

29.4 空间自回归模型

对于时间序列,最常见的建模方式为一阶自回归 AR(1)。但空间自回归的形式更为复杂,因为空间滞后可以来自不同的方向,而且可以双向。对于空间序列 $\{y_i\}_{i=1}^n$ (样本容量为 n),即使假设空间自相关的形式为线性,待估计的参数在理论上最多可达 $(n^2 - n)$ 个(每个区域最多可受 $n - 1$ 个区域影响,共有 n 个区域),显然大大超出了样本容量许可的范围。为此,必须假设空间依赖性服从某种共同的模式,才能简化参数。

作为参照,首先回顾时间序列的一阶自相关模型:

$$y_t = \beta y_{t-1} + \varepsilon_t \quad (t = 2, \dots, n) \quad (29.18)$$

为了示例方便,假设 $n = 4$,而 $y_1 = \varepsilon_1$ 。将这 4 个方程叠放在一起可得:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \beta \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_W \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \equiv \beta W y + \varepsilon \quad (29.19)$$

其中,矩阵 W 的次主对角线元素均为 1,而其他元素都为 0。矩阵 W 可视为“时间滞后矩阵”,它之所以呈现出次对角线元素为 1 而其他元素为 0 的规律性排列,正是由于时间滞后在方向上的单一性。对于空间序列,矩阵 W 则为空间权重矩阵,其元素排列方式表现出多样性(因为空间自相关可以多方向)。为此,引入以下“空间自回归模型”(Spatial Autoregression,简记 SAR):

$$y = \lambda W y + \varepsilon \quad (29.20)$$

其中, W 为已知的空间权重矩阵(非随机),而空间依赖性仅由单一参数 λ 来刻画(待估参数大大减少)。 λ 度量空间滞后 $W y$ 对 y 的影响,称为“空间自回归系数”(spatial autoregressive parameter)。在形式上,此模型与时间滞后模型(29.19)完全相同,只是空间权重矩阵更为复杂。因此,方程(29.20)也称为“空间滞后模型”(Spatial Lag Model)^①。直观来看,相邻地区的被解释变量(比如犯罪率)可能相互依赖,并最终形成一个均衡的结果。又比如,假设以地区税收为被解释变量,则不同地区的政府出于相互竞争或博弈的考虑(比如,竞相吸引 FDI),在制定本地区税收时,会考虑周边地区的税收水平。

作为一个例子,假设 W 为表达式(29.2),则方程(29.20)可写为:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \lambda \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = \lambda \begin{pmatrix} y_2 + y_3 + y_4 \\ y_1 + y_3 \\ y_1 + y_2 + y_4 \\ y_1 + y_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \quad (29.21)$$

显然,这是一个联立方程系统。如果对(29.21)进行 OLS 估计,将会导致联立方程偏差(simultaneity bias),参见第 10、24 章。正是由于空间依赖性,导致变量 $\{y_i\}_{i=1}^n$ 之间互相影响,从而产生内生性。更一般地,可在方程(29.20)中加入自变量:

^① 也称为“混合回归-空间自回归模型”(mixed regressive-spatial autoregressive model)。

$$y = \lambda W y + X \beta + \varepsilon \quad (29.22)$$

其中, X 为 $n \times k$ 数据矩阵(参见第3章), 包括 k 列解释变量; 而 $\beta_{k \times 1}$ 为相应系数。方程(29.22)也称为 SAR 模型, 而方程(29.20)有时称为“纯 SAR 模型”(Pure SAR)。如果 $\lambda = 0$, 则方程(29.22)简化为一般的线性回归模型。因此, 可通过检验原假设 “ $H_0: \lambda = 0$ ” 来考察是否存在空间效应。对于空间自回归模型(29.22), 常使用最大似然估计。首先, 假设扰动项 $\varepsilon \sim N(\mathbf{0}, \sigma^2 I_n)$ 。其次, 方程(29.22)可写为

$$Ay = (I - \lambda W)y = X\beta + \varepsilon \quad (29.23)$$

其中, $A \equiv I - \lambda W$ 。由于雅可比行列式 $J \equiv \left| \frac{\partial \varepsilon}{\partial y} \right| = \left| \frac{\partial (Ay - X\beta)}{\partial y} \right| = \left| \frac{\partial Ay}{\partial y} \right| = |A'| = |A|$ ^①, 故根据多维正态的密度函数公式(参见第2章), 可写出样本的似然函数:

$$L(y | \lambda, \sigma^2, \beta) = (2\pi\sigma^2)^{-n/2} (\text{abs}|A|) \exp \left\{ -\frac{1}{2\sigma^2} (Ay - X\beta)'(Ay - X\beta) \right\} \quad (29.24)$$

其中, $\text{abs}|A|$ 表示行列式 $|A|$ 的绝对值。由此可得对数似然函数为

$$\ln L(y | \lambda, \sigma^2, \beta) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln(\text{abs}|A|) - \frac{1}{2\sigma^2} (Ay - X\beta)'(Ay - X\beta) \quad (29.25)$$

类似于对古典线性回归模型的 MLE 估计(参见第6章), 此最大化问题可分两步进行。第一步, 在给定 λ 的情况下, 选择最优的 β, σ^2 。第二步, 代入第一步的最优 β, σ^2 , 选择最优的 λ 。

在第一步, 选择 β, σ^2 使得 $\ln L(y | \lambda, \sigma^2, \beta)$ 最大。由于 β 只出现于(29.25)式的最后一项, 故这等价于使 $(Ay - X\beta)'(Ay - X\beta)$ 最小, 即 Ay 对 X 进行回归:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X' Ay = (X'X)^{-1} X' (I - \lambda W) y \\ &= (X'X)^{-1} X' y - \lambda (X'X)^{-1} X' Wy \equiv \hat{\beta}_0 - \lambda \hat{\beta}_L \end{aligned} \quad (29.26)$$

其中, $\hat{\beta}_0 \equiv (X'X)^{-1} X' y$ (y 对 X 的回归系数), 而 $\hat{\beta}_L \equiv (X'X)^{-1} X' Wy$ (Wy 对 X 的回归系数)。因此, 只要知道 λ , 即可计算 $\hat{\beta}$ 。在方程(29.25)中, 对 σ^2 求偏导可得 σ^2 的 MLE 估计:

$$\hat{\sigma}^2 = \frac{\mathbf{e}' \mathbf{e}}{n} = \frac{(M_x Ay)'(M_x Ay)}{n} \quad (29.27)$$

其中, \mathbf{e} 为 Ay 对 X 回归的残差向量, $M_x \equiv I_n - X(X'X)^{-1}X'$ 为消灭矩阵(参见第3章)。由于 $Ay = (I - \lambda W)y$, 故

$$\mathbf{e} = M_x Ay = M_x(I - \lambda W)y = M_x y - \lambda M_x Wy \equiv \mathbf{e}_0 - \lambda \mathbf{e}_L \quad (29.28)$$

其中, $\mathbf{e}_0 \equiv M_x y$ (y 对 X 的回归残差), 而 $\mathbf{e}_L \equiv M_x Wy$ (Wy 对 X 的回归残差)。将表达式(29.28)代入(29.27)可得:

$$\hat{\sigma}^2 = \frac{\mathbf{e}' \mathbf{e}}{n} = \frac{(\mathbf{e}_0 - \lambda \mathbf{e}_L)'(\mathbf{e}_0 - \lambda \mathbf{e}_L)}{n} \quad (29.29)$$

因此, 只要知道 λ , 也可以计算 $\hat{\sigma}^2$ 。在第二步, 将 $\hat{\beta}(\lambda), \hat{\sigma}^2(\lambda)$ 代入对数似然函数(29.25), 则可得到“集中对数似然函数”(concentrated log likelihood function), 它只是 λ 的函数。然而, 由于 λ 出现在行列式 $|A| = |I - \lambda W|$, 给计算带来不便。为此, 可利用等式 $|A| = \prod_{i=1}^n (1 - \lambda v_i)$ 来

^① 此处的问题是根据 ε 的概率密度写出 y 的概率密度。在此变换时, 需用到雅可比行列式, 参见 Greene (2012, Appendix B.7.4), 或较为深入的概率统计教材。

计算,其中 v_1, \dots, v_n 为矩阵 A 特征值(eigenvalues)。另外,为了保证扰动项协方差矩阵为正定,还须限制 λ 的取值为 $\frac{1}{v_{\min}} < \lambda < \frac{1}{v_{\max}}$, 其中 v_{\min} 与 v_{\max} 分别为矩阵 A 的最小与最大特征值,而 v_{\min} 一定为负数。

对于 MLE 估计量的渐近协方差矩阵,一般通过信息矩阵来估计,即

$$I(\boldsymbol{\theta})^{-1} \equiv -\left\{ E\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]\right\}^{-1} \quad (29.30)$$

其中, $\boldsymbol{\theta} = (\lambda, \sigma^2, \boldsymbol{\beta})$ 。需要注意的是,对于空间自回归模型(29.22) $y = \lambda W y + X \boldsymbol{\beta} + \boldsymbol{\epsilon}$, 解释变量 X 对 y 的边际效应并非 $\boldsymbol{\beta}$, 因为 X 对 y 产生作用后, y 之间还会相互作用(即不同区域的被解释变量互相影响), 直至达到一个新的均衡。由方程(29.23)可知:

$$y = (I - \lambda W)^{-1} X \boldsymbol{\beta} + (I - \lambda W)^{-1} \boldsymbol{\epsilon} \quad (29.31)$$

容易验证, $(I - \lambda W)^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + \dots$ 。假设 X 中包含 K 个解释变量, 并记第 r 个解释变量为 $x_r = (x_{1r}, x_{2r}, \dots, x_{nr})'$ ($n \times 1$ 列向量), 则 $X \boldsymbol{\beta} = (x_1 \cdots x_K)(\beta_1 \cdots \beta_K)' = \sum_{r=1}^K \beta_r x_r$ 。因此, 方程(29.31)可以写为

$$y = \sum_{r=1}^K \beta_r (I - \lambda W)^{-1} x_r + (I - \lambda W)^{-1} \boldsymbol{\epsilon} \equiv \sum_{r=1}^K S_r(W) x_r + (I - \lambda W)^{-1} \boldsymbol{\epsilon} \quad (29.32)$$

其中, $S_r(W) \equiv \beta_r (I - \lambda W)^{-1}$ 为依赖于 β_r 与 W 的 $n \times n$ 矩阵。将方程(29.32)展开来写:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} S_r(W)_{11} & S_r(W)_{12} & \cdots & S_r(W)_{1n} \\ S_r(W)_{21} & S_r(W)_{22} & \cdots & S_r(W)_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_r(W)_{n1} & S_r(W)_{n2} & \cdots & S_r(W)_{nn} \end{pmatrix} \begin{pmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{pmatrix} + (I - \lambda W)^{-1} \boldsymbol{\epsilon} \quad (29.33)$$

其中, $S_r(W)_{ij}$ 为 $S_r(W)$ 的 (i, j) 元素。根据方程(29.33)可知:

$$\frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij} \quad (29.34)$$

由此可见, 区域 j 的变量 x_{jr} 对任意区域 i 的被解释变量都可能有影响, 这正是空间计量模型的真谛。特别地, 如果 $j = i$, 则有

$$\frac{\partial y_i}{\partial x_{ii}} = S_r(W)_{ii} \quad (29.35)$$

方程(29.35)表明, 区域 i 的变量 x_{ii} 对本区域被解释变量 y_i 的“直接效应”(direct effect)为 $S_r(W)_{ii}$, 即矩阵 $S_r(W)$ 主对角线上的第 i 个元素。因此, 如果将矩阵 $S_r(W)$ 主对角线上的所有元素进行平均, 即可得到变量 x_r 的“平均直接效应”(average direct effect):

$$\text{平均直接效应} = \frac{1}{n} \text{trace}[S_r(W)] \quad (29.36)$$

其中, $\text{trace}[S_r(W)]$ 为矩阵 $S_r(W)$ 的迹, 即主对角线元素之和。另一方面, 假设所有区域的变量 x_r 都变化一个单位, 其对区域 i 被解释变量 y_i 的“总效应”(total effect)为矩阵 $S_r(W)$ 的第 i 行元素之和, 即 $\sum_{j=1}^n S_r(W)_{ij}$ 。如果将对所有区域的总效应进行平均, 则可得到变量 x_r 的“平均总效应”(average total effect), 即矩阵 $S_r(W)$ 所有元素的平均:

$$\text{平均总效应} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n S_r(W)_{ij} = \frac{1}{n} i'_n S_r(W) i_n \quad (29.37)$$

其中, $i_n = (1 \cdots 1)'$ 为 $n \times 1$ 列向量。最后, 可以定义变量 x_i 的“平均间接效应”(average indirect effect)为平均总效应与平均直接效应之差:

$$\text{平均间接效应} = \frac{1}{n} \{ i_n' S_r(\mathbf{W}) i_n - \text{trace}[S_r(\mathbf{W})] \} \quad (29.38)$$

29.5 空间杜宾模型

对于空间效应的另一建模方式是, 假设区域 i 的被解释变量 y_i 依赖于其邻居的自变量:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \quad (29.39)$$

其中, $\mathbf{W}\mathbf{X}\boldsymbol{\delta}$ 表示来自邻居自变量的影响, 而 $\boldsymbol{\delta}$ 为相应的系数向量。例如, 区域 i 的犯罪率不仅依赖于本区域的警力, 还可能依赖于相邻区域的警力。此模型称为“空间杜宾模型”(Spatial Durbin Model), 简记 SDM。由于方程(29.39)不存在内生性, 故可直接进行 OLS 估计; 只是解释变量 \mathbf{X} 与 $\mathbf{W}\mathbf{X}$ 之间可能存在多重共线性。如果 $\boldsymbol{\delta} = \mathbf{0}$, 则方程(29.39)简化为一般的线性回归模型。

将空间杜宾模型与空间自回归模型相结合, 可得

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \quad (29.40)$$

方程(29.40)有时也称为空间杜宾模型。

29.6 空间误差模型

空间依赖性还可能通过误差项来体现。考虑以下“空间误差模型”(Spatial Errors Model, 简记 SEM):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (29.41)$$

其中, 扰动项 \mathbf{u} 的生成过程为

$$\mathbf{u} = \rho \mathbf{Mu} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (29.42)$$

其中, \mathbf{M} 为空间权重矩阵。该模型显示, 扰动项 \mathbf{u} 存在空间依赖性。这意味着, 不包含在 \mathbf{X} 中但对 \mathbf{y} 有影响的遗漏变量存在空间相关性, 或者不可观测的随机冲击存在空间相关性。如果 $\rho = 0$, 则简化为一般的线性回归模型。由方程(29.42)可得:

$$\mathbf{Bu} = (\mathbf{I} - \rho \mathbf{M})\mathbf{u} = \boldsymbol{\varepsilon} \quad (29.43)$$

其中, $\mathbf{B} = \mathbf{I} - \rho \mathbf{M}$ 。对于 SEM 模型(29.41), 尽管扰动项存在自相关, 但由于不存在内生性, 故 OLS 估计是一致的; 但由于忽略了扰动项的自相关而损失了效率(参见第 8 章)。最有效的方法为 MLE 估计。样本的对数似然函数可写为

$$\ln L(\mathbf{y} | \rho, \sigma^2, \boldsymbol{\beta}) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln(\text{abs}|\mathbf{B}|) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (29.44)$$

在最大化上式时, 仍可分两步进行。第一步, 通过 GLS 计算 $\hat{\boldsymbol{\beta}}$, 即最小化上式最后一项 $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$; 然后通过一阶条件得到 $\hat{\sigma}^2$ 。第二步, 将 $\hat{\boldsymbol{\beta}}(\rho), \hat{\sigma}^2(\rho)$ 代入对数似然函数(29.44), 得到集中对数似然函数(仅为 ρ 的函数), 求解最优的 $\hat{\rho}$ 。一个技术性问题是, 在第一步

中,由于目标函数涉及 $\mathbf{B} = \mathbf{I} - \rho \mathbf{M}$,故得不到 $\hat{\boldsymbol{\beta}}(\rho)$ 与 $\hat{\sigma}^2(\rho)$ 的解析表达式。为此,需要在第一步与第二步之间进行迭代,即给定 $\hat{\rho}^{(1)}$,通过第一步得到 $\hat{\boldsymbol{\beta}}(\hat{\rho}^{(1)})$, $\hat{\sigma}^2(\hat{\rho}^{(1)})$ 的具体取值,再通过第二步得到 $\hat{\rho}^{(2)}$,以此类推,直至 $\hat{\rho}$ 收敛。其中,初始值 $\hat{\rho}^{(1)}$ 可通过在第一步进行 OLS 估计,然后代入第二步而得到。

继续以数据集 columbusdata.dta 为例,演示空间自回归与空间误差模型的 Stata 操作。首先,通过命令 spatdiag 诊断是否存在空间效应。为此,须先进行 OLS 回归。

```
. reg crime hoval income
```

Source	SS	df	MS	Number of obs	=	49
Model	7423.32674	2	3711.66337	F(2, 46)	=	28.39
Residual	6014.89281	46	130.758539	Prob > F	=	0.0000
Total	13438.2195	48	279.962907	R-squared	=	0.5524
				Adj R-squared	=	0.5329
				Root MSE	=	11.435

crime	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
hoval	-0.2739315	0.1031987	-2.65	0.011	-.4816597 -.0662033
income	-1.597311	0.3341308	-4.78	0.000	-2.269881 -.9247405
_cons	68.61896	4.735486	14.49	0.000	59.08692 78.151

上表显示,房价(hoval)与家庭收入(income)对于犯罪率均有显著的负作用。然而,如果存在空间效应,则 OLS 估计是有偏差的。

```
. spatdiag, weights(W)
```

其中,必选项“weights(W)”指定空间权重矩阵为 W。有关如何生成此空间权重矩阵,参见本章第3节。

Diagnostic tests for spatial dependence in OLS regression

Fitted model

crime = hoval + income

Weights matrix

Name: W
Type: Imported (binary)
Row-standardized: No

Diagnostics

Test	Statistic	df	p-value
Spatial error:			
Moran's I	1.011	1	0.312
Lagrange multiplier	6.804	1	0.009
Robust Lagrange multiplier	1.759	1	0.185
Spatial lag:			
Lagrange multiplier	13.787	1	0.000
Robust Lagrange multiplier	8.741	1	0.003

上表显示,针对空间误差(Spatial error)的三个检验中,有一个拒绝了“无空间自相关”的原假设;而针对空间滞后(Spatial lag)的两个检验均拒绝了此原假设;这些结果再次表明应进行空间计量分析。

下面,使用命令 `spatreg` 估计空间滞后与空间误差模型。为了运行命令 `spatreg`,除了需要空间权重矩阵 W 外,还需要先计算矩阵 W 的特征值向量。为此,再次使用命令 `spatwmat`。

```
. spatwmat using columbuswm.dta, name(W) eigenval(E)
```

其中,选择项“`eigenval(E)`”表示,计算空间权重矩阵 W 的特征值向量,并记为 E 。

```
The following matrices have been created:
```

1. Imported binary weights matrix W
Dimension: 49x49
2. Eigenvalues matrix E
Dimension: 49x1

上表显示,特征值向量 E 为 49×1 的列向量。

```
. spatreg crime hoval income, weights(W) eigenval(E) model(lag) nolog
```

其中,必选项“`weights(W)`”与“`eigenval(E)`”分别用于指定空间权重矩阵及其特征值向量。选择项“`model(lag)`”表示估计空间自回归模型(SAR)。另外,还可使用选择项“`robust`”得到异方差稳健的标准误^①。

Weights matrix						
Name: W						
Type: Imported (binary)						
Row-standardized: No						
Spatial lag model						
					Number of obs =	49
					Variance ratio =	0.660
					Squared corr. =	0.660
Log likelihood = -180.65071					Sigma =	9.66
crime	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
crime						
hoval	-.2516329	.087319	-2.88	0.004	-.422775	-.0804909
income	-1.154463	.3037289	-3.80	0.000	-1.749761	-.5591653
_cons	51.60482	5.888379	8.76	0.000	40.06381	63.14583
rho	.0545427	.0138575	3.94	0.000	.0273825	.0817029
Wald test of rho=0: chi2(1) = 15.492 (0.000)						
Likelihood ratio test of rho=0: chi2(1) = 13.453 (0.000)						
Lagrange multiplier test of rho=0: chi2(1) = 13.787 (0.000)						
Acceptable range for rho: -1.536 < rho < 1.000						

上表显示,空间自回归系数(记为 ρ)的估计值为 0.055,且在 1% 水平上显著,故存在空间自回归效应。对于空间自回归系数的似然比检验与拉格朗日检验也得到同样的结果。SAR 模型的自变量回归系数与 OLS 模型有所不同。下面估计空间误差模型。

^① 然而,如果存在异方差,则 MLE 估计量不一致;使用异方差稳健的标准误也就意义不大,参见第 6 章。故此处未使用“`robust`”选择项。

```
. spatreg crime hoval income, weights (W) eigenval (E) model (error)
nolog
```

其中,选择项“model(error)”表示估计空间误差模型(SEM)。

Weights matrix						
Name: W						
Type: Imported (binary)						
Row-standardized: No						
Spatial error model						
Number of obs = 49						
Variance ratio = 0.402						
Squared corr. = 0.551						
Log likelihood = -182.26823						
Sigma = 9.98						
crime	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
crime						
hoval	-.2605768	.0917459	-2.84	0.005	-.4403954	-.0807582
income	-1.289794	.3255491	-3.96	0.000	-1.927858	-.6517291
_cons	54.45472	6.149929	8.85	0.000	42.40108	66.50836
lambda	.0361054	.0114345	3.16	0.002	.0136943	.0585165
Wald test of lambda=0: chi2(1) = 9.970 (0.002)						
Likelihood ratio test of lambda=0: chi2(1) = 10.218 (0.001)						
Lagrange multiplier test of lambda=0: chi2(1) = 6.804 (0.009)						
Acceptable range for lambda: -1.536 < lambda < 1.000						

上表显示,误差项的空间自回归系数(记为 λ)的估计值为 0.036,且在 1% 的水平上显著。对于 λ 的似然比检验与拉格朗日检验也得到同样的结果。可以看出,SEM 模型对于自变量回归系数的估计结果与 OLS 十分接近。这是因为,如果 SEM 模型正确,则 OLS 也是一致估计量,只不过 MLE 估计量更有效。

29.7 一般的空间计量模型

以上两节分别考虑了空间滞后与空间误差模型,这两种效应也可能同时发生。更为一般的空间计量模型将空间自回归模型(SAR)与空间误差模型(SEM)结合起来。考虑以下模型:

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (29.45)$$

其中,扰动项 \mathbf{u} 的生成过程为

$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (29.46)$$

其中, \mathbf{W} 与 \mathbf{M} 分别为被解释变量 \mathbf{y} 与扰动项 \mathbf{u} 的空间权重矩阵,二者可以相等。由方程(29.45)与(29.46)构成的模型称为“带空间自回归误差项的空间自回归模型”(Spatial Autoregressive Model with Spatial Autoregressive Disturbances,简记 SARAR)。SAR 模型,SEM 模型以及 SARAR 模型,也统称为“Cliff – Ord 模型”(Cliff and Ord, 1973, 1981; Ord, 1975)。

显然,SAR 模型与 SEM 模型都是 SARAR 模型的特例,分别对应于 $\rho=0$ 与 $\lambda=0$ 的情形。对于 SARAR 模型,也可进行 MLE 估计。可以证明,样本的对数似然函数为

$$\begin{aligned}\ln L(\mathbf{y} | \lambda, \rho, \sigma^2, \boldsymbol{\beta}) = & -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln(\text{abs}|\mathbf{A}|) + \ln(\text{abs}|\mathbf{B}|) \\ & - \frac{1}{2\sigma^2} (\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})\end{aligned}\quad (29.47)$$

在最大化上式时,仍可分两步进行,并进行迭代,与 SEM 模型的情形类似。虽然 Cliff – Ord 模型常用 MLE 进行估计,但 MLE 估计量存在三个缺陷。首先,MLE 估计量可能计算不方便,特别当空间权重矩阵 \mathbf{W} 与 \mathbf{M} 的维度较高时。其次,空间计量模型 MLE 估计量的大样本理论尚不健全。在独立同分布的假设下(但不一定服从正态分布),Lee(2004)导出了 SAR 模型的 QMLE 渐近分布,但未涉及 SEM 模型或更一般的 SARAR 模型^①。再次,如果扰动项 $\boldsymbol{\epsilon}$ 不服从独立同分布的正态分布,则 QMLE 估计量可能不一致。比如,Arraiz et al(2010)通过蒙特卡罗模拟证明,在异方差的情况下,QMLE 估计量是不一致的。

为此,Kelejian and Prucha(1999, 1998, 2004, 2010)提出利用工具变量,通过 GMM 来估计 SARAR 模型,称为“广义空间二段最小二乘法”(Generalized Spatial Two-stage Least Square, 简记 GS2SLS)。Arraiz et al(2010)的模拟结果表明,此 IV 估计量在异方差情况下也是一致的;故更为稳健。而且,GS2SLS 的计算更为方便,并不会因为样本容量太大而产生困难。

GS2SLS 所使用的工具变量为 $\{\mathbf{X}, \mathbf{WX}, \dots, \mathbf{W}^q \mathbf{X}, \mathbf{MX}, \mathbf{MWX}, \dots, \mathbf{MW}^q \mathbf{X}\}$ 中线性独立的列向量,记此工具变量矩阵(matrix of instrumental variables)为 \mathbf{H} 。蒙特卡罗模拟显示,在很多情况下取 $q=2$ 即可。显然,由于 \mathbf{X} 与 \mathbf{y} 相关,而且 \mathbf{X} 为外生变量;故以上 \mathbf{X} 的线性函数为有效工具变量。GS2SLS 法通过以下四个步骤来实现。

第一步,使用工具变量矩阵 \mathbf{H} 对方程(29.45)进行 2SLS 估计。所得估计量 $(\tilde{\lambda}, \tilde{\boldsymbol{\beta}})$ 虽然是致的,但并非最有效,因为它没有考虑扰动项 \mathbf{u} 存在空间自回归。记所得残差为 $\tilde{\mathbf{u}}$ 。

第二步,将第一步的 2SLS 残差 $\tilde{\mathbf{u}}$ 作为 \mathbf{u} 的估计量代入方程(29.46),然后进行 GMM 估计,得到估计量 $\tilde{\rho}$ 。

第三步,使用 $\tilde{\rho}$ 对方程(29.45)进行“空间 Cochrane – Orcutt 变换”(spatial Cochrane – Orcutt transformation),以去掉扰动项的空间自相关。对于时间序列 $\{y_t\}$,如果其扰动项存在一阶自回归,比如 $u_t = \rho u_{t-1} + \varepsilon_t$,则可通过做准差分(即 Cochrane – Orcutt 变换) $(y_t - \rho y_{t-1})$ 来消除此自相关。对于空间序列,也可以做类似的变换。将方程(29.45)两边同时左乘 $(\mathbf{I} - \rho \mathbf{M})$ 可得:

$$(\mathbf{I} - \rho \mathbf{M}) \mathbf{y} = \lambda (\mathbf{I} - \rho \mathbf{M}) \mathbf{W} \mathbf{y} + (\mathbf{I} - \rho \mathbf{M}) \mathbf{X} \boldsymbol{\beta} + \underbrace{(\mathbf{I} - \rho \mathbf{M}) \mathbf{u}}_{=\boldsymbol{\epsilon}} \quad (29.48)$$

其中,根据方程(29.46)可知,新扰动项 $\boldsymbol{\epsilon} = (\mathbf{I} - \rho \mathbf{M}) \mathbf{u}$ 不再存在空间自相关。将第二步估计量 $\tilde{\rho}$ 代入方程(29.48),再次使用工具变量 \mathbf{H} 对变换后的方程(29.48)进行 2SLS 估计,记所得估计量为 $(\hat{\lambda}, \hat{\boldsymbol{\beta}})$,其相应残差为 $\hat{\mathbf{u}}$ 。

第四步,将第三步的 GS2SLS 残差 $\hat{\mathbf{u}}$ 作为 \mathbf{u} 的估计量代入方程(29.46),然后进行 GMM 估

^① 空间计量模型的大样本理论更为复杂。主要困难在于,当样本容量 $n \rightarrow \infty$ 时,随着横截面单位增多,空间权重矩阵 \mathbf{W}_n 也必然随之改变(此处使用下标 n 强调 \mathbf{W}_n 对 n 的依赖性)。为此,必须对 \mathbf{W}_n 的渐近行为进行限制。特别地, $n \rightarrow \infty$ 可通过两种方式来实现,第一种方式通过增加空间规模来增加区域个数,称为“increasing-domain asymptotic”;而第二种方式在给定空间内通过不断细分区域来增加区域个数,称为“infill asymptotic”或“fixed-domain asymptotic”。Lee(2004)表明,在“infill asymptotic”的情形下,SAR 模型的 MLE 估计量有可能不一致。

计,得到估计量 $\hat{\rho}$ 。

根据以上四个步骤所得估计量 $(\hat{\lambda}, \hat{\rho}, \hat{\beta})$ 即为 GS2SLS 估计量。由于该估计量将 2SLS 用于空间数据,而且使用了广义最小二乘法(Cochrane – Orcutt 变换是 GLS 的特例),故称为“广义空间二段最小二乘法”。GS2SLS 法的优点是计算方便且结果稳健(在异方差情况下也成立),但缺点是必须有外生变量 X ,故不适用于没有外生变量的纯 SAR 模型(参见方程(29.20))。

SARAR 模型可通过非官方 Stata 命令 spreg 来实现^①,其下载安装方法为(或通过命令“findit sppack”查找并下载)

```
. net install sppack.pkg
```

此命令将下载一套有关空间计量的命令,由以下三个命令所组成:命令 spmat 用于定义与管理空间权重矩阵,命令 spreg 用于估计 SARAR 模型,命令 spivreg 用于估计含内生解释变量的 SARAR 模型(参见下节)。

继续以数据集 columbusdata.dta 与 columbuswm.dta 为例。首先用命令 spmat 定义空间权重矩阵。

```
. use columbuswm.dta, clear
```

此时,在变量窗口将看到 49 个变量,分别为 a1—a49(即为 49×49 权重矩阵的 49 列)。

```
. spmat dta w1 a1 - a49
```

其中,子命令“dta”表示从 Stata 格式的数据集(扩展名为 dta)来生成空间权重矩阵。上述命令表示根据内存中的变量 a1 – a49 来生成空间权重矩阵 w1。更多有关命令 spmat 的用法说明,参见“help spmat”。

```
. use columbusdata.dta, clear
```

```
. spreg ml crime hoval income, id(id) dpmat(w1) elmat(w1) nolog
```

其中,“ml”表示使用 MLE 估计 SARAR 模型。必选项“id(id)”表示用来确定横截面单位的变量为 id。选择项“dpmat(w1)”表示以 w1 作为空间自回归方程(29.45)中的空间权重矩阵(即上文的 \mathbf{W}),而选择项“elmat(w1)”则表示以 w1 作为空间误差方程(29.46)中的空间权重矩阵(即上文的 \mathbf{M})。

Spatial autoregressive model (Maximum likelihood estimates)		Number of obs = 49				
		Wald chi2(2) = 37.3387				
		Prob > chi2 = 0.0000				
crime	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
crime						
hoval	-.2555749	.0947579	-2.70	0.007	-.4412968	-.0698529
income	-1.175428	.3037636	-3.87	0.000	-1.770794	-.5800625
_cons	52.74999	7.156856	7.37	0.000	38.72281	66.77716
lambda						
_cons	.0507629	.0204347	2.48	0.013	.0107117	.0908142
rho						
_cons	.0061692	.0762303	0.08	0.935	-.1432395	.1555779
sigma2						
_cons	93.32433	18.87086	4.95	0.000	56.33813	.130.3105

^① 此命令有希望日后成为 Stata 的官方命令,因为其作者之一 David Drukker 为 Stata 公司(Stata Corp.)的计量经济学主管(Director of Econometrics)。

上表显示,空间自回归系数 $\lambda = 0.051$,且在 5% 水平上显著。而空间误差系数 ρ 则很不显著,其 p 值高达 0.935。下面以 GS2SLS 法进行估计。

. spreg gs2sls crime hoval income, id(id) dlmat(w1) elmat(w1) het nolog
其中,子命令“gs2sls”表示使用 GS2SLS 法;选择项“het”表示使用异方差稳健的标准误(默认标准误公式假设同方差)。

Spatial autoregressive model (GS2SLS estimates)						Number of obs = 49
	crime	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
crime	hoval	-.2226818	.1531299	-1.45	0.146	-.5228109 .0774472
	income	-1.153734	.4963804	-2.32	0.020	-2.126622 -.1808466
	_cons	48.29302	7.99936	6.04	0.000	32.61457 63.97148
lambda	_cons	.0652404	.016342	3.99	0.000	.0332105 .0972702
rho	_cons	-.0303691	.0759236	-0.40	0.689	-.1791768 .1184385

上表显示,GS2SLS 与 MLE 的估计结果十分类似。由于空间误差系数 ρ 很不显著,下面估计不含空间误差的 SAR 模型。

. spreg gs2sls crime hoval income, id(id) dlmat(w1) het nolog

Spatial autoregressive model (GS2SLS estimates)						Number of obs = 49
	crime	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
crime	hoval	-.2490946	.1596383	-1.56	0.119	-.5619799 .0637907
	income	-1.104052	.5017371	-2.20	0.028	-2.087438 -.1206653
	_cons	49.66803	8.125303	6.11	0.000	33.74273 65.59333
lambda	_cons	.0607515	.0170531	3.56	0.000	.0273281 .0941749

上表显示,SAR 模型的估计结果类似于 SARAR 模型。

29.8 含内生解释变量的 SARAR 模型

使用 GS2SLS 法估计 SARAR 模型的一个优点是,它可以很容易地推广到包含内生解释变量的情形:

$$y = \lambda W y + X \beta + Z \delta + u \quad (29.49)$$

其中, Z 由内生解释变量所组成,而扰动项 u 的生成过程依然为

$$\mathbf{u} = \rho \mathbf{M} \mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (29.50)$$

假设在 \mathbf{X} 之外, 尚有不在方程中的外生变量 \mathbf{X}_e 。定义 $\mathbf{X}_f \equiv (\mathbf{X}, \mathbf{X}_e)$; 如果没有不在方程中的外生变量 \mathbf{X}_e , 则 $\mathbf{X}_f \equiv \mathbf{X}$ 。定义工具变量矩阵 \mathbf{H} 为 $\{\mathbf{X}_f, \mathbf{W}\mathbf{X}_f, \dots, \mathbf{W}^q\mathbf{X}_f, \mathbf{M}\mathbf{X}_f, \mathbf{M}\mathbf{W}\mathbf{X}_f, \dots, \mathbf{M}\mathbf{W}^q\mathbf{X}_f\}$ 中线性独立的列向量, 即可进行 GS2SLS 估计。

延续上节的例子。纯粹为了演示的目的, 假设 income 为内生变量, 而 hoval 为其工具变量, 然后使用命令 spivreg 进行估计。

```
. spivreg crime(income = hoval), id(id) dlm1(w1) elmat(w1) het nolog
```

其中, 选择项“het”表示使用异方差稳健的标准误。

Spatial autoregressive model (GS2SLS estimates)							Number of obs = 49
	crime	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
crime	income	-2.654373	.5777431	-4.59	0.000	-3.786729	-1.522018
	_cons	66.64373	10.06046	6.62	0.000	46.92558	86.36187
lambda	_cons	.0370156	.0152704	2.42	0.015	.0070862	.0669451
rho	_cons	.013882	.0437001	0.32	0.751	-.0717686	.0995325
Instrumented: income Instruments: hoval							

上表显示, 空间自回归系数在 5% 水平上显著, 而空间误差系数很不显著。

29.9 空间面板模型

横截面数据的空间自回归模型与空间误差模型可以推广到面板数据中。首先考察面板的空间自回归模型:

$$\mathbf{y}_{it} = \rho \mathbf{w}'_i \mathbf{y}_i + \mathbf{x}'_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it} \quad (i=1, \dots, n; t=1, \dots, T) \quad (29.51)$$

其中, \mathbf{w}'_i 为空间权重矩阵 \mathbf{W} 的第 i 行, $\mathbf{w}'_i \mathbf{y}_t = \sum_{j=1}^n w_{ij} y_{jt}$, w_{ij} 为空间权重矩阵 \mathbf{W} 的 (i, j) 元素; 而 u_i 为区域 i 的个体效应。如果不考虑空间滞后项 $\rho \mathbf{w}'_i \mathbf{y}_t$, 则方程(29.51)为标准的静态面板模型。进一步, 如果 u_i 与 \mathbf{x}_{it} 相关, 则为固定效应模型; 反之, 则为随机效应模型。

对于固定效应的空间自回归模型, 可以先做组内离差变换(参见第 15 章), 去掉个体效应 u_i ; 然后再使用类似于横截面空间自回归模型的 MLE 估计。对于随机效应的空间自回归模型, 则可先做广义离差变换(参见第 15 章), 然后再进行 MLE 估计。具体推导参见 Elhorst(2010) 与 Lee and Yu(2010)。在决定使用固定效应还是随机效应模型时, 可进行通常的豪斯曼检验。

对于空间面板模型的 MLE 估计, 可通过非官方命令 xsmle 来实现。此命令可估计很一般的空间面板模型:

$$\begin{cases} \mathbf{y}_{it} = \tau \mathbf{y}_{i,t-1} + \rho \mathbf{w}'_i \mathbf{y}_i + \mathbf{x}'_{it} \boldsymbol{\beta} + \mathbf{d}'_i \mathbf{X}_i \boldsymbol{\delta} + u_i + \gamma_t + \varepsilon_{it} \\ \varepsilon_{it} = \lambda \mathbf{m}'_i \boldsymbol{\varepsilon}_i + v_{it} \end{cases} \quad (29.52)$$

其中, $y_{i,t-1}$ 为被解释变量 y_i 的一阶滞后(即动态面板;如果不是动态面板,可令 $\tau=0$); $d'_i X_i \delta$ 表示解释变量的空间滞后, d'_i 为相应空间权重矩阵 D 的第 i 行; γ_i 为时间效应;而 m'_i 为扰动项空间权重矩阵 M 的第 i 行。

由于模型(29.52)太一般,故通常考虑以下特殊情形(每种情形均以命令 xsmle 所使用的术语来命名)。

- (1) 如果 $\lambda=0$, 则为“空间杜宾模型”(Spatial Durbin Model, 简记 SDM)。
- (2) 如果 $\lambda=0$ 且 $\delta=\mathbf{0}$, 则为“空间自回归模型”(Spatial Autoregression Model, 简记 SAR)。
- (3) 如果 $\tau=0$ 且 $\delta=\mathbf{0}$, 则为“空间自相关模型”(Spatial Autocorrelation Model, 简记 SAC); 即上文的 SARAR 模型。
- (4) 如果 $\tau=\rho=0$ 且 $\delta=\mathbf{0}$, 则为“空间误差模型”(Spatial Error Model, 简记 SEM)。

命令 xsmle 的句式为

```
xsmle y x1 x2 x3, wmat (name) emat (name) dmat (name) durbin (varlist)
model (sdm) model (sac) model (sem) re fe robust dlag type (ind) type (time)
type (both) noeffects
```

其中,选择项“wmat (name)”表示空间权重矩阵 W ,选择项“emat (name)”表示扰动项空间权重矩阵 M ,选择项“dmat (name)”表示解释变量的空间权重矩阵 D (默认使用选择项“wmat (name)”所指定的空间权重矩阵);选择项“durbin (varlist)”用于指定空间滞后的解释变量(默认为所有解释变量);选择项“model (sar)”表示估计 SAR 模型(默认选择),选择项“model (sdm)”表示估计 SDM 模型,选择项“model (sac)”表示估计 SAC 模型,选择项“model (sem)”表示估计 SEM 模型;选择项“re”表示估计随机效应模型(默认选择),选择项“fe”表示估计固定效应模型;选择项“robust”表示使用聚类稳健的标准误;选择项“dlag”表示加入被解释变量的一阶滞后作为解释变量(即动态面板),默认为静态面板(即 $\tau=0$)。选择项“type (ind)”表示只有个体效应 u_i (默认选择),选择项“type (time)”表示只有时间效应 γ_t ,而选择项“type (both)”表示既有个体效应,也有时间效应。选择项“noeffects”表示不显示直接效应、间接效应与总效应。

下面以命令 xsmle 附带的数据集 product.dta 以及空间权重矩阵 usaww.spmat 为例,演示 SDM 模型的估计(SAR 模型是 SDM 模型的特例)。该数据集最初为 Munnell(1990) 所使用,包含美国 48 个州 1970–1974 年的如下变量:gsp(州产出,gross state product),pcap(公共资本,public capital),pc(民营资本,private capital),emp(就业量,employment),unemp(失业率,unemployment rate)。

首先下载非官方命令 xsmle^①:

```
. net install xsmle, all from(http://www.econometrics.it/stata)
```

运行此命令后,数据集 product.dta 以及空间权重矩阵 usaww.spmat 将被存入当前工作路径(path of working directory)。

```
. use product.dta, clear
```

由于产出、资本与就业量存在指数增长趋势,故首先将这些变量取对数。

```
. gen lngsp = log(gsp)
. gen lnpcap = log(pcap)
. gen lnpc = log(pc)
```

^① 如果通过命令“ssc install xsmle”下载此命令,则无法下载作为例子的两个附带文件 product.dta 与 usaww.spmat。

```

. gen lnemp = log(emp)
. spmat use usaww using usaww.spmat
此命令表示根据文件“usaww.spmat”来生成空间权重矩阵“usaww”，参见“help spmat”。首先使用聚类稳健的标准误估计随机效应的SDM模型。
. xsmle lngsp lnpcap lnpc lnemp unemp, wmat (usaww) model(sdm) robust nolog

```

SDM with random-effects		Number of obs = 816									
Group variable: state		Number of groups = 48									
Time variable: year		Panel length = 17									
R-sq: within = 0.9460 between = 0.9914 overall = 0.9904											
Log-pseudolikelihood = 1504.9528 (Std. Err. adjusted for 48 clusters in state)											
lngsp	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]						
Main											
lnpcap	.0432197	.0456273	0.95	0.344	-.0462082	.1326476					
lnpc	.2358733	.0818952	2.88	0.004	.0753617	.3963848					
lnemp	.7460606	.0796743	9.36	0.000	.5899019	.9022193					
unemp	-.0012839	.0029046	-0.44	0.658	-.0069768	.004409					
_cons	1.160517	.480234	2.42	0.016	.2192755	2.101758					
Wx											
lnpcap	-.0937203	.0885386	-1.06	0.290	-.2672529	.0798123					
lnpc	-.0215075	.0728551	-0.30	0.768	-.1643009	.121286					
lnemp	-.3895897	.0946901	-4.11	0.000	-.5751789	-.2040006					
unemp	-.0036868	.0038042	-0.97	0.332	-.0111428	.0037692					
Spatial											
rho	.4873069	.071246	6.84	0.000	.3476672	.6269466					
Variance											
lgt_theta	-2.345213	.2519923	-9.31	0.000	-2.839109	-1.851317					
sigma_e	.0010299	.0002446	4.21	0.000	.0005504	.0015094					
Direct											
lnpcap	.0309365	.0429702	0.72	0.472	-.0532834	.1151565					
lnpc	.2540086	.0860114	2.95	0.003	.0854294	.4225877					
lnemp	.7422207	.0789719	9.40	0.000	.5874386	.8970029					
unemp	-.0018148	.0025186	-0.72	0.471	-.0067512	.0031215					
Indirect											
lnpcap	-.1361936	.1765661	-0.77	0.441	-.4822567	.2098696					
lnpc	.1419852	.1107685	1.28	0.200	-.0751171	.3590875					
lnemp	-.0226692	.1240987	-0.18	0.855	-.2658983	.2205599					
unemp	-.0072797	.0057566	-1.26	0.206	-.0185624	.0040029					
Total											
lnpcap	-.105257	.1988103	-0.53	0.597	-.4949181	.2844041					
lnpc	.3959938	.1065194	3.72	0.000	.1872195	.604768					
lnemp	.7195515	.1668882	4.31	0.000	.3924567	1.046646					
unemp	-.0090946	.0061702	-1.47	0.140	-.0211879	.0029988					

从上表可知,空间自回归系数 ρ (rho)在1%水平上显著为正。但变量lnpcap,lnpc与unemp的空间滞后项并不显著,故使用选择项“durbin(varlist)”将这三个变量去掉。上表底部还显示了各变量的直接、间接与总效应(参见本章第4节);为节省空间,下面使用选择项“noeffects”使其不再显示。

```
. xsmle lngsp lnpcap lnpc lnemp unemp, wmat (usaww) model (sdm) durbin  
(lnemp) robust nolog noeffects
```

SDM with random-effects						Number of obs =	816				
Group variable: state						Number of groups =	48				
Time variable: year						Panel length =	17				
R-sq: within = 0.9436 between = 0.9905 overall = 0.9894											
Log-pseudolikelihood = 1493.7191 (Std. Err. adjusted for 48 clusters in state)											
lngsp	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]						
Main											
lnpcap	.0269178	.047324	0.57	0.569	-.0658355	.1196711					
lnpc	.2077165	.0691193	3.01	0.003	.0722452	.3431877					
lnemp	.7700826	.0818839	9.40	0.000	.6095931	.9305721					
unemp	-.0047394	.0027155	-1.75	0.081	-.0100617	.0005829					
_cons	.8293648	.2337675	3.55	0.000	.3711889	1.287541					
Wx											
lnemp	-.4698969	.076081	-6.18	0.000	-.6190129	-.3207809					
Spatial											
rho	.4914654	.0798065	6.16	0.000	.3350474	.6478833					
Variance											
lgt_theta	-2.438141	.2559757	-9.52	0.000	-2.939844	-1.936437					
sigma_e	.0010468	.0002513	4.16	0.000	.0005542	.0015394					

下面进行固定效应的估计。

```
. xsmle lngsp lnpcap lnpc lnemp unemp, wmat (usaww) model (sdm) durbin  
(lnemp) robust nolog noeffects fe
```

SDM with spatial fixed-effects		Number of obs =	816		
Group variable: state		Number of groups =	48		
Time variable: year		Panel length =	17		
R-sq:	within = 0.9457 between = 0.9856 overall = 0.9845				
Mean of fixed-effects =	1.0970				
Log-pseudolikelihood =	1650.1738				
	(Std. Err. adjusted for 48 clusters in state)				
		Robust			
lngsp	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Main					
lnpcap	-.024499	.0475762	-0.51	0.607	-.1177468 .0687487
lnpc	.1775734	.0797904	2.23	0.026	.0211871 .3339597
lnemp	.7326918	.0940438	7.79	0.000	.5483694 .9170143
unemp	-.0037328	.0027181	-1.37	0.170	-.0090602 .0015946
Wx					
lnemp	-.3958431	.0931682	-4.25	0.000	-.5784493 -.2132368
Spatial					
rho	.5184863	.0744086	6.97	0.000	.3726481 .6643245
Variance					
sigma2_e	.0009515	.0002346	4.06	0.000	.0004917 .0014113

究竟应使用随机效应还是固定效应模型，须进行豪斯曼检验。为此，需要存储未使用稳健标准误的随机效应与固定效应估计结果。

```
. qui xsmle lngsp lnpcap lnpc lnemp unemp, wmat (usaww) model (sdm)
durbin(lnemp) nolog noeffects
.est sto re
.qui xsmle lngsp lnpcap lnpc lnemp unemp, wmat (usaww) model (sdm)
durbin(lnemp) nolog noeffects fe
.est sto fe
.hausman fe re
```

	Coefficients		(b-B)	sqrt(diag(V_b-V_B))
	(b) fe	(B) re	Difference	S.E.
lnpcap	-.024499	.0269178	-.0514168	.0073629
lnpc	.1775734	.2077165	-.0301431	.0077758
lnemp	.7326918	.7700826	-.0373908	.0139211
unemp	-.0037328	-.0047394	.0010066	.

b = consistent under Ho and Ha; obtained from xsmle
 B = inconsistent under Ha, efficient under Ho; obtained from xsmle

Test: Ho: difference in coefficients not systematic

```

chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
          = -55.17
  chi2<0 ==> model fitted on these
  data fails to meet the asymptotic
  assumptions of the Hausman test;
  see suest for a generalized test
  
```

上表显示,豪斯曼统计量为负数,故可以接受随机效应的原假设。另外,对空间计量模型的估计,也可以在 Matlab^① 或 Gauss^② 中进行。

29.10 空间计量方法的局限性

上文提及,空间计量模型的主要估计方法之一为 MLE,但此 MLE 的大样本理论尚待完善。空间计量方法的另一局限是,它需要研究者主观设定一个非随机的空间权重矩阵(而非根据数据来估计此矩阵),而此空间权重矩阵很可能无法完全反映不同地区之间复杂的相互关系。

处理区域间互动关系的另一方法是直接将地区间的距离作为变量引入回归模型。比如,在国际贸易领域,常使用以下模型来考察两国间贸易量的决定因素:

$$Export_{ij} = \frac{\alpha Y_i^\beta Y_j^\gamma}{D_{ij}^\delta} \quad (29.53)$$

其中, $Export_{ij}$ 表示 i 国对 j 国的出口额, Y_i, Y_j 分别为 i 国与 j 国的 GDP, 而 D_{ij} 为 i 国与 j 国之间的地理距离(通常以首都间的距离来衡量)。由于公式(29.53)很像牛顿的万有引力公式,故称为“引力模型”(gravity model),最早由丁伯根(Jan Tinbergen)于1962年提出。对方程(29.53)两边取对数,并加上误差项(表示其他遗漏变量与随机扰动的影响):

$$\ln Export_{ij} = \ln \alpha + \beta \ln Y_i + \gamma \ln Y_j - \delta \ln D_{ij} + \varepsilon \quad (29.54)$$

在对引力模型的拓展研究中,经济学家又加入了影响贸易的其他因素,比如地理因素(如岛国、高山)、关税与非关税壁垒、边界效应、文化、宗教、语言、甚至政治制度(比如民主程度),参见余森杰(2013)。

① 参见 <http://www.spatial-econometrics.com/>.

② 参见 <http://web.pdx.edu/~crkl/WISE/SEAUG.htm>.

第30章 久期分析

在实证研究中,有时被解释变量为某种活动持续的时间(duration, spell, time to event),比如病人存活的时间,灯泡报废的时间,失业持续的时间,婚姻持续的时间,罢工持续的时间,战争持续的时间,王朝的寿命,刑满释放犯再次犯罪的时间等。这类数据称为“久期数据”(duration data),其相应的分析方法则称为“久期分析”(duration analysis)。

由于久期分析考察个体从某一状态转换到另一状态所花费的时间,故也称为“转换分析”(transition analysis)或“事件历史分析”(event history analysis)。另外,不同学科对于久期分析的称呼也不同。比如,生物统计领域称其为“生存分析”(survival analysis),运筹学领域称为“报废时间分析”(failure time analysis),人口学领域称为“生命表分析”(life table analysis),而保险领域称为“风险分析”(hazard analysis)。久期分析的许多术语来自生物统计领域。

30.1 久期数据的处理方法

处理久期数据需要特别的方法吗? 久期数据通常为横截面数据。假设数据为 $\{T_i, \mathbf{x}_i\}_{i=1}^n$,其中 T_i 为被解释变量,即个体*i*的持续时间(寿命);而 \mathbf{x}_i 为解释变量。一个自然的想法是用OLS估计以下模型:

$$T_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i \quad (30.1)$$

然而,由于持续时间 $T_i \geq 0$,而由上式得到的预测值 $\hat{T}_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$ 有可能为负数,故这是一个不现实的模型。较为现实的建模方法将 $\ln T_i$ 作为被解释变量(假设 $T_i > 0$):

$$\ln T_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i \quad (30.2)$$

此对数线性模型的最大问题是,在我们抽样获得观测数据时,通常知道个体已经存活了一段时间。而方程(30.2)却总是站在 $T_i = 0$ (即病情刚发作或确诊)的角度,无法纳入“个体已经存活了一段时间”这一信息。事实上,我们常常更为关心给定个体已存活了一段时间的条件下,个体在下个时刻死亡的概率,即下文的风险函数。比如,我们关心已经失业三个月的失业者明天找到工作的概率。另外,如果久期数据存在右归并(参见30.3节),或者随时间而变的解释变量 \mathbf{x}_u ,都很难通过OLS来处理。为此,久期分析常使用基于风险函数的一套特殊方法,形成自成体系的研究领域。

30.2 风险函数

记个体在某种状态中持续的时间(spell)或寿命为 $T \geq 0$,其一个特定取值记为 t 。显然, T 为随机变量。假设 T 为连续型随机变量,并记其概率密度函数与累积分布函数分别为 $f(t)$ 与

$F(t)$, 其中 $F(t)$ 也被称为“失效函数”(failure function)。考虑“病人”存活期超过 t 的概率, 称为“生存函数”(survivor function), 其定义为:

$$S(t) \equiv P(T > t) = 1 - F(t), \quad t \geq 0 \quad (30.3)$$

由上式可知, 生存函数本质上相当于累积分布函数的“反函数”(reverse cumulative distribution function)。由于累积分布函数 $F(t)$ 单调递增, 故生存函数 $S(t)$ 单调递减。假设病人已存活到时刻 t , 则病人在 $[t, t + \Delta t]$ 期间 ($\Delta t > 0$) 死亡的概率为:

$$P(t \leq T < t + \Delta t | T \geq t) = \frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{S(t)} \quad (30.4)$$

进一步, 定义“风险率”(hazard rate)或“风险函数”(hazard function)为病人在时刻 t 的瞬间死亡率:

$$\begin{aligned} \lambda(t) &\equiv \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0^+} \frac{F(t + \Delta t) - F(t)}{\Delta t S(t)} \\ &= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0^+} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{f(t)}{S(t)} \end{aligned} \quad (30.5)$$

由以上定义与表达式可知, 风险函数 $\lambda(t)$ 本质上是在给定存活至时刻 t 条件下的条件密度函数, 故也称为“条件死亡率”(conditional failure rate); 而 $f(t)$ 为无条件密度函数。如果 $f(t) = \phi(t)$ (标准正态的密度函数), 则 $\lambda(t)$ 就是反米尔斯比率(IMR, 参见第14章图14.2)。风险率的可能取值介于0(无死亡风险)与 ∞ (必死无疑)之间。在久期分析中, 风险函数 $\lambda(t)$ 与生存函数 $S(t)$ 比密度函数 $f(t)$ 与累积分布函数 $F(t)$ 更为方便与常用。

也可以从风险函数 $\lambda(t)$ 出发, 反推出生存函数 $S(t)$ 、累积分布函数 $F(t)$ 以及密度函数 $f(t)$ 。首先, 从方程(30.5)可知,

$$\lambda(t) = -\frac{d \ln S(t)}{dt} \quad (30.6)$$

由上式可知 $d \ln S(t) = -\lambda(t) dt$, 然后两边从0到 t 作定积分可得:

$$\ln S(t) = - \int_0^t \lambda(u) du \quad (30.7)$$

其中, u 为积分变量; $S(0) = 1$ (在初始时刻, 所有个体都活着)。因此,

$$F(t) = 1 - S(t) = 1 - \exp \left[- \int_0^t \lambda(u) du \right] \quad (30.8)$$

$$S(t) = \exp \left[- \int_0^t \lambda(u) du \right] \quad (30.9)$$

对方程(30.8)两边求导可得,

$$f(t) = \lambda(t) \exp \left[- \int_0^t \lambda(u) du \right] \quad (30.10)$$

为了度量截止时刻 t 的累积总风险, 定义“累积风险函数”(cumulative hazard 或 integrated hazard)为:

$$\Lambda(t) = \int_0^t \lambda(u) du = - \ln S(t) \quad (30.11)$$

上式中的第二个等号用到了方程(30.7)。累积风险函数的好处在于, 它比风险函数可以更准确地估计。如果知道累积风险函数, 则很容易计算生存函数:

$$S(t) = \exp[-\Lambda(t)] \quad (30.12)$$

从方程(30.12)可以看出累积风险函数 $\Lambda(t)$ 的经济含义, 它完全决定了至时间 t 为止的存活率或概率。

例 假设 T 服从指数分布(exponential distribution), $f(t) = \lambda e^{-\lambda t}$ ($\lambda > 0$), 则 $F(t) = \int_0^t \lambda e^{-\lambda s} ds = \int_0^t e^{-\lambda s} d(\lambda s) = -e^{-\lambda s} \Big|_0^t = 1 - e^{-\lambda t}$, $S(t) = e^{-\lambda t}$, $\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$, $\Lambda(t) = \lambda t$ 。因此, 指数分布的风险函数为常数, 这个性质被称为“无记忆性”(memoryless)。它意味着, 瞬间死亡的概率并不依赖于已存活了多久。进一步, 可以证明, 个体在 $(0, t_2)$ 区间死亡的概率等于在已知个体存活至时刻 t_1 的情况下, 其在 $(t_1, t_1 + t_2)$ 区间死亡的概率(参见习题)。另一方面, 如果风险函数为常数 λ , 则其对应的密度函数为 $f(t) = \lambda \exp\left[-\int_0^t \lambda du\right] = \lambda e^{-\lambda t}$, 一定服从指数分布。根据概率统计知识可知, 指数分布的期望为 $E(T) = \frac{1}{\lambda}$ (风险率的倒数), 方差为 $\text{Var}(T) = \frac{1}{\lambda^2}$ 。

指数分布广泛应用于研究电子元器件的寿命问题。

由于指数分布只有一个参数, 故如果知道期望, 则方差也确定(方差为期望的平方), 缺乏灵活性。另外, 指数分布的无记忆性有时也不现实。它意味着, 一个 20 岁的青年与一个 80 岁的老人不仅瞬间死亡率一样, 而且在未来 10 年内死亡的概率也相同。显然, 无记忆性的假设太强了。为此, 常将指数分布拓展至拥有两个参数的威布尔分布(Weibull distribution)。

定义 如果随机变量 T 的累积分布函数为 $F(t) = 1 - \exp(-\gamma t^p)$, 其中 $\gamma > 0, p > 0$, 则称其服从威布尔分布。

显然, 威布尔分布的生存函数为 $S(t) = 1 - F(t) = \exp(-\gamma t^p)$ 。威布尔分布的密度函数为 $f(t) = F'(t) = \gamma p t^{p-1} \exp(-\gamma t^p)$ 。因此, 威布尔分布的风险函数为

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\gamma p t^{p-1} \exp(-\gamma t^p)}{\exp(-\gamma t^p)} = \gamma p t^{p-1} \quad (30.13)$$

如果 $p = 1$, 则威布尔分布的累积分布函数为 $F(t) = 1 - \exp(-\gamma t)$, 就是指数分布的累积分布函数, 故指数函数是威布尔分布的一个特例。如果 $p > 1$, 则风险函数 $\lambda(t)$ 单调递增, 这意味着, 活得越久则死亡概率越高(或许由于老年化过程), 称为“正向久期依赖”(positive duration dependence); 反之, 如果 $p < 1$, 则风险函数 $\lambda(t)$ 单调递减(或许新生儿死亡概率最高); 被称为“负向久期依赖”(negative duration dependence)。如果 $p = 1$ (即指数分布的情形), 则风险函数 $\lambda(t)$ 为常数(或许死亡由外在的随机因素所造成)。从方程(30.13)可以看出, 参数 p 决定了风险函数 $\lambda(t)$ 的形状, 称为“形状参数”(shape parameter); 而参数 γ 则决定其规模, 称为“规模参数”(scale parameter)。

30.3 久期数据的归并问题

久期数据常常存在“右归并”(right censoring)。一个主要原因是, 当研究结束时, 有些病人可能尚未死亡; 或者有些失业者还未找到工作。此时, 我们观测到的个体存活时间为从 0 开始直至某归并时间(censoring time) C^* 。我们只知道个体的寿命属于区间 (C^*, ∞) , 并不知道其具体取值。导致右归并的原因还包括, 个体中途退出研究(比如, 研究某药物是否能延长病人的寿

命,而病人中途因强烈副作用而提前退出实验),或研究者与个体失去联系,无法继续跟踪调查。

在存在右归并的情形,记个体 i 的真实寿命为 T_i^* (可能不可观测),而归并时间为 C_i^* 。实际观测到的 T_i 或为个体寿命 T_i^* ,或为归并时间 C_i^* ,取决于二者哪个更小:

$$T_i = \min(T_i^*, C_i^*) \quad (30.14)$$

同时,以虚拟变量 d_i 来记录个体 i 的观测记录是否完整:

$$d_i = \mathbf{1}(T_i^* < C_i^*) \quad (30.15)$$

其中, $\mathbf{1}(\cdot)$ 为示性函数,即如果括弧中表达式为真,则取值为 1;反之,则取值为 0。显然,如果 $d_i = 1$,则有完整记录,无归并(uncensored);如果 $d_i = 0$,则存在右归并。

有时也会出现“左归并”(left censoring),即只知道个体的寿命属于区间 $(0, C^*)$,而不知道其具体取值。另一种归并情形为“区间归并”(interval censoring),即只知道个体的寿命属于区间 $[C_a^*, C_b^*]$,也不知道其具体取值。区间归并常因为数据的离散性而发生,比如,研究者只知道失业者在某一周找到工作,而不知道其具体日期。

为了保证久期分析的有效性,常假设“独立归并”(independent censoring)或“无信息归并”(noninformative censoring),即归并时间 C_i^* 的分布不包含任何有关个体寿命 T_i^* 分布的信息。此时,可将表示归并的虚拟变量 d_i 视为外生变量,不必在意归并究竟如何发生,也无须为“归并机制”(censoring mechanism)建模。

另外,在久期样本中,每一个体开始活动(比如,开始生病或失业)的日历时间(calendar time)可以不同。通常将“风险开始”(onset of risk)的时间标准化为 0 时刻,以此度量的时间在 Stata 中称为“分析时间”(analysis time)。久期分析的被解释变量 T_i 正是以分析时间来计算的。

30.4 描述性分析

在引入解释变量之前,我们常希望进行一些粗略的描述性分析,比如根据样本数据来估计生存函数、累积风险函数与风险函数,看看它们的大致形状。在进行这类分析时,既不需要对数据的概率分布作先验假设,也不需要估计参数,故为非参数方法(参见第 27 章)。通俗地说,这是一种撇开理论(不假设变量的概率分布),让数据自己说话的方法(let data speak for themselves)。

1. 生存函数

根据定义,生存函数 $S(t)$ 为个体存活时间超过时刻 t 的概率。如果不存在归并(censoring),则显然可以定义 $S(t)$ 的估计量为,样本中存活时间超过时刻 t 的个体数目 r 占样本容量 n 的比例,即 $\frac{r}{n}$ 。但这个简单方法在存在归并的情况下并不适用。此时,一般使用 Kaplan-Meier 估计量

(Kaplan and Meier, 1958, 简记 KM),它在独立归并(independent censoring)的情况下依然是 $S(t)$ 的一致估计量。

记 $t_1 < t_2 < \dots < t_j < \dots < t_k$ 为样本中观测到的死亡时间。记样本中在区间 $[t_{j-1}, t_j)$ 仍存活而面临危险(at risk)的个体数为 n_j 。到了时间 t_j ,这些 n_j 个体的命运分为三种,即存活、死亡、归并(只知道其死亡时间大于 t_j ,但不再有观测数据)。记在时间 t_j 死亡的人数为 m_j (现实数据一般是离散的,故 m_j 可能大于 1)。给定存活至 t_{j-1} ,能进一步活至 t_j 的概率(频率)为 $\frac{n_j - m_j}{n_j}$ 。一般地,活至 t_j 的无条件概率等于活过之前每一个区间的条件概率之乘积。因此,Kaplan-Meier 估

计量为

$$\hat{S}(t) = \prod_{j|t_j \leq t} \left(\frac{n_j - m_j}{n_j} \right) \quad (30.16)$$

故 Kaplan-Meier 估计量也称为“连乘估计量”(product limit estimator)。显然,如果不存在归并,则 $n_{j+1} = n_j - m_j$,方程(30.16)中的连乘可以错项相约,故 $\hat{S}(t) = r/n$ 。

2. 累积风险函数

根据方程(30.11),累积风险函数 $\Lambda(t) = -\ln S(t)$,将 $\hat{S}(t)$ 的表达式(30.16)代入,即可得到 $\Lambda(t)$ 的估计量。但此估计量的小样本性质不如 Nelson(1972)与 Aalen(1978)所提出的 Nelson-Aalen 估计量(简记 NA):

$$\hat{\Lambda}(t) = \sum_{j|t_j \leq t} \left(\frac{m_j}{n_j} \right) \quad (30.17)$$

其中的每一项 $\frac{m_j}{n_j}$ 为局部的风险率,而上式则为局部风险率的加总。

3. 风险函数

从上面的讨论可知,可以用 $\hat{\lambda}_j = m_j/n_j$ 作为风险率的估计量。但此估计量为不光滑的阶梯函数(step function)。另一方法是对累积风险函数求导,但 $\hat{\Lambda}(t)$ 也是阶梯函数,并不处处可导。在实践中,一般先通过核密度方法将阶梯形的累积风险函数光滑化^①,然后再以此生成风险函数。在 Stata 中可由命令“sts graph,hazard”来直接完成。

30.5 久期模型的最大似然估计

一般来说,(未被归并的)真实持续时间 T^* 的概率分布还可能依赖于某些解释变量 x 。为此,记 T^* 的密度函数、累积分布函数、生存函数与风险函数为 $f(t|x, \theta)$, $F(t|x, \theta)$, $S(t|x, \theta)$ 与 $\lambda(t|x, \theta)$,其中 x 为不随时间而变的解释变量(time-invariant covariates)(以后再考虑随时间而变的解释变量), θ 为待估计的未知参数。比如, T^* 为失业持续的时间, x 包括失业前的教育水平、工作经验、种族、婚姻状况、子女数目,以及政府发放的失业救济金额等;故 T 的分布依赖于 x 。

下面使用 MLE 来估计未知参数 θ 。假设存在右归并,则其 MLE 估计类似于归并回归的 Tobit 模型。对于未被归并的观测值,其对似然函数的贡献为 $f(t_i|x_i, \theta)$ 。另一方面,对于右归并的观测值,我们只知道其持续时间超过 t ,故其对似然函数的贡献为 $P(T^* > t) = S(t|x_i, \theta)$ 。综合起来,个体 i 对似然函数的贡献可写为

$$f(t_i|x_i, \theta)^{d_i} S(t_i|x_i, \theta)^{1-d_i} \quad (30.18)$$

其中, d_i 为右归并虚拟变量,参见(30.15),即 $d_i = 1$ 表示无归并,而 $d_i = 0$ 表示右归并。假设样本为 iid,将似然函数取对数,然后加总可得

$$\ln L(\theta) = \sum_{i=1}^n [d_i \ln f(t_i|x_i, \theta) + (1-d_i) \ln S(t_i|x_i, \theta)] \quad (30.19)$$

上式中的第一项对应于无归并数据,而第二项对应于右归并数据。由于 $f(t) = \lambda(t)S(t)$,故 $\ln f(t) = \ln \lambda(t) + \ln S(t)$,因此方程(30.19)可写为

^① 有关核密度方法,参见第 27 章。

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^n [d_i \ln \lambda(t_i | \mathbf{x}_i, \boldsymbol{\theta}) + \ln S(t_i | \mathbf{x}_i, \boldsymbol{\theta})] \quad (30.20)$$

如果直接对风险函数 $\lambda(t | \mathbf{x}, \boldsymbol{\theta})$ 建模, 则表达式(30.20)可能更为方便。根据 MLE 的一般理论, 如果似然函数正确, 则此 MLE 估计量为一致、有效估计量, 且服从渐近正态分布。然而, 如果似然函数不正确, 则此 MLE 估计量通常是不一致的。一个例外是 T^* 服从指数分布并且不存在归并, 则只要条件期望函数 (conditional mean function) 正确即可保证 MLE 估计量的一致性。然而, 如果存在归并, 即使 T^* 服从指数分布, MLE 估计量也依然不一致。因此, 以上参数回归模型 (parametric regression model) 的最大缺点是缺乏稳健性 (lack of robustness), 对概率分布的具体假设比较敏感 (Tobit 模型也有此缺点)。

如果存在其他形式的归并, 则应对似然函数进行相应的调整。如果观测值存在左归并, 则只知道持续时间小于 t , 故其对似然函数的贡献为 $P(T^* < t) = F(t | \mathbf{x}, \boldsymbol{\theta})$ 。如果观测值存在区间归并, 则只知道持续时间属于区间 $[C_a^*, C_b^*]$, 故其对似然函数的贡献为 $P(C_a \leq T^* < C_b) = S(C_a | \mathbf{x}, \boldsymbol{\theta}) - S(C_b | \mathbf{x}, \boldsymbol{\theta})$ 。

在经济应用中的久期数据常存在区间归并。比如, 失业持续时间经常以周或月来计算, 而指数分布、威布尔分布等都是连续型分布。此时, 通常假设区间归并的影响甚微, 而仍将其作为连续数据来处理。

30.6 比例风险模型

称风险函数 $\lambda(t; \mathbf{x})$ 为“比例风险” (Proportional Hazard, 简记 PH), 如果它可以分解为

$$\lambda(t; \mathbf{x}) = \lambda_0(t) h(\mathbf{x}) \quad (30.21)$$

其中, $h(\cdot) > 0$, 而 $\lambda_0(t)$ 被称为“基准风险” (baseline hazard), 依赖于时间 t , 但不依赖于 \mathbf{x} 。基准风险 $\lambda_0(t)$ 对于总体中的每一个体都相同, 而个体的风险函数则依据 $h(\mathbf{x})$ 与此基准风险 $\lambda_0(t)$ 成正比, 故名“比例风险”。通常, 令 $h(\mathbf{x}) = e^{x'\beta}$, 则

$$\lambda(t; \mathbf{x}) = \lambda_0(t) e^{x'\beta} \quad (30.22)$$

其中, $e^{x'\beta}$ 称为“相对风险” (relative hazard), 故 $x'\beta$ 也被称为“对数相对风险” (log-relative hazard)。根据方程(30.22), 如果所有解释变量都等于 0, 则风险函数 $\lambda(t; \mathbf{x})$ 就等于基准风险 $\lambda_0(t)$ 。对方程(30.22)两边取对数可得

$$\ln \lambda(t; \mathbf{x}) = \mathbf{x}'\beta + \ln \lambda_0(t) \quad (30.23)$$

其中, β 可方便地解释为半弹性, 即某 \mathbf{x} 边际地增加 $\Delta\mathbf{x}$, 将导致风险函数增加百分之 $100\beta \cdot \Delta\mathbf{x}$ 。另一解释是, 如果 \mathbf{x} 增加一单位, 则风险率为 $\lambda(t; \mathbf{x}) = \lambda_0(t) e^{x'\beta + \beta}$, 正好是原来风险率 $\lambda(t; \mathbf{x}) = \lambda_0(t) e^{x'\beta}$ 的 e^β 倍。在 Stata 中, 称 β 为“系数” (coefficient), 而称 e^β 为“风险比率” (Hazard Ratio, 简记 HR), 即 \mathbf{x} 增加一单位, 将使新风险率变为原来风险率的 e^β 倍。在估计模型之后, 可以选择让 Stata 汇报风险比率 (默认值) 或系数。

在方程(30.22)中, 如果令基准风险 $\lambda_0(t) = e^a$, 其中 a 为待估参数 (选择指数形式是为了保证 $e^a > 0$), 则为“指数回归” (exponential regression), 因为指数分布的风险函数为常数。也可以从指数分布的密度函数出发, 即 $f(t) = \lambda e^{-\lambda t}$ ($\lambda > 0$), 而令 $\lambda = e^a e^{x'\beta} = e^{a+x'\beta}$, 其中 \mathbf{x} 不包含常数项 (如果有常数项 e^{β_0} , 可与 e^a 合并; Stata 也不汇报常数项)。将指数分布的风险函数 λ 与生存函数 $S(t) = e^{-\lambda t}$ 代入方程(30.20)可得指数回归的对数似然函数:

$$\ln L(a, \beta) = \sum_{i=1}^n [d_i(a + \mathbf{x}_i' \beta) - e^{a + \mathbf{x}_i' \beta} t_i] \quad (30.24)$$

类似地,如果在方程(30.22)中,令基准风险 $\lambda_0(t) = pt^{p-1}e^a$,其中 $p > 0, a$ 为待估参数,则为“威布尔回归”(Weibull regression),因为威布尔分布的风险函数为 γpt^{p-1} (此处令 $\gamma = e^{a + \mathbf{x}' \beta}$)。将威布尔分布的风险函数 γpt^{p-1} 与生存函数 $S(t) = \exp(-\gamma t^p)$ 代入方程(30.20)可得威布尔回归的对数似然函数:

$$\ln L(a, p, \beta) = \sum_{i=1}^n \{d_i[a + \mathbf{x}_i' \beta + \ln p + (p-1) \ln t_i] - e^{a + \mathbf{x}_i' \beta} t_i^p\} \quad (30.25)$$

由于当 $p = 1$ 时,威布尔分布退化为指数分布,故可通过检验“ $H_0: p = 1$ ”(或 $H_0: \ln p = 0$)来确定应使用威布尔回归还是指数回归。

如果在方程(30.22)中,令基准风险 $\lambda_0(t) = e^{a + \gamma t}$,其中 a, γ 为待估参数,则为“冈珀茨回归”(Gompertz regression),在人口学与保险精算领域应用广泛。与威布尔分布相比,冈珀茨分布的风险函数也是单调函数,即当 $\gamma > 0$ 时单调递增,当 $\gamma < 0$ 时单调递减,而当 $\gamma = 0$ 时为常数(同样退化为指数分布);二者的不同点在于,冈珀茨分布风险函数的变化率为指数形式($e^{\gamma t}$),而威布尔分布风险函数的变化率为幂函数形式(t^{p-1})。由于当 $\gamma = 0$ 时,冈珀茨分布退化为指数分布,故可通过检验“ $H_0: \gamma = 0$ ”来确定应使用冈珀茨回归还是指数回归。

威布尔分布与冈珀茨分布的风险函数都是单调函数,要么单调递增,要么单调递减,这个性质有时也与现实不符。比如,对于人类的死亡率而言,婴儿与老人的死亡率较高,而中青年的死亡率较低,其风险函数并不具有单调性。生物统计学家将人类死亡率的风险函数称为“浴缸风险”(bathtub hazard),因为它的形状就像一个浴缸。在比例风险模型的框架下,一个解决方法是不设定基准风险 $\lambda_0(t)$ 的函数形式,参见下文的 Cox 模型。另一个解决方法是假设基准风险 $\lambda_0(t)$ 为阶梯函数,即 $\lambda_0(t)$ 在每个小区间内都是待估计的常数(正如指数模型的风险率为常数),以便更好地拟合实际的风险函数,这被称为“分段固定风险模型”(piecewise constant hazard model 或 piecewise constant exponential model)。具体来说,分段固定风险模型的风险函数可写为

$$\lambda(t; \mathbf{x}) = \begin{cases} \lambda_1 e^{\mathbf{x}' \beta} & t \in [0, \tau_1) \\ \lambda_2 e^{\mathbf{x}' \beta} & t \in [\tau_1, \tau_2) \\ \vdots & \vdots \\ \lambda_M e^{\mathbf{x}' \beta} & t \in [\tau_{M-1}, \tau_M) \end{cases} \quad (30.26)$$

为了便于估计,上式可写为

$$\lambda(t; \mathbf{x}) = \begin{cases} e^{\ln \lambda_1 + \mathbf{x}' \beta} & t \in [0, \tau_1) \\ e^{\ln \lambda_2 + \mathbf{x}' \beta} & t \in [\tau_1, \tau_2) \\ \vdots & \vdots \\ e^{\ln \lambda_M + \mathbf{x}' \beta} & t \in [\tau_{M-1}, \tau_M) \end{cases} \quad (30.27)$$

从方程(30.27)可以看出,分段固定风险模型等价于在每个时间段内分别引入一个截距项。因此,对应于每个时间段,可以定义 M 个虚拟变量,并将 $(M-1)$ 个虚拟变量作为解释变量引入指数模型(假设模型中包括截距项)。例如,Lalive et al(2006)在研究失业持续时间时,将风险函数每四周分为一段(大多数失业者在 60 周内找到工作)。分段固定风险模型的缺点是,需要估计较多参数,故要求样本容量比较大。

对于比例风险模型 $\lambda(t; \mathbf{x}) = \lambda_0(t) e^{\mathbf{x}' \beta}$,无论其基准风险 $\lambda_0(t)$ 如何设定(指数、威布尔、冈珀茨或分段固定),对于参数 β 的经济含义解释都是一样的,即可将 β 解释为 x 对于风险函数的半

弹性,或将 e^β 解释为风险比率。

30.7 加速失效时间模型

对于比例风险模型,通常分析的重点是解释变量 x 对于风险函数 $\lambda(t; x)$ 的作用,不容易看出 x 对于平均寿命 $E(T)$ 的影响。为此,考虑直接对 $\ln T$ 建模:

$$\ln T = \mathbf{x}'\boldsymbol{\beta} + u \quad (30.28)$$

其中, u 为扰动项。由于 $\ln T$ 可在 $(-\infty, \infty)$ 取值,故 u 可以服从取值于 $(-\infty, \infty)$ 的任何连续型分布。由方程(30.28)可得, $T = e^{\mathbf{x}'\boldsymbol{\beta}}v$, 其中 $v = e^u$ 。记 v 的密度函数、累积分布函数、生存函数与风险函数分别为 $f_v(v), F_v(v), S_v(v)$ 与 $\lambda_v(v)$; 记 T 的密度函数、累积分布函数、生存函数与风险函数分别为 $f_T(t), F_T(t), S_T(t)$ 与 $\lambda_T(t)$, 则

$$S_T(t) = P(T > t) = P[e^{\mathbf{x}'\boldsymbol{\beta}}v > t] = P[v > e^{-\mathbf{x}'\boldsymbol{\beta}}t] = S_v(e^{-\mathbf{x}'\boldsymbol{\beta}}t) \quad (30.29)$$

因此,

$$f_T(t) = F'_T(t) = [1 - S_T(t)]' = -\frac{\partial S_v(e^{-\mathbf{x}'\boldsymbol{\beta}}t)}{\partial t} = f_v(e^{-\mathbf{x}'\boldsymbol{\beta}}t)e^{-\mathbf{x}'\boldsymbol{\beta}} \quad (30.30)$$

由此可知,

$$\lambda_T(t|x) = \frac{f_T(t)}{S_T(t)} = \frac{f_v(e^{-\mathbf{x}'\boldsymbol{\beta}}t)e^{-\mathbf{x}'\boldsymbol{\beta}}}{S_v(e^{-\mathbf{x}'\boldsymbol{\beta}}t)} = \lambda_v(e^{-\mathbf{x}'\boldsymbol{\beta}}t)e^{-\mathbf{x}'\boldsymbol{\beta}} \quad (30.31)$$

由上式可知,如果 $e^{-\mathbf{x}'\boldsymbol{\beta}} > 1$, 则意味着对基准风险 $\lambda_v(t)$ 的加速;而如果 $e^{-\mathbf{x}'\boldsymbol{\beta}} < 1$, 则意味着对基准风险 $\lambda_v(t)$ 的减速。因此,这类模型称为“加速失效时间模型”(Accelerated Failure Time Model, 简记 AFT), 尽管 $e^{-\mathbf{x}'\boldsymbol{\beta}} < 1$ 意味着减速。

如果方程(30.28)中的扰动项 $u \sim N(0, \sigma^2)$, 则称为“对数正态模型”(log-normal model)。如果 u 服从逻辑分布(logistic distribution), 则称为“对数逻辑模型”(log-logistic model)。如果 u 服从广义伽马分布(generalized gamma distribution) $Gamma(\beta_0, \kappa, \sigma)$, 则称为“伽马模型”(gamma model)。其中,参数 κ 读为 kappa, 在 Stata 中也记为 kappa。对数正态模型与对数逻辑模型的风险函数均为非单调函数(non-monotonic), 适用于风险率首先上升然后下降的情形;而包含三个参数的伽马模型,其风险函数形状更为灵活。如果 $\kappa = 1$, 则伽马模型就是威布尔模型;如果 $\kappa = 1$ 而且 $\sigma = 1$, 则伽马模型就是指数模型;如果 $\kappa = 0$, 则伽马模型就是对数正态模型。因此,对原假设 “ $H_0: \kappa = 1$ ”或“ $H_0: \kappa = 0$ ”进行检验,有助于选择模型适用的分布函数。另外,指数模型与威布尔模型也可以写成 AFT 形式,故它们既属于 PH 模型,也属于 AFT 模型;这是指数模型与威布尔模型独有的性质(其他模型,要么为 PH,要么为 AFT)。

对于 AFT 模型,同样可以进行 MLE 估计。但 AFT 模型的参数 $\boldsymbol{\beta}$ 与 PH 模型的参数 $\boldsymbol{\beta}$ 的经济解释不同。在 AFT 模型下,根据方程(30.28), $\boldsymbol{\beta}$ 可以解释为当 x 边际增加时,能使平均寿命增加的百分比(即半弹性)。特别地,对于指数模型而言,在 PH 模型下,风险率为 $\lambda = e^{a+x'\boldsymbol{\beta}}$, 而平均寿命为 $E(T) = \frac{1}{\lambda} = e^{-(a+x'\boldsymbol{\beta})}$, 故其 AFT 模型为 $\ln T = -(a + \mathbf{x}'\boldsymbol{\beta}) + u$ 。由此可见,PH 模型下参数 β 与 AFT 模型下的参数 $\boldsymbol{\beta}$ 正好符号相反,而绝对值相等。在 Stata 中进行指数回归时,默认为 PH 模型;但如果加选择项“time”,则为 AFT 模型;二者的估计系数绝对值相等,但符号正好相反。

由于 PH 模型与 AFT 模型都对风险函数的形式作了具体假设,故属于“参数回归”(parametric regression)。然而,在进行参数回归时,可供选择的分布函数很多(Stata 支持六种分布函数,即指数分布、威布尔、冈珀茨、对数正态、对数逻辑、伽马分布)。如果进行抉择?对于“嵌套模型”(nested models),一个模型是另一模型的特例,可以对相关参数进行 Wald 或似然比检验(比如,指数模型是威布尔模型与冈珀茨模型的特例;威布尔模型与对数正态模型是伽马模型的特例)。对于“非嵌套模型”(non-nested models),不存在这种嵌套关系,一个模型并不是另一模型的特例,则可通过 AIC 信息准则比较模型拟合优度,选择 AIC 最小的那个模型。AIC 准则的定义为

$$AIC = -2 \ln L + 2(K + c) \quad (30.32)$$

其中, $\ln L$ 为对数似然函数, K 为解释变量 x 的维度, c 为概率分布的参数个数(比如,指数分布的 $c=1$;威布尔分布的 $c=2$;伽马分布的 $c=3$)。

30.8 Cox 模型

参数回归对风险函数的具体形式作了假设,如果模型设定正确,则 MLE 是最有效率的。然而,如果风险函数的设定错误,则 MLE 一般将导致不一致的估计。事实上,我们对于风险函数的具体形式常常并无把握。针对参数回归对于分布假定过强的缺点,在比例风险模型(PH)的框架下,Cox(1972,1975)提出了以下广为应用的方法,称为“Cox 模型”或“Cox PH 模型”。

由于比例风险模型 $\lambda(t; x) = \lambda_0(t) e^{x' \beta}$ 在形式上为乘法(multiplicative),故个体 i 与个体 j 的风险函数之比可以写为

$$\frac{\lambda(t; x_i)}{\lambda(t; x_j)} = \frac{\lambda_0(t) e^{x_i' \beta}}{\lambda_0(t) e^{x_j' \beta}} = e^{(x_i - x_j)' \beta} \quad (30.33)$$

显然,个体 i 与个体 j 的风险函数之比不随时间而改变,只与 $(x_i - x_j)$ 有关。这个简单事实使得我们可以不必假设基准风险 $\lambda_0(t)$ 的具体函数形式,而依然得到对 β 的估计。由于构成风险函数 $\lambda_0(t) e^{x' \beta}$ 的前半部分 $\lambda_0(t)$ 不需要估计参数(非参数部分),而后半部分 $e^{x' \beta}$ 需要估计参数(参数部分),故为半参数回归(参见第 27 章)。

1. 失效时间不重叠的情形(no tied failures)

首先考察失效时间均不相同的情形,即每一个体的“死亡”时间都不一样。为了便于理解,先举一个简单的例子^①。假设久期样本只包括以下四个观测值:

subject	t	x
1.	1	2
2.	2	3
3.	3	6
4.	4	12

其中,subject 表示个体,t 表示失效时间(failure time),而 x 为解释变量。记第 j 位个体的失效时间为 t_j ,则四位个体的失效时间分别为 $t_1 = 2, t_2 = 3, t_3 = 6, t_4 = 12$,完全不重叠。记在时间 t_j

^① 此例来自 Cleves et al(2004, p. 137)。

面临失效危险的所有个体所构成的集合为“风险集”(risk set) R_j 。如果个体已经失效或被归并，则不再面临失效危险。

当 $t = 2$ 时，风险集为 $R_1 = \{1, 2, 3, 4\}$ ，实际观测到第 1 位个体失效。

当 $t = 3$ 时，风险集为 $R_2 = \{2, 3, 4\}$ ，实际观测到第 2 位个体失效。

当 $t = 6$ 时，风险集为 $R_3 = \{3, 4\}$ ，实际观测到第 3 位个体失效。

当 $t = 12$ 时，风险集为 $R_4 = \{4\}$ ，实际观测到第 4 位个体失效。

在以上四个失效时间(2, 3, 6, 12)，我们假设必然有一位个体会失效，然后计算在此条件下实际失效的那位个体失效的条件概率，分别记为 P_1, P_2, P_3, P_4 ；则似然函数可写为

$$L(\beta) = P_1 P_2 P_3 P_4 \quad (30.34)$$

即这四个条件概率的乘积。下面分别来计算 P_1, P_2, P_3, P_4 。显然，当 $t = 12$ 时，风险集只包含第 4 位个体，故第 4 位个体失效的条件概率为 1，即 $P_4 = 1$ 。

当 $t = 6$ 时，给定第 3、4 位个体必然有一位失效，在此条件下第 3 位个体失效的条件概率为

$$P_3 = \frac{\lambda(6|x_3)}{\lambda(6|x_3) + \lambda(6|x_4)} = \frac{\lambda_0(6)e^{x_3\beta}}{\lambda_0(6)e^{x_3\beta} + \lambda_0(6)e^{x_4\beta}} = \frac{e^{x_3\beta}}{e^{x_3\beta} + e^{x_4\beta}} \quad (30.35)$$

其中， $\lambda(6|x_3)$ 表示 $t = 6, x = x_3$ 时的风险率。在上式中，由于基准风险 $\lambda_0(6)$ 已被约去，故 P_3 并不依赖于具体的失效时间 $t = 6$ 。

类似地，当 $t = 3$ 时，给定第 2、3、4 位个体必然有一位失效，在此条件下第 2 位个体失效的条件概率为

$$P_2 = \frac{\lambda(3|x_2)}{\lambda(3|x_2) + \lambda(3|x_3) + \lambda(3|x_4)} = \frac{e^{x_2\beta}}{e^{x_2\beta} + e^{x_3\beta} + e^{x_4\beta}} \quad (30.36)$$

同样地，当 $t = 2$ 时，给定第 1、2、3、4 位个体必然有一位失效，在此条件下第 1 位个体失效的条件概率为

$$P_1 = \frac{\lambda(2|x_1)}{\lambda(2|x_1) + \lambda(2|x_2) + \lambda(2|x_3) + \lambda(2|x_4)} = \frac{e^{x_1\beta}}{e^{x_1\beta} + e^{x_2\beta} + e^{x_3\beta} + e^{x_4\beta}} \quad (30.37)$$

因此，似然函数可写为

$$L(\beta) = \prod_{j=1}^4 \left(\frac{e^{x_j\beta}}{\sum_{i \in R_j} e^{x_i\beta}} \right) \quad (30.38)$$

其中， R_j 为对应于 t_j (第 j 位个体失效时间)的风险集。更一般地，如果样本中有 L 个失效时间 $\{t_1, \dots, t_L\}$ 以及多个解释变量 x ，则似然函数可写为

$$L(\beta) = \prod_{j=1}^L \left(\frac{e^{x_j\beta}}{\sum_{i \in R_j} e^{x_i\beta}} \right) \quad (30.39)$$

上式称为“部分似然函数”(partial likelihood function)，因为它并非完整的似然函数。从上面的推导可以看出，部分似然函数只依赖于样本中个体失效时间的排序，而不依赖于个体失效的具体时间。由于后者的信息被忽略了，故它只是“部分似然函数”。该估计量的协方差矩阵可根据 MLE 的一般理论得出。然而，由于同一个体多次出现于不同的风险集中，一般应使用 Lin and Wei(1989) 提出的稳健标准误，在 Stata 中可由选择项“*r*”或“*vce(robust)*”来实现。

可以证明，基于部分似然函数(30.39)的 MLE 估计量是一致且渐近正态的，尽管其有效性不如基于完整似然函数的 MLE 估计量(即前面的参数回归)。但后者的一致性依赖于对风险函数形式的正确设定(我们对此一般无把握)，而前者的效率损失通常不大。这意味着，如果真实

模型为指数回归，则 Cox 模型的估计结果与指数回归接近；如果真实模型为威布尔回归，则 Cox 模型的估计结果与威布尔回归接近；以此类推。由于 Cox 模型的稳健性，它在实践中十分流行，是半参数估计的成功例子。

2. 失效时间有重叠的情形 (tied failures)

如果样本中的失效时间有重叠，则部分似然函数的计算更为复杂。延续上面的简单例子，但假设第 3 位与第 2 位个体同时在 $t = 3$ 时失效，原始数据变为：

subject	t	x
1.	1	2
2.	2	3
3.	3	3
4.	4	12

当 $t = 3$ 时，风险集为 $R_2 = \{2, 3, 4\}$ ，而实际观测到第 2、3 位个体同时失效。此时，应如何计算在给定风险集中两位个体会失效的条件下，第 2、3 位个体失效的条件概率？通常有以下两种处理方法。

处理方法之一是假设第 2、3 位个体的失效时间其实并不完全相同，称为“精确边际计算法”(exact-marginal calculation，相应的 Stata 选择项为“exactm”)。如果失效时间为连续变量，则第 2、3 位个体同时失效的概率为 0。之所以在样本中存在同时失效的个体，只是由于我们的观测时间不够细致(比如，以周或月为分析时间)。基于这个思想，共有两种可能情形。第一种情形是，第 2 位个体先失效，然后第 3 位个体失效，记此情形的条件概率为 P_{23} ；第二种情形是，第 3 位个体先失效，然后第 2 位个体失效，记此情形的条件概率为 P_{32} ；而总的条件概率为 $(P_{23} + P_{32})$ 。

为了简化符号，记 $r_j \equiv e^{x_j \beta}$ ，根据前面的推导逻辑可知：

$$P_{23} = \frac{r_2}{r_2 + r_3 + r_4} \cdot \frac{r_3}{r_3 + r_4} \quad (30.40)$$

$$P_{32} = \frac{r_3}{r_2 + r_3 + r_4} \cdot \frac{r_2}{r_2 + r_4} \quad (30.41)$$

然而，如果多位个体同时失效，则精确边际法可能计算量过大。比如，有 10 位个体同时失效，则共有 $10! = 3628800$ 种情形。为此，Breslow(1974)提出了以下近似计算方法(相应的 Stata 选择项“breslow”为默认选项)，即在计算方程(30.40)与(30.41)时，不调整每一项分母中的风险集(一直使用最初的风险集)。根据 Breslow 的方法，

$$P_{23} \approx P_{32} \approx \frac{r_2 r_3}{(r_2 + r_3 + r_4)^2} \quad (30.42)$$

Efron(1977)提出了另一近似计算方法(相应的 Stata 选择项为“efron”)。Efron 方法比 Breslow 更为精确，但计算费时更长。当 $t = 3$ 时，风险集为 $\{3, 4\}$ 或 $\{2, 4\}$ ，对应的条件概率之分母分别为 $(r_3 + r_4)$ 或 $(r_2 + r_4)$ 。如果使用二者的平均数作为分母，可得

$$P_{23} \approx P_{32} \approx \frac{r_2 r_3}{(r_2 + r_3 + r_4) \left(\frac{r_2 + r_3}{2} + r_4 \right)} \quad (30.43)$$

Breslow 方法与 Efron 方法都是对精确边际法的近似。

处理方法之二是假设第 2、3 位个体确实同时失效，称为“精确部分计算法”(exact-partial

calculation, 相应的 Stata 选择项为“exactp”）。给定风险集 $\{2, 3, 4\}$ 中有两位个体同时失效，则共有三种情形，即第 2、3 位个体同时失效，第 2、4 位个体同时失效，以及第 3、4 位个体同时失效。因此，第 2、3 位个体确实同时失效的条件概率为：

$$P_{23} = \frac{r_2 r_3}{r_2 r_3 + r_2 r_4 + r_3 r_4} \quad (30.44)$$

对比表达式(30.44)与(30.40)、(30.41)可知，精确部分法与精确边际法的计算结果并不相等，但一般差别不大。究竟采取哪种方法，取决于将时间理解为连续型变量（精确边际法及其近似）还是离散型变量（精确部分法）。在实践中，如果风险集较大或同时失效的个体较多，则精确法（精确边际法或精确部分法）的计算时间可能很长，故 Stata 的默认方法为 Breslow 近似法（如果风险集不大或同时失效的个体较少，则可采取精确法）。

3. 估计基准风险函数

虽然 Cox 模型不设定基准风险 $\lambda_0(t)$ 的函数形式，但在估计 β 之后依然可以得到对 $\lambda_0(t)$ 的估计。首先，根据方程(30.12), $S(t) = \exp[-\Lambda(t)]$, 代入比例风险 $\lambda(t; x) = \lambda_0(t)e^{x\beta}$ 可得

$$\begin{aligned} S(t) &= \exp[-\Lambda(t)] = \exp\left[-\int_0^t \lambda(u; x) du\right] = \exp\left[-\int_0^t \lambda_0(u) e^{x\beta} du\right] \\ &= \exp\left[\left(-\int_0^t \lambda_0(u) du\right)(e^{x\beta})\right] = \left[\exp\left(-\int_0^t \lambda_0(u) du\right)\right]^{e^{x\beta}} \equiv S_0(t)^{e^{x\beta}} \end{aligned} \quad (30.45)$$

其中， $S_0(t) \equiv \exp\left(-\int_0^t \lambda_0(u) du\right)$ 为对应于基准风险 $\lambda_0(t)$ 的基准生存函数。根据 Cox 模型，已知 β 的估计量 $\hat{\beta}$ ；根据 KM 估计量，已知 $S(t)$ 的估计量 $\hat{S}(t)$ ；故根据方程(30.45)可得基准生存函数 $S_0(t)$ 的估计量 $\hat{S}_0(t)$ 。其次，累积风险函数可写为

$$\Lambda(t) = \int_0^t \lambda(u) du = \int_0^t \lambda_0(u) e^{x\beta} du = e^{x\beta} \int_0^t \lambda_0(u) du \equiv e^{x\beta} \Lambda_0(t) \quad (30.46)$$

其中， $\Lambda_0(t) \equiv \int_0^t \lambda_0(u) du$ 为对应于基准风险 $\lambda_0(t)$ 的基准累积风险函数。从 KM 估计量 $\hat{S}(t)$ 出发，根据 $\Lambda(t) = -\ln S(t)$ ，可得到 $\Lambda(t)$ 的估计量 $\hat{\Lambda}(t)$ 。故根据方程(30.46)可以得到 $\hat{\Lambda}_0(t)$ 。再次，根据关系式 $\Lambda_0(t) \equiv \int_0^t \lambda_0(u) du$ ，可以计算每个失效时间对于 $\hat{\Lambda}_0(t)$ 的边际贡献，即 $\hat{\lambda}_0(t)$ 。最后，使用核密度方法（参见第 27 章）将此边际贡献光滑化，即得到对 $\lambda_0(t)$ 的估计。

根据对基准风险的估计 $\hat{\lambda}_0(t)$ 以及参数估计 $\hat{\beta}$ ，可以计算在 $x = x^*$ 处的风险函数：

$$\hat{\lambda}(t; x^*) = \hat{\lambda}_0(t) e^{x^* \hat{\beta}} \quad (30.47)$$

通常，令 $x^* = \bar{x}$ ，即计算在解释变量平均值处的风险函数（对应于样本中解释变量都为平均数的假想个体的风险函数，这是 Stata 的默认取值）；也可以选择其他感兴趣的解释变量取值（特别当 x 为二值变量或离散变量时）。

30.9 比例风险模型的设定检验

比例风险模型（包括 Cox 模型）的重要假设是 $\lambda(t; x) = \lambda_0(t) e^{x\beta}$ ，如果这个假设不成立，则比例风险模型不能成立。为此，需要对比例风险模型进行设定检验，在 Stata 中称为“PH-

assumption tests”。

1. 对数 - 对数图 (log-log plot)

根据比例风险假定 $\lambda(t; \mathbf{x}) = \lambda_0(t) e^{\mathbf{x}'\boldsymbol{\beta}}$, 可将累积风险函数写为

$$\Lambda(t) = \int_0^t \lambda(u; \mathbf{x}) du = \int_0^t \lambda_0(u) e^{\mathbf{x}'\boldsymbol{\beta}} du = e^{\mathbf{x}'\boldsymbol{\beta}} \int_0^t \lambda_0(u) du = e^{\mathbf{x}'\boldsymbol{\beta}} \Lambda_0(t) \quad (30.48)$$

其中, $\Lambda_0(t)$ 为对应于 $\lambda_0(t)$ 的累积风险函数。由上式可知,

$$\ln \Lambda(t) = \mathbf{x}'\boldsymbol{\beta} + \ln \Lambda_0(t) \quad (30.49)$$

根据方程(30.11), $\Lambda(t) = -\ln S(t)$, 将此式代入方程(30.49), 且方程两边同乘(-1)可得

$$-\ln[-\ln S(t)] = -\mathbf{x}'\boldsymbol{\beta} - \ln[-\ln S_0(t)] \quad (30.50)$$

其中, $S_0(t)$ 为对应于 $\lambda_0(t)$ 的生存函数。由于 $0 < S(t) < 1$, 故 $\ln S(t) < 0$, $-\ln S(t) > 0$ 。从方程(30.50)可以看出, 函数 $-\ln[-\ln S(t)]$ 的斜率并不依赖于 \mathbf{x} 。因此, 当 \mathbf{x} 取不同值时(比如, \mathbf{x} 为虚拟变量或离散变量), 函数 $-\ln[-\ln S(t)]$ 应为相互平行的曲线, 只是截距项($-\mathbf{x}'\boldsymbol{\beta}$)不同。据此画出的图被称为“对数 - 对数图”(log-log plot)。如果对数 - 对数图中的曲线相互平行, 则支持比例风险假设; 如果不同曲线的斜率相差甚远, 则意味着比例风险假设不成立。对数 - 对数图的缺点是, 如何确定曲线是否平行带有主观性。

2. 观测 - 预测图 (observed versus expected plot)

检验比例风险假设的另一图示法为“观测 - 预测图”。针对每一解释变量 x , 分别画图。假设 x 为离散变量, 首先根据 x 的不同取值水平画其 Kaplan - Meier (KM) 生存函数图, 即观测图(observed plots)。其次, 估计 Cox 模型, 以此计算基准生存函数 $S_0(t)$, 然后代入 x 的不同取值水平, 得到相应的生存函数图, 即预测图(expected plot)。最后, 给定 x 的一个取值水平, 比较其观测图与预测图是否足够接近。如果很接近, 则表明变量 x 满足比例风险假设; 反之, 则比例风险假设不成立。观测 - 预测图的缺点同样是它的主观性。

3. 基于残差的检验 (residual-based tests)

对于像久期分析这样的非线性模型, 有多种残差的定义。其中, 对于检验比例风险假设最有用的是“舍恩菲尔德残差”(Schoenfeld residuals)。对于个体 j 与解释变量 $x_k (k=1, \dots, K)$, 假设共有 K 个解释变量, 可以计算其对应的舍恩菲尔德残差如下:

$$r_{kj} = x_{kj} - \sum_{i \in R_j} \left(x_{ki} \cdot \frac{e^{x_i'\boldsymbol{\beta}}}{\sum_{i \in R_j} e^{x_i'\boldsymbol{\beta}}} \right) \quad (30.51)$$

其中, R_j 为当个体 j 失效时的风险集。直观来看, 舍恩菲尔德残差 r_{kj} 即为失效个体的解释变量观测值 x_{kj} 减去仍处于风险集中个体的解释变量之加权平均, 而权重为“相对风险”(relative hazard) $e^{x_i'\boldsymbol{\beta}}$ 。如果比例风险假设成立, 则舍恩菲尔德残差不应随时间呈现出规律性的变化。针对每个解释变量 x_k , 都可以将其舍恩菲尔德残差与时间画图, 并考察其斜率是否为 0。进一步, 可以把舍恩菲尔德残差对时间回归, 然后检验时间的系数是否为 0。

30.10 分层 Cox 模型

如果比例风险假设不满足, 处理方法之一为“分层 Cox 模型”(stratified Cox model)。不失一般性, 假设变量 sex(性别)不满足比例风险假设, 则可将样本中个体分为两组, 第一组为男性, 第二组为女性; 并假设男性与女性的基准风险函数不同, 但参数 $\boldsymbol{\beta}$ 都相同(此假设也可以检验)。

因此,个体的风险函数可写为

$$\lambda(t; \mathbf{x}_j) = \begin{cases} \lambda_{01}(t) e^{x_j^{\beta}}, & j \text{ 为男性} \\ \lambda_{02}(t) e^{x_j^{\beta}}, & j \text{ 为女性} \end{cases} \quad (30.52)$$

其中, $\lambda_{01}(t)$ 为男性组的基准风险,而 $\lambda_{02}(t)$ 为女性组的基准风险。根据男性组的风险函数 $\lambda_{01}(t) e^{x_j^{\beta}}$,可以计算男性组的部分似然函数 L_1 ;类似地,可以计算女性组的部分似然函数 L_2 。则整个样本的部分似然函数为 $L = L_1 \cdot L_2$ 。显然,由于变量 sex 不直接出现在似然函数中(因为 sex 不满足比例风险假设,故无法将其直接引入 Cox 模型),故无法估计其效应;这是进行分层分析的代价。

进一步,如果某分层变量有 k 个取值水平,则可将样本中个体分为 k 层或 k 组,然后进行分层估计。更一般地,如果有多个分层变量,则应考虑它们的交叉情形。比如,变量 x 有 3 种可能取值,变量 z 有 2 种可能取值,二者都不满足比例风险假设,故同为分层变量;此时,应将样本分为 6 组,使得每组内的 x, z 取值都相同。在 Stata 中,只需指定分层变量(可以为多个分层变量),其余的计算都可自动进行。

分层 Cox 模型的一个重要假定为,不同组的参数 β 都相同。为了检验此假定,可以引入分层变量与留在 Cox 模型中变量的互动项进行分层分析,然后检验所有这些互动项的联合显著性。如果不显著,则支持“不同组参数 β 都相同”的原假设。

30.11 随时间而变的解释变量

如果比例风险假设不满足,处理方法之二为引入随时间而变的解释变量 (Time – Varying Covariates,简记 TVC)。比如,在持续失业的过程中,宏观经济状况可能发生变化,从而影响就业概率。引入 TVC 后,需要注意可能导致内生性。一般来说,不由个体控制的变量(比如宏观经济)是外生的;而个体可以影响的变量则可能为内生的。例如,在考察婚姻持续时间时,如果使用孩子的数量作为 TVC 则可能导致内生性问题。由于稳定的婚姻会增加孩子的数量,故根据数据得出的“孩子越多,婚姻越持久”结论并不可靠。下面,假设 TVC 为外生。

引入 TVC 后,可将风险函数写为

$$\lambda(t; \mathbf{x}(t)) = \lambda_0(t) e^{x(t)\beta} \quad (30.53)$$

其中,解释变量 $\mathbf{x}(t)$ 可以时间而变,而其系数 β 仍为固定。从上式可以看出,一旦解释变量 $\mathbf{x}(t)$ 的取值变化,其风险函数也立即发生相应变化。如果要考虑滞后效应,可引入 $\mathbf{x}(t-1)$ 作为解释变量。

根据方程(30.53),个体 i 与个体 j 的风险函数之比可写为

$$\frac{\lambda(t; \mathbf{x}_i(t))}{\lambda(t; \mathbf{x}_j(t))} = \frac{\lambda_0(t) e^{x_i(t)\beta}}{\lambda_0(t) e^{x_j(t)\beta}} = e^{[x_i(t) - x_j(t)]\beta} \quad (30.54)$$

显然,此风险函数之比是时间 t 的函数,故此模型不再满足比例风险假设,称为“扩展 Cox 模型”(extended Cox model)。但这对于基于似然函数(参数回归)或部分似然函数(Cox 模型)的估计并无实质影响。这是因为,现实数据总是离散的(譬如年度数据、月度数据、日度数据),故可将每位个体的单一记录分为几条记录,对应于几个时间段,使得在每个时间段内 TVC 为常数,然后沿用解释变量不随时间而变的方法进行估计。如果已将数据设为此格式,则无论是否包括

TVC 变量,其 Stata 命令并无区别。

作为一个简单例子^①,假设个体 1 从时间 0 至 T 一直处于失业状态,并在时间 T 找到工作。假设 $0 < t_1 < t_2 < T$ 。共有两个解释变量,其中 x_1 不随时间而变(假设为虚拟变量,比如性别);而 $x_2(t)$ 可以随时间而变,在区间 $[0, t_1], [t_1, t_2], [t_2, T]$ 的取值分别为 $x_2(t_1), x_2(t_2)$ 与 $x_2(T)$,故为阶梯函数。则个体 1 的信息由三条记录构成,参见表 30.1。

表 30.1 个体 1 的相关记录

个体 ID	持续时间	x_1	$x_2(t)$	归并虚拟变量 d
1	t_1	1	$x_2(t_1)$	0
1	t_2	1	$x_2(t_2)$	0
1	T	1	$x_2(T)$	1

由表 14.1 可知,个体 1 的完整信息被分为三个部分,对应于三条记录。在第 1、2 条记录,个体 1 均未失效,而以归并结束,故归并虚拟变量为 0;在第 3 条记录,个体 1 在时间 T 失效,故归并虚拟变量为 1。

扩展 Cox 模型的用途之一是用来检验比例风险假设。回到解释变量不随时间而变的情形,可以将比例风险假设解释为风险函数 $\lambda(t; \mathbf{x}) = \lambda_0(t) e^{x'\beta}$ 的参数 β 不随时间而变。考虑最简单的一维情形。假设 β 随时间而变,并将其写为 $\beta + \gamma g(t)$,其中 $g(t)$ 为时间的函数(比如, t 或 $\ln t$),则风险函数可写为

$$\lambda(t; \mathbf{x}) = \lambda_0(t) e^{x[\beta + \gamma g(t)]} = \lambda_0(t) e^{x\beta + \gamma xg(t)} \quad (30.55)$$

这相当于是在原模型中引入了一个 TVC,即 $x \cdot g(t)$,由不随时间而变的解释变量 x 与时间的函数 $g(t)$ 的乘积构成,是经人为定义而生成的 TVC。如果 $\gamma = 0$,则方程(30.55)简化为标准 Cox 模型的风险函数。因此,可通过检验原假设“ $H_0: \gamma = 0$ ”来检验比例风险假设。通常选择 $g(t)$ 为 t 或 $\ln t$ 。在多维情形下,则引入所有解释变量 \mathbf{x} 与 $g(t)$ 的乘积,然后检验所有这些乘积项的系数联合显著性。在此检验时,由于引入了原模型中所没有的 TVC,故需要调整数据格式,将个体的信息根据需要分为多条记录。幸运的是,在 Cox 模型中,Stata 可自动完成这一操作,只要使用选择项“`tvc(var) texp(exp)`”即可,前者用来指定 \mathbf{x} ,而后者用来指定 $g(t)$ (详见本章最后一节)。

30.12 不可观测的异质性

在进行久期分析时,样本中的个体可能具有“不可观测的异质性”(unobserved heterogeneity),也称为“弱质”(frailty)。比如,在研究病人的存活时间时,病人的体质或弱质并不能完全观察,但却影响病人的死亡风险。在线性模型中,忽略不可观测的异质性相当于遗漏变量;如果此遗漏变量与解释变量不相关,则不影响估计量的一致性。然而,在久期分析这样的非线性模型中,即使不可观测的异质性与解释变量不相关,仍会导致不一致估计。

一个经典的例子是对于失业持续时间的研究。经验数据表明,失业时间越长,则找到工作的概率越低,即风险率随着时间而下降。如果样本中的每位个体完全一样,则意味着负向久期依赖

^① 此例来自 Cameron and Trivedi(2005, p. 598)。

(negative duration dependence)。但事实可能并非如此。假设失业人口可以分为两类,即 F 类 (fast, 其风险函数恒等于 0.4) 与 S 类 (slow, 其风险函数恒等于 0.1), 这两类各占一半。由于 F 类的风险率高于 S 类, 故 F 类平均地比 S 类更快找到工作而退出样本, 导致样本中 F 类占比随时间而下降, 引起整体的风险率下降。具体来说, 假设样本中有 200 人, 其中 100 人为 F 类, 另 100 人为 S 类。在 100 位 F 类人中, 第 1 期将有 40 人找到工作, 第 2 期将有 24 人找到工作, 而第 3 期将有 14.4 人找到工作。另一方面, 在 100 位 S 类人中, 第 1 期将有 10 人找到工作, 第 2 期将有 9 人找到工作, 而第 3 期将有 8.1 人找到工作。因此, 就整个样本而言, 第 1 期的风险率为 $(40+10)/200 = 0.25$, 第 2 期的风险率为 $(24+9)/150 = 0.22$, 第 3 期的风险率为 $(14.4+8.1)/117 = 0.192$, 呈逐渐下降的趋势。显然, 整体风险率的逐渐下降, 只是由于 F 类与 S 类的结构变化而造成, 而个体的风险函数始终为常数。此效应被称为“弱质效应”(frailty effect)。对于威布尔模型, 风险函数为 $\lambda(t) = \gamma p t^{p-1}$, 其中参数 p 决定风险函数的斜率, 表示久期依赖的方向与程度。如果存在不可观测的异质性而被忽略, 则久期依赖参数 p 将被系统地低估, 得不到一致估计。

解决方法是直接将不可观测的异质性引入模型中。假设个体 j 的风险函数为

$$\lambda(t; \mathbf{x}_j, v_j) = \lambda_0(t) e^{\mathbf{x}_j' \boldsymbol{\beta}} v_j, \quad v_j > 0 \quad (30.56)$$

其中, 不可观测的异质性 $v_j > 0$ 以乘积的形式进入风险函数。 v_j 越大, 则个体 j 越弱质, 失效的风险越高。 $\lambda(t; \mathbf{x}_j, v_j)$ 称为“条件风险函数”(conditional hazard function), 因为它是给定 v_j 条件下的风险函数。从方程(30.56)出发, 可以得到生存函数:

$$S(t; \mathbf{x}_j, v_j) = e^{-\Lambda(t)} = \exp\left[-\int_0^t \lambda_0(u) e^{\mathbf{x}_j' \boldsymbol{\beta}} v_j du\right] = \left[\exp\left(-\int_0^t \lambda_0(u) e^{\mathbf{x}_j' \boldsymbol{\beta}} du\right)\right]^{v_j} = [S(t; \mathbf{x}_j)]^{v_j} \quad (30.57)$$

其中, $S(t; \mathbf{x}_j)$ 为不含异质性的生存函数。由于 v_j 不可观测, 故需要假设其概率分布, 然后将 v_j 积分掉, 得到无条件的风险函数与生存函数。假设 v_j 的概率密度函数为 $g(v_j)$, 则无条件风险函数与生存函数分别为

$$\lambda(t; \mathbf{x}_j) = \int_0^\infty \lambda(t; \mathbf{x}_j, v_j) g(v_j) dv_j \quad (30.58)$$

$$S(t; \mathbf{x}_j) = \int_0^\infty S(t; \mathbf{x}_j, v_j) g(v_j) dv_j \quad (30.59)$$

为了使方程(30.58)与(30.59)有解析解, 常假设原模型为威布尔模型(指数模型为其特例), 而 v_j 服从 $Gamma(\delta, \delta)$ 分布或“逆正态分布”(Inverse Gaussian, 简记 IG); 这类模型称为“混合模型”(mixture model)。将 v_j 的期望值标准化为 1(故 $v_j = 1$ 代表平均弱质, average frailty), 并记其方差为 θ (在 Stata 记为 theta)。如果 $\theta = 0$, 则 v_j 退化为常数, 不存在异质性; 故可通过检验 “ $H_0: \theta = 0$ ”来判断是否存在异质性。由于不可观测的异质性不再出现在无条件风险函数(30.58)与生存函数(30.59), 故可对其进行 MLE 估计。

另外, 还可考虑依某变量的取值将个体分组, 然后假设组内个体为同质, 而不同组的个体具有异质性, 这被称为“共享异质性”(shared frailty), 即该异质性为同组内成员所共有。与此对应, 上文的个体异质性则被称为“非共享异质性”(unshared frailty)。

30.13 其他久期分析模型

迄今为止, 我们假设每位个体一旦离开其持续的状态, 就不可能再回去。比如, 考察病人存

活时间时,病人一旦去世,就无法复活。这被称为“单一久期”(single-spell)数据。然而,在考察失业持续时间时,失业者找到工作后,还可能再次失业,导致“多重久期”(multiple-spell)数据。多重久期数据的建模难点在于,不同阶段的失业时间可能相关。

另外,我们一直假设个体只可能存在两种状态,比如生或死,失业或就业。现实中还可能出现在多种状态间转换的情形,比如劳动者的状态可以包括失业、全职就业及兼职就业。此时,失业者可能转换到两种不同的状态中,即全职就业或兼职就业。这类模型称为“竞争风险模型”(competing risk model),需要同时考虑多个风险函数。

进一步学习久期分析的资源包括 Cameron and Trivedi (2005, Chapter 17~19), Cleves et al (2004), Hosmer and Lemeshow (1999), Kleinbaum and Klein (2011), Lancaster (1990) 与 van den Berg (2001)。

30.14 久期分析的 Stata 命令及实例

Stata 称久期数据为“生存时间”数据(survival time, 简记 st),故所有久期分析的命令都以 st 开头(正如所有面板命令都以 xt 开头)。在进行久期分析之前,首先要告诉 Stata 我们的数据为久期数据,其命令格式为

```
stset duration_var,failure(failvar) id(idvar)
```

其中,“duration_var”表示持续时间,即久期分析的被解释变量;选择项“failure(failvar)”表示归并虚拟变量(失效 = 1, 归并 = 0);选择项“id(idvar)”用来确定个体 ID(适用于一位个体有多条记录的情形)^①。有关命令 stset 的更多选择项,参见“help stset”。

运行 stset 命令后,Stata 将自动生成四个新变量_t0,_t,_d,_st,分别表示记录开始的时间,记录结束的时间,归并虚拟变量(失效 = 1, 归并 = 0),以及有效观测值虚拟变量(与当前分析有关的观测值 = 1, 其他 = 0)。Stata 的后续计算将基于此四个变量,而忽略数据集中原有的时间或事件变量。因此,如果中途改变原始数据,则应再次运行 stset 命令,以更新变量_t0,_t,_d,_st 的取值。

stdes (描述久期数据)

stsum (显示久期数据的特征)

sts graph,survival cumhaz hazard ci by(varlist) (画图,描述性分析)

其中,选择项“survival”为默认选项,表示画 Kaplan-Meier 生存函数;选择项“cumhaz”表示画累积风险函数;选择项“hazard”表示画风险函数;选择项“ci”表示同时画出置信区间;选择项“by(varlist)”表示根据某变量的不同取值,划分子样本,然后分别画图。

如果需要将一条记录分为几条记录,可以用以下命令:

sts split newvar,at(numlist) every(#) (在指定的时间点分割数据)

sts split,at(failures) (在失效时间分割数据)

如果进行参数回归,可使用以下命令:

streg x1 x2 x3,r distribution(dist_var) frailty(frailty_dist) shared

^① 有时同一个体的持续时间被分时段记录在案,对于这种多记录数据(multiple-record data),必须指明 ID 以识别每位个体。

```
(varname) strata(strata_var) nolog noshow time
```

其中，“`x1 x2 x3`”为解释变量。命令 `streg` 的独特之处在于，不需要指定被解释变量 `y`，因为在运行 `stset` 命令时，已经确定了被解释变量。选择项“`r`”表示使用稳健标准误。选择项“`distribution(dist_var)`”用来指定概率分布，备选分布包括 `exponential`, `weibull`, `gompertz`, `loglogistic`, `lognormal`, `gamma`。选择项“`frailty(frailty_dist)`”表示使用包含不可观测的异质性模型，备选的异质性分布包括 `gamma` 与 `invgaussian`。选择项“`shared(varname)`”表示界定“共享异质性”(shared frailty)的变量。选择项“`strata(strata_var)`”表示进行分层分析，并指定分层变量。选择项“`nolog`”表示不显示 MLE 计算的迭代过程，选择项“`noshow`”表示不显示有关久期数据设定(`stset`)的信息(默认显示)。选择项“`time`”表示估计加速失效模型，该选项只适用于 `exponential` 和 `weibull` 分布，因为只有这两者既属 AFT 又属 PH 模型。

如果估计半参数 Cox 模型，可使用以下命令：

```
stcox x1 x2 x3,r frailty(frailty_dist) shared(varname) strata(strata_var) breslow efron exactm exactp nolog noshow
```

命令 `stcox` 的结构与选择项与 `streg` 大致相同。选择项“`breslow`, `efron`, `exactm`, `exactp`”分别表示使用 Breslow 方法(默认值)、Efron 方法、精确边际法、精确部分法来处理失效时间重叠(ties)的情形。

使用命令 `streg` 或 `stcox` 后，可用以下命令来得到相关函数的直观图形：

```
stcurve,survival cumhaz hazard alpha1 at(varname1 = # varname2 = #)
```

其中，选择项“`survival`”表示画生存函数，选择项“`cumhaz`”表示画累积风险函数，选择项“`hazard`”表示画风险函数。在存在不可观测异质性的情况下，选择项“`alpha1`”表示根据给定异质性的“条件异质性模型”(conditional frailty model)进行画图；默认值为根据“无条件异质性模型”(unconditional frailty model)进行画图。选择项“`at (varname1 = # varname2 = #)`”表示，在变量 `varname1 = #`，变量 `varname2 = #`，而其他变量等于均值处画图；默认值为在所有变量的均值处画图。

检验比例风险假定的命令包括：

```
stphplot,by(varname)
```

```
stcoxkm,by(varname)
```

```
estat phtest,detail plot(varname)
```

其中，命令“`stphplot,by(varname)`”(表示 st proportional hazard plot)根据变量 `varname` 画对数 - 对数图。命令“`stcoxkm,by(varname)`”(表示 st cox Kaplan - Meier) 根据变量 `varname` 画观测 - 预测图；此命令可单独使用，无须先运行 `stcox` 命令。在运行命令 `stcox` 之后，可使用命令“`estat phtest`”进行基于舍恩菲尔德残差的检验，选择项“`detail`”表示对每个解释变量都分别进行检验(默认为仅进行联合检验)，而选择项“`plot(varname)`”表示将变量 `varname` 的检验结果画图(这两个选择项不能并用)。

下面以数据集 `recid.dta` 为例^①。该数据集用来研究刑满释放犯再次被捕的时间，记为 `durat`(单位为月)。首先，对在 1977 年 7 月 1 日至 1978 年 6 月 30 日期间释放的罪犯进行随机抽样，然后于 1984 年 4 月考察这些刑满释放犯的再被捕记录。因此，1984 年 4 月为共同的归并日

^① 该数据集来自 Wooldridge(2010)，最早为 Chung et al(1991) 所使用。

期,但由于每位犯人开始自由的时间不尽相同,故归并时间 C_i^* 仍可不同。解释变量包括: workprg(是否参加狱中劳动项目), priors(前科次数), tserved(狱中关押月数), felon(是否犯重罪), alcohol(是否酗酒), drugs(是否有吸毒史), black(是否黑人), married(在狱中时是否已婚), educ(教育年限), age(年龄), 其中 workprg 为我们主要关心的变量。另外,归并虚拟变量为 cens(1 = 右归并)。

(1) 首先,调用此数据集,定义为久期数据,并考察数据特征。

```
. use recid.dta, clear
. stset durat, failure(cens = 0)
```

其中,由于归并虚拟变量的定义为“1 = 右归并,0 = 失效”(与通常的定义相反),故使用选择项“failure(cens = 0)”,而不使用“failure(cens)”。

failure event: cens == 0	
obs. time interval: (0, durat]	
exit on or before: failure	
<hr/>	
1445 total obs.	
0 exclusions	
<hr/>	
1445 obs. remaining, representing	
552 failures in single record/single failure data	
80013 total analysis time at risk, at risk from t = 0	
earliest observed entry t = 0	
last observed exit t = 81	

上表显示,失效事件以“cens == 0”来定义,在 1 445 个有效观测值中,552 位个体失效,而其他个体被归并。下面,看一下久期数据的基本特征。

```
. stdes
```

Category	total	per subject			
		mean	min	median	max
no. of subjects	1445				
no. of records	1445	1	1	1	1
(first) entry time	0	0	0	0	0
(final) exit time	55.37232	1	71	71	81
subjects with gap	0				
time on gap if gap	0				
time at risk	80013	55.37232	1	71	81
failures	552	.3820069	0	0	1

(2) 其次,进行描述性分析,即画生存函数、累积风险函数与风险函数。

```
. sts graph,ci
```

该命令的默认值为画样本生存函数,即 Kaplan-Meier 估计量,结果参见图 30.1。

由于我们关心参加狱中劳动项目的效应,故以 workprg 的取值将全样本一分为二,考察两个子样本的生存函数是否有区别。

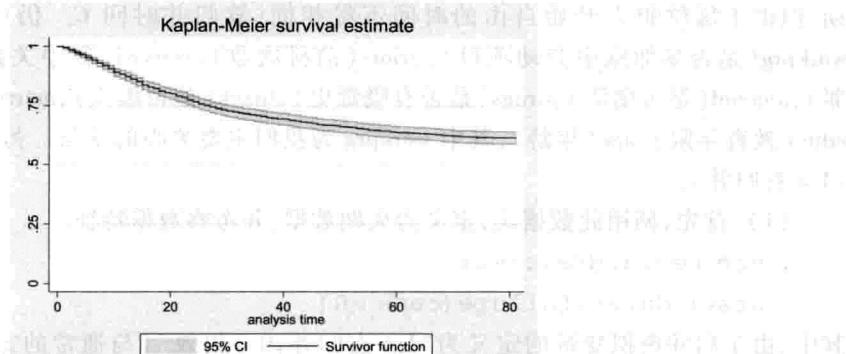


图 30.1 Kaplan-Meier 生存函数估计量

```
. sts graph,by(workprg) plot2opts(lp("-"))
```

其中,选择项“plot2opts(lp(" - "))”表示第二个图以虚线表示,便于与第一个图的默认实线相区别。结果参见图 30.2。

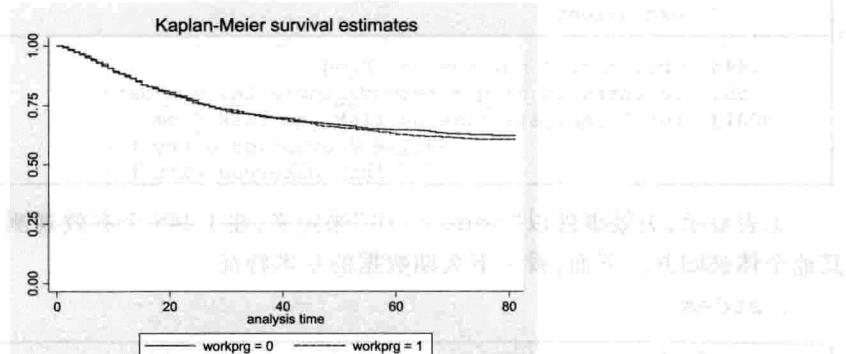


图 30.2 根据 workprg 分组的 Kaplan-Meier 生存函数

从图 30.2 可知,是否参加狱中劳动项目,对于生存函数并无明显影响。这预示着,狱中劳动项目对于罪犯再次被捕的时间可能无显著影响。

```
. sts graph,cumhaz ci
```

其中,选择项“cumhaz”表示画累积风险函数,结果参见图 30.3。

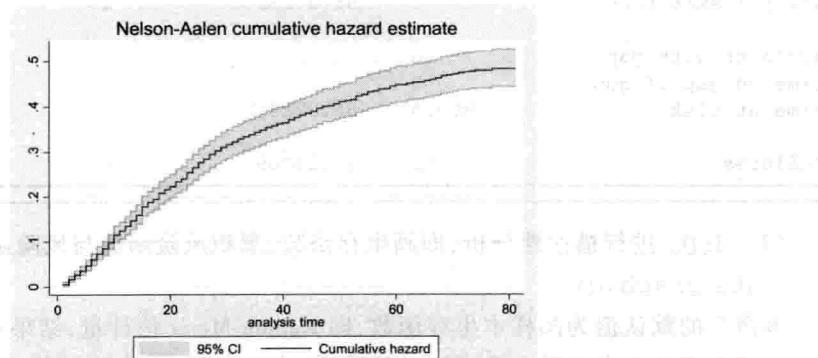


图 30.3 Nelson-Aalen 累积风险函数估计量

```
. sts graph,hazard ci
```

其中,选择项“hazard”表示画风险函数,结果参见图 30.4。

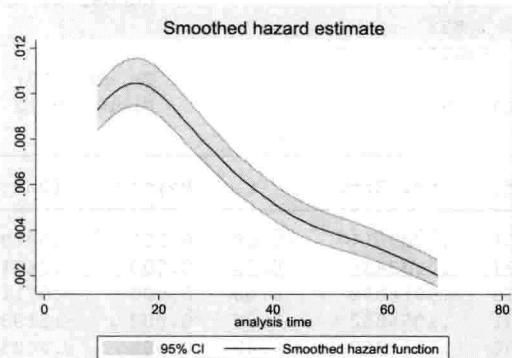


图 30.4 风险函数估计量

从图 30.4 可知,在罪犯出狱后再次被捕的风险率将首先上升,然后才下降。下面,根据 workprg 的取值分组画风险函数,结果参见图 30.5。

```
. sts graph,hazard by (workprg) plot2opts(lp("-"))
```

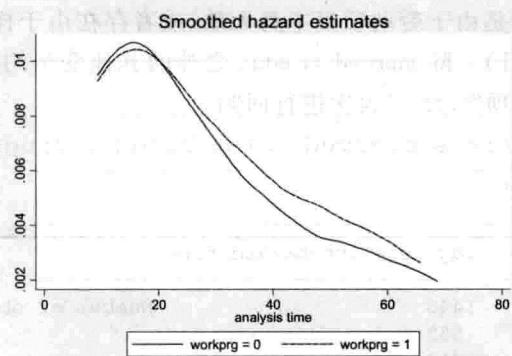


图 30.5 根据 workprg 分组的风险函数

从图 30.5 可知,是否参加狱中劳动项目对于风险函数影响不大。甚至在相当大的区间内,参加劳动项目者的风险率反而更高。当然,由于未引入解释变量,以上分析只是对被解释变量 durat 无条件分布的初步分析。

(3) 进行参数回归。首先进行最简单的单参数指数回归。

```
. streg workprg priors tserved felon alcohol drugs black married educ  
age,d(e) nohr nolog noshow
```

其中,选择项“nohr”表示显示回归系数,不显示风险比率(hazard ratio),默认显示风险比率。

Exponential regression -- log relative-hazard form						
No. of subjects =	1445	Number of obs	=	1445		
No. of failures =	552					
Time at risk =	80013			LR chi2(10) =		180.28
Log likelihood =	-1649.7531			Prob > chi2 =		0.0000
_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	.0955801	.0905511	1.06	0.291	-.0818968	.2730569
priors	.091337	.0133434	6.85	0.000	.0651845	.1174896
tserved	.0144009	.0016599	8.68	0.000	.0111475	.0176543
felon	-.3122241	.1056552	-2.96	0.003	-.5193045	-.1051437
alcohol	.4676706	.1057752	4.42	0.000	.2603551	.6749862
drugs	.2941648	.0978584	3.01	0.003	.1023658	.4859637
black	.4756749	.0881961	5.39	0.000	.3028137	.6485362
married	-.1519512	.1092578	-1.39	0.164	-.3660925	.0621901
educ	-.0242123	.0194075	-1.25	0.212	-.0622503	.0138256
age	-.0039112	.0005282	-7.40	0.000	-.0049464	-.002876
_cons	-4.169213	.2756276	-15.13	0.000	-4.709433	-3.628993

上表显示,参加狱中劳动项目(workprg = 1)反而会增大再次被捕的风险,尽管统计上并不显著(p 值为0.291)。这可能是由于劳动项目本身无效,或者存在由于样本选择问题(表现恶劣的罪犯被优先安排参加此项目)。除 married 与 educ 之外的其他变量均很显著。为了方便解释各变量的效应,下面去掉选择项“nohr”再次进行回归。

```
. streg workprg priors tserved felon alcohol drugs black married educ
age,d(e) nolog noshow
```

Exponential regression -- log relative-hazard form						
No. of subjects =	1445	Number of obs	=	1445		
No. of failures =	552					
Time at risk =	80013			LR chi2(10) =		180.28
Log likelihood =	-1649.7531			Prob > chi2 =		0.0000
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	1.100297	.0996331	1.06	0.291	.921367	1.313975
priors	1.095638	.0146195	6.85	0.000	1.067356	1.12467
tserved	1.014505	.001684	8.68	0.000	1.01121	1.017811
felon	.7318175	.0773203	-2.96	0.003	.5949342	.9001952
alcohol	1.596272	.1688459	4.42	0.000	1.297391	1.964006
drugs	1.342005	.1313265	3.01	0.003	1.107789	1.625741
black	1.6091	.1419164	5.39	0.000	1.353662	1.912739
married	.8590302	.0938557	-1.39	0.164	.6934387	1.064165
educ	.9760784	.0189432	-1.25	0.212	.9396476	1.013922
age	.9960964	.0005261	-7.40	0.000	.9950658	.9971282
_cons	.0154644	.0042624	-15.13	0.000	.0090099	.0265429

是否使用选择项“nohr”不影响回归的结果，只是决定汇报系数(Coef.)还是风险比率(Haz. Ratio)而已。例如，上表显示，alcohol 的风险比率约为 1.60，这意味着酗酒者比不酗酒者再次被捕的风险高 60%。由于指数回归的风险函数为常数，假定太强，故下面放松此假定，进行威布尔回归。

```
. streg workprg priors tserved felon alcohol drugs black married educ  
age,d(w) nolog noshow
```

Weibull regression -- log relative-hazard form						
No. of subjects =	1445	Number of obs =	1445			
No. of failures =	552					
Time at risk =	80013					
Log likelihood =	-1633.0325				LR chi2(10) =	165.48
					Prob > chi2 =	0.0000
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	1.095148	.0992728	1.00	0.316	.9168814	1.308074
priors	1.092848	.014683	6.61	0.000	1.064445	1.122008
tserved	1.013655	.0017037	8.07	0.000	1.010321	1.017
felon	.7412054	.0785485	-2.83	0.005	.6021898	.9123128
alcohol	1.564179	.165389	4.23	0.000	1.271406	1.92437
drugs	1.325064	.1296765	2.88	0.004	1.093791	1.605237
black	1.574149	.1390031	5.14	0.000	1.32398	1.871587
married	.8593436	.0938794	-1.39	0.165	.6937084	1.064527
educ	.9769709	.0189724	-1.20	0.230	.9404845	1.014873
age	.9962823	.000523	-7.09	0.000	.9952577	.997308
_cons	.0333035	.0100249	-11.30	0.000	.0184613	.0600781
/ln_p	-.2158398	.0389149	-5.55	0.000	-.2921115	-.1395681
p	.8058644	.0313601			.7466852	.8697338
1/p	1.240904	.0482896			1.149777	1.339252

从上表可知，参加劳动项目(workprg)的效应依然为正且不显著。从上表倒数第三行还可知，原假设“ $H_0: \ln p = 0$ ”对应的 p 值为 0.000，故强烈拒绝指数回归，认为应使用威布尔回归。进一步，根据上表倒数第二行， $\hat{p} = 0.806 < 1$ ，故风险函数随时间而递减，存在负向久期依赖。更直观地，可以通过画图来看风险函数，结果参见图 30.6。

```
. stcurve,hazard
```

下面，进行指数回归的另一推广，即冈珀茨回归。

```
. streg workprg priors tserved
```

```
felon alcohol drugs black married educ age,d(gom) nolog noshow
```

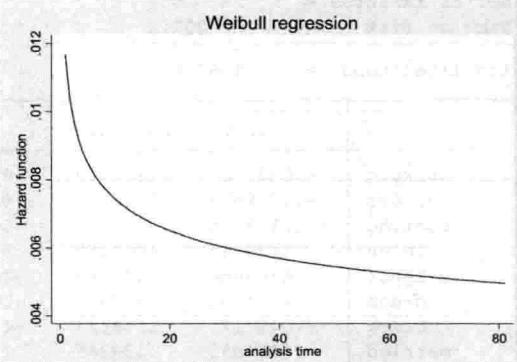


图 30.6 威布尔回归的风险函数

Gompertz regression -- log relative-hazard form						
No. of subjects	=	1445	Number of obs	=	1445	
No. of failures	=	552				
Time at risk	=	80013	LR chi2(10)	=	154.39	
Log likelihood	=	-1595.5837	Prob > chi2	=	0.0000	
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	1.089068	.0988625	0.94	0.347	.9115602	1.301142
priors	1.090805	.0147222	6.44	0.000	1.062328	1.120045
tserved	1.012968	.001712	7.62	0.000	1.009618	1.016329
felon	.7516398	.0798318	-2.69	0.007	.6103845	.9255845
alcohol	1.536163	.1623417	4.06	0.000	1.248769	1.889697
drugs	1.313292	.1285117	2.79	0.005	1.084095	1.590945
black	1.54159	.1362263	4.90	0.000	1.296434	1.833105
married	.8581703	.0937313	-1.40	0.161	.6927928	1.063025
educ	.9783779	.0190234	-1.12	0.261	.9417942	1.016383
age	.9964432	.0005198	-6.83	0.000	.995425	.9974624
_cons	.0270245	.0075192	-12.98	0.000	.0156648	.0466221
/gamma	-.021703	.002192	-9.90	0.000	-.0259992	-.0174068

上表最后一行显示,原假设“ $H_0: \gamma = 0$ ”的 p 值为 0.000,故强烈拒绝指数回归,认为应选择冈珀茨回归。至于在威布尔回归与冈珀茨回归之间进行选择,则可根据 AIC 准则进行。由于这两个回归的解释变量个数与参数个数均相等,故 AIC 准则等价于最大化对数似然函数,参见公式(30.32)。由于威布尔回归的对数似然值为 -1 633,而冈珀茨回归的对数似然值为 -1 596,故应选择冈珀茨回归。

其他备选的参数回归还包括属于加速失效时间模型(AFT)的对数正态回归与对数逻辑回归。首先进行对数正态回归。

```
. streg workprg priors tserved felon alcohol drugs black married educ  
age, d(logn) nolog noshow
```

Lognormal regression -- accelerated failure-time form						
No. of subjects	=	1445	Number of obs	=	1445	
No. of failures	=	552				
Time at risk	=	80013	LR chi2(10)	=	166.74	
Log likelihood	=	-1597.059 <th>Prob > chi2</th> <td>=</td> <td>0.0000</td> <td></td>	Prob > chi2	=	0.0000	
_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	-.0625714	.1200369	-0.52	0.602	-.2978394	.1726965
priors	-.1372528	.0214587	-6.40	0.000	-.179311	-.0951946
tserved	-.0193305	.0029779	-6.49	0.000	-.0251671	-.0134939
felon	.4439944	.1450865	3.06	0.002	.1596302	.7283586
alcohol	-.6349088	.1442165	-4.40	0.000	-.9175681	-.3522496
drugs	-.2981599	.1327355	-2.25	0.025	-.5583168	-.0380031
black	-.5427175	.1174427	-4.62	0.000	-.772901	-.312534
married	.3406835	.139843	2.44	0.015	.0665962	.6147707
educ	.0229195	.0253974	0.90	0.367	-.0268584	.0726975
age	.0039103	.0006062	6.45	0.000	.0027221	.0050984
_cons	4.099386	.3475349	11.80	0.000	3.41823	4.780542
/ln_sig	.5935861	.0344122	17.25	0.000	.5261395	.6610327
sigma	1.810469	.0623022			1.692386	1.936791

从上表可知,对数正态回归的系数符号正好与比例风险模型的系数符号相反,因为二者的经济意义不同。加速失效时间模型中的系数表示该变量对平均寿命的半弹性,而比例风险模型中的系数表示对风险率的半弹性。显然,风险率越高,则平均寿命越短;故二者的作用方向正好相反。对数正态回归的对数似然值为 -1 597,略逊于冈珀茨回归。

下面,进行对数逻辑回归。

```
. streg workprg priors tserved felon alcohol drugs black married educ  
age,d(logl) nolog noshow
```

Loglogistic regression -- accelerated failure-time form							
No. of subjects =	1445	Number of obs =	1445				
No. of failures =	552						
Time at risk =	80013						
			LR chi2(10) =		178.92		
			Prob > chi2 =		0.0000		
Log likelihood = -1610.9431							
_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
workprg	-.0664501	.1195644	-0.56	0.578	-.300792	.1678918	
priors	-.148287	.0230937	-6.42	0.000	-.1935498	-.1030243	
tserved	-.0191425	.0029906	-6.40	0.000	-.025004	-.013281	
felon	.4108081	.1436511	2.86	0.004	.1292572	.6923591	
alcohol	-.6325083	.1427802	-4.43	0.000	-.9123523	-.3526643	
drugs	-.3493944	.129943	-2.69	0.007	-.6040781	-.0947107	
black	-.5612106	.1161105	-4.83	0.000	-.7887831	-.3336381	
married	.2708632	.1387666	1.95	0.051	-.0011143	.5428407	
educ	.0314568	.0253634	1.24	0.215	-.0182546	.0811682	
age	.0045069	.0006628	6.80	0.000	.0032079	.005806	
_cons	3.817714	.3509807	10.88	0.000	3.129805	4.505624	
/ln_gam	.0347687	.0375148	0.93	0.354	-.0387589	.1082964	
gamma	1.03538	.0388421			.9619826	1.114378	

从上表可知,对数逻辑回归的对数似然值为 -1 610,也低于冈珀茨回归。最后,进行最具灵活性的三参数广义伽马回归。

```
. streg workprg priors tserved felon alcohol drugs black married educ  
age,d(gam) nolog noshow
```

Gamma regression -- accelerated failure-time form						
No. of subjects =	1445	Number of obs	1445			
No. of failures =	552					
Time at risk	80013					
Log likelihood	-1588.8622			LR chi2(10)	=	144.06
				Prob > chi2	=	0.0000
_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	-.0286748	.1215211	-0.24	0.813	-.2668519	.2095023
priors	-.1403375	.0235349	-5.96	0.000	-.186465	-.0942099
tserved	-.0209601	.0036622	-5.72	0.000	-.0281378	-.0137824
felon	.5244066	.1532514	3.42	0.001	.2240394	.8247738
alcohol	-.6328808	.148593	-4.26	0.000	-.9241178	-.3416439
drugs	-.1892097	.1400017	-1.35	0.177	-.4636079	.0851885
black	-.4965192	.1206089	-4.12	0.000	-.7329084	-.26013
married	.4761969	.1433174	3.32	0.001	.1953	.7570938
educ	.0086858	.0264798	0.33	0.743	-.0432136	.0605852
age	.0029722	.0005984	4.97	0.000	.0017994	.0041451
cons	3.969667	.3481397	11.40	0.000	3.287326	4.652009
/ln_sig	.7303791	.0375734	19.44	0.000	.6567367	.8040215
/kappa	-.8128316	.2073818	-3.92	0.000	-1.219292	-.4063708
sigma	2.075867	.0779973			1.928489	2.234509

从上表倒数第二行可知,参数 κ 的 95% 置信区间为 (-1.22, -0.41), 故显著地不等于 0(对应于对数正态回归)或 1(对应于威布尔回归),故拒绝对数正态回归与威布尔回归,认为应使用广义伽马回归。下面,画出根据广义伽马回归计算的风险函数,结果参见图 30.7。

. stcurve,hazard

根据图 30.7 可知,广义伽马模型的风险函数首先上升然后才下降,这与威布尔模型单调递减的风险函数大不相同(参见图 30.6)。最后,使用 AIC 信息准则在广义伽马回归与冈珀茨回归之间进行选择。首先计算广义伽马回归的 AIC。

. dis -2 * e(11) + 2 * e(k)

3203.7245

其中, $e(11)$ 为对数似然函数,而 $e(k)$ 为模型中参数(包含解释变量系数)的个数。

然后计算冈珀茨回归的 AIC。

. qui streg workprg priors tserved felon alcohol drugs black married
educ age,d(gom) nolog noshow

. dis -2 * e(11) + 2 * e(k)

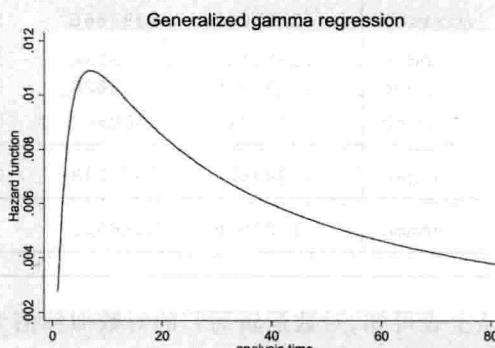


图 30.7 广义伽马回归的风险函数

3215.1673

由以上计算可知,广义伽马回归的 AIC 小于冈珀茨回归,在上述所有参数回归中为最小。总之,如果严格按照 AIC 准则来选择参数回归,则应使用广义伽马回归。

(4) 以上分析未考虑每位罪犯的个体异质性,而此异质性可能影响回归结果。为了与 Wooldridge(2010)的结果对比,将不可观测的异质性引入威布尔回归,并假设异质性服从伽马分布。

```
. streg workprg priors tserved felon alcohol drugs black married educ  
age,d(w) frailty(gamma) nolog noshow
```

Weibull regression -- log relative-hazard form						
Gamma frailty						
No. of subjects =	1445					
No. of failures =	552					
Time at risk =	80013					
Log likelihood =	-1584.9172					
	Number of obs = 1445					
	LR chi2(10) = 143.82					
	Prob > chi2 = 0.0000					
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	1.00741	.2053883	0.04	0.971	.6755623	1.502267
priors	1.275214	.0537558	5.77	0.000	1.17409	1.385049
tserved	1.035554	.0072673	4.98	0.000	1.021408	1.049896
felon	.4534124	.1208835	-2.97	0.003	.2688784	.7645939
alcohol	3.233478	.9070623	4.18	0.000	1.865903	5.60339
drugs	1.329452	.2968761	1.28	0.202	.8582091	2.059453
black	2.163173	.4409171	3.79	0.000	1.450748	3.225453
married	.4467732	.1151877	-3.13	0.002	.2695437	.7405342
educ	.9732451	.0436997	-0.60	0.546	.8912559	1.062777
age	.9947974	.0009922	-5.23	0.000	.9928546	.996744
_cons	.0045453	.0032737	-7.49	0.000	.0011079	.0186482
/ln_p	.5352553	.0951206	5.63	0.000	.3488225	.7216882
/ln_the	1.790243	.1788498	10.01	0.000	1.439703	2.140782
p	1.707884	.1624549			1.417398	2.057904
1/p	.5855198	.055695			.4859312	.7055184
theta	5.990906	1.071472			4.219445	8.506084
Likelihood-ratio test of theta=0: chibar2(01) = 96.23 Prob>chibar2 = 0.000						

上表最后一行显示,原假设“ $H_0: \theta = 0$ ”(θ 为异质性 v_j 的方差)的 p 值为 0.000,故强烈拒绝无异质性的原假设,认为存在异质性。下面将看到,引入异质性对于风险函数有很大影响。

```
. stcurve,hazard alpha1
```

其中,选择项“hazard alpha1”表示给定异质性 $v_j = 1$ (即平均异质性)条件下的风险函数,结果参见图 30.8。

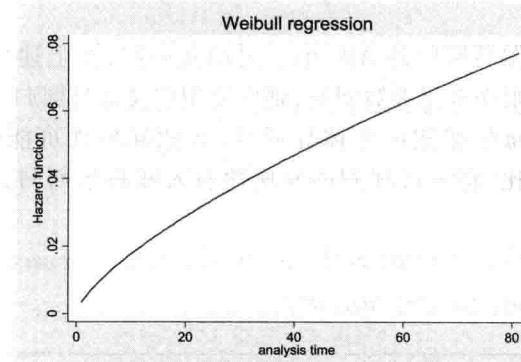


图 30.8 给定异质性的条件威布尔风险函数 (Conditional Weibull hazard)

从图 30.8 可知, 给定个体异质性, 则风险函数单调递增, 与不考虑异质性的威布尔风险函数大相径庭(参见图 30.6)。

```
. stcurve,hazard  
(option unconditional assumed)
```

如果不含选择项“alpha1”, 则默认设置为将个体异质性 v_j 积分平均掉, 然后画无条件 (unconditional) 的风险函数, 结果参见图 30.9。此无条件分布称为“伯尔分布” (Burr distribution)。

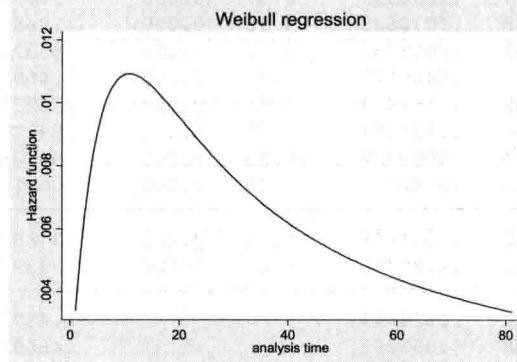


图 30.9 无条件的伯尔风险函数 (Unconditional Burr hazard)

从图 30.9 可知, 无条件的伯尔风险函数呈驼峰形 (hump shape)。无条件的风险函数之所以先升后降, 可能与样本中个体结构的变化有关。虽然条件风险函数不断上升, 但由于高风险个体首先退出样本, 故剩下的个体越来越多地为低风险个体。这两种相反力量的共同作用形成了驼峰形的无条件风险函数。

(5) 由于我们对于参数回归的具体分布始终没有把握, 故下面进行半参数 Cox 回归。

```
. stcox workprg priors tserved felon alcohol drugs black married educ  
age,r nolog noshow
```

Cox regression -- Breslow method for ties						
No. of subjects	=	1445	Number of obs	=	1445	
No. of failures	=	552				
Time at risk	=	80013				
					Wald chi2(10) =	148.14
					Prob > chi2 =	0.0000
Log pseudolikelihood =	-3816.3799					

<u>t</u>	Haz. Ratio	Robust Std. Err.	t	P> z	[95% Conf. Interval]	
					[95% Conf. Interval]	[95% Conf. Interval]
workprg	1.087709	.1026332	0.89	0.373	.9040565	1.308668
priors	1.091542	.0180384	5.30	0.000	1.056754	1.127476
tserved	1.013035	.0018621	7.05	0.000	1.009392	1.016691
felon	.7537191	.0804309	-2.65	0.008	.6114712	.9290583
alcohol	1.538265	.1671288	3.96	0.000	1.243226	1.903321
drugs	1.317338	.1308915	2.77	0.006	1.08423	1.600564
black	1.541249	.1369638	4.87	0.000	1.294882	1.834489
married	.8565166	.0925895	-1.43	0.152	.6929803	1.058646
educ	.9789062	.0188643	-1.11	0.269	.9426224	1.016587
age	.9964248	.0006028	-5.92	0.000	.9952439	.997607

从上表可知,半参数 Cox 回归的结果在性质上与前面的参数回归类似,但不依赖于具体的分布假设,故更为稳健,在实践中十分流行。下面,根据 Cox 模型画风险函数,结果参见图 30.10。

. stcurve,hazard

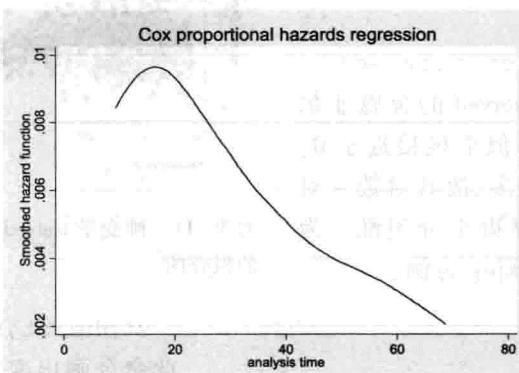


图 30.10 基于 Cox 模型的风险函数

当然,进行 Cox 回归的前提是比例风险假定,为此需进行相应的检验。首先,进行基于舍恩菲尔德残差的比例风险检验。

. estat phtest,detail

Test of proportional-hazards assumption

Time: Time

	rho	chi2	df	Prob>chi2
workprg	0.05246	1.67	1	0.1967
priors	-0.07237	3.42	1	0.0645
tserved	-0.09395	4.14	1	0.0420
felon	0.00497	0.01	1	0.9108
alcohol	-0.04469	1.17	1	0.2802
drugs	-0.01081	0.07	1	0.7963
black	0.04290	1.02	1	0.3128
married	0.03940	0.90	1	0.3430
educ	-0.04669	1.11	1	0.2925
age	0.04627	2.11	1	0.1463
global test		15.26	10	0.1228

note: robust variance-covariance matrix used.

从上表最后一行可知,整体来看,可以接受比例风险的假定。然而,变量 tserved(狱中关押月数)可能违反比例风险假定, p 值为 0.042。更直观地,画变量 tserved 的残差与时间的拟合图,考察其斜率是否为 0,结果参见 30.11。

```
. estat phtest, plot(tserved)
```

根据图 30.11,变量 tserved 的舍恩菲尔德残差对时间的回归斜率似乎很接近于 0。由于变量 tserved 的取值较多,故其对数 - 对数图与观测 - 预测图均显得十分凌乱。为了演示的目的,以变量 workprg 为例。

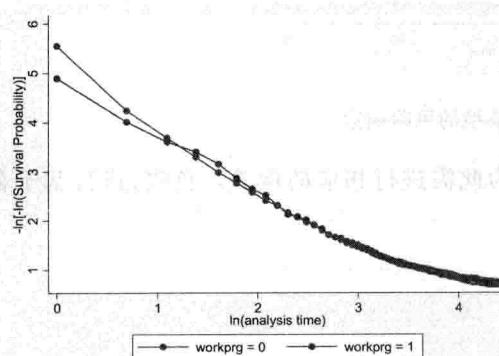


图 30.12 基于变量 workprg 的对数 - 对数图

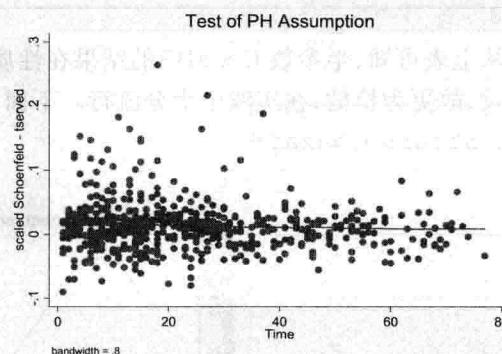


图 30.11 画变量 tserved 的舍恩菲尔德残差与时间的拟合图

```
. stphplot, by(workprg)
```

此命令画出变量 workprg 的对数 - 对数图,结果参见图 30.12。

尽管图 30.12 中的两条曲线有相交,二者的斜率依然大致相等,似乎不违背比例风险假定。

```
. stcoxkm, by(workprg)
```

此命令画出变量 workprg 的观测 - 预测图,结果参见图 30.13。

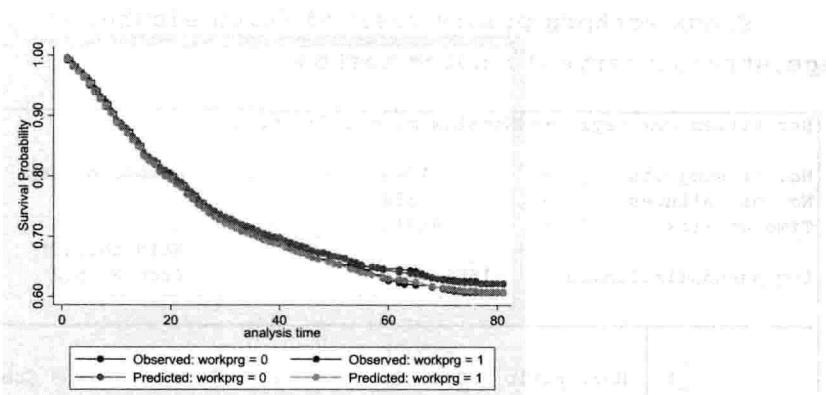


图 30.13 基于变量 workprg 的观测 - 预测图

从图 30.13 可知, 生存函数的观测值与预测值相距不远, 似乎与比例风险假定相一致。

最后, 我们通过引入时变解释变量(TVC)的方法来检验比例风险模型。具体来说, 引入变量 tserved 与时间 t 的互动项, 进行 Cox 回归, 然后检验此互动项的显著性。

```
. stcox workprg priors tserved felon alcohol drugs black married educ  
age, tvc(tserved) r nolog noshow
```

Cox regression -- Breslow method for ties						
	Robust					
	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
main						
workprg	1.086906	.1021223	0.89	0.375	.9040987	1.306677
priors	1.091509	.0180037	5.31	0.000	1.056787	1.127372
tserved	1.016055	.0023921	6.77	0.000	1.011378	1.020755
felon	.7614949	.0811258	-2.56	0.011	.6179941	.9383173
alcohol	1.533541	.1659029	3.95	0.000	1.240537	1.895749
drugs	1.316638	.1303001	2.78	0.005	1.084495	1.598471
black	1.530788	.1358424	4.80	0.000	1.286409	1.821591
married	.8584437	.0925918	-1.42	0.157	.6948662	1.060529
educ	.9797185	.0187874	-1.07	0.285	.9435792	1.017242
age	.9964386	.0005998	-5.93	0.000	.9952636	.997615
tvc						
tserved	.999833	.0000945	-1.77	0.077	.9996479	1.000018

Note: variables in tvc equation interacted with _t

上表倒数第二行显示, 互动项 $tserved \times t$ 的 p 值为 0.077, 故可在 5% 的水平上认为该互动项不显著, 支持比例风险假定。

(6) 上面的分析显示, 变量 tserved 可能违反了比例风险假定。解释方法之一为以变量 tserved 分层变量, 估计分层 Cox 模型。

. stcox workprg priors tserved felon alcohol drugs black married educ
age, strata(tserved) r nolog noshow

Stratified Cox regr. -- Breslow method for ties						
No. of subjects	=	1445	Number of obs	=	1445	
No. of failures	=	552				
Time at risk	=	80013				
Log pseudolikelihood =	-1668.4461		Wald chi2(9)	=	95.27	
			Prob > chi2	=	0.0000	
_t	Haz. Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
workprg	1.007903	.0985513	0.08	0.936	.832127	1.22081
priors	1.091654	.0184789	5.18	0.000	1.05603	1.12848
tserved	1.039889					
felon	.6944598	.0807721	-3.13	0.002	.5528974	.8722676
alcohol	1.524112	.1771505	3.63	0.000	1.213612	1.914051
drugs	1.273412	.1307645	2.35	0.019	1.041263	1.557318
black	1.456758	.1311006	4.18	0.000	1.221192	1.737766
married	.8387165	.0943829	-1.56	0.118	.6727091	1.04569
educ	.9839777	.0194734	-0.82	0.414	.9465412	1.022895
age	.9969553	.0005942	-5.12	0.000	.9957914	.9981205

Stratified by tserved

上表显示,分层 Cox 模型的估计结果与标准 Cox 模型类似,也说明变量 tserved 违反比例风险假定的嫌疑不严重。

习题

30.1 假设寿命服从指数分布。证明个体在 $(0, t_2)$ 区间死亡的概率等于在已知个体存活至时刻 t_1 的情况下,其在 $(t_1, t_1 + t_2)$ 区间死亡的概率。

30.2 使用数据集 kennan - 1985.dta 进行久期分析。Kennan (1985) 使用此数据集研究美国制造业 1968—1976 年的罢工持续时间,主要变量包括 t(罢工持续时间)与 prod(产量对数,已作季节调整并去掉时间趋势)。

30.3 使用数据集 ema1996.dta 分析失业持续时间(spell)的决定因素。McCall(1996) 使用此数据集分析失业保险制度(失业者可获得一定上限的兼职收入而不影响领取失业救济)对失业时间的影响。

第31章 贝叶斯估计简介

31.1 贝叶斯估计的思想

在统计学中有两派,占主流的是“频率学派”(frequentist school),也称“古典学派”(Classical school),即通常数理统计课程中的那些内容;另外一派则是“贝叶斯学派”(Bayesian school),由18世纪英国统计学家贝叶斯(Thomas Bayes)创立。二者的主要区别在于,频率学派假设总体服从某个分布,比如 $f(x;\theta)$,其中 θ 为待估计、未知、给定的参数(或参数向量);而贝叶斯学派则认为,既然 θ 有不确定性,则应将 θ 本身也视为随机变量(或随机向量),并用概率分布来描述。由于 θ 的这个分布在研究者看到数据之前就有,故称为“先验分布”(prior distribution)。得到样本数据后,可以根据贝叶斯定理(Bayes' Theorem)将先验分布更新(update)为后验分布(posterior distribution),并以后验分布作为统计推断的依据。

由于先验分布的主观性(subjectivity),有些学者对贝叶斯学派持有保留意见。然而,在后面我们将看到,在大样本中,先验分布的作用将变得很小。另外,可以使用“相对不含信息”(relatively uninformative)的先验分布,并对后验分布对于先验分布的依赖性进行“敏感度分析”(sensitivity analysis)。

贝叶斯估计的主要优点包括以下几点。

(1) 古典学派一般通过最优化(比如,MLE,OLS)进行参数估计,但有时不易求得最优解。而贝叶斯学派只要反复使用贝叶斯定理即可,不需要进行最优化计算。虽然贝叶斯分析也常常没有解析解,但随着近几十年计算方法的迅猛发展,这已经基本不成问题。

(2) 古典学派需要用不同的统计量来估计期望、方差、中位数、分位数等,而贝叶斯学派可以直接得到参数的整个后验分布。从这个后验分布,可以容易地算出其各阶矩。

(3) 对于古典学派的统计量,常常不易找出其“精确的有限样本分布”(exact finite-sample distribution),故只能退而求其次,推导大样本渐近分布。而贝叶斯学派一般可以直接计算精确的有限样本分布,不需要渐近理论。

31.2 贝叶斯定理

贝叶斯估计的实质就在于,反复使用贝叶斯定理^①,将先验分布与样本数据综合为后验分布。对于随机事件 A 与 B ,有如下的贝叶斯公式:

① 其应用极为广泛,可以说“就靠这一招,打遍天下”。

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (31.1)$$

其中,第一个等号为条件概率的定义,而第二个等号使用了概率的乘法公式。如果将 $P(A)$ 视为先验概率,将 B 视为样本数据,则贝叶斯公式给出了在看到样本数据 B 后,如何将先验概率 $P(A)$ 更新为后验概率 $P(A|B)$ 的规则。

更一般地,对于随机向量 θ (视为参数)与随机向量 y (视为样本数据),根据贝叶斯定理(Bayes' Theorem)可知,

$$f(\theta|y) = \frac{f(\theta, y)}{f(y)} = \frac{f(y|\theta)\pi(\theta)}{f(y)} \quad (31.2)$$

其中, $f(\theta|y)$ 为看到数据 y 之后 θ 的条件分布密度(即后验分布), $\pi(\theta)$ 为参数 θ 的先验分布密度, $f(\theta, y)$ 为 θ 与 y 的联合分布, $f(y|\theta)$ 为给定参数 θ 时 y 的密度函数,而 $f(y)$ 为 y 的边缘分布密度。在联合分布 $f(\theta, y)$ 中将随机参数 θ 积分掉,就可得到 y 的边缘分布密度:

$$f(y) = \int f(\theta, y) d\theta = \int f(y|\theta)\pi(\theta) d\theta \quad (31.3)$$

在贝叶斯分析中,一般把后验分布 $f(\theta|y)$ 记为 $p(\theta|y)$ (p 表示 posterior),而把 y 的密度函数 $f(y|\theta)$ 记为似然函数 $L(\theta; y)$ (将其视为 θ 的函数,而把数据 y 视为给定)。在公式(31.2)中,分母为边缘分布 $f(y)$,不包含 θ ,故可以将其视为常数^①。因此,后验分布与该公式的分子成正比:

$$\propto \bar{h}(\theta - \bar{\mu})^2 \quad (31.4)$$

其中,“ \propto ”表示“成正比”。省略常数可以简化后验分布的推导,而被省略的常数可在以后加上^②。省去常数的密度函数被称为“密度核”(density kernel)。因此, $L(\theta; y)\pi(\theta)$ 就是后验分布 $p(\theta|y)$ 的密度核。

31.3 贝叶斯估计的一个例子

记随机样本为 $y = (y_1 y_2 \cdots y_n)'$,其中 $y_i \sim N(\theta, \sigma^2)$,方差 σ^2 已知,而均值 θ 未知。古典学派通常选择 θ 使似然函数最大化,得到 $\hat{\theta}_{MLE} = \bar{y} \equiv \frac{1}{n} \sum_{i=1}^n y_i$ 。贝叶斯学派则要额外地设定 θ 的先验分布。为了计算方便,选择先验正态分布^③,即 $\theta \sim N(\mu, \tau^2)$,其中先验均值 μ 与先验方差 τ^2 为已知常数。如果 τ^2 较大,就表示先验分布的不确定性较大。我们的目标是求出后验分布 $p(\theta|y)$ 。可以把此计算过程分为以下三步。

第一步 写出先验分布密度 $\pi(\theta)$ 。

在贝叶斯分析中,使用方差的倒数有时更为方便。定义 θ 的“精确度”(precision)为

$$h \equiv 1/\tau^2 \quad (31.5)$$

显然,精确度 h 越大,方差 τ^2 就越小,表明我们对随机变量 θ 知道得越精确。类似地,由于样本均值 \bar{y} 的方差为 (σ^2/n) ,记 \bar{y} 的精确度为 $h^* \equiv n/\sigma^2$ 。 θ 的先验密度为

① 得到样本数据 $y = (y_1 y_2 \cdots y_n)'$ 后, y 已经确定,可以视为常数。

② 它本质上是一个积分常数,即保证密度函数的积分为 1。

③ 有关如何选择先验分布,参见下文。

$$\pi(\theta) = (2\pi\tau^2)^{-1/2} \exp\{-(\theta-\mu)^2/2\tau^2\} \quad (\text{一元正态密度})$$

$$\propto \exp\{-h(\theta-\mu)^2/2\} \quad (\text{去掉不含 } \theta \text{ 的常数项, 代入 } h=1/\tau^2)$$

第二步 写出样本数据 y 的联合密度, 即似然函数:

$$L(\theta; y) = \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\{-(y_i - \theta)^2/2\sigma^2\} \quad (n \text{ 个一元正态密度函数相乘})$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{-\sum_{i=1}^n (y_i - \theta)^2/2\sigma^2\right\} \quad (\text{合并同类项})$$

$$\propto \exp\left[-\sum_{i=1}^n (y_i - \bar{y} + \bar{y} - \theta)^2/2\sigma^2\right] \quad (\text{去掉常数项, 加减 } \bar{y})$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (y_i - \bar{y})^2 + 2\sum_{i=1}^n (y_i - \bar{y})(\bar{y} - \theta) + \sum_{i=1}^n (\bar{y} - \theta)^2\right]\right\} \quad (\text{展开平方项})$$

$$\propto \exp\left\{-\sum_{i=1}^n (\bar{y} - \theta)^2/2\sigma^2\right\}$$

$$\left(\text{去掉常数 } \sum_{i=1}^n (y_i - \bar{y})^2, \text{ 而 } \sum_{i=1}^n (y_i - \bar{y})(\bar{y} - \theta) = (\bar{y} - \theta) \sum_{i=1}^n (y_i - \bar{y}) = 0\right)$$

$$\propto \exp\{-n(\bar{y} - \theta)^2/2\sigma^2\} \quad (n \text{ 个相同的数相加变为乘法})$$

$$\propto \exp\{-h^*(\bar{y} - \theta)^2/2\} \quad (\text{定义 } h^* = n/\sigma^2)$$

注意到, 似然函数的形式与先验密度相似, 这为后面的计算提供了方便。

第三步 根据贝叶斯定理, 写出后验分布的密度核。

$$L(\theta; y)\pi(\theta) \propto \exp\{-h^*(\bar{y} - \theta)^2/2\} \cdot \exp\{-h(\theta - \mu)^2/2\}$$

$$\propto \exp\left\{-\frac{1}{2}[h^*(\bar{y} - \theta)^2 + h(\theta - \mu)^2]\right\} \quad (\text{指数相加})$$

其中, 指数项中的方括弧项可以简化为

$$h^*(\bar{y} - \theta)^2 + h(\theta - \mu)^2$$

$$= h^*(\bar{y}^2 - 2\bar{y}\theta + \theta^2) + h(\theta^2 - 2\mu\theta + \mu^2) \quad (\text{展开平方项})$$

$$= (h + h^*)\theta^2 - 2(h\mu + h^*\bar{y})\theta + h^*\bar{y}^2 + h\mu^2 \quad (\text{根据 } \theta \text{ 合并同类项})$$

$$= \bar{h}\theta^2 - 2\bar{h}\bar{\mu}\theta + h^*\bar{y}^2 + h\mu^2 \quad \left(\text{定义 } \bar{h} = h + h^*, \bar{\mu} = \frac{h\mu + h^*\bar{y}}{\bar{h}}\right)$$

$$= \bar{h}(\theta^2 - 2\bar{\mu}\theta + \bar{\mu}^2) + h^*\bar{y}^2 + h\mu^2 \quad (\text{提取公因子 } \bar{h})$$

$$= \bar{h}(\theta^2 - 2\bar{\mu}\theta + \bar{\mu}^2) - \bar{h}\bar{\mu}^2 + h^*\bar{y}^2 + h\mu^2 \quad (\text{配方})$$

$$\propto \bar{h}(\theta - \bar{\mu})^2 \quad (\text{去掉不含 } \theta \text{ 的常数项})$$

根据以上推导可知,

$$L(\theta; y)\pi(\theta) \propto \exp\left\{-\frac{1}{2}[\bar{h}(\theta - \bar{\mu})^2]\right\} \quad (31.6)$$

显然, 这还是一个正态分布的密度核, 故后验分布仍是正态分布。后验分布的精确度提高为

$$\bar{h} = h + h^* \quad (31.7)$$

即先验精确度 h 与样本精确度 h^* 之和。而后验分布的期望值调整为

$$\bar{\mu} = (h\mu + h^*\bar{y})/\bar{h} \quad (31.8)$$

即先验均值 μ 与样本均值 \bar{y} 之加权平均, 权重为各自的精确度。如果先验信息不精确, 即 h 较小, 则先验均值对于后验均值的影响就较小。方程(31.7)与(31.8)给出了如何将“先验信息”与“样本信息”综合成“后验信息”的规则, 即“贝叶斯更新规则”(Bayesian updating rule)。

由于 $\bar{h} = h + h^* = h + \frac{n}{\sigma^2}$, 故当 $n \rightarrow \infty$ 时, 后验精确度 $\bar{h} \rightarrow \infty$ (后验方差趋于 0); 而且样本精确

度对后验精确度的贡献越来越大,即 $h^*/\bar{h} \rightarrow 1$ 。因此,当 $n \rightarrow \infty$ 时,后验均值 $\bar{\mu} \rightarrow \bar{y}$,即完全由样本数据决定,不受先验分布的影响。直观来看^①,当 $n \rightarrow \infty$ 时,后验分布 $\theta | y \xrightarrow{d} N(\bar{y}, \sigma^2/n)$ 。对比古典学派的结果则为 $\bar{y} \xrightarrow{d} N(\theta, \sigma^2/n)$ 。

总之,当样本容量越来越大时,先验分布所起的作用越来越小。这个结论在一般情况下也成立。回到后验分布密度核的一般表达式:

$$\begin{aligned} p(\boldsymbol{\theta} | \mathbf{y}) &\propto L(\boldsymbol{\theta}; \mathbf{y}) \pi(\boldsymbol{\theta}) \\ &\propto L(\boldsymbol{\theta}; y_1, \dots, y_n) \pi(\boldsymbol{\theta}) \\ &\propto L(\boldsymbol{\theta}; y_1) \cdots L(\boldsymbol{\theta}; y_n) \pi(\boldsymbol{\theta}) \quad (\text{假设样本为独立同分布的}) \end{aligned}$$

因此,后验分布密度核的对数为

$$\ln p(\boldsymbol{\theta} | \mathbf{y}) \propto \ln \pi(\boldsymbol{\theta}) + \sum_{i=1}^n \ln L(\boldsymbol{\theta}; y_i) \quad (31.9)$$

当样本容量增大时,对数先验密度 $\ln \pi(\boldsymbol{\theta})$ 始终保持不变,而对数似然函数之和 $\sum_{i=1}^n \ln L(\boldsymbol{\theta}; y_i)$ 包含的项数越来越多,故越来越以后者为主。这个结论有助于冲淡人们对于贝叶斯估计结果依赖于主观先验分布的顾虑。另外,得到后验分布之后,可以把它视为“修正的先验分布”(revised prior),并作为未来的先验分布,如此反复,不断更新我们对世界的认识。

31.4 基于后验分布的统计推断

贝叶斯分析的主要结果为后验分布。有了后验分布,就可以进行一系列的统计推断。

1. 边缘后验分布

假设参数 $\boldsymbol{\theta}$ 为向量, $\boldsymbol{\theta} = (\theta_1 \cdots \theta_q)'$ 。在知道 $\boldsymbol{\theta}$ 的后验分布 $p(\boldsymbol{\theta} | \mathbf{y})$ 之后,可以求得单个参数 θ_k ($1 \leq k \leq q$) 的“边缘后验分布”(marginal posterior):

$$p(\theta_k | \mathbf{y}) = \int p(\boldsymbol{\theta} | \mathbf{y}) d\theta_1 \cdots d\theta_{k-1} d\theta_{k+1} \cdots d\theta_q \quad (31.10)$$

即只要把其余的参数 $(\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_q)$ 积分掉就行。边缘后验分布常常既不对称也非“单峰”(unimodal),这与古典学派的统计量不同。

2. 后验分布的各阶矩

根据后验分布,可以计算其各阶矩,比如均值、中位数、方差等。

3. 点估计

在贝叶斯分析中,未知参数 $\boldsymbol{\theta}$ 被视为随机变量,而非一个固定的点,故点估计(point estimation)在贝叶斯分析中不那么重要,关注的重点是 $\boldsymbol{\theta}$ 的整个后验分布。另一方面,后验均值(posterior mean)或后验中位数(posterior median)常常被作为点估计来汇报。

4. 区间估计

知道参数 $\boldsymbol{\theta}$ 的整个后验分布后,很容易找到置信区间或置信区域(confidence interval or region)。但置信度为 $(1 - \alpha)$ (比如, $\alpha = 5\%$) 的置信区间并不唯一。一个简单的做法是选择分位数 $\alpha/2$ 与分位数 $(1 - \alpha/2)$ 之间的区间。对贝叶斯置信区间的解释与频率派的解释完全不同。

^① 这只是一种不严格的直观写法,因为当 $n \rightarrow \infty$ 时, $\sigma^2/n \rightarrow 0$ 。要得到非退化的渐近分布,则要乘 \sqrt{n} 。

假设 θ 的 95% 后验置信区间为 $(1, 3)$, 则可以说 θ 落在区间 $(1, 3)$ 的概率为 95%。而对于频率派而言, 如果 θ 的 95% 置信区间为 $(1, 3)$, 则只能说, 如果进行 100 次同样的随机抽样, 大约有 95 次, 这些随机的置信区间能覆盖真实的参数值 θ 。

5. 假设检验

由于贝叶斯分析认为参数不是固定的, 故像 “ $H_0 : \theta = \theta_0$ ” 这类的假设检验通常也就没有意义(假设 θ 为连续变量)。但可以检验参数 θ 是否属于参数空间 Θ 的某个子集 Θ_1 , 比如检验原假设 “ $H_0 : \theta \in \Theta_1 \subset \Theta$ ”, 而替代假设为 “ $H_1 : \theta \in \Theta_2 \subset \Theta$ ”, 其中 Θ_1 与 Θ_2 可以有交集, 即允许“非嵌套式”(non-nested)检验。在古典学派中, “ $H_0 : \theta \in \Theta_1$ ” 成立的概率要么为 0(伪命题), 要么为 1(真命题)。而贝叶斯学派则直接计算 “ $H_0 : \theta \in \Theta_1$ ” 成立的后验概率, 即

$$P(H_0 | y) = \frac{P(y | H_0) P(H_0)}{P(y)} \quad (31.11)$$

其中, y 为样本数据。同样地, 替代假设 “ $H_1 : \theta \in \Theta_2$ ” 成立的后验概率为

$$P(H_1 | y) = \frac{P(y | H_1) P(H_1)}{P(y)} \quad (31.12)$$

两个假设的后验概率之比称为“后验几率比”(posterior odds ratio):

$$\underbrace{\frac{P(H_0 | y)}{P(H_1 | y)}}_{\text{后验几率比}} = \underbrace{\frac{P(y | H_0)}{P(y | H_1)}}_{\text{贝叶斯因子}} \cdot \underbrace{\frac{P(H_0)}{P(H_1)}}_{\text{先验几率比}} \quad (31.13)$$

其中, $B_{01} = \frac{P(y | H_0)}{P(y | H_1)}$ 称为“贝叶斯因子”(Bayes factor), 而 $\frac{P(H_0)}{P(H_1)}$ 称为“先验几率比”(prior odds ratio)。上式表明, 后验几率比等于贝叶斯因子与先验几率比之乘积。如果后验几率比大于 1, 则数据显示我们更应相信 H_0 而不是 H_1 。总之, 贝叶斯假设检验的实质是, 计算不同假设的后验概率, 然后进行比较。这一点与古典假设检验的小概率事件的“反证法”思想很不相同。

31.5 先验分布的选择

1. 无信息先验分布

与古典学派相比, 贝叶斯分析面临的一个主要挑战是如何确定先验分布(specification of the prior)。由于先验分布是贝叶斯分析中主观性的来源, 为了减少主观性, 一种方法是选择基本不带信息的先验分布, 即“无信息先验分布”(uninformative prior), 以减少先验分布对后验分布的影响^①。

如果参数空间 Θ 是有界的, 则可以使用均匀分布来作为无信息先验分布, 即 $\pi(\boldsymbol{\theta}) = c > 0$, $\forall \boldsymbol{\theta} \in \Theta$, 这被称为“均匀先验分布”(uniform prior)。但如果参数空间 Θ 是无界的, 则均匀先验分布就是“非正常密度函数”(improper density), 因为 $\int \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \infty$ 而不是 1。这可能导致后验分布也是非正常密度函数。

均匀先验分布的另一缺点是, 在作参数变换(reparameterization)时, 不满足不变性(invariance)。比如, 假设一维参数 $\theta > 0$ 服从均匀先验分布, 即 $\pi(\theta) = c > 0$ 。考虑参数变换 $\gamma =$

^① 另一降低主观性的方法是增大样本容量。

$\ln \theta$, 则 $-\infty < \gamma < \infty$ 。根据随机变量函数的概率分布公式可知, γ 的先验密度为 $\pi^*(\gamma) = \pi(\theta) \left| \frac{d\theta}{d\gamma} \right| = c e^\gamma$, 故不再是均匀分布。如此看来, 似乎无信息的参数 θ 经过参数转换变成 γ 后, 却带有明显的信息了。

为了避免均匀先验分布的非正常密度函数而同时维持很少的信息, 可以假设先验分布的方差很大。比如, $\theta \sim N(\mu, \tau^2)$, 其中 τ^2 很大。这种先验分布被称为“模糊先验分布”或“弥漫先验分布”(vague or diffuse prior), 但它在参数变换下也不满足不变性。

在实践中, 较常用的无信息先验分布是 Jeffreys (1946) 提出的如下“杰佛里先验分布”(Jeffreys' prior):

$$\pi(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2} = \left| -E\left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right) \right|^{1/2} \quad (31.14)$$

从上式可知, 杰佛里先验分布的密度核是信息矩阵 $I(\boldsymbol{\theta})$ 的行列式的平方根, 其中 $L \equiv L(\boldsymbol{\theta}; \mathbf{y})$ 为似然函数。杰佛里先验分布的优点是, 它具有不变性, 即无论作什么参数变换, 先验分布的形式不变。下面以一维情形为例来证明其不变性。

证明: 对于参数 θ , 其杰佛里先验分布为 $\pi(\theta) \propto |I(\theta)|^{1/2}$ 。假设参数变换为 $\gamma = h(\theta)$, 则需要证明, γ 的杰佛里先验分布为 $\pi^*(\gamma) \propto |I(\gamma)|^{1/2}$ 。根据随机变量函数的分布公式,

$$\pi^*(\gamma) = \pi(\theta) \left| \frac{\partial \theta}{\partial \gamma} \right| \propto |I(\theta)|^{1/2} \left| \frac{\partial \theta}{\partial \gamma} \right| \quad (31.15)$$

因此, 只要证明 $|I(\gamma)|^{1/2} = |I(\theta)|^{1/2} \left| \frac{\partial \theta}{\partial \gamma} \right|$ 即可。为此, 计算 $I(\gamma) = -E\left[\frac{\partial^2 \ln L}{\partial \gamma^2}\right]$ 。

$$\frac{\partial \ln L}{\partial \gamma} = \frac{\partial \ln L}{\partial \theta} \cdot \frac{\partial \theta}{\partial \gamma} \quad (\text{使用求导连锁法则求一阶导数})$$

$$\frac{\partial^2 \ln L}{\partial \gamma^2} = \frac{\partial^2 \ln L}{\partial \theta^2} \cdot \left(\frac{\partial \theta}{\partial \gamma} \right)^2 + \frac{\partial \ln L}{\partial \theta} \cdot \frac{\partial^2 \theta}{\partial \gamma^2} \quad (\text{使用求导连锁法则求二阶导数})$$

其中, $\frac{\partial \ln L}{\partial \theta}$ 为得分函数, 其期望值为 $E\left[\frac{\partial \ln L}{\partial \theta}\right] = 0$ (参见第6章)。因此,

$$I(\gamma) = -E\left[\frac{\partial^2 \ln L}{\partial \gamma^2}\right] = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right] \left(\frac{\partial \theta}{\partial \gamma}\right)^2 = I(\theta) \left(\frac{\partial \theta}{\partial \gamma}\right)^2 \quad (31.16)$$

对方程(31.16)两边取行列式^①, 然后开平方可得

$$|I(\gamma)|^{1/2} = |I(\theta)|^{1/2} \left| \frac{\partial \theta}{\partial \gamma} \right| \quad (31.17)$$

因此, $\pi^*(\gamma) \propto |I(\gamma)|^{1/2}$, 故杰佛里先验分布具备不变性。

例 正态分布的杰佛里先验分布。假设 $y \sim N(\mu, \sigma^2)$ 。

情形一, μ 未知而 σ^2 已知, 则 $I(\mu) = -E\left[\frac{\partial^2 \ln L}{\partial \mu^2}\right] = \frac{n}{\sigma^2}$ (参见习题)。由于 σ^2 已知, 故杰佛里先验分布为 $|I(\mu)|^{1/2} \propto c$, 其中 c 为常数。此时, 杰佛里先验分布为非正常密度函数。

情形二, σ^2 未知而 μ 已知, 则 $I(\sigma^2) = -E\left[\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}\right] = \frac{n}{2\sigma^4}$ (参见习题), 故杰佛里先验分布为 $|I(\sigma^2)|^{1/2} \propto \sigma^{-2}$ 。

① 常数的行列式就是常数本身。

情形三, μ 与 σ^2 皆未知, 则 $I(\mu, \sigma^2) = -E \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \mu^2} & \frac{\partial^2 \ln L}{\partial \mu \partial (\sigma^2)} \\ \frac{\partial^2 \ln L}{\partial \mu \partial (\sigma^2)} & \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} = \frac{n^2}{2\sigma^6}$ (参见习题), 故杰佛里先验分布为 $|I(\mu, \sigma^2)|^{1/2} \propto \sigma^{-3}$ 。

在此例中, 在第一种情形下, 杰佛里先验分布为均匀分布, 但在后两种情形下都不是均匀分布。因此, 究竟杰佛里先验分布在何种意义上是无信息先验分布并不清楚。

2. 自然共轭先验分布

为了计算方便, 对于不同的样本密度函数(sample density)常常选择合适的先验分布, 使得样本密度函数(即似然函数)、先验分布与后验分布都“属于同一族的密度函数”(in the same class of densities), 即具有同样的函数形式。这种先验分布被称为“自然共轭先验分布”(natural conjugate prior), 而这样的先验分布与样本密度函数被称为“自然共轭对”(natural conjugate pair)。

比如, 在前面的例子中, 方差已知而均值未知的正态分布, 其自然共轭先验分布就是正态分布。此时, 由于样本密度函数与先验分布的密度核在形式上都是指数函数, 故把它们相乘时, 得到的后验分布仍然具有指数函数的形式。

除了计算方便(可能得到解析解), 使用共轭先验分布的另一好处是, 由于后验分布与先验分布的形式相同, 故容易把后验分布作为下一轮估计的先验分布。

常见的共轭先验分布参见 Cameron and Trivedi (2005, p. 428)。

3. 敏感度分析

由于贝叶斯分析的结果依赖于先验分布, 而我们对先验分布的具体形式可能并无把握, 故常需要进行敏感度分析, 即考虑使用不同的先验分布对后验分布的影响有多大; 或考虑使用不同的样本密度函数的影响。

31.6 多元回归的贝叶斯分析

考虑以下的多元回归模型:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (31.18)$$

其中, \mathbf{X} 为 $n \times K$ 满列秩数据矩阵, 扰动项 $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ 。因此, 给定 $(\mathbf{X}, \boldsymbol{\beta}, \sigma^2)$, 条件分布 $\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ 。

以无信息的杰佛里先验分布为例。根据前面的例子, 对于 $y \sim N(\mu, \sigma^2)$, 当 μ 未知而 σ^2 已知时, μ 的杰佛里先验分布为常数; 而当 σ^2 未知而 μ 已知时, σ^2 的杰佛里先验分布与 $(1/\sigma^2)$ 成正比。推广到多元回归的情形, 则意味着 $\pi(\beta_j) \propto c, j = 1, \dots, K$, 而 $\pi(\sigma^2) \propto 1/\sigma^2$ 。假设 $\boldsymbol{\beta}$ 与 σ^2 相互独立, 则 $(\boldsymbol{\beta}, \sigma^2)$ 的联合先验分布(joint prior)为

$$\pi(\boldsymbol{\beta}, \sigma^2) \propto 1/\sigma^2 \quad (31.19)$$

样本数据的似然函数可以写为

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (31.20)$$

定义最小二乘估计量为 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, 残差向量 $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ 。由于 $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$, 故 $\mathbf{y} - \mathbf{X}\boldsymbol{\beta} =$

$X\hat{\beta} + e - X\beta = e - X(\beta - \hat{\beta})$ 。由此可得

$$\begin{aligned} L(\beta, \sigma^2; y, X) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} [e' - (\beta - \hat{\beta})'X'][e - X(\beta - \hat{\beta})]\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} [e'e + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]\right\} \quad (\text{因为 } X'e = 0) \\ &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]\right\} \\ &\quad (\text{其中, 样本方差 } s^2 \equiv e'e/(n-K)) \end{aligned}$$

因此, 联合后验分布满足

$$\begin{aligned} p(\beta, \sigma^2 | y, X) &\propto (1/\sigma^2)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]\right\} \underbrace{(1/\sigma^2)}_{\text{先验}} \\ &\propto (1/\sigma^2)^{\frac{n}{2}+1} \exp\left\{-\frac{1}{2\sigma^2} [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]\right\} \\ &\propto (1/\sigma^2)^{\frac{n}{2}+1} \exp\left\{-\frac{(n-K)s^2}{2\sigma^2} - \frac{1}{2}(\beta - \hat{\beta})'(\sigma^2(X'X)^{-1})^{-1}(\beta - \hat{\beta})\right\} \\ &\propto (1/\sigma^2)^{n/2} \exp\left\{-\frac{1}{2}(\beta - \hat{\beta})'(\sigma^2(X'X)^{-1})^{-1}(\beta - \hat{\beta})\right\} \\ &\quad \times (1/\sigma^2)^{\frac{n-K}{2}+1} \exp\{- (n-K)s^2/2\sigma^2\} \quad (\text{拆成两部分, 第二部分不含 } \beta) \end{aligned}$$

从上式可以看出, 给定 σ^2 , 则 β 的条件后验分布 (conditional posterior) 为

$$p(\beta | \sigma^2, y, X) \propto (1/\sigma^2)^{n/2} \exp\left\{-\frac{1}{2}(\beta - \hat{\beta})'(\sigma^2(X'X)^{-1})^{-1}(\beta - \hat{\beta})\right\} \quad (31.21)$$

这是一个 K 维正态分布, 其期望为 β , 方差为 $\sigma^2(X'X)^{-1}$ 。要得到 β 的边缘后验分布 (marginal posterior), 则需要将联合后验分布 $p(\beta, \sigma^2 | y, X)$ 中的 σ^2 积分积掉:

$$\begin{aligned} p(\beta | y, X) &\propto \int_0^\infty (1/\sigma^2)^{\frac{n}{2}+1} \exp\left\{-\frac{1}{2\sigma^2} [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]\right\} d\sigma^2 \\ &\propto \int_\infty^0 z^{\frac{n}{2}-1} \exp\left\{-\frac{z}{2} [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]\right\} (-z^{-2}) dz \\ &\quad (\text{作积分变换 } z \equiv 1/\sigma^2, \sigma^2 = 1/z, d\sigma^2 = -(1/z^2) dz) \\ &\propto \int_0^\infty z^{\frac{n}{2}-1} \exp\left\{-\frac{z}{2} [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]\right\} dz \end{aligned}$$

上面这个表达式在形式上非常接近于 Γ 函数的定义式, 即 $\Gamma(c) \equiv \int_0^\infty z^{c-1} e^{-z} dz$ 。可以证明^①,

$$\int_0^\infty z^c e^{-az} dz = \Gamma(c+1)/a^{c+1} \quad (31.22)$$

在方程(25.22)中, 令 $c = \frac{n}{2} - 1, a = -\frac{1}{2}[(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]$, 可得

$$\begin{aligned} p(\beta | y, X) &\propto \Gamma(n/2) \left\{ -\frac{1}{2} [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})] \right\}^{-n/2} \\ &\propto [(n-K)s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})]^{-n/2} \quad (\text{去掉常数项}) \end{aligned}$$

① 作积分变换 $t \equiv az$, 则 $\int_0^\infty z^c e^{-az} dz = \int_0^\infty (t/a)^c e^{-t} \frac{1}{a} dt = \frac{1}{a^{c+1}} \int_0^\infty t^c e^{-t} dt = \frac{\Gamma(c+1)}{a^{c+1}}$ 。

$$\propto \left\{ (n-K)s^2 \left[1 + \frac{1}{(n-K)} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' (s^2 (X'X)^{-1})^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right] \right\}^{-(n/2)} \quad (\text{提取因子})$$

$$\propto \left[1 + \frac{1}{(n-K)} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' (s^2 (X'X)^{-1})^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right]^{-\frac{n-K+K}{2}} \quad (\text{去掉常数项})$$

而这也正是 K 维 t 分布的密度核, 其中心在 $\hat{\boldsymbol{\beta}}$, 自由度为 $(n-K)$, 而协方差矩阵为 $s^2 (X'X)^{-1}$ 。这个结果与古典多元回归类似。单个参数 β_j 的后验分布则为一维 t 分布。

最后, 要得到 σ^2 的边缘后验分布 (marginal posterior), 只要将联合后验分布 $p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X})$ 中的 $\boldsymbol{\beta}$ 积分积掉即可。回到联合后验分布的表达式:

$$p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto (1/\sigma^2)^{K/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' (\sigma^2 (X'X)^{-1})^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}$$

$$\times (1/\sigma^2)^{\frac{n-K}{2}+1} \exp \{ -(n-K)s^2/2\sigma^2 \}$$

其中, 第二部分不含 $\boldsymbol{\beta}$, 而第一部分为 K 维正态分布的密度核, 积分为 1, 故

$$p(\sigma^2 | \mathbf{y}, \mathbf{X}) \propto (1/\sigma^2)^{\frac{n-K}{2}+1} \exp \{ -(n-K)s^2/2\sigma^2 \} \quad (31.23)$$

这是“ Γ 分布的平方根之倒数的密度核” (kernel of an inverted square-root gamma density)。

总之, 使用无信息先验分布得到的结果与古典多元回归相类似。有关使用“有信息的先验分布” (informative prior) 的结果, 参见 Cameron and Trivedi (2005, p. 437)。这些结果启示我们, 虽然贝叶斯估计与古典估计在方法论上迥异, 有限样本的结论也不同, 但在大样本下, 却常有异曲同工之妙。

31.7 马尔可夫链蒙特卡罗法

古典学派的核心为点估计。在贝叶斯估计中, 则对应于后验均值的计算:

$$E(\boldsymbol{\theta} | \mathbf{y}) = \int \boldsymbol{\theta} p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (31.24)$$

其中, $p(\boldsymbol{\theta} | \mathbf{y})$ 为后验密度。但 $p(\boldsymbol{\theta} | \mathbf{y})$ 通常并无解析表达式 (可能包括复杂的积分), 故难以得到 (31.24) 式的解析式。为此, 常使用蒙特卡罗积分方法 (参见第 19 章)。此处的一个技术难题是, 如果 $p(\boldsymbol{\theta} | \mathbf{y})$ 无解析式, 如何从后验分布 $p(\boldsymbol{\theta} | \mathbf{y})$ 中获得随机样本?

“吉布斯抽样法” (Gibbs sampler) 的基本思想是, 通过更简单的条件分布来抽样。比如, 从上节可知, 联合后验分布 $p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X})$ 的表达式较复杂, 但条件后验分布 $p(\boldsymbol{\beta} | \sigma^2, \mathbf{y}, \mathbf{X})$ 较简单 (为 K 维正态)。一般地, 考虑从二维联合分布 $f(x, y)$ 进行随机抽样, 但由于 $f(x, y)$ 的表达式太复杂而无法直接抽样。另一方面, 假设条件分布 $f(x | y)$ 与 $f(y | x)$ 相对简单, 可从中进行一维随机抽样。为此, 可使用以下迭代法得到联合分布 $f(x, y)$ 的随机样本。

吉布斯抽样法的步骤:

- (1) 选择 x_0 (只要在条件分布 $f(x | y)$ 的取值范围即可);
- (2) 从条件分布 $f(y | x_0)$ 中随机抽取 y_0 ;
- (3) 从条件分布 $f(x | y_0)$ 中随机抽取 x_1 ;
- (4) 从条件分布 $f(y | x_1)$ 中随机抽取 y_1 ;
- (5) 从条件分布 $f(x | y_{t-1})$ 中随机抽取 x_t ;
- (6) 从条件分布 $f(y | x_t)$ 中随机抽取 y_t 。

重复以上步骤(5)与(6)几千次之后,即可得到联合分布 $f(x, y)$ 的随机抽样(通常去掉前面几千个抽样值,以避免初始值的影响,称为“burn in”)。吉布斯抽样的理论基础在于,当 $t \rightarrow \infty$ 时, (x_t, y_t) 的极限分布(limiting distribution)就是 $f(x, y)$ 。使用类似方法,可以得到更高维分布(比如 $f(x, y, z)$)的随机样本。

然而,吉布斯抽样并非真正的随机抽样,因为每次抽样都是上次抽样的函数,只不过上次抽样也是随机的。这种性质表明,吉布斯抽样所得到的序列其实是马尔可夫链。

定义:考虑时间为离散的随机过程 $\{x_t\}_{t=0}^{\infty}$ 。如果对于任何 t ,都有 $P(x_{t+1} \leq x | x_t, x_{t-1}, \dots, x_0) = P(x_{t+1} \leq x | x_t)$, 则称 $\{x_t\}$ 为“马尔可夫链”(Markov chain)。

显然,在给定整个历史的情况下, x_{t+1} 的条件分布仅与 x_t 有关。换言之,马尔可夫链的未来状态(x_{t+1})仅与现在状态(x_t)有关,而与历史状态(x_{t-1}, \dots, x_0)无关。由于吉布斯抽样法使用了蒙特卡罗抽样,而得到的序列又是马尔可夫链,故也称为“马尔可夫链蒙特卡罗法”(Markov Chain Monte Carlo, 简记 MCMC), 在贝叶斯分析中应用广泛。

有关贝叶斯估计的进一步知识,参见 Zellner (1971), Poirier (1995), Koop (2003), Lancaster (2004), Cameron and Trivedi (2005) 以及 Geweke (2005)。遗憾的是,目前还(几乎)没有专门针对贝叶斯估计的 Stata 官方命令;但读者可搜到少量的非官方 Stata 命令,比如在 Stata 中输入命令“findit Bayesian”。

习 题

31.1 一名贝叶斯学派的研究者从总体 $y \sim N(\mu, 100)$ 中抽了一个样本容量为 4 的随机样本, 得到样本均值为 56。该学者得出结论, μ 的后验分布为 $N(52, 100)$ 。该学者对 μ 的先验分布是什么?

31.2 (正态分布的杰佛里先验分布)假设 $y \sim N(\mu, \sigma^2)$ 。

(1) 假设 μ 未知而 σ^2 已知, 证明 $I(\mu) = n/\sigma^2$;

(2) 假设 σ^2 未知而 μ 已知, 证明 $I(\sigma^2) = n/2\sigma^4$;

(3) 假设 μ 与 σ^2 皆未知, 证明 $I(\mu, \sigma^2) = n/2\sigma^6$ 。

第32章 如何做规范的实证研究

在本书中,我们已经学了不少的计量经济学知识。到底应该如何使用这些计量知识,以及在做规范的“实证研究”(empirical research)^①时,一般有哪些标准的步骤与注意事项,这些是本章关注的内容^②。

有人把计量经济学理解为概率统计的一门分支,将注意力集中于定理的推导与证明。事实上,计量知识只是一种手段,而最终目的是通过计量知识的指导来做好实证研究。比如,计量知识告诉我们,在何种情况下,应该使用何种计量方法,其假设与局限性又是什么,如何检验,等等。

32.1 计量理论与现实数据

计量经济学的理论总是建立于一些理想化的假定基础之上^③,而现实的经济数据常常或多或少地不符合这个或那个假定。因此,尽管计量理论是可以严格证明的一门科学,但实证研究的实践本身在一定程度上却是一门艺术。

Kennedy(2003,p.390)认为,“理论计量经济学家”(econometric theorists)与“计量实践者”(applied econometricians)常常缺乏交流,而“不使用真实数据的计量理论”在某种意义上更容易些。其结果是,计量实践者在处理现实数据时有时不知所措,甚至盲目进行。

事实上,无论多么高深的计量方法,如果原始数据的质量有问题,也只能是“垃圾进去,垃圾出来”(garbage in,garbage out)。Griliches(1994)在给美国经济学会做的主席演讲(presidential address)中指出,由于经济学家不够注意数据的来源及产生过程,经常错误地解释数据,导致研究的进展缓慢;如果不提高数据质量,计量理论方面的重大进展将无用武之地。当然,对于大多数从事应用研究的学者而言,主要是使用别人(比如,统计局)提供的数据。即便如此,也应该对数据的质量有一个清醒的判断,并使用相应的计量方法(至少在做出实证研究的结论时,应考虑到数据质量的影响)。

另一方面,对于计量的初学者来说,与真实数据打交道是加深对计量经济学理解的重要途径。只有弄脏你的手,才能学会实证研究(So let's get our hands dirty)。

① “实证研究”的另一含义是“positive analysis”,即研究世界实际上是怎么样的;与此相对的是“规范研究”(normative analysis),即研究世界应该是怎样的(比如,福利经济学)。本书中的“实证研究”指的是使用数据进行的经验分析,即“empirical research”。

② 本章包括了一些个人见解,不代表实证研究应该这样做或必须这样做,不当之处请老师与同学们指正。

③ 主要是为了推导与证明的方便。

32.2 实证研究的主要步骤

计量经济学是实证研究的工具。因此,只有站在实证研究的框架上才能更好地回答“如何使用计量知识”这一问题。一般来说,规范的实证分析包括以下几个步骤^①。

1. 提出研究问题

对于刚开始进行研究的人来说,常常不知该如何选择研究课题。研究者通常知道自己想要研究的领域(比如,经济增长),但这还是一个具体的研究问题(research question)。对实证分析而言,研究问题通常是有关“ x 对 y 有何作用”之类的问题。如果想研究“金融发展对经济增长的作用”,就更具体了,此处 x 指的是“金融发展”(可以包含多个变量),而 y 指的是“经济增长”^②。

研究问题可以来源于理论(比如,检验资产定价模型CAPM是否成立),也可以来自对经济现象的观察(比如,媒体的热点问题报道,社会调研),可以研究某个政策的效应(比如,新劳动法对失业率的影响),也可以对文献中已有论文进行改进。

提出好的研究问题,也许你的研究就成功了一半。什么是好的研究问题呢?总的来说,研究问题越具体、越有趣、越新颖、越有可行性,则越好!

(1) **具体:**简单来说,在以上“ x 对 y 有何作用”这样的句型中,应能明确 x 与 y 具体是什么。

(2) **有趣:**你研究的问题别人会感兴趣吗?为什么我们要在乎你的问题(Why should we care)?知道你的问题的答案后,能影响人们对世界的某个方面的看法吗?你的研究问题为什么重要?

(3) **新颖:**论文的核心价值在于它的创新性,即做出了文献中所没有的边际贡献。这种边际贡献可以是研究了新的现象、使用了新的(更好的)计量方法或使用了新的数据集。做研究的过程是创造新知识的过程,在本质上不同于学习已有(旧)知识的过程。

(4) **可行:**即使你的研究问题很具体、很有趣、很新颖,如果找不到相应的数据,则不可行。

有了潜在的研究问题后,就可以进行初步的探索性研究(exploratory study),看看它是否满足以上标准。比如,通过查文献,看别人是否已经做过类似的研究,或大致了解数据是否可得。显然,备选的研究问题或想法越多越好,因为能“存活”下来的课题通常不多。

2. 文献回顾

论文贵在有新意。假设你找到了一个具体、有趣而可行的研究问题,那么它究竟有多少新颖性,这就不可避免地需要查阅文献,看看文献中是否已有类似的研究。对于中文论文,可在CNKI中搜索;对于英文论文,可在JSTOR或EconLit中搜索^③。某些工作论文则可通过百度或Google搜索。

对于刚开始进行研究的研究生们而言,似乎自己能够想得到的研究题目,都已经被别人做过了。其实未必如此。要想做出新的边际贡献,当然可以在前人的基础上,继续拓展与改进(改进

^① 如果是纯粹的理论文章(不使用数据),则不必遵循此步骤。

^② 事实上,文献中充斥着“某某因素对经济增长的作用”这类实证论文。

^③ 这里仅列出了主要的论文数据库。

计量方法,增加变量,使用新数据等)^①。也可以完全撇开前人,去研究全新的现象。比如,20世纪70年代末中国农村实行了家庭联产承包责任制改革,这是史无前例的。到了1990年左右,就涌现出一批研究农村改革对中国农业产出影响的论文^②。又比如,2008年中国开始实施“新劳动法”。假以时日(有了足够的数据后),就可以研究新劳动法对劳动力市场的影响。即使是前人已经研究过的现象,也常常可以用新的眼光、新的视角去观察。重要的是,要有敏锐的观察力,并“像经济学家那样去思考”(Think like an economist)。当然,对于刚起步的研究者,应尽量避免已经被研究得很烂或过于有挑战性的题目。

什么时候开始看文献,即看文献的时机,也很重要。如果从一开始就大量地看文献,则可能被文献所淹没,望洋兴叹,自觉渺小。更好的方法是,只有当自己已经有了一定的想法后,再去系统地看文献。这样,才会知道自己究竟想看什么,也更能带着批判性的眼光去看。

阅读文献的态度十分重要。虚心地从经典论文中汲取营养,才能“站在巨人的肩膀上”(Standing on the shoulders of giants),但仍应带着某种批判性的眼光。事实上,由于经济现象的复杂性,任何论文一定都有缺点(比如,忽略了某些可能重要的因素),也都有可以改进的空间(甚至可能推翻作者的结论),不必太迷信“权威”。既然经济学还不是科学,获得诺贝尔奖的经济学家可能持有相反的观点,那么又哪来的权威呢?重要的是,使用逻辑与实证的方法对不同的观点进行甄别与质疑。如果认为前人所做的研究已经十全十美,你都赞同,那么,你怎么可能做出新的边际贡献呢?

3. 建立理论模型

虽然实证研究可以没有理论模型,但如果有好的理论模型作为基础,则更有说服力。具体来说,回归分析一般只能说明变量之间的相关性,要对变量之间的因果关系做出判断,常常需要依赖于经济理论。因此,即使无法提供完整的理论模型,也应该进行一定的理论分析。最理想的情形是,从理论模型中推导出计量模型(econometric model),即待估计的回归方程。

4. 收集数据

在正式开始研究之前,应大致知道你所需要的数据是存在的,而且可以得到^③。总的来说,数据要么是别人提供的(比如,统计局),要么是自己收集的(比如,问卷调查)。如果使用别人提供的数据,要注意其数据是如何定义、计算的(统计口径),是否是与你的理论模型中的变量相对应。本章附录中列出了常用的数据来源。

如果确实不知道该从哪里找数据,可以看文献中同类研究的数据来源,并追溯其本源。近年来,《美国经济评论》(American Economic Review)与《应用计量经济学杂志》(Journal of Applied Econometrics)都在其网站上公开了发表论文中使用的数据集。

对于非电子版的数据,则需要自己耐心地手工输入数据,并注意检查,防止出错。

5. 使用适当的计量方法

有了数据之后,就可以根据数据的类型与特点,来决定使用合适的计量方法。比如,被解释变量为二元变量,则可以使用probit或logit;如果是非负整数,则可以使用泊松回归;如果是面板数据,则可以考虑固定效应、随机效应、时间效应、动态面板等;如果是时间序列,则应先判断是否

^① 如果你的新贡献只是增加了一两个控制变量,恐怕不足以作为一篇论文发表。

^② 包括林毅夫老师的论文,Lin(1992)。

^③ 有些时候,虽然数据存在,但可能是保密的内部数据,或获得的代价过高。

含单位根^①,再决定使用相应的计量方法。

对于一般的数据,通常先作 OLS,看看结果,作为一个参照系。做完 OLS 后,可以画残差图(residual plot),大致看看扰动项是否符合经典假定,然后进行严格的检验。如果有所违背(比如,存在异方差、自相关),则做相应的处理。对于时间序列,还可以检验是否存在结构变动(邹检验)。另外,应该对数据的质量进行检验,判断是否存在多重共线性、极端值、弱工具变量等,并做相应的调整。

对于回归函数的具体形式,可以考虑线性、对数(如果解释变量只取正数且有指数增长趋势)、双对数、非线性(边际效应不是常数)等。

由于受数据可获得性的限制,遗漏变量几乎不可避免。因此,很有必要在实证论文中对此进行讨论。不外乎两种情况。第一,存在遗漏变量,但与解释变量不相关(需要说明为什么不相关),故可以不做处理。第二,存在遗漏变量,且与解释变量相关,则必须进行处理,处理方法包括增加控制变量、寻找代理变量、使用工具变量、使用面板数据等。

另一常见问题是内生解释变量。此时,一般需要找到有效的工具变量才能得到一致的估计。

由于面板数据可以在一定程度上克服遗漏变量问题,故比横截面数据或时间序列更有说服力。因此,如果可以获得面板数据,则应尽力争取。

大多数的实证论文都希望说明 x 对 y 的因果作用。而从回归分析的相关关系到因果关系,是一个很大的飞跃。

总之,在这部分应该说明的是,为什么你所使用的计量方法是最恰当的。

32.3 实证论文的结构

实证论文的一般结构如下。

- (1) 引言(Introduction)
- (2) 文献回顾(Literature review)
- (3) 理论框架(Theoretical framework)
- (4) 计量模型(Econometric model)
- (5) 数据说明(Data description)
- (6) 计量方法(Estimation method)
- (7) 回归结果(Regression results)
- (8) 结论(Conclusion)

对于初学者,要注意观察经典论文的文章结构与风格(包括如何用表格来汇报回归结果),以便模仿。比如,中文论文可以模仿《经济研究》或《经济学(季刊)》,而英文论文则可参照《美国经济评论》等。

^① 如果样本容量太小(比如小于 20),则可以不进行单位根检验,因为单位根检验是大样本检验,不适用于小样本。

32.4 计量实践的十诫

Kennedy (2003, p. 390) 提出了以下“计量实践的十诫”(The Ten Commandments of Applied Econometrics),作为指导实证研究的十个准则。

1. 运用常识与理论(Use common sense and economic theory)

在进行模型设定时,应尽量使用常识与经济理论。比如,将“人均变量”(如人均消费)与“人均变量”(如人均 GDP)相匹配;使用实际汇率来解释实际进出口;避免使用带时间趋势的变量来解释不带时间趋势的变量。

2. 避免第Ⅲ类错误(Avoid Type III Error)

在作统计检验时,要尽量减少第Ⅰ类错误与第Ⅱ类错误。而第Ⅲ类错误则指的是,对错误的问题给出了正确的答案(the right answer to a wrong question)。比如,考虑 FDI 对经济增长的作用。由于 FDI 起作用需要时间,如果仅把当年的增长率对当年的 FDI 进行回归,则可能没有太大意义。比较适当的做法是,考虑期初的 FDI 对随后 5 年(或若干年)经济增长率的作用。

3. 熟悉经济现象(Know the context)

实证研究者应该知道所研究现象的历史、制度与文化背景,以及数据的来源与产生过程。换言之,你应该知道得比计算机更多。比如,如果你研究中国的失业问题,那么你应该知道,中国的失业率指的是“城镇登记失业率”,其统计口径与标准教科书中以及西方国家的失业率概念很不相同。

4. 查看数据(Inspect the data)

研究者的一个常见误区是,只知道按计算机键盘,而不去熟悉原始数据(raw data)。查看数据的方法包括,计算数据的主要统计量(summary statistics),比如,均值、最大最小值、标准差、相关系数等;画图;以及增加对数据的感觉(get a feel for the data)。在查看数据的过程中,还要注意观察数据中是否存在不一致的(inconsistent)地方,比如,出现了不可能、不现实或可疑的取值。如果发现,则要进行处理(比如,可能是数据输入错误),这被称为“数据清理”(data cleaning)。

5. 模型应尽量简洁(Keep It Sensibly Simple, KISS)

尽量不要把模型搞得太复杂。“从小到大”(specific-to-general)的建模方法简单易行,但可能偏差较大(因为存在遗漏变量)。而“从大到小”(general-to-specific)的建模方法偏差小,但却不易执行。通常的折中方案是,选择简单而有解释力的模型。

6. 理解计量结果(Use the interocular trauma test)

计算机输出的计量结果常常很多页,包含密密麻麻的表格与数字。如果你尝试了各种不同的计量方法与控制变量(alternative specifications),则结果就会更复杂。如何看这些结果?只能用双眼一个字一个字地看,直到看明白为止。比如,系数的符号对吗,大小合理吗,主要变量在统计上显著吗,与经济理论矛盾吗?如果把你的结果告诉外行,对方会发笑吗?

7. 理解数据挖掘的成本与收益(Understand the costs and benefits of data mining)

不好的“数据挖掘”(data mining)指的是,根据数据的特征来设定计量模型。这种数据挖掘之所以不好,是因为模型设定是由数据产生的,如果再用这个数据去检验由它产生的模型,那么就不是客观的检验。比如,进行回归之后,发现主要变量的系数不显著,有些研究者就想使用“何种方法”,能使得它变得显著。事实上,“系数不显著”本身也是一种结果,也可以在论文中汇

报。如果数据本身没有问题,则它或许预示着,实证数据与理论相矛盾,故可能需要对理论进行修正。

好的“数据挖掘”指的是,对数据进行各种“实验”,以期发现数据中的某种规律性,发现模型设定的错误,以此改进理论。事实上,一定程度的数据挖掘是不可避免的,而这两种形式的数据挖掘的界限并不清晰,正如 Heckman(2000)指出,“尽管使用数据来检验受到该数据启发的理论存在严重的问题,但如果拒绝从数据中学习并修改理论,则会导致更严重的问题”^①。

8. 权衡理论与现实 (Be prepared to compromise)

在一般情况下,计量实践中的问题与理论计量经济学中最接近的情形会有一定的距离。比如,使用的代理变量,是否满足教科书中对代理变量的要求?内生解释变量问题是否严重?单位根检验的结果是否可信?因此,通常需要在现实与理论之间找到一个适当的妥协。

9. 区分统计显著性与经济显著性 (Do not confuse statistical significance with meaningful magnitude)

通常,样本容量越大,则估计值的标准误越小,该系数的“统计显著性”(statistical significance)就越高。但如果该系数估计值为0.001,则可能不具有“经济显著性”(economic significance),即这个解释变量的变化对被解释变量的影响不大,尽管这种影响的幅度被估计得很精确。

10. 进行敏感度分析 (Report a sensitivity analysis)

为了得到论文的结果,研究者通常需要做一系列的假定。问题是,论文的主要结果是否对这些假定很敏感?为此,有必要放松论文的假定,看结果是否稳健(基本不变),这被称为“敏感度分析”。比如,改变样本区间、函数形式、控制变量等。只有稳健的结果才是有说服力的。

32.5 结束语

在很多情况下,经济理论并不能给我们确切的答案。比如,有关“中国的内需不足问题”,学界与媒体已经提出不下十种解释,莫衷一是。究竟哪个(些)因素是导致内需不足的主要原因呢?要真正回答这个问题,则必须仔细地收集数据,进行深入的实证分析。在一定意义上,或许可以说:“理论是灰色的,而实证之树常青。”

有关如何做研究、写论文以及经济学方法论方面的讨论,可进一步参考 Friedman (1953), Stigler and Becker (1977), 林毅夫(1995,2013), Varian (2001,2009), Cochrane (2005), 李子奈(2011), 傅十和、佩雷拉(2012)等。

习 题

32.1 学习实证研究的好方法是复制(replicate)经典论文或最新工作论文的结果。挑选你感兴趣的几篇论文,下载其数据集,并复制其结果。

^① While there are serious problems in using the data that suggest a theory to test that theory, even more important problems arise from refusing to learn from the data in revising economic models.

附录：常用数据来源^①

【中国数据】

国家统计局：www.stats.gov.cn

中国人民银行：www.pbc.gov.cn

财政部：www.mof.gov.cn

国研网(国务院发展研究中心)：www.drcnet.com

中经网(国家信息中心)：www.cei.gov.cn

中宏网(国家发改委)：www.macrochina.com.cn

万得(Wind)金融终端：www.wind.com.cn

国泰安csmar数据库：www.gtarsc.com

CCER经济金融研究数据库：www.ccerdata.com

EPS数据库：www.epsnet.com.cn

CEIC数据库：www.ceicdata.com

China Data Center(Univ. of Michigan)：<http://chinadatacenter.org/newcdc/>

中国调查数据网(CCER and Univ. of Michigan)：www.chinasurveycenter.org

China Health and Nutrition Survey(CHNS)：www.cpc.unc.edu/projects/china

China Human Capital Project(Univ. of Pennsylvanian)：www.ssc.upenn.edu/china/index.htm

经济发展论坛：www.fed.org.cn/

各类统计年鉴。

【国际数据】

Bloomberg Data：http://about.bloomberg.com/product_data.html

BREAD data in development economics：<http://ipl.econ.duke.edu/bread/index.htm>

Economic Time Series Page：www.economagic.com

IMF：www.imf.org

NBER online data：www.nber.com/data_index.html

Panel Study of Income Dynamics：www.umich.edu/~psid

Penn World Table：http://pwt.econ.upenn.edu/php_site/pwt_index.php

OECD：www.oecd.org

Standard & Poor's Compustat：<http://www.compustat.com>

Thomson Datastream：www.datastream.com

US Federal Reserve Bank：www.federalreserve.gov

^① 某些数据库为收费数据库，可向你所在的大学或单位询问是否已购买。

US Bureau of Labor Statistics: www.bls.gov

US Bureau of Economic Analysis: www.bea.gov

United Nations: www.un.org

World Bank: www.worldbank.org

BVD 财经系列数据库(欧洲): www.bvdinfo.com (包括 Bankscope)

Inter-University Consortium for Political and Social Research:

www.icpsr.umich.edu/icpsrweb/ICPSR

Journal of Applied Econometrics Data Archive: www.econ.queensu.ca/jae

KOF Index of Globalization: <http://globalization.kof.ethz.ch/>

Daron Acemoglu's datasets: <http://economics.mit.edu/faculty/acemoglu/data>

Robert Barro's datasets: www.economics.harvard.edu/faculty/barro/data_sets_barro.html

Charles Jones' datasets: www.stanford.edu/~chadj/datasets.html

其他学者的个人网站。

【历史数据】

Maddison, Angus. *The World Economy: Volume 1: A Millennial Perspective and Volume 2: Historical Statistics*. Paris: OECD, 2001 – 2003.

Economic History Service: <http://eh.net>

NBER Macrohistory Database: www.nber.org/databases/macrohistory/contents/index.html

Historical Statistics (Princeton): <http://firestone.princeton.edu/econlib/history1.htm>

Historical Financial Statistics: www.centerforfinancialstability.org

Integrated Network for Societal Conflict Research (INSCR): www.systemicpeace.org/inscr/inscr.htm

【精英项目】

【精英项目】

参 考 书 目

- [1] 艾春荣,陈小红.计量经济学:半参数计量经济学方法.北京:北京大学出版社,2000.
- [2] 白仲林.面板数据的计量经济分析.天津:南开大学出版社,2008.
- [3] 陈海燕.面板数据模型的检验方法.北京:经济科学出版社,2012.
- [4] 陈家鼎,刘婉如,汪仁官.概率统计讲义.3 版.北京:高等教育出版社,2004.
- [5] 陈强.气候冲击、王朝周期与游牧民族的征服.经济学(季刊),2014,即将发表.
- [6] 阿德里安·达内尔.计量经济学辞典.钱晓明,译.上海:上海财经大学出版社,2006.
- [7] 沃尔特·恩德斯.应用计量经济学:时间序列分析.2 版.杜江,谢志超,译.北京:高等教育出版社,2006.
- [8] 傅十和,佩雷拉.如何撰写并发表英文经济学论文.上海:格致出版社、上海人民出版社,2012.
- [9] 高铁梅.计量经济分析方法与建模:Eviews 应用及实例.2 版.北京:清华大学出版社,2009.
- [10] 洪永淼.金融计量的新近发展.经济学季刊,2002 (2).
- [11] 洪永淼.计量经济学的地位、作用和局限.经济研究,2007(5).
- [12] 洪永淼.高级计量经济学.赵西亮,吴吉林,译.北京:高等教育出版社,2011.
- [13] 胡咏梅.计量经济学基础与 STATA 应用.北京:北京师范大学出版社,2010.
- [14] 姜近勇,潘冠中.金融计量学.北京:中国财政经济出版社,2011.
- [15] 斯云汇,金赛男.高级计量经济学(上册).北京:北京大学出版社,2007.
- [16] 斯云汇,金赛男.高级计量经济学(下册).北京:北京大学出版社,2011.
- [17] 劳伦斯·汉密尔顿.应用 Stata 做统计分析.5 版.郭志刚,等,译.重庆:重庆大学出版社,2008.
- [18] 兰草.截面、面板数据分析与 STATA 应用.武汉:武汉大学出版社,2012.
- [19] 李宝仁.高级计量经济学.北京:经济科学出版社,2012.
- [20] 李春涛,张璇.随机模拟与金融数据处理 Stata 教程.北京:中国金融出版社,2009.
- [21] 李宏彬,孟岭生,施新政,吴斌珍.父母的政治资本如何影响大学生在劳动力市场的表现:基于中国高校应届毕业生就业调查的经验研究.经济学季刊,2012(3).
- [22] 李贤平.概率论基础.2 版.北京:高等教育出版社,1997.
- [23] 李雪松.高级经济计量学.北京:中国社会科学出版社,2008.
- [24] 李子奈.计量经济学模型方法论.北京:清华大学出版社,2011.
- [25] 李子奈,叶阿忠.高级应用计量经济学.北京:清华大学出版社,2012.
- [26] 李子奈,叶阿忠.高等计量经济学.北京:清华大学出版社,2000.
- [27] 李忠民,张世英,袁学民.经济计量学教程.2 版.天津:天津大学出版社,2009.
- [28] 林光平.计算计量经济学.杨大勇,译.北京:清华大学出版社,2003.
- [29] 林光平,龙志和,吴梅.我国地区经济收敛的空间计量实证分析:1978—2002 年.经济学季刊,2005(4).
- [30] 林少宫.微观计量经济学要义:问题与方法探讨.武汉:华中科技大学出版社,2003.
- [31] 林毅夫.本土化、规范化、国际化——庆祝《经济研究》创刊 40 周年.经济研究,1995(10).
- [32] 林毅夫.本体与常无:经济学方法论对话.北京:北京大学出版社,2013.
- [33] 陆懋祖.高等时间序列经济计量学.上海:上海人民出版社,1999.
- [34] 栾惠德,张晓峒.季节调整中的春节模型.经济学(季刊),2007(2).
- [35] 赫尔穆特·鲁克波尔,马库斯·克莱茨希.应用时间序列计量经济学.易行健,邓可斌,等,译.北京:机械工业出版社,2008.

- [36] 茹诗松,王静龙,濮晓龙.高等数理统计.2版.北京:高等教育出版社,2006.
- [37] 潘省初.计量经济学中级教程.北京:清华大学出版社,2009.
- [38] 任燕燕.平行数据模型及其在经济分析中的应用.北京:经济科学出版社,2006.
- [39] 沈体雁,冯等田,孙铁山.空间计量经济学.北京:北京大学出版社,2010.
- [40] 盛骤,谢式千,潘承毅.概率论与数理统计.4版.北京:高等教育出版社,2008.
- [41] 苏良军.高等数理统计.北京:北京大学出版社,2007.
- [42] 孙敬水.中级计量经济学.上海:上海财经大学出版社,2009.
- [43] 童恒庆.理论计量经济学.北京:科学出版社,2005.
- [44] 王群勇.Stata在统计与计量分析中的应用.天津:南开大学出版社,2007.
- [45] 王群勇.Stata使用指南与应用案例.北京:中国财政经济出版社,2008.
- [46] 王少平,杨继生,欧阳志刚.计量经济学.北京:高等教育出版社,2011.
- [47] 王升.计量经济学导论.北京:清华大学出版社,2006.
- [48] 王天夫,李博柏.Stata实用教程.北京:中国人民大学出版社,2008.
- [49] 王文博.计量经济学.西安:西安交通大学出版社,2004.
- [50] 王燕.应用时间序列分析.北京:中国人民大学出版社,2005.
- [51] 王志刚.面板数据模型及其在经济分析中的应用.北京:经济科学出版社,2008.
- [52] 吴可杰,秦宛顺,赵德滋.经济计量学——理论与方法.南京:南京大学出版社,1986.
- [53] 谢识予,朱弘鑫.高级计量经济学.上海:复旦大学出版社,2005.
- [54] 杨菊华.数据管理与模型分析:STATA软件应用.北京:中国人民大学出版社,2012.
- [55] 叶阿忠.非参数计量经济学.天津:南开大学出版社,2003.
- [56] 叶阿忠.非参数和半参数计量经济模型理论.北京:科学出版社,2008.
- [57] 余森杰.国际贸易学:理论、政策与实证.北京:北京大学出版社,2013.
- [58] 赵国庆.高级计量经济学——理论与方法.北京:中国人民大学出版社,2014.
- [59] 张车伟,薛欣欣.国有部门与非国有部门工资差异及人力资本贡献.经济研究,2008(4).
- [60] 张成思.金融计量学:时间序列分析视角.大连:东北财经出版社,2008.
- [61] 张定胜.计量经济学.武汉:武汉大学出版社,2005.
- [62] 张鹏伟,李嫣怡.Stata统计分析与应用.北京:电子工业出版社,2011.
- [63] 张晓峒.计量经济分析.修订版.北京:经济科学出版社,2000.
- [64] 周绍杰,张俊森,李宏彬.中国城市居民的家庭收入、消费和储蓄行为:一个基于组群的实证研究.经济学季刊,2009,8(4).
- [65] 朱建平,胡朝霞,王艺明,林光平.高级计量经济学导论.北京:北京大学出版社,2009.
- [66] 朱平芳.现代计量经济学.上海:上海财经大学出版社,2004.
- [67] Aalen, O., 1978. "Nonparametric Inference for a Family of Counting Processes," *Annals of Statistics*, 6, 701 – 726.
- [68] Abadie, A., and G. W. Imbens, 2002. "Simple and Bias-corrected Matching Estimators," Technical report, Department of Economics, University of California, Berkeley.
- [69] Abadie, A., D. Drukker, J. L. Herr, and G. W. Imbens, 2004. "Implementing Matching Estimators for Average Treatment Effects in Stata," *The Stata Journal*, 4, 290 – 311.
- [70] Abadie, A., and G. W. Imbens, 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235 – 267.
- [71] Abadie, A., and G. W. Imbens, 2008. "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 76, 1537 – 1557.
- [72] Abadie, A., and G. W. Imbens, 2011. "Bias Corrected Matching Estimators for Average Treatment Effects," *Journal of Business and Economic Statistics*, 29, 1 – 11.
- [73] Abadie, A., and G. W. Imbens, 2012. "Matching on the Estimated Propensity Score," Harvard University working

paper.

- [74] Acemoglu, D., S. Johnson, and J. Robinson, 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369 – 1401.
- [75] Acemoglu, D., S. Johnson, J. Robinson, and P. Yared, 2008. "Income and Democracy," *American Economic Review*, 98, 808 – 842.
- [76] Acemoglu, D., S. Johnson, and J. Robinson, 2012. "The Colonial Origins of Comparative Development: An Empirical Investigation: Reply," *American Economic Review*, 102, 3077 – 3110.
- [77] Aigner, D., K. Lovell, and P. Schmidt, 1977. "Formulation and Estimation of Stochastic Frontier Production Models," *Journal of Econometrics*, 6, 21 – 37.
- [78] Akaike, H., 1974. "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716 – 722.
- [79] Albouy, D., 2012. "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment," *American Economic Review*, 102, 3059 – 3076.
- [80] Altonji, J., T. Elder, and C. Taber, 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151 – 184.
- [81] Amemiya, T., 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- [82] Amemiya, T. and T. MaCurdy, 1986. "Instrumental Variable Estimation of An Error Component Model," *Econometrica*, 54, 869 – 880.
- [83] Amisano, G., and C. Giannini. 1997. *Topics in Structural VAR Econometrics*. 2nd ed. Heidelberg: Springer.
- [84] Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- [85] Anselin, L., 2010. "Thirty Years of Spatial Econometrics," *Papers in Regional Science*, 89, 3 – 25.
- [86] Anderson, T. and C. Hsiao, 1981. "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598 – 606.
- [87] Andrews, D., and E. Zivot, 1992. "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251 – 270.
- [88] Angrist, J., 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 80, 313 – 336.
- [89] Angrist, J. and A. Krueger, 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979 – 1014.
- [90] Angrist, J., and V. Lavy, 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533 – 575.
- [91] Angrist, J. and J. Pischke, 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, New Jersey: Princeton University Press.
- [92] Arbia, G., 2006. *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer-Verlag, Berlin.
- [93] Arellano, M. and S. Bond, 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equation," *Review of Economic Studies*, 58, 277 – 298.
- [94] Arellano, M. and O. Bover, 1995. "Another Look at Instrumental Variable Estimation of Error Components Models," *Journal of Econometrics*, 68, 29 – 51.
- [95] Arraiz, I., D. M. Drukker, H. H. Kelejian, and I. R. Prucha, 2010. "A Spatial Cliff-Ord-type Model with Heteroskedastic Innovations: Small and Large Sample Results," *Journal of Regional Science*, 50, 592 – 614.
- [96] Baba, Y., R. F. Engle, D. F. Kraft, and K. F. Kroner, 1990. "Multivariate Simultaneous Generalized ARCH," *mimeo*, Department of Economics, University of California, San Diego.
- [97] Ball, R., and P. Brown, 1968. "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting*

- Research*, 6, 159 – 178.
- [98] Baltagi, B., 2005. *Econometric Analysis of Panel Data*, 3rd edition. Chichester, West Sussex, England; John Wiley & Sons Ltd.
- [99] Baltagi, B. and S. Khanti-Akom, 1990. "On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables," *Journal of Applied Econometrics*, 5, 401 – 406.
- [100] Banerjee, A., M. Marcellino, and C. Osbat, 2005. "Testing for PPP: Should We Use Panel Methods?" *Empirical Economics*, 30, 77 – 91.
- [101] Basmann, R. L., 1957. "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, 25, 77 – 83.
- [102] Battese, G. E., and T. J. Coelli, 1992. "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India," *Journal of Productivity Analysis*, 3, 153 – 169.
- [103] Baum, C., 2006. *An Introduction to Modern Econometrics Using Stata*. College Station, Texas: Stata Press.
- [104] Baum, C., M. Schaffer, and S. Stillman, 2007. "Enhanced Routines for Instrumental Variables/GMM Estimation and Testing," *The Stata Journal*, 7, 465 – 506.
- [105] Beckett, S., 2013. *Introduction to Time Series Using Stata*. College Station, Texas: Stata Press.
- [106] Belman, J., Z. Hrubec, P. Taubman, and T. Wales, 1980. *Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment, and Schooling*, Amsterdam; North-Holland.
- [107] Bernanke, B., 1986. "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, North-Holland, Amsterdam.
- [108] Berndt, E. R., 1996. *The Practice of Econometrics: Classic and Contemporary*. New York: Addison-Wesley.
- [109] Berndt, E., B. Hall, R. Hall and J. Hausman, 1974. "Estimation and Inference in Nonlinear Structural Model," *Annals of Economics and Social Measurement*, 3/4, 653 – 655.
- [110] Black, S., 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114, 577 – 599.
- [111] Blanchard, O., and D. Quah, 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79, 655 – 673.
- [112] Blundell, R. and S. Bond, 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115 – 143.
- [113] Bollerslev, T., 1986. "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307 – 327.
- [114] Bollerslev, T., 1990. "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72, 498 – 505.
- [115] Bollerslev, T., R. F. Engle, and J. M. Wooldridge, 1988. "A Capital Asset Pricing Model with Time-varying Covariances," *Journal of Political Economy*, 96, 116 – 131.
- [116] Bound, J., D. Jaeger, and R. Baker, 1995. "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443 – 450.
- [117] Box, G., G. Jenkins and G. Reinsel, 1994. *Time Series Analysis: Forecasting and Control*, 3rd edition. 影印本. 北京: 人民邮电出版社, 2005.
- [118] Box, G. and D. Pierce, 1970. "Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509 – 1526.
- [119] Breitung, J. 2000. "The Local Power of Some Unit Root Tests for Panel Data," In *Advances in Econometrics, Volume 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. B. H. Baltagi, 161 – 178. Amsterdam: JAI Press.

- [120] Breitung, J., and S. Das, 2005. "Panel Unit Root Tests under Cross-sectional Dependence," *Statistica Neerlandica*, 59, 414 – 433.
- [121] Breslow, N. E., 1974. "Covariance Analysis of Censored Survival Data," *Biometrics*, 30, 89 – 99.
- [122] Breusch, T., 1978. "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17, 334 – 355.
- [123] Breusch, T. and A. Pagan, 1979. "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287 – 1294.
- [124] Breusch, T. and A. Pagan, 1980. "The LM Test and Its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47, 239 – 254.
- [125] Brewer, D., E. Eide, and R. Ehrenberg, "Does It Pay to Attend an Elite Private College? Cross-cohort Evidence on the Effects of College Type on Earnings," *Journal of Human Resources*, 1999, 34, 104 – 123.
- [126] Browning, M., A. Deaton and M. Irish, 1985. "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53, 503 – 543.
- [127] Bruno, G., 2005. "Estimation and Inference in Dynamic Unbalanced Panel-data Models with a Small Number of Individuals," *The Stata Journal*, 5, 473 – 500.
- [128] Buchinsky, M., 1994. "Changes in the U. S. Wage Structure, 1963 – 1987: Application of Quantile Regression," *Econometrica*, 62, 405 – 458.
- [129] Buckles, K. and D. Hungerman, 2012. "Season of Birth and Later Outcomes: Old Questions, New Answers," *Review of Economics and Statistics*, forthcoming.
- [130] Cameron, A. C. and P. K. Trivedi, 1998. *Regression Analysis of Count Data*. Econometric Society Monograph No. 30, Cambridge University Press, New York, NY.
- [131] Cameron, A. C. and P. K. Trivedi, 2005. *Microeconomics: Methods and Applications*. Cambridge University Press, New York, NY.
- [132] Cameron, A. C. and P. K. Trivedi, 2010, *Microeconomics using Stata*, revised edition. College Station, Texas: Stata Press.
- [133] Campbell, J., A. Lo, and C. MacKinlay, 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.
- [134] Card, D. and A. Krueger, 1994. "Minimum Wages and Employment: A Case Study of the Fast Food Industry," *American Economic Review*, 84, 772 – 793.
- [135] Casella, G. and R. Berger, 2002. *Statistical Inference*, 2nd edition. Pacific Grove, CA: Duxbury.
- [136] Che, Y., Y. Lu, Z. Tao, and P. Wang, 2013, "The Impact of Income on Democracy Revisited," *Journal of Comparative Economics*, forthcoming.
- [137] Chen, Q., 2014. "Climate Shocks, State Capacity, and Peasant Uprisings in North China during 25 – 1911 CE," *Economica*, forthcoming.
- [138] Cheng, S., and J. S. Long, 2007. "Testing for IIA in the Multinomial Logit Model," *Sociological Methods Research*, 35, 583 – 600.
- [139] Choi, I., 2001. "Unit Root Tests for Panel Data," *Journal of International Money and Finance*, 20, 249 – 272.
- [140] Chow, G., 1960. "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591 – 605.
- [141] Christensen, L. and W. Greene, 1976. "Economies of Scale in U. S. Electric Power Generation," *Journal of Political Economy*, 84, 655 – 676.
- [142] Chung, C. -F., P. Schmidt, and A. D. Witte, 1991. "Survival Analysis: A Survey," *Journal of Quantitative Criminology*, 7, 59 – 98.
- [143] Cheung, Y. -W., and K. S. Lai, 1995. "Lag Order and Critical Values of a Modified Dickey-Fuller Test," *Oxford*

- Bulletin of Economics and Statistics*, 57, 411 – 419.
- [144] Cleveland, W., 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829 – 836.
- [145] Cleves, M. A., W. W. Gould, and R. G. Gutierrez, 2004. *An Introduction to Survival Analysis Using Stata*. Texas: Stata Press.
- [146] Cliff, A. D., and J. K. Ord, 1973. *Spatial Autocorrelation*. London: Pion.
- [147] Cliff, A. D., and J. K. Ord, 1981. *Spatial Processes, Models and Applications*. London: Pion.
- [148] Cochrane, D. and G. Orcutt, 1949. "Application of Least Square Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, 44, 32 – 61.
- [149] Cochrane, J., 2005. "Writing Tips for Ph. D. Students," available online at <http://www.bus.lsu.edu/hill/writing/cochrane.pdf>.
- [150] Coelli, T. J. 1995. "Estimators and Hypothesis Tests for A Stochastic Frontier Function: A Monte Carlo Analysis," *Journal of Productivity Analysis*, 6, 247 – 268.
- [151] Cox, D. R., 1972. "Regression Models and Life Tables (with Discussion)," *Journal of the Royal Statistical Society, B*, 34, 187 – 220.
- [152] Cox, D. R. 1975. "Partial Likelihood," *Biometrika*, 62, 269 – 276.
- [153] Cragg, J. G., 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, 39, 829 – 844.
- [154] Creel M., 2008. *Econometrics*. University Autonoma de Barcelona, mimeo.
- [155] Davidson, R. and J. MacKinnon, 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- [156] Davidson, R. and J. MacKinnon, 2004. *Econometric Theory and Methods*. New York: Oxford University Press.
- [157] D'Agostino, R. B., A. Balanger, and R. B. D'Agostino, Jr., 1990. "A Suggestion for Using Powerful and Informative Tests of Normality," *American Statistician*, 44, 316 – 321.
- [158] Deaton, A., 1985. "Panel Data from Time Series of Cross-Sections," *Journal of Econometrics*, 30, 109 – 126.
- [159] Dehejia, R. H., and S. Wahba, 1999. "Causal Effects in Nonexperimental Studies: Reevaluation of the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053 – 1062.
- [160] Dell, Melissa, 2010. "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78, 1863 – 1903.
- [161] Dickey, D. and W. Fuller, 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427 – 431.
- [162] Dickey, D. and W. Fuller, 1981. "Likelihood Ratio Tests for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057 – 1072.
- [163] DiNardo, J. and J. L. Tobias, 2001. "Nonparametric Density and Regression Estimation," *Journal of Economic Perspectives*, 15, 11 – 28.
- [164] Dougherty, Christopher, 2002. *Introduction to Econometrics*, 2nd edition. Oxford, UK: Oxford University Press.
- [165] Drukker, D. M., 2002. "Bootstrapping a Conditional Moments Test for Normality after Tobit Estimation," *The Stata Journal*, 2, 125 – 139.
- [166] Durbin, J., 1954. "Errors in Variables," *Review of the International Statistical Institute*, 22, 23 – 32.
- [167] Durbin, J. and G. Watson, 1950. "Testing for Serial Correlation in Least-Square Regression," *Biometrika*, 37, 409 – 428.
- [168] Efron, B., 1977. "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, 72, 557 – 565.
- [169] Efron, B., 1979. "Bootstrapping Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1 – 26.
- [170] Eicker, F., 1967. "Limit Theorems for Regression with Unequal and Dependent Errors," In *Proceedings of the*

- Fifth Berkeley Symposium in Mathematical Statistics*, Vol. 1. Berkeley: University of California Press.
- [171] Elhorst, J. P., 2010. "Spatial Panel Data Models," in Fischer, M. and A. Getis (eds), *Handbook of Applied Spatial Analysis*. Berlin, Heidelberg and New York: Springer, 377 – 407.
- [172] Elliot, G., T. Rothenberg and J. Stock, 1996. "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813 – 836.
- [173] Engle, R., 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987 – 1008.
- [174] Engle, R. F., 2002. "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models," *Journal of Business and Economic Statistics*, 20, 339 – 350.
- [175] Engle, R. and C. Granger, 1987. "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251 – 276.
- [176] Engle, R. F. and K. F. Kroner, 1995. "Multivariate Simultaneous Generalized GARCH," *Econometric Theory*, 11, 122 – 150.
- [177] Engle, R., D. Lilien and R. Robins, 1987. "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391 – 407.
- [178] Fairlie, R., and R. London, 2012. "The Effects of Home Computers on Educational Outcomes: Evidence from a Field Experiment with Community College Students," *Economic Journal*, 122, 727 – 753.
- [179] Fama, E., L. Fisher, M. Jensen, and R. Roll, 1969. "The Adjustment of Stock Prices to New Information," *International Economic Review*, 10, 1 – 21.
- [180] Fan, J., 1992. "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998 – 1004.
- [181] Frankel, J. and D. Romer, 1999. "Does Trade Cause Growth," *American Economic Review*, 89, 379 – 399.
- [182] Frees, E. W., 1995. "Assessing Cross-sectional Correlations in Panel Data," *Journal of Econometrics*, 64, 393 – 414.
- [183] Frees, E. W., 2004. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge, UK: Cambridge University Press.
- [184] Friedman, M., 1937. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, 32, 675 – 701.
- [185] Friedman, M., 1953. "The Methodology of Positive Economics," *Essays in Positive Economics*, Chicago University Press.
- [186] Geary, R., 1954. "The Contiguity Ratio and Statistical Mapping," *The Incorporated Statistician*, 5, 115 – 145.
- [187] Getis, A., and J. Ord, 1992. "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis*, 24, 189 – 206.
- [188] Geweke, J., 2005. *Contemporary Bayesian Econometrics and Statistics*. New York: John Wiley and Sons.
- [189] Glosten, L., R. Jagannathan and D. Runkle, 1993. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779 – 1801.
- [190] Godfrey, L., 1978. "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1293 – 1302.
- [191] Godfrey, L., 1988. *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. Cambridge, UK: Cambridge University Press.
- [192] Goldberger, A., 1991. *A Course in Econometrics*. Cambridge, MA: Harvard University Press.
- [193] Granger, C., 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424 – 438.
- [194] Granger, C. and P. Newbold, 1974. "Spurious Regressions in Econometrics," *Journal of Econometrics*,

- 2,111 – 120.
- [195] Greene, W. H., 2000. *Econometric Analysis*, 4th ed. Upper Saddle River, New Jersey: Prentice Hall.
 - [196] Greene, W. H., 2008. *Econometric Analysis*, 6th ed. Upper Saddle River, New Jersey: Prentice Hall.
 - [197] Greene, W. H., 2012. *Econometric Analysis*, 7th ed. Upper Saddle River, New Jersey: Prentice Hall.
 - [198] Greene, W. H., and D. Hensher, 1995. "Multinomial Logit and Discrete Choice Models," in *LIMDEP version 7.0*, ed. W. Greene. Bellport, NY: Econometric Software, Inc. CDA.
 - [199] Griliches, Z., 1994. "Productivity, R&D and the Data Constraint," *American Economic Review*, 84, 1 – 23.
 - [200] Griliches, Z., and W. Mason, 1972. "Education, Income, and Ability," *Journal of Political Economy*, 80, S74 – S103.
 - [201] Groves, T., Y. Hong, J. McMillan, and B. Naughton, 1994. "Autonomy and Incentives in Chinese State Enterprises," *Quarterly Journal of Economics*, 109, 183 – 209.
 - [202] Guo, S., and M. Fraser, 2009. *Propensity Score Analysis: Statistical Methods and Applications*. SAGE Publications, Inc.
 - [203] Gutierrez, R. G., S. Carter, and D. M. Drukker. 2001. "sg160: On boundary-value likelihood-ratio tests," *Stata Technical Bulletin* 60; 15 – 18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 269 – 273. College Station, TX: Stata Press.
 - [204] Hadri, K., 2000. "Testing for Stationarity in Heterogeneous Panel Data," *The Econometrics Journal*, 3, 148 – 161.
 - [205] Hahn, J., P. Todd, and W. van der Klaauw, 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201 – 209.
 - [206] Hall, B., Z. Griliches, and J. Hausman, 1986. "Patents and R&D: Is There a Lag?" *International Economic Review*, 27, 265 – 283.
 - [207] Hamilton, J., 1994. *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.
 - [208] Hannan, E. and Quinn, G., 1979. "The Determination of the Order of an Autoregression," *Journal of Royal Statistical Society, B*, 41, 190 – 195.
 - [209] Hansen, B., 1999. "Threshold Effects in Non-dynamic Panels: Estimation, Testing and Inference," *Journal of Econometrics*, 93, 345 – 368.
 - [210] Hansen, B., 2000. "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575 – 603.
 - [211] Hansen, L., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029 – 1054.
 - [212] Harris, R. D. F., and E. Tzavalis, 1999. "Inference for Unit Roots in Dynamic Panels where the Time Dimension Is Fixed," *Journal of Econometrics*, 91, 201 – 226.
 - [213] Hartwig, J., 2009. "A Panel Granger-causality Test of Endogenous vs. Exogenous Growth," KOF Working papers 09 – 231, KOF Swiss Economic Institute, ETH Zurich.
 - [214] Hartwig, J., 2010. "Is Health Capital Formation Good for Long-term Economic Growth? -Panel Granger-causality Evidence for OECD Countries," *Journal of Macroeconomics*, 32, 314 – 325.
 - [215] Hayashi, F., 2000. *Econometrics*. Princeton, New Jersey: Princeton University Press.
 - [216] Hausman, J., 1978. "Specification Tests in Econometrics," *Econometrica*, 46, 1251 – 1271.
 - [217] Hausman, J., and W. Taylor, 1981. "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377 – 1398.
 - [218] Hausman, J. and D. McFadden, 1984. "Specification Tests for the Multinomial Logit Model," *Econometrica*, 52, 1219 – 1240.
 - [219] Heckman, J., 1979. "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153 – 161.
 - [220] Heckman, J., 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth-Century Retrospective,"

- Quarterly Journal of Economics*, 115, 45 – 97.
- [221] Heckman, J. J., H. Ichimura, and P. E. Todd, 1997, "Matching As An Econometric Evaluation Estimator; Evidence from Evaluating a Job Training Programme", *Review of Economic Studies*, 64, 605 – 654.
- [222] Heckman, J. J., H. Ichimura, and P. E. Todd, 1998, "Matching as an Econometric Evaluation Estimator", *Review of Economic Studies*, 65, 261 – 294.
- [223] Hildreth, C. and J. Lu, 1960. "Demand Relations with Autocorrelated Disturbances," Technical Bulletin No. 276, Michigan State University.
- [224] Holtz-Eakin, D., W. Newey, and H. S. Rosen, 1988. "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371 – 1395.
- [225] Horowitz, J., 1997. "Bootstrap Methods in Econometrics: Theory and Numerical Performance," in D. Kreps and K. Wallis (ed.), *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, Vol. 3, 188 – 222. Cambridge, UK: Cambridge University Press.
- [226] Horowitz, J., 2001. "The Bootstrap," in J. Heckman and E. Leamer (ed.), *Handbook of Econometrics*, Vol. 5, 3159 – 3228. Amsterdam: North-Holland.
- [227] Hosmer, D., and S. Lemeshow, 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. New York: John Wiley & Sons, Inc.
- [228] Hoxby, C., 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 90, 1209 – 1238.
- [229] Hoxby, C., 2007. "Does Competition Among Public Schools Benefit Students and Taxpayers? Reply," *American Economic Review*, 97, 2038 – 2055.
- [230] Hsiao, C., 2003. *Analysis of Panel Data*, 2nd edition. Cambridge, UK: Cambridge University Press.
- [231] Huber, P., 1967. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics*, Vol. 1. Berkeley: University of California Press, 1967.
- [232] Im, K., M. Pesaran and Y. Shin, 2003. "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, 53 – 74.
- [233] Imbens, G., and J. Angrist, 1994. "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467 – 476.
- [234] Imbens, G., and T. Lemieux, 2008. "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615 – 635.
- [235] Imbens, G., and K. Kalyanaraman, 2009. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," NBER Working Paper 14726.
- [236] Imbens, G., and J. Wooldridge, 2009. "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47, 5 – 86.
- [237] James, C., 1987. "Some Evidence on the Uniqueness of Bank Loans," *Journal of Financial Economics*, 19, 217 – 235.
- [238] Jarque, C. and A. Bera, 1987. "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163 – 172.
- [239] Jeffreys, H., 1946. "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Statistical Society of London*, 186, 453 – 461.
- [240] Johansen, S., 1988. "Statistical Analysis of Cointegrated Vectors," *Journal of Economic Dynamics and Control*, 12, 231 – 254.
- [241] Johnson, R. and D. Wichern, 2002. *Applied Multivariate Statistical Analysis*, 5th edition. Upper Saddle River, New Jersey: Prentice Hall.

- [242] Johnston, J. and J. Dinardo, 1996. *Econometric Methods*, 4th edition. McGraw-Hill/Irvin.
- [243] Judson, R. A. and A. L. Owen, 1999. "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists," *Economics Letters*, 65, 9 – 15.
- [244] Kaplan, E. L., and P. Meier, P., 1958. "Nonparametric Estimation from Incomplete Observations," *Journal of American Statistical Association*, 53, 457 – 481.
- [245] Kelejian, H. H., and I. R. Prucha, 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17, 99 – 121.
- [246] Kelejian, H. H., and I. R. Prucha, 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40, 509 – 533.
- [247] Kelejian, H. H., and I. R. Prucha, 2004. "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations," *Journal of Econometrics*, 118, 27 – 50.
- [248] Kelejian, H. H., and I. R. Prucha, 2010. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, 140, 53 – 130.
- [249] Kennan, J., 1985. "The Duration of Contract Strikes in U. S. Manufacturing," *Journal of Econometrics*, 28, 5 – 28.
- [250] Kennedy, P., 2003. *A Guide to Econometrics*. Cambridge, MA: MIT Press.
- [251] King, G., and L. Zeng, 2001a. "Explaining Rare Events in International Relations," *International Organization*, 55, 693 – 715.
- [252] King, G., and L. Zeng, 2001b. "Logistic Regression in Rare Events Data," *Political Analysis*, 9, 137 – 163.
- [253] Kiviet, J. F., 1995. "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53 – 78.
- [254] Kleinbaum, D., and M. Klein, 2011. *Survival Analysis: A Self-Learning Text*, 3rd edition. New York: Springer.
- [255] Koenker, R., 1981. "A Note on Studentizing a Test for Heteroskedasticity," *Journal of Econometrics*, 17, 107 – 112.
- [256] Koenker, R., 2004. "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74 – 89.
- [257] Koenker, R. and G. Bassett, 1978. "Regression Quantiles," *Econometrica*, 46, 107 – 112.
- [258] Koenker, R. and Z. Xiao, 2002. "Inference on the Quantile Regression Process," *Econometrica*, 81, 1583 – 1612.
- [259] Koenker, R., and Z. Xiao, 2004. "Unit Root Quantile Regression Inference," *Journal of the American Statistical Association*, 99, 775 – 787.
- [260] Koenker, R. and Z. Xiao, 2006. "Quantile Autoregression," *Journal of the American Statistical Association*, with discussion and rejoinder, 475, 980 – 1006.
- [261] Koop, G., 2003. *Bayesian Econometrics*. New York: John Wiley and Sons.
- [262] Krugman, P., 1991. "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483 – 499.
- [263] Kumbhakar, S., and K. Lovell. *Stochastic Frontier Analysis*. New York: Cambridge University Press, 2000.
- [264] Kwiatkowski, D., P. Phillips, P. Schmidt and Y. Shin, 1992. "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159 – 178.
- [265] Lalive, R., J. van Ours, and J. Zweimuller, 2006. "How Changes in Financial Incentives Affect the Duration of Unemployment," *Review of Economic Studies*, 73, 1009 – 1038.
- [266] Lalonde, R. J., 1986. "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review*, 76, 604 – 620.
- [267] Lancaster, T., 1990. *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.

- [268] Lancaster, T., 2004. *An Introduction to Modern Bayesian Inference*. Oxford; Oxford University Press.
- [269] Lee, L. F., 2004. "Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72, 1899 – 1925.
- [270] Lee, L. F., and J. Yu, 2010. "Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects," *Journal of Econometrics*, 154, 165 – 185.
- [271] Leigh, D., and A. Gill, 1997, "Labor Market Returns to Community Colleges: Evidence from Returning Adults," *Journal of Human Resources*, 32, 334 – 353.
- [272] Levin, A., C. Lin and C. Chu, 2002. "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties," *Journal of Econometrics*, 108, 1 – 24.
- [273] Levitt, S., 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, 87, 270 – 290.
- [274] Levitt, S., 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Reply," *American Economic Review*, 92, 1244 – 1250.
- [275] Li, H. and L. Zhou, 2005. "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China," *Journal of Public Economics*, 89, 1743 – 1762.
- [276] Lin, D. Y., and L. J. Wei. 1989. "The Robust Inference for the Cox Proportional Hazards Model," *Journal of the American Statistical Association*, 84, 1074 – 1078.
- [277] Lin, J., 1992. "Rural Reforms and Agricultural Growth in China," *American Economic Review*, 82, 34 – 51.
- [278] Liu, R., 1988. "Bootstrap Procedures under Some Non-iid Models," *Annals of Statistics*, 16, 1696 – 1708.
- [279] Ljung and Box, 1979. "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 66, 265 – 270.
- [280] Long, J. Scott, and Jeremy Freese, 2006. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd edition. College Station, TX: Stata Press.
- [281] Lutkepohl, H., 2005. *New Introduction to Multiple Time Series Analysis*. Berlin; Springer.
- [282] Maddala, G. S., 2001. *Introduction to Econometrics*, 3rd Edition. Chichester, West Sussex, England: John Wiley & Sons Ltd.
- [283] Maddala, G. S., 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- [284] Mankiw, N. G., D. Romer and D. Weil, 1992. "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 407 – 437.
- [285] McCall, B. P., 1996. "Unemployment Insurance Rules, Joblessness, and Part-Time Work," *Econometrica*, 64, 647 – 682.
- [286] McCrary, J., 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment," *American Economic Review*, 92, 1236 – 1243.
- [287] McCrary, J., 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698 – 714.
- [288] McFadden, D., 1974. "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers of Econometrics*, P. Zarembka (Ed.), New York, Academic Press.
- [289] McFadden, D., 1978. "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, 75 – 96. A. Karlquist, L. Lundquist, F. Snickars, and J. W. Weibull et al. (Eds.), Amsterdam, New York, North-Holland.
- [290] Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, 2004. "Economic Shocks and Civil Conflict: An Instrumental Variable Approach," *Journal of Political Economy*, 112, 725 – 753.
- [291] Moran, P., 1950. "Notes on Continuous Stochastic Phenomena," *Biometrika*, 37, 17 – 23.
- [292] Morgan, S., and C. Winship, 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*.

- Research. Cambridge University Press.
- [293] Munnell, A. H., 1990. "Why Has Productivity Growth Declined? Productivity and Public Investment," *New England Economic Review*, January/February, 3 – 22.
 - [294] Munnell, A. H., G. Tootell, L. Browne, and J. McEneaney, 1996. "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review*, 86, 25 – 53.
 - [295] Nadaraya, E., 1964. "On Estimating Regression," *Theory of Probability and Its Applications*, 9, 141 – 142.
 - [296] Nelson, W., 1972. "Theory and Application of Hazard Plotting for Censored Failure Data," *Technometrics*, 14, 945 – 965.
 - [297] Nelson, C. R., and C. I. Plosser, 1982. "Trends versus Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139 – 162.
 - [298] Newey, W. K. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Journal of Econometrics*, 36, 231 – 250.
 - [299] Newey, W. and D. McFadden, 1994. "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, Vol. IV, edited by R. Engle and D. McFadden, 2112 – 2245, Elsevier Science.
 - [300] Newey, W. and K. West, 1987. "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703 – 708.
 - [301] Ng, S., and P. Perron, 1995. "Unit Root Tests in ARMA Models with Data-dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268 – 281.
 - [302] Ng, S., and P. Perron, 2000. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519 – 1554.
 - [303] Nerlove, M., 1963. "Returns to Scale in Electricity Supply," in C. Christ (ed.), *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*. Stanford: Stanford University Press.
 - [304] Nickell, S. J., 1981. "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417 – 1426.
 - [305] O'Connell, P. G. J., 1998. "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44, 1 – 19.
 - [306] Ord, J. K., 1975. "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120 – 126.
 - [307] Pesaran, M. H., 2004. "General Diagnostic Tests for Cross Section Dependence in Panels," Cambridge Working Papers in Economics, 0435, University of Cambridge.
 - [308] Phillips, P., 1986. "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311 – 340.
 - [309] Phillips, P., 1987. "Time Series Regressions with a Unit Root," *Econometrica*, 55, 277 – 301.
 - [310] Phillips, P. and P. Perron, 1988. "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335 – 346.
 - [311] Pindyck, R. S., and D. L. Rubinfeld. 1998. *Econometric Models and Economic Forecasts*. 4th ed. New York: McGraw-Hill.
 - [312] Poirier, D. J., 1980. "Partial Observability in Bivariate Probit Models," *Journal of Econometrics*, 12, 209 – 217.
 - [313] Poirier, D. J., 1995. *Intermediate Statistics and Econometrics*. Cambridge, MA: MIT Press.
 - [314] Powell, J. L., 1984. "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303 – 325.
 - [315] Prais, S. and C. Winsten, 1954. "Trend Estimation and Serial Correlation," Cowles Commission Discussion Paper No. 383.
 - [316] Pregibon, D., 1979. "Data Analytic Methods for Generalized Linear Models," PhD dissertation, University of

Toronto.

- [317] Pregibon, D., 1980. "Goodness of Link Tests for Generalized Linear Models," *Applied Statistics*, 29, 15–24.
- [318] Ramsey, J. B., 1969. "Tests for Specification Errors in Classical Linear Least-squares Regression Analysis," *Journal of the Royal Statistical Society, Series B*, 31, 350–371.
- [319] Rivers, D., and Q. H. Vuong, 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39, 347–366.
- [320] Robinson, P., 1988. "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.
- [321] Romer, C. and D. Romer, 2004. "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.
- [322] Roodman, D., 2009. "How to Do xtabond2: An Introduction to Difference and System GMM in Stata," *The Stata Journal*, 9, 86–136.
- [323] Rosenblatt, M., 1956. "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 832–837.
- [324] Rosenbaum, P. R., 2010. *Observational Studies*. Springer.
- [325] Rosenbaum, P. R., and D. B. Rubin, 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.
- [326] Rosenbaum, P. R. and D. B. Rubin, 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *The American Statistician*, 39, 33–38.
- [327] Rosenzweig, M. and K. Wolpin, 2000. "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*, 38, 827–874.
- [328] Rothstein, J., 2007. "Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)," *American Economic Review*, 97, 2026–2037.
- [329] Rubin, D. B., 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.
- [330] Schwarz, G., 1978. "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.
- [331] Schwert, G., 1989. "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–159.
- [332] Shapiro, M., and M. Watson, 1988. "Sources of Business Cycle Fluctuations," *NBER Macroeconomics Annual*, 3, 111–156.
- [333] Shea, J., 1997. "Instrumental Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics*, 79, 348–352.
- [334] Shen, Y. and Y. Yao, 2008. "Does Grassroots Democracy Reduce Income Inequality in China?" *Journal of Public Economics*, 92, 2182–2198.
- [335] Sims, C., 1980. "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- [336] Sims, C., 1981. "An Autoregressive Index Model for the U. S. 1948–1975", in J. Kmenta & J. B. Ramsey (eds), *Large-Scale Macro-Econometric Models*, North-Holland, Amsterdam, pp. 283–327.
- [337] Sims, C., 1986. "Are Forecasting Models Usable for Policy Analysis?" *Quarterly Review, Federal Reserve Bank of Minneapolis*, 10, 2–16.
- [338] Sims, C., 1992, "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36, 975–1011.
- [339] Skeels, C. L., and F. Vella, 1999. "A Monte Carlo Investigation of the Sampling Behavior of Conditional Moment Tests in Tobit and Probit Models," *Journal of Econometrics*, 92, 275–294.
- [340] Small, K., and C. Hsiao, 1985. "Multinomial Logit Specification Tests," *International Economic Review*, 26, 619–627.

- [341] Stevenson, R. E., 1980. "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics*, 13, 57 - 66.
- [342] Stigler, G. and G. Becker, 1977. "De Gustibus Non Est Disputandum," *American Economic Review*, 67, 76 - 90.
- [343] Stock, J. and M. Watson, 1993. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783 - 820.
- [344] Stock, J. and M. Watson, 2001. "Vector Autoregressions," *Journal of Economic Perspectives*, 15, 101 - 115.
- [345] Stock, J. and M. Watson, 2011. *Introduction to Econometrics*, 3rd edition, Boston: Addison-Wesley.
- [346] Stock, J. and M. Yogo, 2005. "Testing for Weak Instruments in Linear IV Regression," in D. Andrew and J. Stock (ed.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80 - 108. Cambridge: Cambridge University Press.
- [347] Swamy, P., 1970. "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38, 311 - 323.
- [348] Theil, H., 1953. "Repeated Least Squares Applied to Complete Equation System," *The Hague, Central Planning Bureau*, mimeo.
- [349] Thistlethwaite, D., and D. T. Campbell, 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51, 309 - 51, 317.
- [350] Tobin, J., 1958. "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24 - 36.
- [351] Tobler, W., 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, 46, 234 - 240.
- [352] Tse, Y. K., and A. K. C. Tsui, 2002. "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-varying Correlations," *Journal of Business and Economic Statistics*, 20, 351 - 362.
- [353] Tukey, J. W., 1949. "One Degree of Freedom for Non-additivity," *Biometrics*, 5, 232 - 242.
- [354] Van den Berg, G., 2001. "Duration Models: Specification, Identification, and Multiple Durations," in *Handbook of Econometrics*, Volume 5, ed. J. J. Heckman and E. Leamer. Amsterdam: North Holland, 3381 - 3460.
- [355] Van der Klaauw, W., 2008. "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics," *Labour*, 22, 219 - 245.
- [356] Varian, H., 2001. "What I Have Learned about Writing Economics," available online at <http://people.ischool.berkeley.edu/~hal/Papers/writing-economics.html>.
- [357] Varian, H., 2009. "How to Build an Economic Model in Your Spare Time," available online at <http://people.ischool.berkeley.edu/~hal/Papers/how.pdf>.
- [358] Verbeek, M., 2004. *A Guide to Modern Econometrics*, 2nd edition. Chichester, West Sussex, England: John Wiley & Sons Ltd.
- [359] Vuong, Q. 1989. "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses." *Econometrica*, 57, 307 - 334.
- [360] Waldman, M., S. Nicholson, and N. Adilov, 2006. "Does Television Cause Autism?" Cornell University, working paper.
- [361] Waldman, M., S. Nicholson, N. Adilov, and J. Williams, 2008. "Autism Prevalence and Precipitation Rates in California, Oregon, and Washington Counties," *Archives of Pediatrics & Adolescent Medicine*, 162, 1026 - 1034.
- [362] Wang, Q., and N. Wu, 2012. "Menu-driven X-12-ARIMA Seasonal Adjustment in Stata," *The Stata Journal*, 12, 214 - 241.
- [363] Watson, G., 1964. "Smooth Regression Analysis," *Sankhya A*, 26, 359 - 372.

- [364] Windmeijer, F. 2005. "A Finite Sample Correction for the Variance of Linear Efficient Two-step GMM Estimators," *Journal of Econometrics*, 126, 25 - 51.
- [365] White, H., 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817 - 838.
- [366] Wooldridge, J., 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA : MIT Press.
- [367] Wooldridge, J., 2003. *Introductory Econometrics: A Modern Approach*, 2nd edition. South-Western, a division of Thomson Learning.
- [368] Wooldridge, J., 2009. *Introductory Econometrics: A Modern Approach*, 4th edition. South-Western, a part of Cengage Learning.
- [369] Wooldridge, J., 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd edition. Cambridge, MA : MIT Press.
- [370] Wu, C., 1986. "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics*, 14, 1261 - 1295.
- [371] Wu, D., 1973. "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrics*, 41, 733 - 775.
- [372] Xiao, Z., 2012. "Time Series Quantile Regression," *Handbook of Statistics*, Vol. 30, Elsevier.
- [373] Zellner, A., 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias," *Journal of the American Statistical Association*, 58, 977 - 992.
- [374] Zellner, A., 1971. *Introduction to Bayesian Inference in Econometrics*. New York : John Wiley and Sons.
- [375] Zellner, A., and N. S. Revankar. 1969. "Generalized Production Functions," *Review of Economic Studies*, 36, 241 - 250.
- [376] Zellner, A. and H. Theil, 1962. "Three Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica*, 30, 63 - 68.

数 学 符 号

数学符号	中文名称	英文名称
β	参数 β 的真实值	True value of parameter β
$\hat{\beta}$	参数 β 的估计值	Estimated value of parameter β
$\tilde{\beta}$	参数 β 的假想值	Hypothetical value of parameter β
\equiv	定义	Definition
$\mathbf{A} \otimes \mathbf{B}$	克罗内克尔乘积	Kronecker product
∞	无穷大	Infinity
\propto	成正比	Proportional to
\xrightarrow{d}	依分布收敛	Converge in distribution
\xrightarrow{p}	依概率收敛	Converge in probability
$ x $	x 的绝对值	Absolute value of x
$\mathbf{1}(x)$	示性函数	Indicator Function
$\exp(x)$	指数函数	Exponential Function
\mathbf{A}' 或 \mathbf{A}^T	矩阵 \mathbf{A} 的转置	Transpose of matrix \mathbf{A}
\mathbf{A}^{-1}	矩阵 \mathbf{A} 的逆矩阵	Inverse of matrix \mathbf{A}
$ \mathbf{A} $	矩阵 \mathbf{A} 的行列式	Determinant of matrix \mathbf{A}
Rank(\mathbf{A})	矩阵 \mathbf{A} 的秩	Rank of matrix \mathbf{A}
Trace(\mathbf{A})	矩阵 \mathbf{A} 的迹	Trace of matrix \mathbf{A}
Δ	差分算子	Difference operator
L	滞后算子	Lag operator
$\sum_{t=1}^T$	求和算子	Summation operator
$\prod_{t=1}^T$	求积算子	Product operator
\int_a^b	定积分	Definite integral
$\frac{dy}{dx}$	y 对 x 的导数	Differentiation of y with respect to x
$\frac{\partial y}{\partial x}$	y 对 x 的偏导数	Partial differentiation of y with respect to x
$f'(x)$	函数 $f(x)$ 的一阶导数	First order differentiation of $f(x)$
$f''(x)$	函数 $f(x)$ 的二阶导数	Second order differentiation of $f(x)$
Var	方差(矩阵)	Variance (Covariance matrix)
Avar	渐近方差(矩阵)	Asymptotic variance (matrix)

续表

数学符号	中文名称	英文名称
Bias	偏差	Bias
Corr	相关系数	Correlation coefficient
Cov	协方差(矩阵)	Covariance (matrix)
E	(数学)期望	(Mathematical) Expectation
max	最大值	maximize
min	最小值	minimize
MSE	均方误差	Mean Squared Error
P(A)	事件A发生的概率	Probability of event A
$A \cap B$	事件A与B的交集	Intersection of event A and B
$A \cup B$	事件A与B的并集	Union of event A and B
$P(A B)$	给定事件B发生的条件下,事件A发生的概率	Probability of event A, given event B
plim	依概率收敛	Converge in probability
lim	收敛(微积分)	Converge (as in Calculus)
s. t.	受约束	Subject to
vech	半向量化算子	half vectorization

英 文 缩 写

英文缩写	中文名称	英文名称
2SLS	二阶段最小二乘法	Two Stage Least Square
3SLS	三阶段最小二乘法	Three Stage Least Square
ACF	自相关函数	Autocorrelation Function
ADF	ADF 检验	Augmented Dickey-Fuller Test
ADL	自回归分布滞后模型	Autoregressive Distributed Lag Model
AFT	加速失效时间模型	Accelerated Failure Time Model
AIC	赤池信息准则	Akaike Information Criterion
AR(p)	p 阶自回归过程	Autoregressive Process of order p
ARCH	自回归条件异方差过程	Autoregressive Conditional Heteroskedasticity Process
ARCH-M	“ARCH-in-Mean”模型	ARCH-in-Mean Model
ARMA	自回归移动平均过程	Autoregressive Moving Average Process
ARIMA	自回归单整移动平均过程	Autoregressive Integrated Moving Average Process
AS	绝对值可加总	Absolutely Summable
ATE	平均处理效应	Average Treatment Effect
ATT	参与者平均处理效应	Average Treatment Effect on the Treated
ATU	非参与者平均处理效应	Average Treatment Effect on the Untreated
BG	BG 自相关检验	Breusch-Godfrey Test for Autocorrelation
BIC	贝叶斯信息准则	Bayesian Information Criterion
BLUE	最佳线性无偏估计	Best Linear Unbiased Estimator
BP	BP 异方差检验	Breusch-Pagan Test for Heteroskedasticity
BUE	最佳无偏估计	Best Unbiased Estimator
cdf	累积分布函数	cumulative distribution function
CIRF	累积脉冲响应函数	Cumulative Impulse Response Function
CLRM	古典线性回归模型	Classical Linear Regression Model
CO	Cochrane-Orcutt 估计法	Cochrane-Orcutt estimation
CV	交叉核实法	Cross Validation
DF	DF 检验	Dickey-Fuller Test
DF-GLS	DF-GLS 单位根检验	DF-GLS Unit Root Test
DGP	数据生成过程	Data Generating Process
DD	双重差分估计量	Differences-in-differences Estimator
DDD	三重差分估计量	Differences-in-differences-in-differences Estimator

续表

英文缩写	中文名称	英文名称
DPD	动态面板数据	Dynamic Panel Data
DVECH	对角半向量化 GARCH	Diagonal Vech GARCH
DW	DW 统计量	Durbin-Watson Statistic
DWH	德宾 - 吴 - 豪斯曼检验	Durbin-Wu-Hausman Test
ECM	误差修正模型	Error Correction Model
EG-ADF	EG-ADF 检验	EG-ADF Test
EGARCH	指数 GARCH	Exponential GARCH
FE	固定效应模型	Fixed Effects Model
FEVD	预测误差方差分解	Forecast-error Variance Decomposition
FGLS	可行广义最小二乘法	Feasible Generalized Least Square
FIML	全信息最大似然估计法	Full Information Maximum Likelihood Estimation
FWLS	可行加权最小二乘法	Feasible Weighted Least Square
GARCH	广义自回归条件异方差过程	Generalized Autoregressive Conditional Heteroskedasticity Process
GLS	广义最小二乘法	Generalized Least Square
GMM	广义矩估计	Generalized Method of Moments
H_0	原假设、零假设	Null Hypothesis
H_1	替代假设、备择假设	Alternative Hypothesis
HAC	异方差自相关稳健的标准误	Heteroskedasticity and Autocorrelation Consistent Standard Error
HQIC	汉南 - 昆信息准则	Hannan-Quinn Information Criterion
HT	HT 面板单位根检验	Harris and Tzavalis (1999) Panel Unit Root Test
iid	独立同分布	independently identically distributed
ILS	间接最小二乘法	Indirect Least Square
IMR	反米尔斯比率	Inverse Mill's Ratio
IMSE	积分均方误差	Integrated Mean Squared Error
IPS	IPS 面板单位根检验	Im, Pesaran and Shin(2002) Panel Unit Root Test
IRF	脉冲响应函数	Impulse Response Function
IV	工具变量(法)	Instrumental Variable (Estimation)
JB	雅克 - 贝拉正态检验	Jarque-Bera Test for Normality
KPSS	KPSS 平稳性检验	Kwiatkowski, Phillips, Schmidt and Shin (1992) Stationarity Test
LAD	最小绝对离差估计量	Least Absolute Deviation Estimator
LHS	方程左边	Left Hand Side
LIML	有限信息最大似然估计法	Limited Information Maximum Likelihood Estimation
LLC	LLC 面板单位根检验	Levin, Lin and Chu (2002) Panel Unit Root Test

续表

英文缩写	中文名称	英文名称
LM	拉格朗日乘子检验	Lagrange Multiplier Test
Lowess	局部加权散点光滑估计量	Locally weighted scatterplot smoothing
LPM	线性概率模型	Linear Probability Model
LR	似然比检验	Likelihood Ratio Test
LSDV	最小二乘虚拟变量模型	Least Square Dummy Variable Model
LSDVC	偏差校正 LSDV	Biased-corrected LSDV
MA(q)	q 阶移动平均过程	Moving Average Process of Order q
MC	蒙特卡罗方法	Monte Carlo Methods
MCMC	马尔可夫链蒙特卡罗法	Markov Chain Monte Carlo Methods
MDS	鞅差分序列	Martingale Difference Sequence
MGARCH	多维 GARCH	Multivariate GARCH
MLE	最大似然估计法	Maximum Likelihood Estimation
MM	矩估计	Method of Moments
MSE	均方误差	Mean Squared Error
MSL	最大模拟似然法	Maximum Simulated Likelihood
MSM	模拟矩估计	Method of Simulated Moments
NLS	非线性最小二乘法	Nonlinear Least Square
OIM	观测信息矩阵	Observed Information Matrix
OIRF	正交化的脉冲响应函数	Orthogonalized Impulse Response Function
OLS	普通最小二乘法	Ordinary Least Square
OPG/BHHH	梯度向量外积	Outer Product of Gradients
OS	一可加总	One-summable
PA	总体平均估计量	Population-averaged Estimator
PSM	倾向得分匹配	Propensity Score Matching
PACF	偏自相关函数	Partial Autocorrelation Function
pdf	概率密度函数	probability density function
PH	比例风险模型	Proportional Hazard Model
PL	部分线性模型	Partially Linear Model
PP	Phillips-Perron 单位根检验	Phillips-Perron Unit Root Test
PW	Prais-Winsten 估计法	Prais-Winsten Estimation
QLR	匡特似然比统计量	Quandt Likelihood Ratio
QMLE	准最大似然估计	Quasi Maximum Likelihood Estimation
QQ plot	分位数 - 分位数图	Quantile-quantile plot
QR	分位数回归	Quantile Regression
R^2	拟合优度、可决系数	Goodness of Fit, Coefficient of Determination
\bar{R}^2	校正拟合优度	Adjusted Goodness of Fit

续表

英文缩写	中文名称	英文名称
R_{ue}^2	非中心 R^2	Uncentered R^2
RDD	断点回归设计	Regression Discontinuity Design
RE	随机效应模型	Random Effects Model
RHS	方程右边	Right Hand Side
RLS	约束最小二乘法	Restricted Least Square
SAR	空间自回归	Spatial Autoregression
SARAR	带空间自回归误差项的空间自回归模型	Spatial Autoregressive Model with Spatial Autoregressive Disturbances
SIC	施瓦茨信息准则	Schwarz Information Criterion
SSR	残差平方和	Sum of Squared Residuals
SUR	似不相关回归	Seemingly Unrelated Regression Estimation
TARCH	门限 GARCH	Threshold GARCH
VAR	向量自回归	Vector Autoregression
vech	半向量化算子	half vectorization
VIF	方差膨胀系数	Variance Inflation Factor
VMA	向量移动平均过程	Vector Moving Average Process
Wald	沃尔德检验	Wald Test
WLS	加权最小二乘法	Weighted Least Square
WN	白噪声	White Noise
ZINB	零膨胀负二项回归	Zero-inflated Negative Binomial Regression
ZIP	零膨胀泊松回归	Zero-inflated Poisson Regression