

# 第10章多重共线性: 回归元相关会怎样?

教师 黄光辉  
hgh@cqu.edu.cn

- 多重共线性的含义
- 多重共线性的后果和判断方法
- 解决多重共线性的经验方法

- 多重共线性的含义
- 多重共线性的后果和判断方法
- 解决多重共线性的经验方法

# 共线性和矩阵的逆

线性回归模型的矩阵表示:

$$Y = X\beta + U \quad (1)$$

# 共线性和矩阵的逆

线性回归模型的矩阵表示:

$$Y = X\beta + U \quad (1)$$

系数OLS估计的正规方程:

$$(X'X)\beta = X'Y \quad (2)$$

# 共线性和矩阵的逆

线性回归模型的矩阵表示:

$$Y = X\beta + U \quad (1)$$

系数OLS估计的正规方程:

$$(X'X)\beta = X'Y \quad (2)$$

方程可解的条件是:  $(X'X)^{-1}$  存在, 即  $|X| \neq 0$

# 共线性和矩阵的逆

线性回归模型的矩阵表示:

$$Y = X\beta + U \quad (1)$$

系数OLS估计的正规方程:

$$(X'X)\beta = X'Y \quad (2)$$

方程可解的条件是:  $(X'X)^{-1}$  存在, 即  $|X| \neq 0$

若观测具有共线性, 即观测矩阵各列具有:

- ①  $\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_K X_K = 0$  具有非零解, 称为**完全共线性**.
- ②  $\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_K X_K + V = 0$  具有非零解. 其中,  $V$  随机, 称为**不完全共线性**.

- 多重共线性的含义
- 多重共线性的后果和判断方法
- 解决多重共线性的经验方法



# 多重共线性的后果和判断方法

多重共线性的后果:

- ① 若完全共线性, 则回归系数不能确定;

# 多重共线性的后果和判断方法

多重共线性的后果:

- ① 若完全共线性, 则回归系数不能确定;
- ② 若高度共线性, 则参数可以估计, 但方差太大, 无法作出高精度的预测.

# 多重共线性的后果和判断方法

多重共线性的后果:

- ① 若完全共线性, 则回归系数不能确定;
- ② 若高度共线性, 则参数可以估计, 但方差太大, 无法作出高精度的预测.

识别共线性的经验方法:

- ① 高度共线性最明显的信号,  $R^2$  异常高, 但是有若干回归系数  $t$  检验不显著.

# 多重共线性的后果和判断方法

多重共线性的后果:

- ① 若完全共线性, 则回归系数不能确定;
- ② 若高度共线性, 则参数可以估计, 但方差太大, 无法作出高精度的预测.

识别共线性的经验方法:

- ① 高度共线性最明显的信号,  $R^2$ 异常高, 但是有若干回归系数 $t$ 检验不显著.
- ② 仅有两个自变量时, 可采用零阶相关系数判断各列相关情况.

# 多重共线性的后果和判断方法

多重共线性的后果:

- ① 若完全共线性, 则回归系数不能确定;
- ② 若高度共线性, 则参数可以估计, 但方差太大, 无法作出高精度的预测.

识别共线性的经验方法:

- ① 高度共线性最明显的信号,  $R^2$  异常高, 但是有若干回归系数  $t$  检验不显著.
- ② 仅有两个自变量时, 可采用零阶相关系数判断各列相关情况.
- ③ 多于两个自变量时, 零阶相关系数很低, 也可能有多重相关性, 可用偏相关系数判断.

# 多重共线性的后果和判断方法

多重共线性的后果:

- ① 若完全共线性, 则回归系数不能确定;
- ② 若高度共线性, 则参数可以估计, 但方差太大, 无法作出高精度的预测.

识别共线性的经验方法:

- ① 高度共线性最明显的信号,  $R^2$ 异常高, 但是有若干回归系数 $t$ 检验不显著.
- ② 仅有两个自变量时, 可采用零阶相关系数判断各列相关情况.
- ③ 多于两个自变量时, 零阶相关系数很低, 也可能有多重相关性, 可用偏相关系数判断.
- ④ 若 $R^2$ 很高, 但是偏相关系数很低, 可能存在多重共线性, 一般有一个或多个**多余变量**;  
但是若 $R^2$ 偏高, 偏相关系数也偏高, 那么多重共线性不易识别.

# 多重共线性的后果和判断方法

多重共线性的后果:

- ① 若完全共线性, 则回归系数不能确定;
- ② 若高度共线性, 则参数可以估计, 但方差太大, 无法作出高精度的预测.

识别共线性的经验方法:

- ① 高度共线性最明显的信号,  $R^2$ 异常高, 但是有若干回归系数 $t$ 检验不显著.
- ② 仅有两个自变量时, 可采用零阶相关系数判断各列相关情况.
- ③ 多于两个自变量时, 零阶相关系数很低, 也可能有多重相关性, 可用偏相关系数判断.
- ④ 若 $R^2$ 很高, 但是偏相关系数很低, 可能存在多重共线性, 一般有一个或多个**多余变量**;  
但是若 $R^2$ 偏高, 偏相关系数也偏高, 那么多重共线性不易识别.
- ⑤ 自变量 $X$ 中, 取每个 $X_i$ 和其他变量做线性回归, 若 $R_i^2$ 很高, 则该变量可去除. 但是存在误删风险.

- 多重共线性的含义
- 多重共线性的后果和判断方法
- 解决多重共线性的经验方法



# 解决多重共线性的经验方法

没有消除多重共线性的一般方法. 仅有若干经验方法:

- ① 利用外部信息或先验信息;

# 解决多重共线性的经验方法

没有消除多重共线性的一般方法.仅有若干经验方法:

- ① 利用外部信息或先验信息;
- ② 横截面数据和时间序列数据共用;

# 解决多重共线性的经验方法

没有消除多重共线性的一般方法.仅有若干经验方法:

- ① 利用外部信息或先验信息;
- ② 横截面数据和时间序列数据共用;
- ③ 剔除高度相关变量;

# 解决多重共线性的经验方法

没有消除多重共线性的一般方法,仅有若干经验方法:

- ① 利用外部信息或先验信息;
- ② 横截面数据和时间序列数据共用;
- ③ 剔除高度相关变量;
- ④ 数据变换,例如一阶差分;

# 解决多重共线性的经验方法

没有消除多重共线性的一般方法.仅有若干经验方法:

- ① 利用外部信息或先验信息;
- ② 横截面数据和时间序列数据共用;
- ③ 剔除高度相关变量;
- ④ 数据变换,例如一阶差分;
- ⑤ 补充新数据.

# 解决多重共线性的经验方法

没有消除多重共线性的一般方法.仅有若干经验方法:

- ① 利用外部信息或先验信息;
- ② 横截面数据和时间序列数据共用;
- ③ 剔除高度相关变量;
- ④ 数据变换,例如一阶差分;
- ⑤ 补充新数据.

多重共线性时,回归模型用于预测很难保证精度.

# 解决多重共线性的经验方法

没有消除多重共线性的一般方法.仅有若干经验方法:

- ① 利用外部信息或先验信息;
- ② 横截面数据和时间序列数据共用;
- ③ 剔除高度相关变量;
- ④ 数据变换,例如一阶差分;
- ⑤ 补充新数据.

多重共线性时,回归模型用于预测很难保证精度.

讨论数据表10.7, 预测被雇佣的人数。

变量说明:

$Y$  = 被雇佣人数 (千人),  $X_1$  = GNP价格缩减指数,

$X_2$  = GNP (百万美元),  $X_3$  = 失业人数 (千人),

$X_4$  = 军队人数,  $X_5$  = 14岁以上的非编制人口,

$X_6$  = 年份, 1947年到1962年分别取1, 2,  $\dots$ , 16.