

计量经济学基础

第四章 多重共线性

主要内容

- 一、多重共线性的概念
- 二、多重共线性的来源与后果
- 三、多重共线性的检验
- 四、多重共线性的修正方法

第一节 多重共线性的概念

对于模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad i=1,2,\dots,n$$

其基本假设之一是解释变量相互独立

如果某两个或多个解释变量之间出现了相关性，则称为模型中解释变量间存在多重共线性(Multicollinearity)。

第一节 多重共线性的概念

若有 $c_1 + c_2 X_{2i} + c_3 X_{3i} + \dots + c_k X_{ki} = 0$ $i=1,2,\dots,n$ 。其中： c_i 不全为0，则称解释变量间存在完全多重共线性

若存在： $c_1 + c_2 X_{2i} + c_3 X_{3i} + \dots + c_k X_{ki} \approx 0$ $i=1,2,\dots,n$ 。其中： c_i 不全为0，则称为解释变量间存在近似多重共线性。

完全共线性的情况并不多见，一般出现的是在一定程度上的共线性，即近似共线性。

第二节 多重共线性的来源与后果

一、来源

1、许多经济变量具有共同变动趋势

时间序列样本：经济繁荣时期，各基本经济变量（收入、消费、投资、价格）都趋于增长；衰退时期，又同时趋于下降。

横截面数据：生产函数中，资本投入与劳动力投入往往出现高度相关情况，大企业二者都大，小企业都小。

第二节 多重共线性的来源与后果

2、滞后变量的引入

在经济计量模型中，往往需要引入滞后经济变量来反映真实的经济关系。

例如，消费= f (当期收入, 前期收入)

显然，两期收入间有较强的线性相关性。

3、样本资料的限制

由于完全符合理论模型所要求的样本数据较难收集，特定样本可能存在某种程度的多重共线性（数据不完备）。

一般经验：

时间序列数据样本：简单线性模型，往往存在多重共线性；**截面数据**样本：问题不那么严重，但多重共线性仍然是存在的。

第二节 多重共线性的来源与后果

二、后果

1、完全共线性下OLS参数估计量不存在

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

的OLS估计量为：

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

如果存在完全共线性，则 $(\mathbf{X}'\mathbf{X})^{-1}$ 不存在，无法得到参数的估计量。

第二节 多重共线性的来源与后果

2、近似共线性下OLS估计量的方差较大

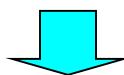
近似共线性下，可以得到OLS参数估计量，但参数估计量方差的表达式为

$$\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

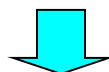
由于 $|X'X| \approx 0$ ，引起 $(X'X)^{-1}$ 主对角线元素较大，使参数估计量的方差较大，使得在参数的显著性检验中增加了接受原假设的可能，从而舍去对被解释变量有显著影响的解释变量。

3、参数的显著性检验失去意义

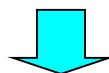
存在多重共线性时



参数估计量的方差（标准差）较大



容易使通过样本计算的 t 值小于临界值，
误导作出参数显著为0的推断



可能将重要的解释变量排除在模型之外

第二节 多重共线性的来源与后果

4、参数估计值不稳定，经济含义不合理

样本观测值稍有变动、增加或减少解释变量等都会使参数估计值发生较大变化，甚至出现符号错误，从而不能正确反映解释变量对被解释变量的影响。

5、模型的预测功能失效

较大的方差容易使预测区间变大，从而使预测失去意义

注意：只要模型满足经典假设，则在近似多重共线性情况下，OLS估计量仍然满足无偏性、线性性和有效性。但此时，无偏性并不意味着对某一给定样本，其参数估计值就等于真实值。有效性也不意味着参数估计量的方差一定很小。

第三节 多重共线性的检验

1、对两个解释变量的模型，采用简单相关系数法

求出 X_1 与 X_2 的简单相关系数 r ，若 $|r|$ 接近1，则说明两变量存在较强的多重共线性。也可以建立两变量之间的辅助线性回归模型，根据拟合优度的高低进行判断。

2、对多个解释变量的模型，采用辅助回归法

分别以一个解释变量为因变量，其他解释变量为自变量，建立多元线性回归模型并进行OLS估计，拟合优度最大且接近1时，说明这个变量与其他所有解释变量间存在共线性。

第三节 多重共线性的检验

辅助回归法中的方差膨胀因子:

对于多元线性回归模型: $Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$

为判断诸自变量间是否存在多重共线性, 进行如下辅助回归:

$$X_{ji} = \alpha_1 + \alpha_2 X_{2i} + \dots + \alpha_{j-1,i} X_{j-1,i} + \alpha_{j+1,i} X_{j+1,i} + \dots + \alpha_k X_{ki} + v_i, j = 2, \dots, k$$

若上述辅助回归的可决系数为 R_j^2 , 则定义自变量 X_j 的方差膨胀因子为:

$$VIF_j = \frac{1}{1 - R_j^2}$$

经验表明, 当 $VIF_j \geq 10$ (即 $R_j^2 \geq 0.9$) 时, 表示自变量 X_j 与其余自变量之间存在着严重的多重共线性问题。

第三节 多重共线性的检验

3、参数估计值的经济意义检验

考察OLS估计中参数估计值的符号和大小，若出现异常，则可能存在多重共线性。

4、参数估计值的稳定性检验

增加或减少模型中的解释变量或样本观测值，若参数估计值变化明显，则可能存在多重共线性。

5、参数估计值的统计检验

若拟合优度较大而t统计量的绝对值很小，则说明可能存在多重共线性问题。

第四节 多重共线性的修正方法

- 1、若多重共线性对参数估计值没有严重影响，可以不修正；
- 2、若多重共线性只影响某些不重要解释变量对应的参数估计值，则只须在模型中剔除这些不重要的解释变量即可；
- 3、在满足经典假定的前提下，由于近似多重共线性并不破坏 OLS 估计量的 BLUE 性质，因此修正多重共线性的思路并不是改变估计方法，而是改变模型本身。包括：

- 1) 增加样本观测值
- 2) 略去不重要的解释变量

3) 用被解释变量的滞后值代替解释变量的滞后值

若多重共线性是由解释变量的现期值与其过去值高度相关引起的，则

可用被解释变量的滞后值代替解释变量的滞后值，避免多重共线性。

例如个人消费 Y_t 取决于现期收入 X_t 和过去收入 X_{t-1}, X_{t-2}, \dots , 则: $Y_t = \beta_1 +$

$$\beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + \dots + u_t$$

通常 $X_t, X_{t-1}, X_{t-2}, \dots$ 高度相关，用 Y_{t-1} 代替 X_{t-1}, X_{t-2}, \dots 则可得新模

$$\text{型: } Y_t = \beta_1 + \beta_2 X_t + \rho Y_{t-1} + u_t^*$$

其中 Y_{t-1} 和 X_t 之间的相关性通常较弱

4、利用参数之间的关系进行参数代换

如果多重共线性是由某些解释变量引起的，根据经济理论和实际分析又知道他们对应的参数之间满足一定的关系，则可通过参数代换减少和避免多重共线性。

例：若产出量取决于资本 K 和劳动投入量 L ，且模型为

$$\ln Y_i = \ln A + \alpha \ln L_i + \beta \ln K_i + u_i$$

通常 L 和 K 高度相关。若已知该 CD 生产函数规模报酬不变，即满足 $\alpha + \beta = 1$ ，则将 $\alpha = 1 - \beta$ 代入原模型可得：

$$\ln \frac{Y_i}{L_i} = \ln A + \beta \ln \frac{K_i}{L_i} + u_i$$

新模型只有一个解释变量，自然也就不存在多重共线性问题

5、变换模型的形式

如果某些解释变量之间高度相关，根据研究目的的实际情况，可通过改变模型形式来避免多重共线性。

例：某产品的销售量 Y 取决于其出厂价格 X_1 、市场价格 X_2 和市场总供应量 X_3 ，模型为：

$$\ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \mu_i$$

通常， X_1 与 X_2 是高度相关的，如果研究目的是为了预测某厂该产品的销售量，则可以用相对价格

$$\frac{X_1}{X_2} \text{来代替 } X_1、X_2 \text{ 对 } Y \text{ 的影响，得：} \ln Y_i = \beta_0 + \beta_1 \ln \frac{X_{1i}}{X_{2i}} + \beta_2 \ln X_{3i} + \mu_i$$

从而避免了多重共线性。

6、修正Frisch法

不仅可检验多重共线性，也是处理多重共线性问题的有效方法。步骤：

(1) 用被解释变量分别对每个解释变量进行回归，根据经济理论和统计检验中选择一个最合适的回归方程作为基本回归方程，通常选择拟合优度最大的回归方程。

6、修正Frisch法

在基本回归方程中**逐个增加其他解释变量**，重新进行线性回归，如新增的解释变量提高了回归方程的拟合优度，并且回归方程中的其他参数在统计上仍然显著，就在模型中保留该解释变量；如果新增解释变量没有提高回归方程的拟合优度，则不在模型中保留该解释变量；

如新增解释变量提高了基本回归方程的拟合优度，并且在方程中某些参数的数值和符号受到显著影响，说明模型中存在多重共线性，对该解释变量同与之相关的其它解释变量进行比较，在模型中保留对被解释变量影响比较大的，略去影响较小的。