变量是相互依赖的。

- 因果关系：
  - ✓ 意味着一个特定的行为会导致一个特定的、可测量的结果。

相关关系与因果关系

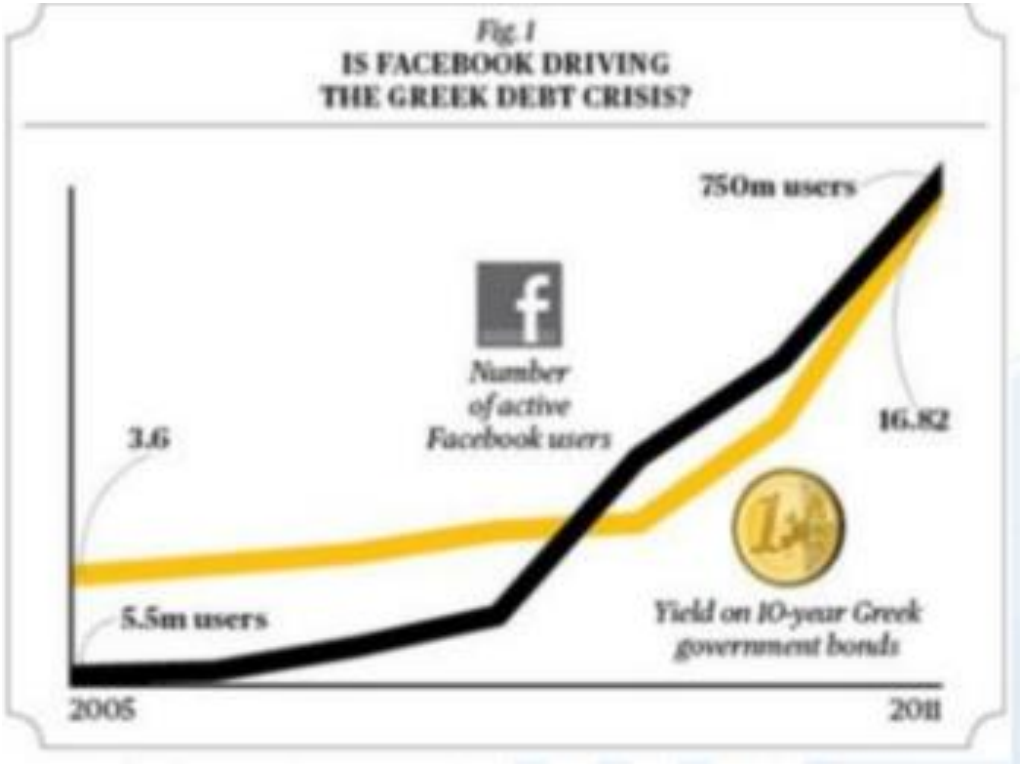


Fig.1
IS FACEBOOK DRIVING
THE GREEK DEBT CRISIS?
750m users
Number of active Facebook users
3.6
16.82
5.5m users
Yield on 10-year Greek government bonds
2005
2011

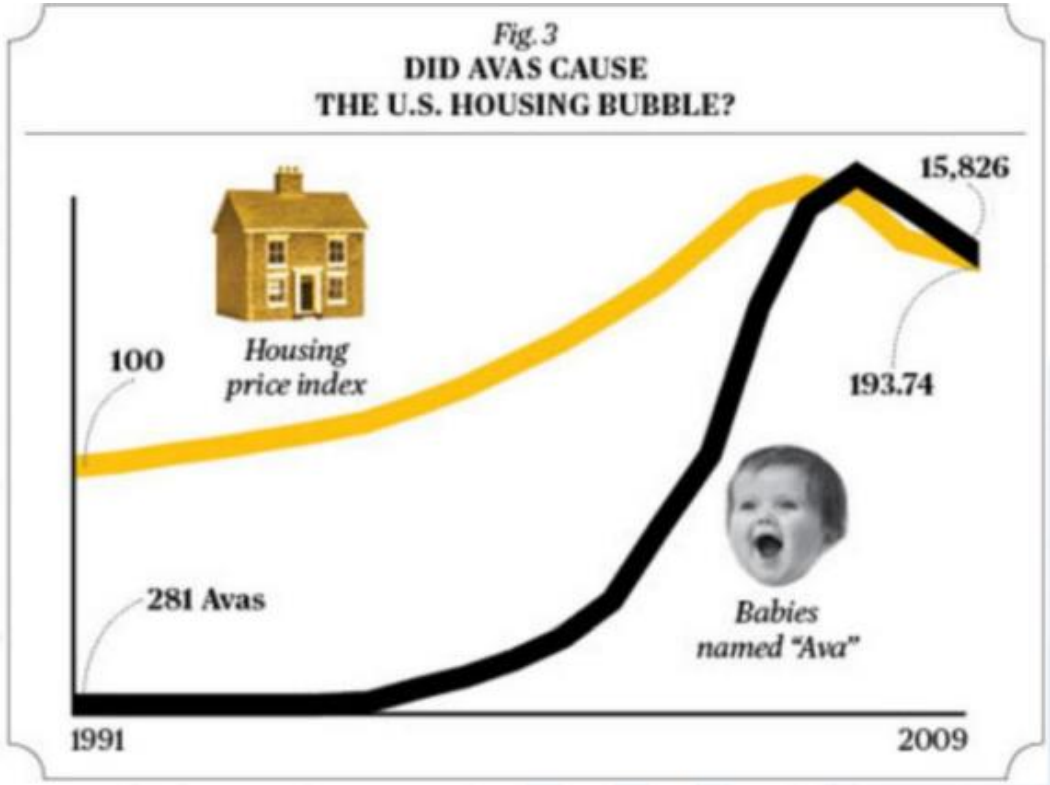


Fig.3
DID AVAS CAUSE
THE U.S. HOUSING BUBBLE?
15,826
100
Housing price index
193.74
281 Avas
Babies named "Ava"
1991
2009

相关关系与因果关系

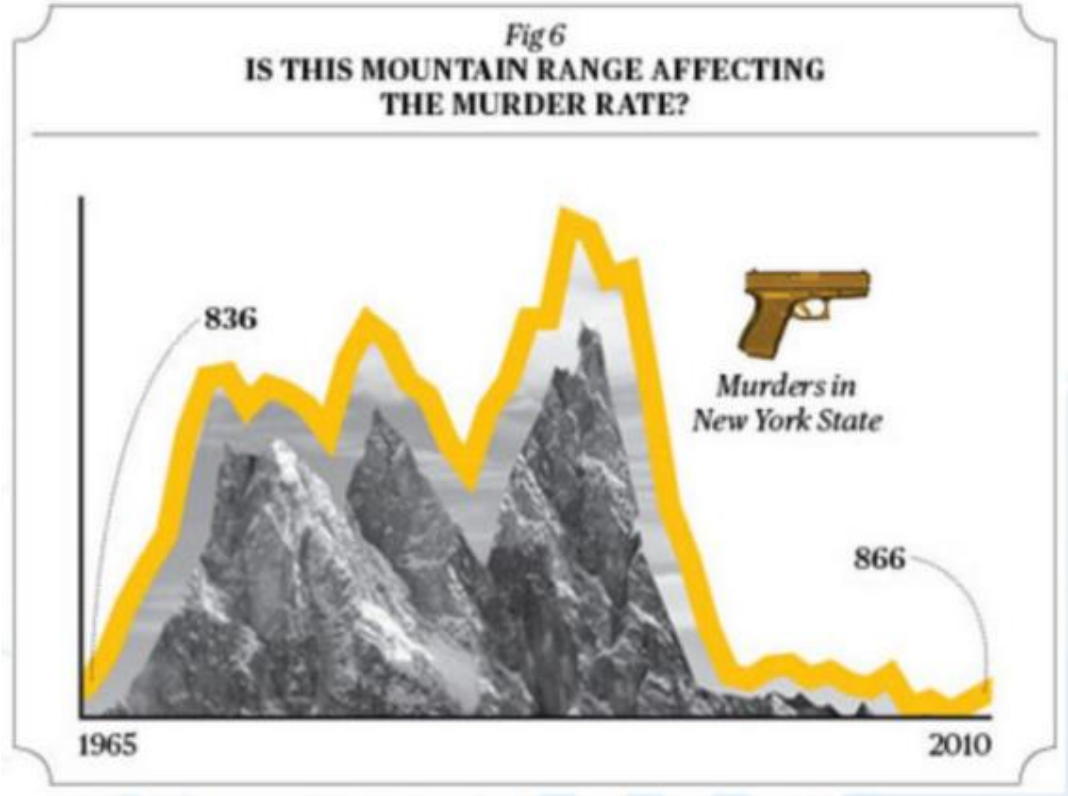


Fig 6
IS THIS MOUNTAIN RANGE AFFECTING
THE MURDER RATE?
836
Murders in
New York State
866
1965
2010

相关关系与因果关系

- 例(相关关系) 你看到街上的人们带雨伞，于是预测今天要下雨。这只是相关关系，“人们带伞”并不导致“下雨”。

- 例(相关关系) 根据与流感相关的海量词条搜索记录，谷歌公司通过分析大数据(big data)，可以很快地预测流行病的地域传播。这也只是相关关系，上网搜索流感信息并不导致流感传播。

相关关系的定义

- 协方差
- $cov(X,Y) = E[X - \mu_X][Y - \mu_Y]$
- $S_{XY} = \frac{1}{n-1}\sum(x_i - \bar{x})(y_i - y)$

- 相关系数
- $Corr(X,Y) = \frac{cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

相关关系与因果关系

- 如果只对预测感兴趣，则相关关系就足够了。
- 对于经济学分析而言，仅仅建立变量之间的相关关系是不够的，我们希望更进一步确定因果关系。
  - 雨量与雨伞数量的关系
  - 教育的回报率
  - 贸易与增长

相关关系与因果关系

- 由于实验数据的缺乏，计量经济学常常不足以确定经济变量之间的因果关系。
- 但大多数实证分析的目的恰恰正是要确定变量之间的因果关系 (即X 是否导致Y)，而非相关关系。
- 如果要推断变量之间的因果关系，则计量分析必须建立在经济理论的基础之上，即在理论上存在X 导致Y 的作用机制。
- 但即使有理论基础，因果关系常常依然不好分辨。
  - 逆向因果
  - 遗漏变量

逆向因果关系

- 首先，可能存在“逆向因果关系”(reverse causality)或“双向因果关系”。
- **例** FDI（外商直接投资）促进经济增长，但 FDI 也被吸引到快速增长的地区。
- **例** 收入增加引起消费增长，而消费增长也拉动收入增加。
- **例** 经济萧条可能引起内战，但内战也会导致经济停滞。

遗漏变量

- 其次， 被遗漏的第三个变量(Z)也可能对这两个变量(X, Y)同时起作用。

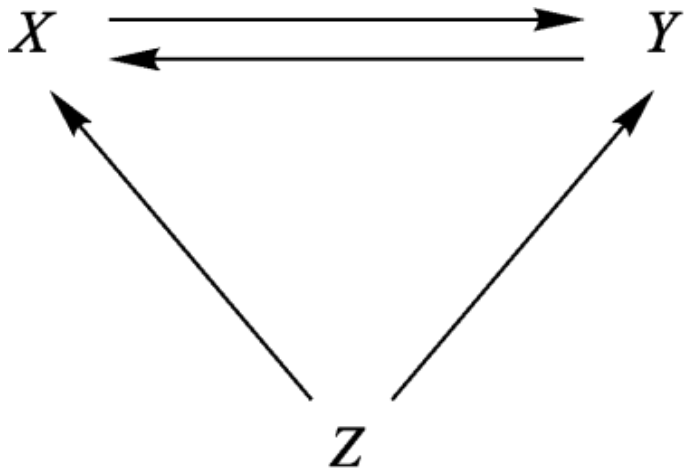


图 1.1　可能的因果关系

遗漏变量

- **例** 某外星人来到地球，发现人类会死亡，十分不解。于是开始在全球广泛观察死亡现象，并收集了大量的数据。

- 结果发现，许多人类躺在医院病床(X)之后死去(Y)，故推断医院病床是死亡的原因。
- 外星人认为，由于躺在医院病床上，总是发生于死亡之前，故不可能存在逆向因果关系。
- 外星人于是将研究报告投稿发表于某顶尖经济学期刊，并在文末给出政策建议“珍爱生命，远离病床”。

遗漏变量

- **例**(遗漏变量)　考虑决定教育投资回报率(returns to schooling) 的因素:

$$\ln w_i = \alpha + \beta s_i + \varepsilon_i$$

- $\ln w_i$ (工资对数)：被解释变量(dependent variable)
- s_i (schooling, 教育年限)：解释变量 (explanatory variable, regressor)或自变量(independent variable)。
- ε_i ：不可观测(unobservable)的“误差项”(error term)或“随机扰动项”(stochastic disturbance)，包括所有除s_i以外对$\ln w_i$有影响的因素，以及人类行为的随机性。
- 下标 i：表示第 i 个观测值（即个体 i）。

- 截距项α与斜率β为待估参数。
- β的经济含义为教育投资的回报率，即多上一年学，未来工资能增加百分之几。

- 如果用数据估计此回归方程，其结果一般会显示，对数工资与教育年限显著正相关，且教育投资回报率β还挺高。
- 然而，工资收入也与能力有关，但能力一般不能直接观测，而能力高的人通常选择接受更多教育。
- 在此简单回归中，教育的高回报其实包含了对能力的回报。

- 影响工资收入的因素还可能包括工作经验、毕业学校、人种、性别、外貌等。
- 需尽可能多地引入“控制变量”(control variables)，也就是多元回归的方法，才能较准确地估计我们“感兴趣的参数”(parameters of interest)，即本例的教育投资回报率β。
- 如果我们控制了足够多的其他变量，那么估计的其他条件不变效应通常可以被认为是因果关系。
- 现实中总有某些相关的变量无法观测，即存在“遗漏变量”(omitted variables)，而遗漏变量统统被纳入到随机扰动项ε_i中。
- 随机扰动项ε_i中还可能包含哪些其他因素呢？如果真实模型(true model)为

$$\ln w_i = a + \beta s_i + \gamma s_i^2 + \varepsilon_i$$

- 那么γs_i^2也被纳入到扰动项中了(可以视为广义的遗漏变量)。

- 如果变量测量得不准确，则测量误差也被放入扰动项中了。
- 扰动项就像是一个“垃圾桶”，所有你不想要、无法把握的东西都往里面扔。
- 但我们又希望扰动项拥有很好的性质。在很多情况下，这是自相矛盾的。
- 计量经济学的很多玄妙之处就在于扰动项。如果真正理解扰动项，就加深了对计量经济学的理解。
- 近年来在因果识别方面，计量经济学有许多新方法，例如DID，RDD等。

数据分类

- **不同来源数据：**实验数据(experimental data)，观测数据(observational data)

- 实验数据：来自于设计的实验，用以评价处理效应或因果效应
- 观测数据来源于直接观测非预设实验的结果。
- 计量经济学分析的数据多为观测数据，这些数据往往是多个因素混杂的结果。
- 经济变量原则上都是随机变量。

- **经济数据按照其性质分类**：

- 横截面数据(cross-sectional data，简称截面数据)
- 时间序列数据(time series data)
- 面板数据(panel data)

数据分类：横截面数据（Cross-sectional Data）

- 指的是多个经济个体的变量在同一时点上的取值。

例如：$x_i, i = 1,2, \dots, N$，其中N表示横截面个体数。

数据形式为：$[x_1, x_2, \dots, x_N]'$

- 特点：
  - 非自然排序
  - 随机抽样

表 1.1　2023 年中国分省 GDP (亿元)

省份	GDP	省份	GDP
北京	43760.7	湖北	55803.6
天津	16737.3	湖南	50012.9
河北	43944.1	广东	135673.2
山西	25698.2	广西	27202.4
内蒙古	24627	海南	7551.2
辽宁	30209.4	重庆	30145.8
吉林	13531.2	四川	60132.9
黑龙江	15883.9	贵州	20913.3
上海	47218.7	云南	30021.1
江苏	128222.2	西藏	2392.7
浙江	82553.2	陕西	33786.1
安徽	47050.6	甘肃	11863.8
福建	54355.1	青海	3799.1
江西	32200.1	宁夏	5315
山东	92068.7	新疆	19125.9
河南	59132.4		

数据分类：时间序列数据（Time series data）

- 指的是某个经济个体的变量在不同时点上的取值。

例如，$x_t, t = 1, \ldots, T$

数据形式：$[x_1, x_2, \ldots, x_T]'$

- 特点：
  - 自然排序（按照时间先后顺序排序）
  - 非随机抽样

表 1.2　2014-2023 年北京市 GDP (亿元)

年份	GDP
2014	22926
2015	24779.1
2016	27041.2
2017	29883
2018	33106
2019	35445.1
2020	35943.3
2021	41045.6
2022	41540.9
2023	43760.7

数据分类：面板数据（Panel data）

- 指的是多个经济个体的变量在不同时点上的取值。
- 强调同一组个体

例如，$x_{it}; i = 1,2, \dots, N; t = 1, \dots, T.$

数据形式：$[x_{11}, \dots, x_{1T}, x_{21}, \dots, x_{2T}, \dots, x_{N1}, \dots, x_{NT}]'$

- 面板数据中，固定截面单位可以得到一个时间序列，固定时间单位可以得到一个横截面数据。
- 混合截面数据 （Pooled Cross-sectional Data）与面板数据有所不同，不同时期观察的截面可以不同。

数据分类：面板数据（Panel data）

- 长面板数据（long panel），时间周期T>N观测值数量
- 短面板数据（short panel），时间周期T<N观测值数量

- 平衡面板（Balanced pane）：给定时间点t，所观测到的个体数一致
- 非平衡面板（Unbalanced panel）：每个时期观测到的个体不完全相同
- 若非平衡面板数据缺失是由随机原因造成的，它和平衡面板的处理方法并没有区别
- 若数据缺失是由于非随机因素造成的，则必须考虑缺失的原因
- 造成数据缺失的原因：如企业倒闭、个体死亡、退出调查、新个体加入调查等

表 1.3　2014-2023 年中国分省 GDP (亿元)

省份	年份	GDP
北京	2014	22926
北京	2015	24779.1
……	……	……
北京	2022	41540.9
北京	2023	43760.7
天津	2014	10640.6
天津	2015	10879.5
……	……	……
天津	2022	16132.2
天津	2023	16737.3
……	……	……
新疆	2014	9264.5
新疆	2015	9306.9
……	……	……
新疆	2022	18042.7
新疆	2023	19125.9

课程概览

1. 计量经济学基础
2. 一元线性回归
3. 多元线性回归
4. 大样本OLS
5. 异方差
6. 自相关
7. 模型设定与数据问题
8. 工具变量法
9. 二值选择模型
10. 面板数据
11. 实证研究初步

谢谢！