第四章

经典单方程计量经济学模型

放宽基本假定的模型

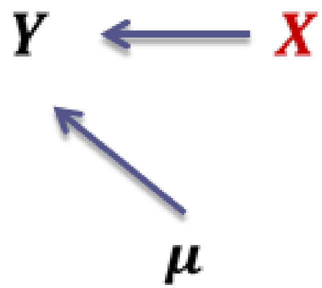
第四章 放宽基本假定的模型

4.3 内生解释变量问题

- 线性计量经济学模型中有一个重要假设——**随机干扰项具有给定条件下的零均值。**
- 如果该假设成立，则称解释变量是外生解释变量或者具有严格的外生性。
- 如违背该假设，则称解释变量为内生解释变量或解释变量具有内生性。

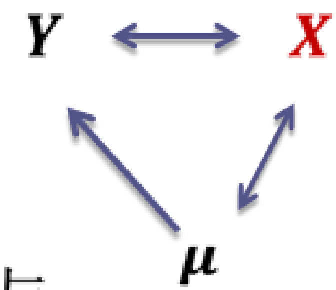
外生变量

- 由系统外因素决定的变量
- 与误差项无关的变量
- 解释变量X和μ独立影响被解释变量Y



内生变量

- 由系统内部因素决定的变量
- 与误差项相关的变量
- μ影响解释变量X，进而间接影响Y，X与Y存在双向因果关系



一、内生解释变量的概念和类型

- **内生解释变量的概念**

对于模型

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i$$

基本假设要求X_1、X_2、…、X_k是严格外生变量

如果存在一个或多个随机变量是内生解释变量，则

称模型存在内生解释变量问题。

- 从假设角度来说，随机干扰项零均值假设

$$E(\mu_i|X_{i1},X_{i2},\dots,X_{ik})=0$$

它的推论为

$$\boxed{Cov(\mu_i,X_i)=E(\mu_iX_i)-E(\mu_i)E(X_i)=E(\mu_iX_i)=E(\mu_i)E(X_i)=0}$$

- 若$Cov(X_{ij},\mu_i)=0$，即解释变量与随机误差项是线性无关的，这一假设被称为外生性假设。

- 若$Cov(X_{ij},\mu_i)\neq 0$，即解释变量与随机误差项存在某种程度相关性，即认为模型存在内生性问题，与随机误差项相关的解释变量就被称为内生解释变量。

- **内生解释变量的类型（以X_2为内生解释变量为例）**

第一类：内生解释变量与随机干扰项同期无关，但异期相关。

$$Cov(X_{i2},\mu_i)=E(X_{i2}\mu_i)=0$$

但

$$Cov(X_{i2},\mu_{i-s})=E(X_{i2}\mu_{i-s})\neq 0 \qquad s\neq 0$$

【注意】截面数据中几乎不存在该类型内生解释变量问题。

第二类：内生解释变量与随机干扰项同期相关。

即

$$Cov(X_{i2},\mu_i)=E(X_{i2}\mu_i)\neq 0$$

这时称内生解释变量为同期内生变量。

【注意】截面数据模型中，内生解释变量问题主要表现为内生解释变量与随机干扰项同期相关性上。

二、产生内生解释变量问题的原因

实际经济问题中，同期内生变量问题往往出现在下面三种情况中：

（一）被解释变量与解释变量具有联立因果关系。

（二）模型中遗漏了重要解释变量，所遗漏变量与模型中一个或多个解释变量具有同期相关性。

（三）滞后变量的引入。

（一）被解释变量与解释变量具有联立因果关系

【例】

为了考察企业引进外资是否真正提高了企业的效益，以企业资金利润率LR为被解释变量，企业资产中外资所占比例WR和其他外生变量X为解释变量，建立如下模型

$$LR_i = \alpha_0 + \alpha_1 WR_i + \beta X_i + \mu_i$$

- 实际计算中发现，效益好的企业更容易引进外资，效益差的企业很难吸引外资进入。
- 模型中WR既影响LR，又被LR所影响，LR_i与μ_i具有同期相关性，导致WR_i与μ_i具有同期相关性。

（二）模型中遗漏了重要解释变量

【例】

劳动者的工资$wage$由劳动者受教育程度$educ$、工作经验$exper$、个人能力$abil$等因素决定，建立模型

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 abil_i + \varepsilon_i$$

估计模型时发现，劳动者个人能力很难测度，因此无法引入变量$abil$，只能置于随机扰动项ε_i中，即实际应用的模型为

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \mu_i$$

$$\mu_i = abil_i + \varepsilon_i$$

个人能力水平往往与受教育程度有很密切的联系，因而导致了$educ_i$与随机干扰项μ_i出现了同期相关性。

（三）滞后变量的引入

【例】

消费C_t由收入Y_t和消费习惯（上一期的消费）C_{t-1}共同决定：

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 C_{t-1} + \mu_t \quad , \quad t = 1, 2, \ldots, T$$

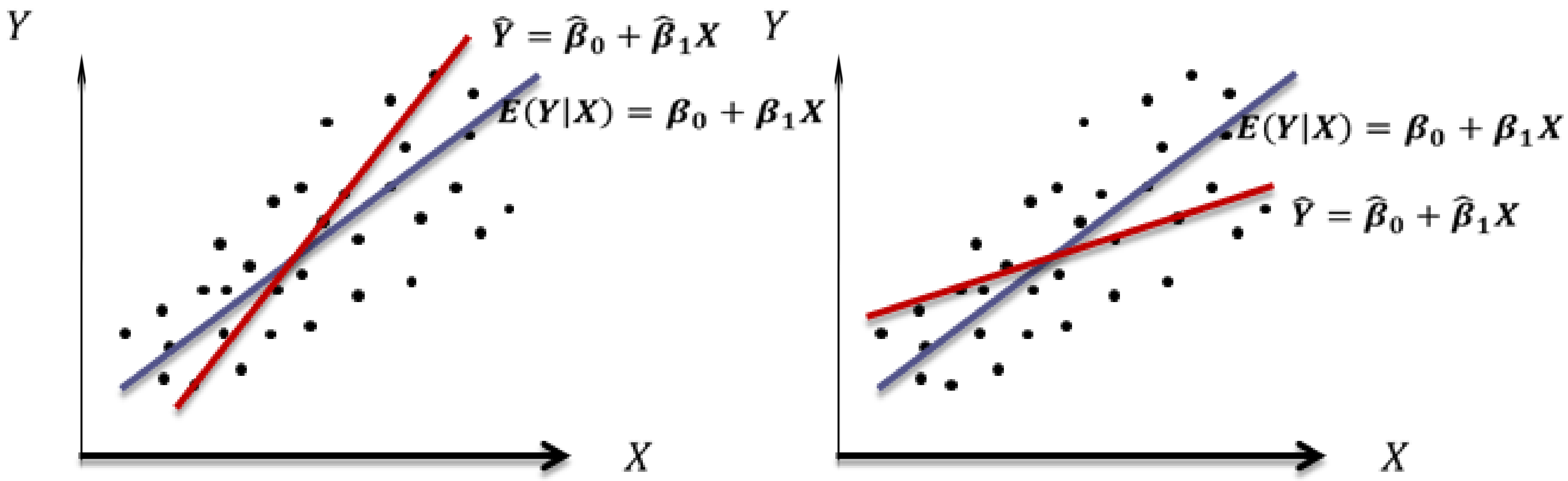
这是一个滞后因变量作为解释变量的模型。

模型中的解释变量C_{t-1}与μ_{t-1}相关，但是与μ_t不相关，因而属于同期不相关但异期相关情况。

三、内生解释变量问题的后果

计量经济学模型一旦出现内生解释变量，且与随机误差项相关的话，如果仍采用OLS法估计模型参数，不同性质的内生解释变量问题会产生不同的后果。

以下以一元线性回归模型为例说明。



- 若 X 与 μ 呈现正相关，样本回归线则可能低估截距项、高估斜率项。
- 若X与μ呈现正相关负相关，样本回归线则可能高估截距项、低估斜率项。

对一元线性回归模型来说：

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

其普通最小二乘估计

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \beta_1 + \frac{\sum x_i}{\sum x_i^2}\mu_i$$

如果X_i与μ_i相关，令$\sum k_i = \frac{\sum x_i}{\sum x_i^2}$，则容易由上式得到

$$E(\hat{\beta}_1) = \beta_1 + E(k_i \mu_i) \neq \beta_1$$

因而参数估计量**不再是无偏估计量**。

从一致性角度来分析：

$$P\lim_{n\to\infty}\left(\beta_1+\frac{\sum x_i\mu_i}{\sum x_i^2}\right)=\beta_1+\frac{P\lim_{n\to\infty}\left(\frac{1}{n}\sum x_i\mu_i\right)}{P\lim_{n\to\infty}\left(\frac{1}{n}\sum x_i^2\right)}$$

$P\lim_{n\to\infty}\left(\frac{1}{n}\sum x_i\mu_i\right)$是$X$与$\mu$的样本协方差的概率极限，等于总体协方差$Cov(x_i,\mu_i)$，$P\lim_{n\to\infty}\left(\frac{1}{n}\sum x_i^2\right)$是$X$样本方差的概率极限，等于$X$的总体方差，因而

$$P\lim_{n\to\infty}\left(\beta_1+\frac{\sum x_i\mu_i}{\sum x_i^2}\right)=\beta_1+\frac{Cov(x_i,\mu_i)}{Var(x_i)}\neq\beta_1$$

即**参数估计量是不一致的**。

对于多元线性回归模型来说

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\mu}$$

可以得到

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X\beta} + \boldsymbol{\mu}) \\ &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mu}\end{aligned}$$

因而

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}E(\boldsymbol{X}'\boldsymbol{\mu})$$

由于$E(\hat{\beta}) = \beta + (X'X)^{-1}E(X'\mu)$，因而内生解释变量问题的后果，即估计量的优良由$E(X'\mu)$决定，也就是说，随机解释变量带来什么后果，取决于它与随机误差项是否相关以及相关到什么程度。

- **如X与μ相互独立，得到的参数估计量是无偏、一致的。**
- **如X与μ同期不相关，但是异期相关，得到的参数估计量有偏、但却是一致的。**
- **如X与μ同期相关，得到的参数估计量有偏、且非一致。**

四、内生解释变量的克服——工具变量法（IV）

工具变量法： 工具变量法（Instrumental Variable Method）是指在模型估计过程中被作为工具使用，以替代与随机干扰项相关的内生解释变量。

方法原理： 利用工具变量Z将内生解释变量X分解成与μ相关和不相关的两个部分，然后利用两阶段最小二乘法进行估计。

（一）工具变量的选取原则

工具变量的目的是替代模型中与随机误差项相关的内生解释变量，因此必须满足以下条件：

1. 与所替代的随机解释变量高度相关；$Cov(Z_i, X_j) \neq 0$

2. 与随机干扰项不相关； $Cov(Z, \mu) = 0$ （依靠常识和经济理论判断）

3. 与模型中其他解释变量不高度相关，避免出现多重共线性。

（二）两阶段最小二乘法

对于模型

内生变量

外生变量

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \mu_i$$

其中，X_i为内生解释变量，Z_i为外生解释变量，为其找到工具变量Z_1

此时，工具变量法可以等价的分解成两步OLS回归

第一步：用普通最小二乘方法进行内生解释变量（X_i）关于工具变量（ Z_{i1} ）和外生变量（Z_i）的回归

$$\hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i + \hat{\alpha}_2 Z_{i1}$$

利用普通最小二乘法估计该模型

若选择了多个工具变量，且不想损失这些工具变量提供的信息时，可以全部作为解释变量引入回归中，如选择两个工具变量Z_1和Z_2，可做回归如下

$$\hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i + \hat{\alpha}_2 Z_{i1} + \hat{\alpha}_3 Z_{i2}$$

第二步：记录通过第一步得到的$\hat{X}_i$拟合值，将其做为解释变量，替代原有的X_i，再次对以下模型进行普通最小二乘估计

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 Z_i + \mu_i$$

● 使用工具变量法的注意事项

1. 在小样本下，工具变量法估计量仍然是有偏的。

2. 工具变量并没有替代模型中的解释变量，只是在估计过程中

   作为“工具”被使用。

3. 如果模型中有两个以上的随机解释变量与随机误差项相关，

   就必须找到两个以上的工具变量。

4. OLS可以看作工具变量法的一种特殊情况。

五、内生性的检验——豪斯曼检验

检验目的：检验随机解释变量是否是**同期外生变量**。

【例】以二元回归模型为例

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_{i1} + \mu_i$$

其中，Z_1为外生变量，怀疑X为同期内生变量，对其进行豪斯曼检验

第一步：用普通最小二乘法进行内生解释变量X关于工具变量Z_2的回归，同时引入原模型中包含的外生变量，记录估计结果中的残差序列v_i

$$X_i = \hat{\alpha}_0 + \boxed{\hat{\alpha}_1 Z_{i1} + \hat{\alpha}_2 Z_{i2}} + \boxed{v_i}$$

无关部分，外生

有关部分，内生

第二步：引入上一步中的残差序列v_i，将其作为解释变量引入原模型进行回归，对其回归结果中的参数进行显著性检验，若其参数δ显著不为0，则变量存在内生性问题

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_{1i} + \delta \widehat{v}_i + \varepsilon_i$$

【注意】该检验等价于判断$\mu_i = \delta \widehat{v}_i + \varepsilon_i$中$\mu$和$v$是否同期相关

案例分析

采用普通最小二乘法估计中国居民人均消费函数：

$$CONSP = \beta_0 + \beta_1 GDPP + \mu$$

其中：*CONSP*为居民人均消费支出，*GDPP*为人均国内生产总值

初步分析：

由于*CONSP*与*GDPP*存在相互影响，容易判断*CONSP*与μ存在同期相关（往往是正相关），OLS估计量有偏且非一致（低估截距项而高估斜率项 ）。

OLS估计结果：

$$\widehat{CONSP} = 201.11 + 0.3862\,GDPP$$

(13.51)　　(53.47)

$$R^2 = 0.9927 \quad F = 2859.23 \quad D.W. = 0.5503$$

如果用$GDPP_{t-1}$为工具变量，可得如下工具变量法估计结果：

$$\widehat{CONSP} = 212.45 + 0.3817\,GDPP$$

(14.84)　　(56.04)

$$R^2 = 0.9937 \quad F = 3140.58 \quad D.W. = 0.6619$$