

计量经济学I (Econometrics I)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2019-10-15

西北农林科技大学

模块2：结构方程模型（SEM）

Chapter 18. 为什么要关心联立方程模型？

Chapter 19. 联立方程模型的识别问题

Chapter 20. 联立方程模型的估计方法

Chapter 19. 联立方程模型的识别问题

19.1 识别问题

19.2 识别规则

19.3 内生性检验

19.4 外生性检验

19.1 识别问题

识别问题的定义

识别问题 (identification problem), 是指能否从所估计的约简型系数求出一个结构方程参数的估计值。

识别的情形:

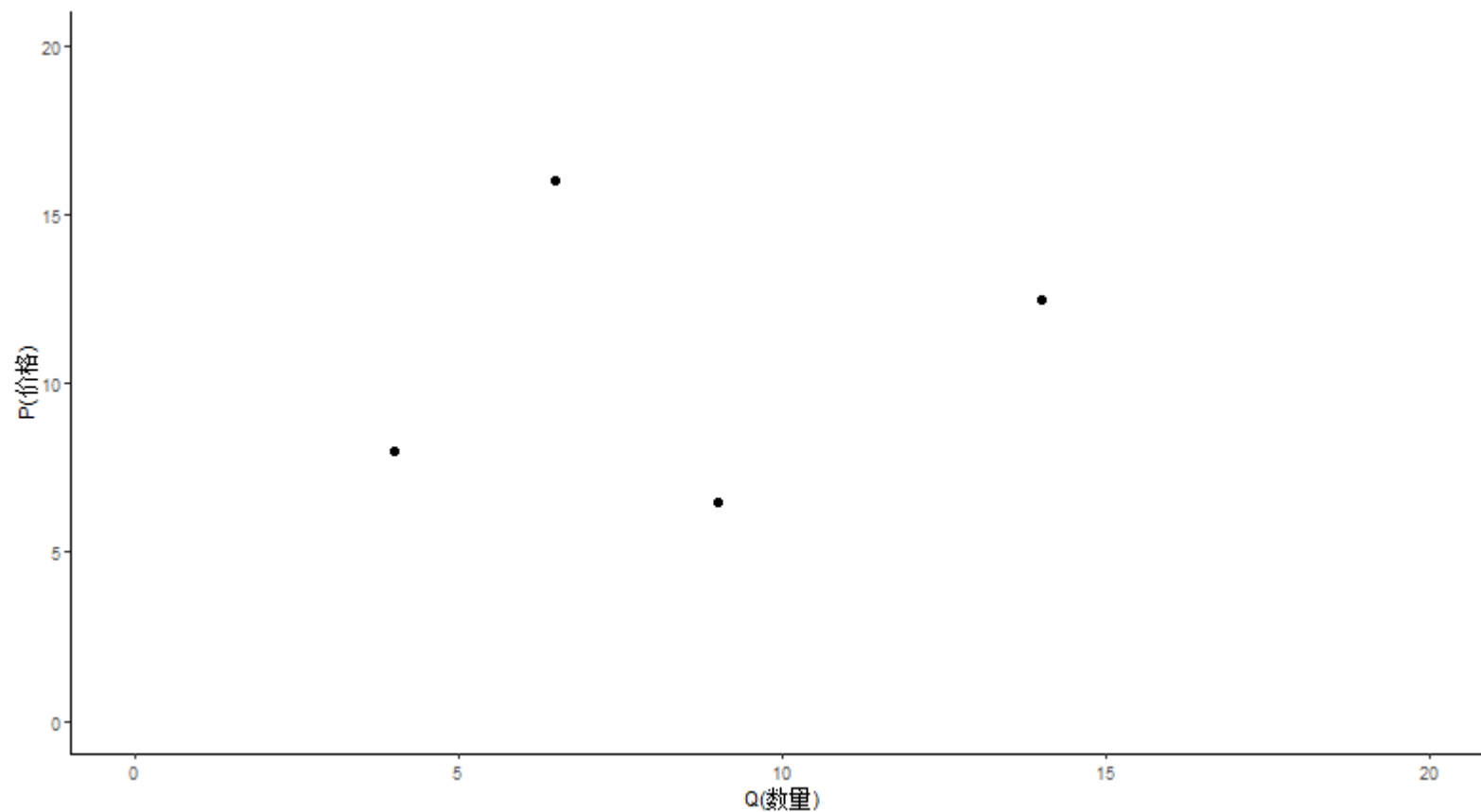
- 如果能够求出估计值, 就说该方程是**可以识别的**, 具体又分为:
 - **恰好识别的** (充分或刚好识别的, Just/Exact Identification) : 是指能够得到结构参数的唯一数值。
 - **过度识别的**(Overidentification): 是指可获得结构参数不止一个的值。
- 如果不能求出估计值, 就说所考虑的方程是**不可识别的或识别不足的** (Underidentification)。

识别问题的演示：联立方程

考虑如下的供需联立方程：

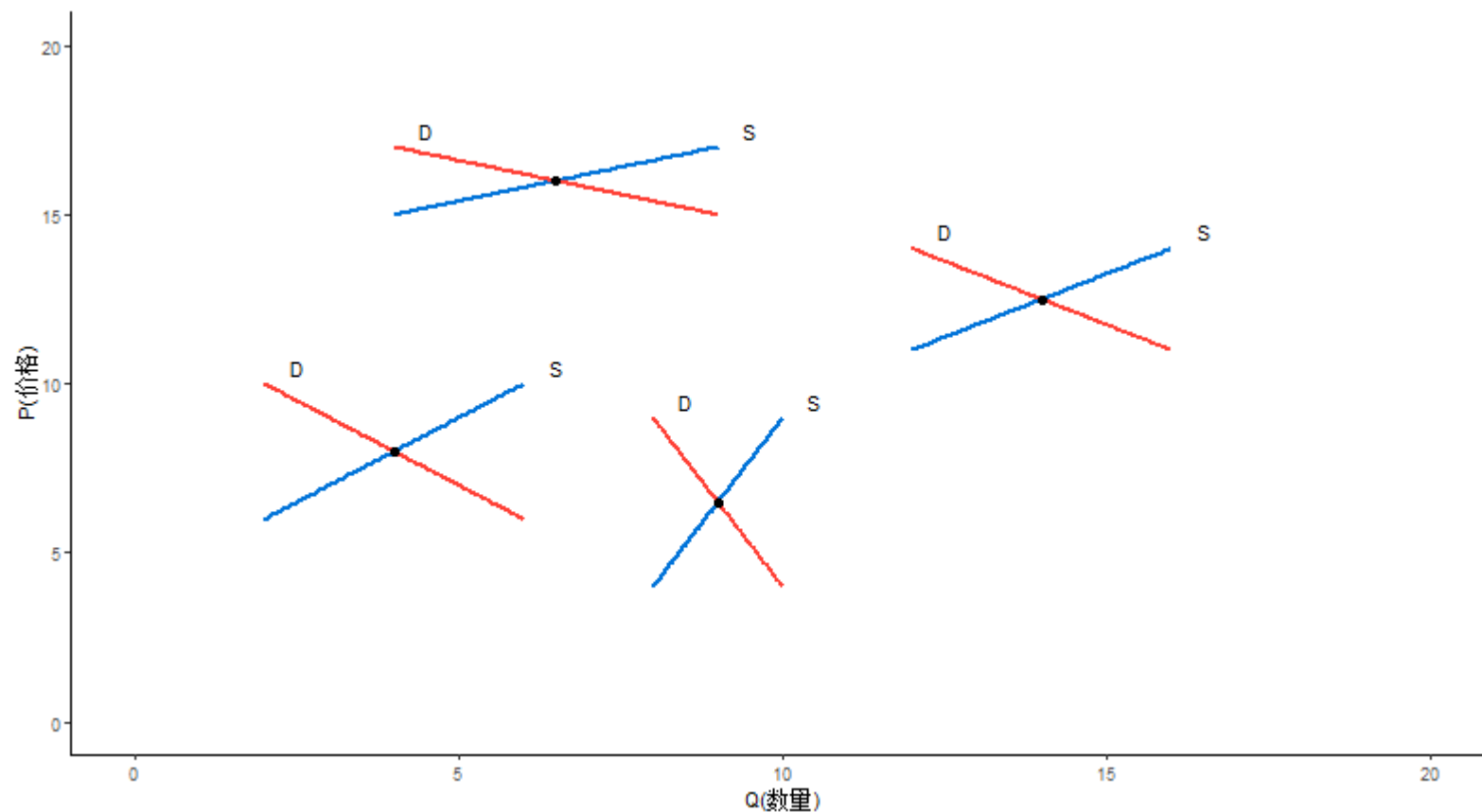
$$\begin{array}{lll} Q = \alpha_0 + \alpha_1 P_t + u_{1t} & (\alpha_1 < 0) & \text{(需求函数)} \\ Q = \beta_0 + \beta_1 P_t + u_{2t} & (\beta_1 > 0) & \text{(供给函数)} \end{array}$$

识别问题的演示：散点图



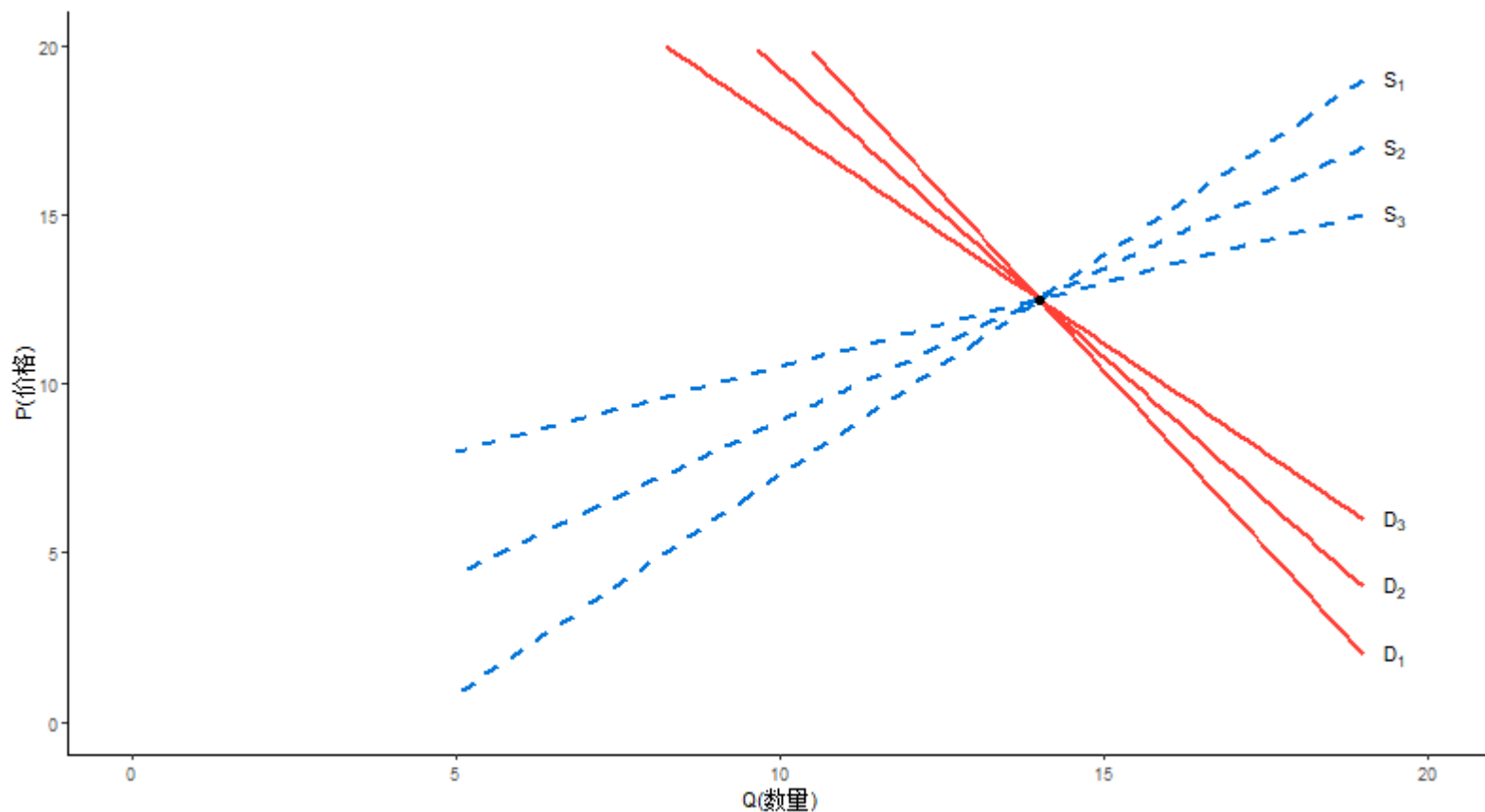
数量Q和价格P的散点分布

识别问题的演示：供需均衡点



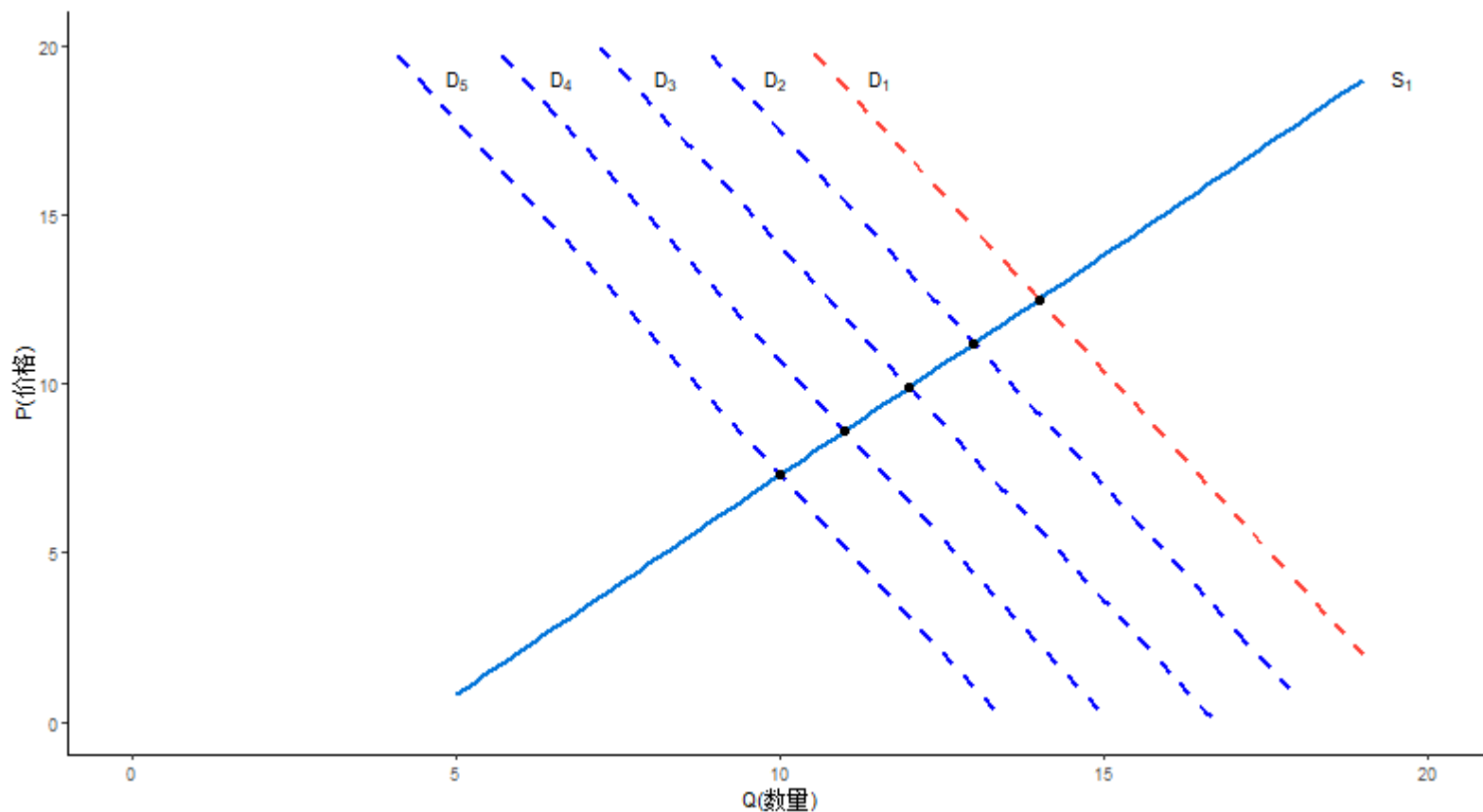
每一散点代表一条需求曲线和一条供给曲线的交点。

识别问题的演示：不可识别



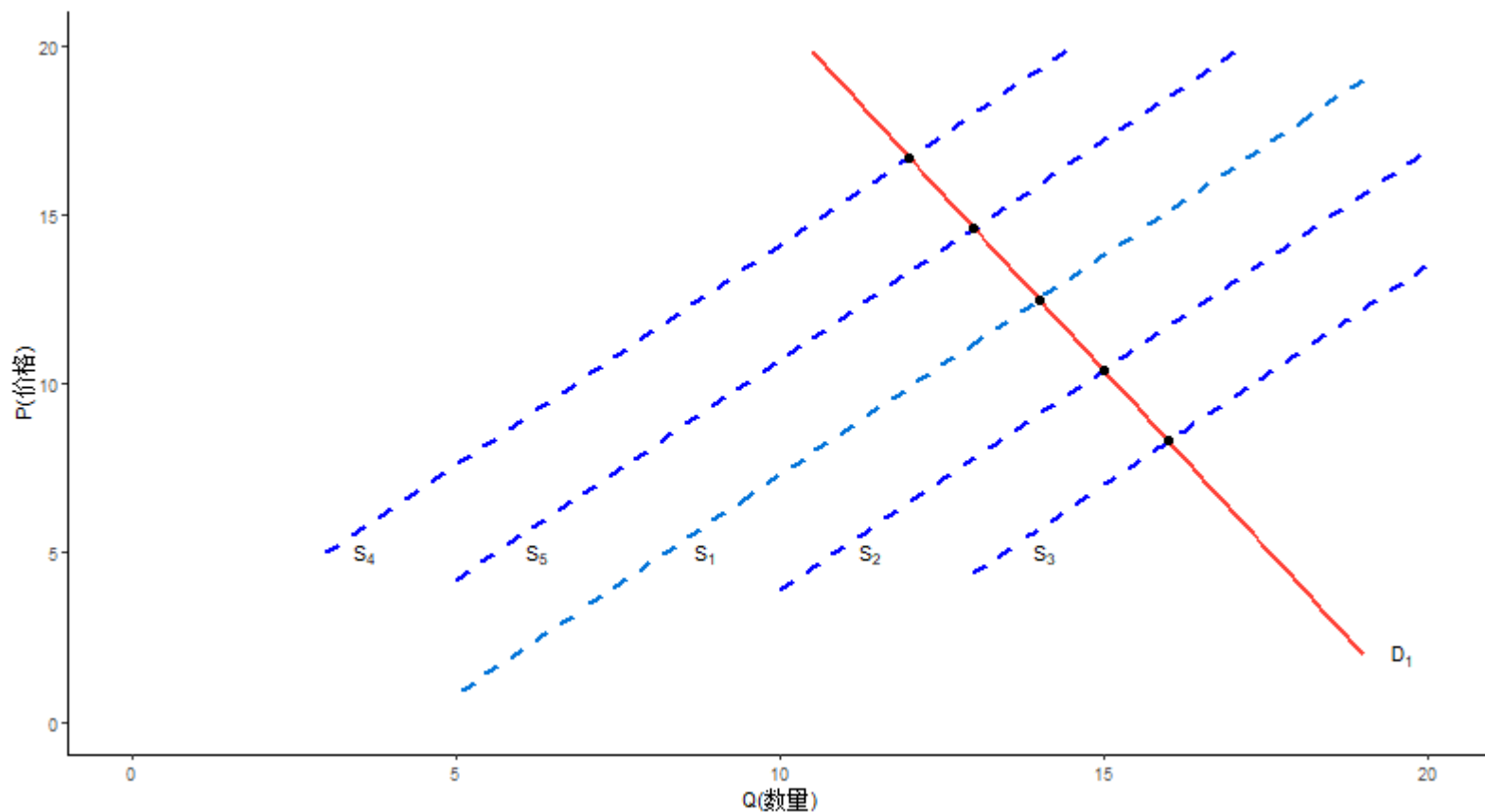
- 我们无法肯定这个点是由图中整个(供求)曲线族中的哪对供求曲线产生的。这时，我们就说供需曲线是**不可识别**的。

识别问题的演示：能识别出供给曲线



- 如果由于收入、嗜好等等的变化，需求曲线随时间而移动，而供给曲线则保持相对稳定，则散点将展现出一条**供给曲线**。这时，我们就说供给曲线是**可识别的**。

识别问题的演示：能识别出需求曲线



- 如果由于气候条件的变化或其他外部因素的变化，供给曲线随时间而移动，但需求曲线保持相对稳定，则散点将展现出一条**需求曲线**。这时，我们就说需求曲线是**可识别的**。

识别不足：结构方程和约简方程

给定联立方程：

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + u_{1t} & (\alpha_1 < 0) & \text{(需求函数)} \\ Q_t = \beta_0 + \beta_1 P_t + u_{2t} & (\beta_1 > 0) & \text{(供给函数)} \end{cases}$$

则容易得到约简方程形式：

$$\begin{cases} P_t = \pi_{11} + v_{1t} \\ Q_t = \pi_{21} + v_{2t} \end{cases}$$

显然，前述的结构方程是**不可识别的**！

- 结构参数个数？
- 约简参数个数？

其中，约简参数和结构参数的关系为：

$$\begin{cases} \pi_{11} = -\frac{\alpha_0 - \beta_0}{\alpha_1 - \beta_1} \\ \pi_{21} = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} \\ v_{1t} = -\frac{u_{1t} - u_{2t}}{\alpha_1 - \beta_1} \\ v_{2t} = \frac{\alpha_1 u_{2t} - \alpha_0 u_{1t}}{\alpha_1 - \beta_1} \end{cases}$$

识别不足：虚构的一个“混杂”方程

针对结构方程，我们可以变换得到：

$$\begin{cases} \lambda Q = \lambda\alpha_0 + \lambda\alpha_1 P_t + \lambda u_{1t} & (\alpha_1 < 0) & \text{(需求函数变换)} \\ (1 - \lambda)Q = (1 - \lambda)\beta_0 + (1 - \lambda)\beta_1 P_t + (1 - \lambda)u_{2t} & (\beta_1 > 0) & \text{(供给函数变换)} \end{cases}$$

进一步，我们可以构造一个如下的“混杂”方程：

$$Q_t = \lambda\alpha_0 + (1 - \lambda)\beta_0 + (\lambda\alpha_1 + (1 - \lambda)\beta_1)P_t + \lambda u_{1t} + (1 - \lambda)u_{2t} \quad \text{(混杂方程)}$$

并记为：

其中：

$$Q_t = \gamma_0 + \gamma_1 P_t + w_t$$

$$\begin{cases} \gamma_0 = \lambda\alpha_0 + (1 - \lambda)\beta_0 \\ \gamma_1 = \lambda\alpha_1 + (1 - \lambda)\beta_1 \\ w_t = \lambda u_{1t} + (1 - \lambda)u_{2t} \end{cases}$$

- 该虚构的**混杂方程**与结构方程中的任何一个都是无差异的！
- 因此表明原来的结构方程是**不可识别的**！

恰好识别：结构方程和约简方程

给定**结构方程**方程：

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} & (\alpha_1 < 0, \alpha_2 > 0) & \text{(需求函数)} \\ Q_t = \beta_0 + \beta_1 P_t + u_{2t} & (\beta_1 > 0) & \text{(供给函数)} \end{cases}$$

得到**约简方程**形式：

$$\begin{cases} P_t = \pi_{11} + \pi_{12} I_t + v_{1t} \\ Q_t = \pi_{21} + \pi_{22} I_t + v_{2t} \end{cases}$$

其中：

$$\begin{aligned} \pi_{11} &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}; & \pi_{12} &= -\frac{\alpha_2}{\alpha_1 - \beta_1}; & v_{1t} &= \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} \\ \pi_{21} &= \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1}; & \pi_{22} &= -\frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1}; & v_{2t} &= \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} \end{aligned}$$

恰好识别：只有一个方程能恰好识别

前述结构方程能否识别么？

答案：

- 约简参数个数？
 - 结构参数个数？
- 只有4个约简参数： $\pi_0, \pi_1, \pi_2, \pi_3$
 - 但是有5个结构参数： $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1$
 - 因此不可能完全求解全部5个结构参数

然而，**供给方程**的结构参数是**恰好识别**的，因为：

$$\beta_0 = \pi_{21} - \beta_1 \pi_{11}; \quad \beta_1 = \frac{\pi_{22}}{\pi_{12}}$$

过度识别：结构方程

给定**结构方程**：

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} & (\alpha_1 < 0, \alpha_2 > 0) & \text{(需求函数)} \\ Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} & (\beta_1, \beta_2 > 0) & \text{(供给函数)} \end{cases}$$

提问：请变换得到约简方程？

过度识别：约简方程

得到**约简方程**形式：

$$\begin{cases} P_t = \pi_{11} + \pi_{12}I_t + \pi_{13}R_t + \pi_{14}P_{t-1} + v_{1t} \\ Q_t = \pi_{21} + \pi_{22}I_t + \pi_{23}R_t + \pi_{24}P_{t-1} + v_{2t} \end{cases}$$

$$\begin{cases} \pi_{11} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \\ \pi_{12} = -\frac{\alpha_2}{\alpha_1 - \beta_1} \\ \pi_{13} = -\frac{\alpha_3}{\alpha_1 - \beta_1} \\ \pi_{14} = \frac{\beta_2}{\alpha_1 - \beta_1} \\ v_{1t} = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} \end{cases}$$

$$\begin{cases} \pi_{21} = -\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \\ \pi_{22} = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1} \\ \pi_{23} = -\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1} \\ \pi_{24} = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} \\ v_{2t} = \frac{\alpha_1u_{2t} - \beta_1u_{1t}}{\alpha_1 - \beta_1} \end{cases}$$

过度识别：存在多个满足条件的解

前述结构方程能否识别么？

- 约简参数个数？
- 结构参数个数？

答案：

- 有8个**约简参数**：
 $\pi_{11}, \pi_{12}, \pi_{13}, \pi_{14}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{24}$
- 只有7个**结构参数**：
 $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2$
- 因此方程数多于未知系数个数，不能求解
得出全部7个结构系数的唯一估计值。

结论：结构方程存在多个满足条件的解，是**过度识别**的。

19.2 识别规则

符号约定

首先给出结构方程中**变量个数**的记号约定：

- M = 结构方程中全部内生变量的个数
- m = 某个方程中内生变量的个数
- K = 结构方程中全部前定变量的个数（包括截距项）
- k = 某定方程中前定变量的个数

可识别的阶条件：规则

阶条件规则1：

- 在一个含有M个方程的结构模型中，一个方程能被识别，它必须**排除**——在联立模型中出现而它自己不出现的——至少M-1个变量（包括**内生或前定**）。也即， $M + K - (k + m) \geq (M - 1)$
 - 如果它恰好有M-1个（内生或前定）变量没有出现，则该方程是**恰好识别的**。也即， $M + K - (k + m) = (M - 1)$
 - 如果它超过M-1个（内生或前定）变量没有出现，则该方程是**过度识别的**；也即， $M + K - (k + m) > (M - 1)$

阶条件规则2：

- 在一个含有M个方程的结构模型中，一个方程能被识别，该方程所**排除的**——在联立模型中出现而它自己不出现——**前定变量**个数必须多于它所含的内生变量的个数减1，也即 $K - k \geq m - 1$ 。
 - 如果 $K - k = m - 1$ ，则该方程是**恰好识别的**；
 - 如果 $K - k > m - 1$ ，则该方程是**过度识别的**；

可识别的阶条件：示例1（都不可识别）

有如下的**结构方程**：

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + u_{1t} & (\alpha_1 < 0) & \text{(需求函数)} \\ Q_t = \beta_0 + \beta_1 P_t + u_{2t} & (\beta_1 > 0) & \text{(供给函数)} \end{cases}$$

阶条件规则1的结论：

- 结构方程的全部变量有 $(M+K)=2+0=2$ 个，且 $(M-1)=2-1=1$
- 对于第1个方程，因为其全部变量数有 $(m+k)=2+0=2$ 个，所以 $(M+K)-(m+k)=2-2=0$ ，显然 $(M+K)-(m+k) < (M-1)$ 。因此第一个方程（需求方程）是**不可识别的**
- 对于第二个方程，因为 $(m+k)=2+0=2$ ，所以 $(M+K)-(m+k)=2-2=0$ ，显然 $(M+K)-(m+k) < (M-1)$ 。因此第二个方程（供给方程）也是**不可识别的**

阶条件规则2的结论：

- 结构方程中前定变量个数 $K=0$
- 对于第1个方程，前定变量个数 $k=0$ ，则 $(K-k)=0$ ；第1个方程的内生变量个数 $m=1$ ，则 $(m-1)=1$ ；显然 $(K-k) < (m-1)$ 。因此第1个方程（需求方程）是**不可识别的**

可识别的阶条件：示例2（不可识别+恰好识别）

给定**结构方程**方程：

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} & (\alpha_1 < 0, \alpha_2 > 0) & \text{(需求函数)} \\ Q_t = \beta_0 + \beta_1 P_t + u_{2t} & (\beta_1 > 0) & \text{(供给函数)} \end{cases}$$

阶条件规则1的结论：

- 结构方程中 $(M+K)=2+1=3$ ，且 $(M-1)=2-1=1$
- 对于第1个方程，因为 $(m+k)=2+1=3$ ，所以 $(M+K)-(m+k)=3-3=0$ ，显然 $(M+K)-(m+k)<(M-1)$ 。第1个方程（需求方程）是**不可识别的**
- 对于第2个方程，因为 $(m+k)=2+0=2$ ，所以 $(M+K)-(m+k)=3-2=1$ ，显然 $(M+K)-(m+k)=(M-1)$ 。第2个方程（供给方程）是**恰好识别的**

阶条件规则2的结论：

- 结构方程中前定变量个数 $K=1$
- 对于第1个方程，前定变量个数 $k=1$ ，则 $(K-k)=0$ ；第一个方程的内生变量个数 $m=2$ ，则 $(m-1)=1$ ；显然 $(K-k)<(m-1)$ 。因此第一个方程（需求方程）是**不可识别的**

可识别的阶条件：示例3（恰好识别+过度识别）

给定**结构方程**：

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} & (\alpha_1 < 0, \alpha_2 > 0) & \text{(需求函数)} \\ Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} & (\beta_1, \beta_2 > 0) & \text{(供给函数)} \end{cases}$$

阶条件规则1的结论：

- 结构方程中 $(M+K)=2+3=5$ ，且 $(M-1)=2-1=1$
- 对于第1个方程，因为 $(m+k)=2+2=4$ ，所以 $(M+K)-(m+k)=5-4=1$ ，显然 $(M+K)-(m+k)=(M-1)$ 。第1个方程（需求方程）是**恰好识别的**
- 对于第二个方程，因为 $(m+k)=2+1=3$ ，所以 $(M+K)-(m+k)=5-3=2$ ，显然 $(M+K)-(m+k)>(M-1)$ 。第2个方程（供给方程）是**过度识别的**

阶条件规则2的结论：

- 结构方程中前定变量个数 $K=3$
- 对于第1个方程，前定变量个数 $k=2$ ，则 $(K-k)=1$ ；第1个方程的内生变量个数 $m=2$ ，则 $(m-1)=1$ ；显然 $(K-k)=(m-1)$ 。因此第1个方程（需求方程）是**恰好识别的**

可识别的阶条件：总结

总结起来：

- 阶条件规则是识别的**必要而非充分条件**，即使它得到满足，方程也会出现不能识别的情形。
例如：

$$Q = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \quad (\alpha_1 < 0, \alpha_2 > 0) \quad (\text{需求函数})$$

$$Q = \beta_0 + \beta_1 P_t + u_{2t} \quad (\beta_1 > 0) \quad (\text{供给函数})$$

根据阶条件规则，我们判断第二个方程（供给方程）是**恰好可识别的**。但实际上，我们还需要确保第一个方程中的收入 I_t 前面的系数 $\alpha_2 \neq 0$ 。也就是说，收入变量不仅仅有可能进入而且**确实**进入了需求函数。

- 阶条件规则1和阶条件规则2是等价的。
- 利用阶条件规则判断识别问题，适合于比较简单的联立方程情形。

可识别的秩条件

秩条件规则：

- 在一个包含 M 个内生变量的 M 个方程的联立模型中，一个方程可识别的**充分必要条件**是，我们能从模型所含而该方程所不含的（内生或前定）变量系数矩阵中构造出**至少**1个 $(M - 1) * (M - 1)$ 阶的非零行列式来。

主要操作步骤：

1. 把结构模型移项，写出**代数式**
2. 按方程写出对应的**系数列表**
3. 找到该方程不含的所有**列**
4. 从这些列中随意找出1个 $(M - 1) * (M - 1)$ 的矩阵
5. 判断并得出结论：如果能找到任何1个行列式为0的矩阵，则该方程就是**不可识别的**

可识别的秩条件：示例（写出代数式2；写出系数列表）

$$\begin{cases} Y_{1t} - \beta_{10} & -\beta_{12}Y_{2t} - \beta_{13}Y_{3t} - \gamma_{11}X_{1t} & = u_{1t} \\ Y_{2t} - \beta_{20} & -\beta_{23}Y_{3t} - \gamma_{21}X_{1t} - \gamma_{22}X_{2t} & = u_{2t} \\ Y_{3t} - \beta_{30} - \beta_{31}Y_{1t} & -\gamma_{31}X_{1t} - \gamma_{32}X_{2t} & = u_{3t} \\ Y_{4t} - \beta_{40} - \beta_{41}Y_{1t} - \beta_{42}Y_{2t} & -\gamma_{43}X_{3t} & = u_{4t} \end{cases}$$

可以将上述结构方程的参数写成如下的形式：

<i>eq</i>		Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
1	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
2	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
3	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
4	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

可识别的秩条件：示例（分析第1个方程的变量数;找到缺失列）

$$\begin{aligned}
 Y_{1t} - \beta_{10} & \qquad \qquad -\beta_{12}Y_{2t} - \beta_{13}Y_{3t} - \gamma_{11}X_{1t} & = u_{1t} \\
 Y_{2t} - \beta_{20} & \qquad \qquad -\beta_{23}Y_{3t} - \gamma_{21}X_{1t} - \gamma_{22}X_{2t} & = u_{2t} \\
 Y_{3t} - \beta_{30} - \beta_{31}Y_{1t} & \qquad \qquad -\gamma_{31}X_{1t} - \gamma_{32}X_{2t} & = u_{3t} \\
 Y_{4t} - \beta_{40} - \beta_{41}Y_{1t} - \beta_{42}Y_{2t} & \qquad \qquad -\gamma_{43}X_{3t} & = u_{4t}
 \end{aligned}$$

- 第1个方程不包含的内生变量有： Y_{4t} ； 不包含的前定变量有： X_{2t}, X_{3t}

<i>eq</i>	1	Y_1	Y_2	Y_3	$[Y_4]$	X_1	$[X_2]$	$[X_3]$
1	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
2	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
3	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
4	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

可识别的秩条件：示例（尝试找到1个目标矩阵；计算其行列式）

---	---	---	---	---	---	---	---	---
<i>eq</i>	1	Y_1	Y_2	Y_3	$[Y_4]$	X_1	$[X_2]$	$[X_3]$
---	---	---	---	---	---	---	---	---
1	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
2	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
3	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
4	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

$$A = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix}$$

$$\det(A) = \begin{vmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{vmatrix} = 0$$

这意味着，矩阵的秩 $rank(A) = \rho(A) < 3$ 。因此，**第一个方程**不满足秩条件，**不能识别**。

可识别的秩条件：总结

总结起来，秩条件识别规则的步骤如下：

- 把**结构方程**写成**代数式**
- 把**系数**写成表格形式
- 划掉被考虑的方程所在行
- 再划掉被考虑的方程中**非零系数**对应的列
- 余下的系数将构成一个**系数矩阵**
- 利用**系数矩阵**，构造任意的 $(M - 1) * (M - 1)$ 阶方阵，并计算方阵的行列式
- 如果能找到**至少一个**行列式不等于零的 $(M - 1) * (M - 1)$ 阶方阵，也即该方阵的秩为M-1，则表明被考虑方程是**可识别的**（恰好或多度识别）
- 如果所有可能的 $(M - 1) * (M - 1)$ 阶方阵，它们的行列式全等于零，也即所有可能 $(M - 1) * (M - 1)$ 阶方阵的秩都**小于**M-1，则表明被考虑方程是**不可识别的**

可识别的秩条件：示例（练习，其他3个方程是否可识别？）

eq	1	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
1	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
2	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
3	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
4	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

答案：

- 第1个方程：不可识别
- 第2个方程：？
- 第3个方程：？
- 第4个方程：？

可识别的秩条件：与阶条件判断法进行比较

$$\begin{aligned} Y_{1t} - \beta_{10} & - \beta_{12}Y_{2t} - \beta_{13}Y_{3t} - \gamma_{11}X_{1t} &= u_{1t} \\ Y_{2t} - \beta_{20} & - \beta_{23}Y_{3t} - \gamma_{21}X_{1t} - \gamma_{22}X_{2t} &= u_{2t} \\ Y_{3t} - \beta_{30} - \beta_{31}Y_{1t} & - \gamma_{31}X_{1t} - \gamma_{32}X_{2t} &= u_{3t} \\ Y_{4t} - \beta_{40} - \beta_{41}Y_{1t} - \beta_{42}Y_{2t} & - \gamma_{43}X_{3t} &= u_{4t} \end{aligned}$$

下面我们再运用**识别的阶条件**对所有方程进行判断，可以得到下面结果：

eq	m	k	M	K	K-k	m-1	结论
1	3	1	4	3	2	2	恰好识别
2	2	2	4	3	1	1	恰好识别
3	2	2	4	3	1	1	恰好识别
4	3	1	4	3	2	2	恰好识别

识别规则的总结

下面对识别规则进行一个全面总结：

- 阶条件是识别的**必要但非充分条件**，即使它得到满足，方程也会出现不能识别的情形。
- 秩条件是识别的**充分必要条件**
- **秩条件**告诉我们所考虑的方程**是否可识别**，而**阶条件**告诉我们它是**恰好识别**还是**过度识别**
- 严格地，我们需要先分析秩条件，确定方程是否可识别；然后再利用阶条件，判断是恰好识别还是过度识别！

19.3 内生性检验

联立性检验（内生性检验）：概念与定义

联立性检验在本质上是检验一个（内生）回归元是否与误差项（ u_t ）相关。

- 如果相关，就存在联立性问题，从而需要找到不同于OLS的估计方法；
- 如果不相关，则认为不存在联立性问题，可以继续使用OLS方法

豪斯曼设定误差检验 (Hausman specification error test)的**一种形式**可用于检验联立性问题。

豪斯曼联立性检验又被称为**豪斯曼内生性检验** (Hausman test of endogeneity)

联立性检验（内生性检验）：检验的原理

给定联立方程：

$$\begin{aligned} Q &= \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} & (\alpha_1 < 0, \alpha_2 > 0) & \quad (\text{需求函数}) \\ Q &= \beta_0 + \beta_1 P_t + u_{2t} & (\beta_1 > 0) & \quad (\text{供给函数}) \end{aligned}$$

其中：P=价格；I=收入；R=财富

很容易得到约简方程：

$$\begin{aligned} P_t &= \pi_{11} + \pi_{12} I_t + \pi_{13} R_t + v_{1t} & (\text{约简方程1}) \\ Q_t &= \pi_{21} + \pi_{22} I_t + \pi_{23} R_t + v_{2t} & (\text{约简方程2}) \end{aligned}$$

联立性检验（内生性检验）：检验的原理

我们可以直接用OLS方法估计约简方程1（价格方程），得到：

$$\begin{aligned} P_t &= \hat{\pi}_{11} + \hat{\pi}_{12}I_t + \hat{\pi}_{13}R_t + \hat{v}_{1t} && \text{(价格方程的OLS估计)} \\ &= \hat{P}_t + \hat{v}_{1t} \end{aligned}$$

进而可以将价格方程的OLS估计结果带入如下**检验方程**，并做如下的OLS估计：

$$Q_t = \beta_0 + \beta_1 \hat{P}_t + \beta_1 \hat{v}_{1t} + u_{2t} \quad \text{(豪斯曼检验方程)}$$

$$Q_t = \beta_0 + \beta_1 P_t + \beta_1 \hat{v}_{1t} + u_{2t} \quad \text{(平狄克检验方程)}$$

豪斯曼联立性检验下：

- 原假设 H_0 ：无联立性问题，要求 \hat{v}_{1t} 和 u_{2t} 不相关；
- 备择假设 H_1 ：有联立性问题，要求 \hat{v}_{1t} 和 u_{2t} 相关；

因此，只需要**检验方程**参数 β_1 进行检验：

- 如果 β_1 统计上**不显著**，则不能拒绝原假设 H_0 ，认为**无联立性问题**
- 如果 β_1 统计**显著**，则拒绝原假设 H_0 ，接受备择假设 H_1 ，认为**存在联立性问题**

联立性检验（内生性检验）

因此，豪斯曼检验的步骤包括：

- step 1: 利用OLS方法，做 $P_t = \hat{\pi}_{11} + \hat{\pi}_{12}I_t + \hat{\pi}_{13}R_t + \hat{v}_{1t}$ 的回归，得到 \hat{v}_{1t}
- step 2: 利用OLS方法，做 $Q_t = \beta_0 + \beta_1\hat{P}_t + \beta_1\hat{v}_{1t} + u_{2t}$ 的回归，并对 \hat{v}_{1t} 的系数进行t检验，并作出联立性问题的判断。需要注意的是，为了更有效地估计，平狄克 (Pindyck) 和鲁宾费尔德 (Rubinfeld) 建议对 $Q_t = \beta_0 + \beta_1P_t + \beta_1\hat{v}_{1t} + u_{2t}$ 的回归。

联立性检验：松露案例（结构和约简方程）

我们仍旧以松露案例模型为例，给定如下联立方程：

$$Q_i = \alpha_0 + \alpha_1 P_i + \alpha_2 PS_i + \alpha_3 DI_i + u_{1i} \quad (\text{需求函数})$$

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 PF_i + u_{2i} \quad (\text{供给函数})$$

很快可以得到约简方程：

$$P_i = \pi_{11} + \pi_{12} PS_i + \pi_{13} DI_i + \pi_{14} PF_i + v_{1i}$$

$$Q_i = \pi_{21} + \pi_{22} PS_t + \pi_{23} DI_t + \pi_{24} PF_i + v_{2i}$$

联立性检验：松露案例（结构和约简系数）

其中**约简系数**和**结构系数**的关系为：

$$\left\{ \begin{array}{l} \pi_{11} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \\ \pi_{12} = -\frac{\alpha_2}{\alpha_1 - \beta_1} \\ \pi_{13} = -\frac{\alpha_3}{\alpha_1 - \beta_1} \\ \pi_{14} = \frac{\beta_2}{\alpha_1 - \beta_1} \\ v_{1i} = \frac{u_{2i} - u_{1i}}{\alpha_1 - \beta_1} \end{array} \right.$$

$$\left\{ \begin{array}{l} \pi_{21} = -\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \\ \pi_{22} = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1} \\ \pi_{23} = -\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1} \\ \pi_{24} = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} \\ v_{2i} = \frac{\alpha_1u_{2i} - \beta_1u_{1i}}{\alpha_1 - \beta_1} \end{array} \right.$$

联立性检验：松露案例（OLS估计价格方程）

按照豪斯曼联立性检验的**第一个步骤**，利用OLS方法回归如下价格**约简方程**，并得到 \hat{P}_i, \hat{v}_{1i}

$$\begin{aligned} P_i &= \hat{\pi}_{11} + \hat{\pi}_{12}PS_i + \hat{\pi}_{13}DI_i + \hat{\pi}_{14}PF_i + \hat{v}_{1i} \\ &= \hat{P}_i + \hat{v}_{1i} \end{aligned}$$

联立性检验：松露案例（OLS估计价格方程；R原始报告）

约简方程1（价格方程）的原始回归结果报告如下：

```
Call:
lm(formula = Hausman_models$mod.P, data = truffles)

Residuals:
    Min       1Q   Median       3Q      Max
-20.48  -3.59   0.28   4.53  12.92

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -32.512     7.984   -4.07  0.00039 ***
PS             1.708     0.351    4.87  4.8e-05 ***
DI             7.602     1.724    4.41  0.00016 ***
PF             1.354     0.299    4.54  0.00011 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.6 on 26 degrees of freedom
Multiple R-squared:  0.889,    Adjusted R-squared:  0.876
F-statistic: 69.2 on 3 and 26 DF,  p-value: 1.6e-12
```

联立性检验：松露案例（OLS估计价格方程；精简报告）

约简方程1（价格方程）的精简回归报告如下：

$$\begin{array}{rcccc} \hat{P} = & -32.51 & +1.71PS & +7.60DI & +1.35PF \\ (t) & (-4.0721) & (4.8682) & (4.4089) & (4.5356) \\ (se) & (7.9842) & (0.3509) & (1.7243) & (0.2985) \\ (fitness) & n=30; & R^2=0.8887; & \bar{R}^2=0.8758 & \\ & F^*=69.19; & p=0.0000 & & \end{array}$$

请解释回归结果？

联立性检验：松露案例（OLS估计价格方程；估计得到的变量）

进而可以分别估计得到 \hat{P}_i, \hat{v}_{1i}

估计价格方程并得到hat.Pi和hat.v1i

P	Q	PS	DI	PF	hat.Pi	hat.vi
29.64	19.89	19.97	2.103	10.52	31.8304	-2.1904
40.23	13.04	18.04	2.043	19.67	40.4658	-0.2358
34.71	19.61	22.36	1.87	13.74	38.5011	-3.7911
41.43	17.13	20.87	1.525	17.95	39.033	2.397
53.37	22.55	19.79	2.709	13.71	40.449	12.921
38.52	6.37	15.98	2.489	24.95	47.4863	-8.9663

Showing 1 to 6 of 30 entries

联立性检验：松露案例（联立性检验模型）

下面，我们继续进行豪斯曼联立性检验的**第二个步骤**（采用平狄克和鲁宾费尔德的建议），采用OLS方法估计如下**豪斯曼检验方程**和**平狄克检验方程**：

$$Q_i = \beta_0 + \beta_1 \hat{P}_i + \beta_1 \hat{v}_{1i} + u_{2i} \quad (\text{豪斯曼检验方程})$$

$$Q_i = \beta_0 + \beta_1 P_i + \beta_1 \hat{v}_{1i} + u_{2i} \quad (\text{平狄克检验方程})$$

因此，我们马上可以得到**第二个步骤的检验结果**。

联立性检验：松露案例（豪斯曼检验；原始报告1）

如果采用**豪斯曼检验方程**，则有原始报告结果如下：

```
Call:
lm(formula = Hausman_models$mod.Q.Hausman, data = truffles_Hausman)

Residuals:
    Min       1Q   Median       3Q      Max
-7.476 -2.892  0.277  3.394  5.380

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.946     2.604    4.59 9.2e-05 ***
hat.Pi          0.104     0.040    2.60  0.0151 *
hat.vi          0.338     0.113    2.99  0.0059 **
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.8 on 27 degrees of freedom
Multiple R-squared:  0.367,    Adjusted R-squared:  0.32
F-statistic: 7.84 on 2 and 27 DF,  p-value: 0.00207
```

联立性检验：松露案例（豪斯曼检验；精简报告1）

如果采用豪斯曼检验方程，上述精简报告结果如下：

$$\begin{array}{rcll} \hat{Q} = & + 11.95 & + 0.10\hat{P}i & + 0.34\hat{v}i \\ (t) & (4.5880) & (2.5952) & (2.9901) \\ (se) & (2.6037) & (0.0400) & (0.1130) \\ (fitness) & n = 30; & R^2 = 0.3673; & \bar{R}^2 = 0.3205 \\ & & F^* = 7.84; & p = 0.0021 \end{array}$$

- 豪斯曼联立性检验的结论：
 - \hat{v}_{1i} 前的系数等于0.34, $t^* = 2.9901 > t(\alpha/2, n - 3) = 2.47$ 。(给定 $\alpha = 0.01$)
 - 因此, \hat{v}_{1i} 前系数的t检验是**极显著的**(给定 $\alpha = 0.01$ 水平下), 从而**拒绝原假设** H_0 (无联立性问题), 接受备择假设 H_1 , 认为**存在联立性问题**。

联立性检验：松露案例（平狄克检验；原始报告2）

如果采用**平狄克检验方程**，则R原始报告结果如下：

```
##
## Call:
## lm(formula = Hausman_models$mod.Q.Pindyck, data = truffles_Hausman)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.476 -2.892  0.277  3.394  5.380
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.946      2.604   4.59 9.2e-05 ***
## P               0.104      0.040   2.60  0.015 *
## hat.vi         0.234      0.120   1.95  0.061 .
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.8 on 27 degrees of freedom
## Multiple R-squared:  0.367,    Adjusted R-squared:  0.32
## F-statistic: 7.84 on 2 and 27 DF,  p-value: 0.00207
```


联立性检验：松露案例（平狄克检验；精简报告2）

如果采用**平狄克检验方程**，可以得到如下精简报告结果：

$$\begin{array}{rcll} \hat{Q} = & + 11.95 & + 0.10P & + 0.23\hat{v}_{1i} \\ (t) & (4.5880) & (2.5952) & (1.9529) \\ (se) & (2.6037) & (0.0400) & (0.1199) \\ (fitness) & n = 30; & R^2 = 0.3673; & \bar{R}^2 = 0.3205 \\ & & F^* = 7.84; & p = 0.0021 \end{array}$$

- 联立性问题检验结论：
 - \hat{v}_{1i} 前的系数等于0.23, $t^* = 1.9529 > t(\alpha/2, n - 3) = 1.7$ 。(给定 $\alpha = 0.1$)
 - 因此, \hat{v}_{1i} 前系数的t检验是**显著的**(给定 $\alpha = 0.1$ 水平下), 从而**拒绝原假设** H_0 (无联立性问题), 接受备择假设 H_1 , 认为**存在联立性问题**

19.4 外生性检验

外生性检验

略

要点与结论

1. 识别问题的考虑应先于估计问题。
2. 识别问题是问我们能否从约简系数估计值求出结构系的唯一数值估计值。
3. 如果能做到，就说联立方程组中的某个方程是可识别的；如果做不到，该方程就是不可识别或识别不足的。
4. 一个可识别的方程可以是恰好识别的或过度识别的。在前一种情形中，可以得到的唯一值；而在后一种情形中，可以得到不止一个估计值。
5. 识别问题之所以出现，是因为同样的数据集适合于不同的 的模型。

要点与结论

1. 要判断一个结构方程的可识别性，我们可以应用约简型方程的技术，把一个内生变量表达为前定变量的函数。
2. 这种耗时的程序由于**阶条件**或**秩条件**的利用而得以避免。阶条件虽然易用应用，但它仅是可识别性的**必要而非充分**条件；秩条件则是可识别性的**充分必要条件**。
3. 当出现联立性问题时，OLS一般而言是不适用的。但如果我们仍想用它，则必须明确地进行联立性检验，为此，可利用豪斯曼内生性检验方法。
4. 虽然在实践中一个变量是内生或外生的，是凭判断而决定的，但我们可以用豪斯曼设定检验判定一个或一组变量是内生的还是外生的。
5. 因果关系和外生性虽属于同一类问题，但它们的概念却是不同的。其中一个概念并不蕴含另一个概念。在实践中，仍然是把这两个概念区分开来为好。

本章结束

