

第 1 章 绪 论

1.1 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。

计量经济学常不足以确定经济变量间的因果关系(由于实验数据的缺乏);

多数实证分析正是要确定变量间的因果关系(X 是否导致 Y), 而非仅仅是相关关系。

【例】看到街上人们带伞，可预测今天要下雨。这是相关关系；“人们带伞”并不造成“下雨”。

计量分析须建立在经济理论上。但即使有理论，因果关系依然不好分辨。

首先，可能存在“逆向因果”(reverse causality)。

【例】FDI 促进经济增长，但 FDI 也可能被吸引到高增长地区。

其次，可能是被遗漏的第三个变量(Z)对这两个变量(X , Y)同时起作用。

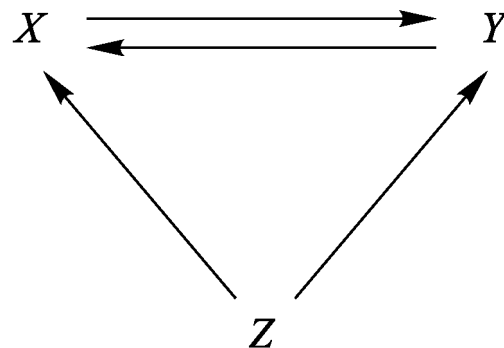


图 1.1 可能的因果关系

例：决定教育投资回报率(returns to schooling)的因素

$$\ln W_i = \alpha + \beta S_i + \varepsilon_i$$

其中， $\ln W$ (工资对数)为“被解释变量” (dependent variable),
 S (教育年限)为“解释变量” (explanatory variable, regressor), ε
为“随机扰动项”(stochastic disturbance)或“误差项”(error term);

下标 i 表示第 i 个观测值(个体 i)； α 与 β 为待估参数。

用数据估计此一元回归会发现，工资与受教育年限显著正相关，而且教育投资回报率 β 还挺高。

但工资收入也与能力有关；能力无法观测，而能力高的人通常选择接受更多教育。教育的高回报率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。

须尽可能多地引入“控制变量”(control variables)，即多元回归的方法，才能准确估计“感兴趣的参数”(parameters of interest)，即本例的教育投资回报率 β 。

现实中总有某些相关变量无法观测，即“遗漏变量”(omitted variables)，都被纳入随机扰动项 ε_i 中。

如果真实模型为

$$\ln W_i = \alpha + \beta S_i + \gamma S_i^2 + \varepsilon_i$$

则 γS_i^2 被纳入到扰动项中了(可视为遗漏变量)。

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“垃圾桶”，所有不想要、无法把握的东西都往里面扔。但又希望扰动项有很好的性质，常导致自相矛盾。

“The devil is in the details.” \Rightarrow “The devil is in the error term.”

计量经济学的很多玄妙之处就在于扰动项。

1.2 经济数据的特点与类型

经济学通常无法像自然科学那样做“控制实验”(controlled experiment)，故经济数据一般不是“实验数据”(experimental data)，而是自然发生的“观测数据”(observational data)。

由于个人行为的随机性，经济变量原则上都是随机变量。

本科教学中，有时假设解释变量是非随机的、固定的(fixed regressors)。

这只是教学法上的权宜之计。如果解释变量为非随机，则无法考虑其与扰动项的相关性。

在本研究生课程中，所有变量都是随机的（即使非随机的常数，也可视为退化的随机变量）。

经济数据按照其性质，可大致分成三种类型：

- 横截面数据(cross-sectional data，简称截面数据)：多个经济个体的变量在同一时点上的取值。比如，2012 年中国各省的 GDP。
- 时间序列数据(time series data)：某个经济个体的变量在不同时点上的取值。比如，在 1978—2012 年山东省每年的 GDP。
- 面板数据(panel data)：多个经济个体的变量在不同时点上的取值。比如，在 1978—2012 年中国各省每年的 GDP。