

Probit regression		Number of obs = 3197				
		Wald chi2(11) = 366.94				
		Prob > chi2 = 0.0000				
Log pseudolikelihood = -1933.4275		Pseudo R2 = 0.0946				
ins		Robust				
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
linc	.3466893	.0402173	8.62	0.000	.2678648	.4255137
female	-.0815374	.0508549	-1.60	0.109	-.1812112	.0181364
age	.1162879	.1151924	1.01	0.313	-.109485	.3420608
age2	-.0009395	.0008568	-1.10	0.273	-.0026187	.0007397
educyear	.0464387	.0089917	5.16	0.000	.0288153	.0640622
married	.1044152	.0636879	1.64	0.101	-.0204108	.2292412
hisp	-.3977334	.1080935	-3.68	0.000	-.6095927	-.1858741
white	-.0418296	.0644391	-0.65	0.516	-.168128	.0844687
chronic	.0472903	.0186231	2.54	0.011	.0107897	.0837909
adl	-.0945039	.0353534	-2.67	0.008	-.1637953	-.0252125
hstatusg	.1138708	.0629071	1.81	0.070	-.0094248	.2371664
_cons	-5.744548	3.871615	-1.48	0.138	-13.33277	1.843677

上表显示, linc 的系数为 0.35, 且在 1% 水平上显著, 即家庭收入越高, 越可能购买私人健康保险。但怀疑 linc 为内生变量, 因为可能存在同时影响 ins 与 linc 的遗漏变量。为此, 使用数据集中另一变量 retire(本人是否退休)作为工具变量。是否退休显然与收入 linc 相关, 满足工具变量的相关性; 另一方面, 假设是否退休不直接影响购买私人健康保险的决策, 故满足工具变量的外生性。下面使用此工具变量进行 IV Probit 估计。

```
. ivprobit ins female age age2 educyear married hisp white chronic adl  
hstatusq (linc = retire),r nolog
```

Probit model with endogenous regressors				Number of obs = 3197		
Log pseudolikelihood = -5408.3856				Wald chi2(11) = 417.68		
				Prob > chi2 = 0.0000		
<hr/>						
	Robust				[95% Conf. Interval]	
	Coef.	Std. Err.	z	P> z		
linc	-.6199245	.358139	-1.73	0.083	-1.321864	.0820151
female	-.1424789	.0476242	-2.99	0.003	-.2358205	-.0491373
age	.298707	.1237695	2.41	0.016	.0561234	.5412907
age2	-.002233	.0009006	-2.48	0.013	-.0039981	-.000468
educyear	.1185139	.0208859	5.67	0.000	.0775783	.1594495
married	.7571106	.2162909	3.50	0.000	.3331881	1.181033
hisp	-.5103837	.1032417	-4.94	0.000	-.7127337	-.3080337
white	.174971	.097886	1.79	0.074	-.016882	.3668241
chronic	.0014631	.026898	0.05	0.957	-.0512561	.0541823
adl	-.1362883	.03404	-4.00	0.000	-.2030055	-.069571
hstatusg	.2419553	.066904	3.62	0.000	.1108258	.3730847
_cons	-10.25689	3.977975	-2.58	0.010	-18.05358	-2.460204
/athrho	.7576717	.3626733	2.09	0.037	.0468452	1.468498
/lnsigma	-.3313108	.0233359	-14.20	0.000	-.3770484	-.2855732
rho	.6397035	.2142599			.0468109	.8992906
sigma	.717982	.0167548			.6858829	.7515833

上表底部提供了对外生性原假设“ $H_0 : \rho = 0$ ”的沃尔德检验结果,其  $p$  值为 0.0367,故可在 5% 的水平上认为 linc 为内生变量。根据模型估计结果, $u_i$  与  $v_i$  的相关系数为  $\hat{\rho} = 0.64$ ,这表明,未度量的遗漏变量在增加家庭收入的同时,也会提高购买保险的倾向。相关系数高达 0.64 也意味着 ivprobit 的估计结果会与 probit 大不相同。事实上,使用 ivprobit 时,linc 的估计系数为 -0.62,不仅改变了符号,而且在 10% 水平上显著( $p$  值为 0.083)。这表明,如果使用一般的 Probit 模型进行估计,由于忽略了 linc 的内生性,将高估家庭收入对于保险购买倾向的正作用。

下面进行两步法估计。

```
. ivprobit ins female age age2 educyear married hisp white chronic adl
hstatusg (linc = retire),first twostep
```

Checking reduced-form model...						
First-stage regression						
Source	SS	df	MS	Number of obs = 3197		
Model	1172.11143	11	106.555584	F( 11, 3185) =	205.93	
Residual	1648.04737	3185	.517440303	Prob > F =	0.0000	
Total	2820.15879	3196	.882402626	R-squared =	0.4156	
				Adj R-squared =	0.4136	
				Root MSE =	.71933	
linc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
retire	-.0955818	.0286254	-3.34	0.001	-.1517078	-.0394557
female	-.1065852	.0282552	-3.77	0.000	-.1619853	-.051185
age	.262005	.0626898	4.18	0.000	.1390887	.3849214
age2	-.0018737	.0004643	-4.04	0.000	-.002784	-.0009635
educyear	.0945358	.00435	21.73	0.000	.0860068	.1030649
married	.7647985	.0312142	24.50	0.000	.7035966	.8260004
hisp	-.2340063	.0523453	-4.47	0.000	-.3366401	-.1313725
white	.2329204	.0347781	6.70	0.000	.1647307	.30111
chronic	-.0386546	.0100859	-3.83	0.000	-.0584301	-.0188791
adl	-.0741595	.0173479	-4.27	0.000	-.1081738	-.0401453
hstatusg	.1739715	.0338516	5.14	0.000	.1075984	.2403446
_cons	-7.505571	2.114278	-3.55	0.000	-11.65106	-3.360086
Two-step probit with endogenous regressors						
				Number of obs =	3197	
				Wald chi2(11) =	207.35	
				Prob > chi2 =	0.0000	
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
linc	-.8064908	.6526256	-1.24	0.217	-2.085613	.4726318
female	-.1853647	.084243	-2.20	0.028	-.3504779	-.0202515
age	.3886148	.2094068	1.86	0.063	-.0218149	.7990446
age2	-.0029052	.0015308	-1.90	0.058	-.0059056	.0000952
educyear	.1541854	.0617338	2.50	0.013	.0331895	.2751814
married	.9849852	.503158	1.96	0.050	-.0011864	1.971157
hisp	-.6640124	.1983783	-3.35	0.001	-1.052827	-.2751979
white	.2276309	.1709828	1.33	0.183	-.1074892	.562751
chronic	.0019054	.0339542	0.06	0.955	-.0646436	.0684544
adl	-.1773114	.0621765	-2.85	0.004	-.299175	-.0554478
hstatusg	.3147822	.1360959	2.31	0.021	.0480392	.5815251
_cons	-13.34419	6.427051	-2.08	0.038	-25.94097	-.7473972
Instrumented:	linc					
Instruments:	female age age2 educyear married hisp white chronic adl hstatusg retire					
Wald test of exogeneity:	chi2(1) =	4.37		Prob > chi2 =	0.0366	

上表显示,两步法的估计系数与 MLE 类似,但前者的绝对值比后者约大 25%,这与 MLE 对  $\rho$  的估计相一致,因为  $\sqrt{1 - \rho^2} = \sqrt{1 - 0.6397035^2} \approx 0.77$ 。上表底部对于 linc 外生性原假设的 Wald 检验结果表明,  $p$  值为 0.0366,故可在 5% 水平上认为 linc 为内生变量。另外,第一步回归的结果显示,工具变量 retire 对于内生变量 linc 具有较强的解释力。

## 11.7 双变量 Probit 模型(选读)

有时候需要对两个虚拟变量同时考虑其发生的概率。比如,被解释变量之一是病人是否去看医生,而被解释变量之二是病人是否住院。如果对这两个被解释变量分别进行 Probit 建模,则可能损失效率(但依然为一致估计),因为“看医生”与“住院”这两件事通常是相关的,即两个 Probit 方程的扰动项之间可能存在相关性。具体考察以下模型:

$$\begin{cases} y_1^* = \mathbf{x}'_1 \boldsymbol{\beta}_1 + \varepsilon_1 \\ y_2^* = \mathbf{x}'_2 \boldsymbol{\beta}_2 + \varepsilon_2 \end{cases} \quad (11.35)$$

其中,  $y_1^*$  与  $y_2^*$  为不可观测的潜变量,扰动项  $(\varepsilon_1, \varepsilon_2)$  服从二维联合正态分布,期望为 0,方差为 1,而相关系数为  $\rho$ ,即

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\} \quad (11.36)$$

可观测变量  $y_1$  与  $y_2$  由以下方程决定:

$$y_1 = \begin{cases} 1 & \text{若 } y_1^* > 0 \\ 0 & \text{若 } y_1^* \leq 0 \end{cases} \quad (11.37)$$

$$y_2 = \begin{cases} 1 & \text{若 } y_2^* > 0 \\ 0 & \text{若 } y_2^* \leq 0 \end{cases} \quad (11.38)$$

如果  $\mathbf{x}_1 = \mathbf{x}_2$  (两个方程的解释变量完全相同),则方程(11.35)–(11.38)称为“双变量 Probit”;反之,如果  $\mathbf{x}_1 \neq \mathbf{x}_2$  (两个方程的解释变量不完全相同),则称为“似不相关双变量 Probit”(seemingly unrelated bivariate probit),因为这两个方程的唯一联系是扰动项的相关性<sup>①</sup>。显然,如果  $\rho = 0$ ,则此模型等价于两个单独的 Probit 模型。当  $\rho \neq 0$  时,可写下  $(y_1, y_2)$  的取值概率,然后进行最大似然估计。比如:

$$\begin{aligned} p_{11} &\equiv P(y_1 = 1, y_2 = 1) = P(y_1^* > 0, y_2^* > 0) \\ &= P(\varepsilon_1 > -\mathbf{x}'_1 \boldsymbol{\beta}_1, \varepsilon_2 > \mathbf{x}'_2 \boldsymbol{\beta}_2) \\ &= P(\varepsilon_1 < \mathbf{x}'_1 \boldsymbol{\beta}_1, \varepsilon_2 < \mathbf{x}'_2 \boldsymbol{\beta}_2) \\ &= \int_{-\infty}^{\mathbf{x}'_1 \boldsymbol{\beta}_1} \int_{-\infty}^{\mathbf{x}'_2 \boldsymbol{\beta}_2} \phi(z_1, z_2, \rho) dz_1 dz_2 \\ &= \Phi(\mathbf{x}'_1 \boldsymbol{\beta}_1, \mathbf{x}'_2 \boldsymbol{\beta}_2, \rho) \end{aligned} \quad (11.39)$$

其中,  $\phi(z_1, z_2, \rho)$  与  $\Phi(z_1, z_2, \rho)$  分别为标准化的二维正态分布 (standardized bivariate normal) 的概率密度函数与累积分布函数,期望为 0,方差为 1,而相关系数为  $\rho$ 。类似地,可计算  $p_{10}, p_{01}$ ,

<sup>①</sup> 有关“似不相关回归”(seemingly unrelated regression),参见第 23 章。

$p_{00}$ , 将这些概率取对数后加总, 即得到对数似然函数。最后, 对原假设“ $H_0 : \rho = 0$ ”进行检验, 可判断有无必要使用双变量 Probit 模型, 或估计两个单独的 Probit 模型。在理论上, 双变量 Probit 模型可推广到多变量 Probit(multivariate probit), 但实际操作上较难计算高维积分, 故 Stata 尚未提供多变量 Probit 的命令。

双变量 Probit 的 Stata 命令为

```
. biprobit y1 y2 x1 x2 x3, r          (解释变量完全相同的情形)
.biprobit (y1 = x1 x2) (y2 = x1 x3), r  (解释变量不完全相同的情形)
```

其中, 选择项“r”表示使用稳健标准误。

以 Pindyck and Rubin (1998) 所提供的数据集 school.dta 为例。该数据集包括了 1973 年美国密西根州特洛伊市(Troy, Michigan)增加房产税以支持公立学校议案的 95 名投票人信息。主要变量包括: vote(是否投票赞成), private(是否有孩子在私立学校上学), logptax(房产税对数), loginc(家庭年收入对数), school(投票人是否教师), years(在特洛伊市居住的年数)。假设同时关心是否送孩子上私立学校以及是否投票赞成, 而这两件事可能相关(如果孩子上私立学校, 则可能不愿增加房产税以支持公立学校), 故使用双变量 Probit 模型进行估计。

```
. biprobit (private = loginc logptax school) (vote = loginc logptax
years school), r nolog
```

Seemingly unrelated bivariate probit		Number of obs = 95			
		Wald chi2(7) = 611.22			
		Prob > chi2 = 0.0000			
<b>Log pseudolikelihood = -84.695294</b>					
		Robust			
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
private					
loginc	.3911055	.5826198	0.67	0.502	-.7508083 1.533019
logptax	-.1877972	.758957	-0.25	0.805	-1.675326 1.299731
school	-4.889164	.2553132	-19.15	0.000	-5.389568 -4.388759
_cons	-3.795514	6.355104	-0.60	0.550	-16.25129 8.660261
vote					
loginc	1.262054	.4367689	2.89	0.004	.4060021 2.118105
logptax	-1.254895	.535599	-2.34	0.019	-2.30465 -.2051403
years	-.0138503	.0177383	-0.78	0.435	-.0486168 .0209162
school	1.724164	.4377039	3.94	0.000	.8662806 2.582048
_cons	-3.543213	4.549227	-0.78	0.436	-12.45953 5.373108
/athrho	-.193618	.2324816	-0.83	0.405	-.6492736 .2620376
rho	-.1912343	.2239796			-.5711808 .2562004
Wald test of rho=0:			chi2(1) = .693608	Prob > chi2 = 0.4049	

上表显示, 送孩子上私立学校(private)与教师身份(school)显著负相关; 而投票赞成(vote)与家庭收入(loginc)、教师身份(school)显著正相关, 与房地产税(logptax)显著负相关, 而与居住年数(years)无关。上表最后一行对原假设“ $H_0 : \rho = 0$ ”的沃尔德检验显示,  $p$  值高达 0.40, 故可认为  $\rho = 0$ , 没必要使用双变量 Probit 模型。

为了比较, 下面对变量 vote 进行单独的 Probit 估计。

```
.probit vote loginc logptax years school, r nolog
```

Probit regression		Number of obs = 95 Wald chi2(4) = 22.07 Prob > chi2 = 0.0002 Pseudo R2 = 0.1324				
		Log pseudolikelihood = -54.693088				
vote	Coef.	Robust				
		Std. Err.	z	P> z	[95% Conf. Interval]	
loginc	1.260286	.4369633	2.88	0.004	.403854	2.116719
logtax	-1.23849	.5281871	-2.34	0.019	-2.273717	-.2032619
years	-.0130624	.0177873	-0.73	0.463	-.0479248	.0217999
school	1.72503	.4363312	3.95	0.000	.8698368	2.580224
_cons	-3.646993	4.479862	-0.81	0.416	-12.42736	5.133376

从上表可知,单独的 Probit 估计结果(系数与标准误)与双变量 Probit 很接近,这也印证了没必要使用双变量 Probit。

## 11.8 部分可观测的双变量 Probit 模型(选读)

对于双变量 Probit 模型,有时无法同时观测到  $y_1$  与  $y_2$ ,而只能看到  $y_1$  与  $y_2$  都取值为 1 的情形。比如,估计实习生在某公司实习后留任的概率。显然,“留任”既取决于该公司是否向该实习生发聘书(公司的二值选择),也取决于该实习生是否愿意在此公司工作(实习生的二值选择)。而我们只能观测到该实习生是否留任了。如果留任,则  $y_1$  与  $y_2$  都取值为 1;如果未留任,则只知道  $y_1$  与  $y_2$  至少有一个为 0。在此情况下,公司的二值选择与实习生的二值选择可能相关,故适用于双变量 Probit 模型的框架。但对于  $y_1$  与  $y_2$ ,只有其乘积  $y_1y_2$  可观测,要么  $y_1y_2 = 1$ ,要么  $y_1y_2 = 0$ ;称为“部分可观测的双变量 Probit”(bivariate probit with partial observability),最早由 Poirier (1980) 提出。作为另一例子,由两名委员组成的委员会匿名对一个议案投票,而该议案只有全票通过才生效。对于外人来说,只能看到该议案通过与否;如果没有通过,则无法知道谁投了否决票。

定义  $z = y_1y_2$ ,可计算  $z$  的取值概率,并写出似然函数进行 MLE 估计。

$$P(z=1) = P(y_1=1, y_2=1) = \Phi(\mathbf{x}'_1 \boldsymbol{\beta}_1, \mathbf{x}'_2 \boldsymbol{\beta}_2, \rho) \quad (11.40)$$

$$P(z=0) = 1 - P(y_1=1, y_2=1) = 1 - \Phi(\mathbf{x}'_1 \boldsymbol{\beta}_1, \mathbf{x}'_2 \boldsymbol{\beta}_2, \rho) \quad (11.41)$$

部分可观测的双变量 Probit 模型的 Stata 命令为

```
. biprobit (y1 = x1 x2) (y2 = x1 x3), r partial difficult
```

其中,选择项“partial”表示估计部分可观测的双变量 Probit,但加上此选择项后可能使得数值计算不收敛。如果出现不收敛的情形,可使用选择项“difficult”,它表示在最大化的目标函数存在非凹区域(nonconcave regions)的情况下,使用另一迭代方法。

在使用命令“biprobit,partial”进行估计时,只用到  $y_1y_2$  的信息,与  $y_1, y_2$  的分别取值无关。对于现实中的部分可观测的双变量 Probit 模型,其实只知道  $y_1y_2$ ,记  $z_1 = y_1y_2$ 。为此,可先定义另一个被解释变量  $z_2 = z_1$ ,然后再进行估计。显然,由于  $z_1 = z_2$  都是虚拟变量,故  $z_1z_2 = z_1$ 。需要注意的是,我们无法令  $z_2 = 1$ (尽管也满足  $z_1z_2 = z_1$ ),因为 Stata 不允许 Probit 模型中出现所有取值都是 1 的被解释变量。具体操作命令如下:

```
. gen z2 = z1
.biprobit (z1 = x1 x2) (z2 = x1 x3), r partial difficult
```

为了演示的目的,对上节的 school.dta 数据集进行部分可观测的双变量 Probit 估计(假设只能观测到 vote 与 private 的乘积)。

```
. biprobit (private = loginc logptax school) (vote = loginc logptax years school), r partial difficult nolog
```

Partial observability bivariate probit						Number of obs = 95
						Wald chi2(6) = .
						Prob > chi2 = .
Log pseudolikelihood = -14.126282						
<hr/>						
		Robust				
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
<hr/>						
private						
loginc	.3416941	.9233022	0.37	0.711	-1.467945	2.151333
logptax	-.8713551	1.01297	-0.86	0.390	-2.85674	1.11403
school	-6.106526	.5756618	-10.61	0.000	-7.234802	-4.978249
_cons	1.587501	6.879705	0.23	0.818	-11.89647	15.07148
<hr/>						
vote						
loginc	21.08716	1.287221	16.38	0.000	18.56425	23.61006
logptax	-30.21983	2.239179	-13.50	0.000	-34.60854	-25.83112
years	-14.56743	.8514695	-17.11	0.000	-16.23628	-12.89858
school	-4.1137	.7031372	-5.85	0.000	-5.491824	-2.735577
_cons	56.3888	11.57035	4.87	0.000	33.71134	79.06626
<hr/>						
/athrho	-3.430044	.0173811	-197.34	0.000	-3.464111	-3.395978
<hr/>						
rho	-0.9979046	.0000728			-0.9980424	-0.997757
<hr/>						
Wald test of rho=0:		chi2(1) = 38944.3				Prob > chi2 = 0.0000

上表显示,部分可观测的双变量 Probit 估计结果与(全信息)双变量 Probit 在估计系数上有较大差别。对于前者而言,由于信息较少,故损失了效率。上表最后一行的 Wald 检验  $p$  值为 0.000 0,故强烈拒绝“ $\rho = 0$ ”的原假设(当然,这是基于信息不完全情况下作出的统计推断;根据上节的结果,如果同时知道 private 与 vote,则可接受“ $\rho = 0$ ”)。

## 习题

11.1 对于二值变量  $y_i$ ,考虑线性概率模型  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ,其中  $P(y_i = 1 | x_i) = \beta_0 + \beta_1 x_i$ 。

- (1) 证明  $E(\varepsilon_i | x_i) = 0$ 。
- (2) 求  $\text{Var}(\varepsilon_i | x_i)$  的表达式。
- (3) 扰动项  $\varepsilon_i$  是否存在异方差?为什么?

11.2 假定第  $i$  个家庭选择消费水平  $c_i$  与捐赠水平  $q_i$ ,来求解以下预算约束条件下的效用最大化问题:

$$\begin{aligned} \max_{c_i, q_i} U_i &= c_i + a_i \ln(1 + q_i) \\ \text{s. t. } c_i + p_i q_i &\leq m_i, c_i, q_i \geq 0 \end{aligned}$$

其中, $m_i$  为收入, $p_i$  为捐赠的“价格”且  $p_i < 1$ (因为捐赠可以抵税), $a_i$  表示捐赠的边际效用。将  $m_i$  与  $p_i$  均视为外生变量。可以证明,如果  $a_i \leq p_i$ ,则最优捐赠  $q_i = 0$ ;反之,如果  $a_i > p_i$ ,则最优捐赠  $q_i = (a_i/p_i) - 1$ 。定义二值变量  $y_i = \begin{cases} 1, & \text{若 } q_i > 0 \\ 0, & \text{若 } q_i = 0 \end{cases}$ ,假设家庭的捐赠偏好  $a_i$  取决于  $a_i = \exp(x'_i \boldsymbol{\beta} + \varepsilon_i)$ ,其中  $x_i$  为可观测的家庭特征,而

$\varepsilon_i$  为不可观测的扰动项。假定  $\varepsilon_i$  独立于  $(x_i, m_i, p_i)$ , 且  $\varepsilon_i \sim N(0, \sigma^2)$ 。求  $P(y_i = 1 | x_i, m_i, p_i)$  的表达式。这是一个 Probit 模型吗?

11.3 数据集 `hmda_short.dta`<sup>①</sup> 包含了美国波士顿有关房贷申请的横截面数据,包括以下变量:虚拟变量 `deny`(房贷申请被拒),`piratio`(`payment to income ratio`, 支出占收入比重),`score`(信用分数,1 最高,6 最低),`black`(是否为黑人)。考虑以下模型

$$\text{deny}_i = \beta_0 + \beta_1 \text{piratio}_i + \beta_2 \text{score}_i + \beta_3 \text{black}_i + \varepsilon_i \quad (11.42)$$

- (1) 用 OLS 估计线性概率模型；
  - (2) 估计 Probit 模型，计算边际效应及正确预测的比率；
  - (3) 估计 Logit 模型，计算边际效应及正确预测的比率；
  - (4) 估计条件异方差 Probit 模型，并确定是否存在条件异方差。

**11.4** 数据集 gss\_cdvm.dta 来自 2000 年与 2002 年的 General Social Survey, 包含以下变量: 虚拟变量 trust (是否信任大多数人), www(是否用互联网), educate(教育年限), income(家庭收入), age(年龄), male(是否男性)。

- (1) 估计 Probit 模型, 考察 `educate`, `income`, `age`, `male` 对 `trust` 的影响。
  - (2) 估计 Probit 模型, 考察 `educate`, `income`, `age`, `male` 对 `www` 的影响。
  - (3) 估计双变量 Probit 模型, 同时考察 `trust` 与 `www` 的决定。是否有必要使用双变量 Probit 模型?

<sup>①</sup> 此数据集来自 Stock and Watson (2011)。

# 第 12 章 多值选择模型

## 12.1 多项 Logit 与多项 Probit

个体面临的选择有时是多值的,而不仅仅是二值的。比如,交通工具的选择,职业的选择,手机消费品牌的选择。假设可供个体选择的方案为  $y = 1, 2, \dots, J$ , 其中  $J$  为正整数,即共有  $J$  种互相排斥的选择。如果  $J=2$ , 即为二值选择模型(参见第 11 章)。

使用随机效用法,假设个体  $i$  选择方案  $j$  所能带来的随机效用为

$$U_{ij} = \mathbf{x}'_i \boldsymbol{\beta}_j + \varepsilon_{ij} \quad (i=1, \dots, n; j=1, \dots, J) \quad (12.1)$$

其中,解释变量  $\mathbf{x}_i$  只随个体  $i$  而变,不随方案  $j$  而变。比如,个体的性别、年龄、收入等特征。这种解释变量被称为“只随个体而变”(case-specific)或“不随方案而变”(alternative-invariant)。系数  $\boldsymbol{\beta}_j$  表明,  $\mathbf{x}_i$  对随机效用  $U_{ij}$  的作用取决于方案  $j$ 。

显然,个体  $i$  选择方案  $j$ ,当且仅当方案  $j$  带来的效用高于所有其他方案,故个体  $i$  选择方案  $j$  的概率可写为

$$\begin{aligned} P(y_i=j | \mathbf{x}_i) &= P(U_{ij} \geq U_{ik}, \forall k \neq j) \\ &= P(U_{ik} - U_{ij} \leq 0, \forall k \neq j) \\ &= P(\varepsilon_{ik} - \varepsilon_{ij} \leq \mathbf{x}'_i \boldsymbol{\beta}_j - \mathbf{x}'_i \boldsymbol{\beta}_k, \forall k \neq j) \end{aligned} \quad (12.2)$$

假设  $\{\varepsilon_{ij}\}$  为 iid 且服从 I 型极值分布(type I extreme value distribution),则可证明:

$$P(y_i=j | \mathbf{x}_i) = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_j)}{\sum_{k=1}^J \exp(\mathbf{x}'_i \boldsymbol{\beta}_k)} \quad (12.3)$$

显然,选择各项方案的概率之和为 1,即  $\sum_{j=1}^J P(y_i=j | \mathbf{x}_i) = 1$ 。方程(12.3)是二值选择 Logit 模型向多值选择模型的自然推广。需要注意的是,无法同时识别所有的系数  $\boldsymbol{\beta}_k, k=1, \dots, J$ 。这是因为,如果将  $\boldsymbol{\beta}_k$  变为  $\boldsymbol{\beta}_k^* = \boldsymbol{\beta}_k + \boldsymbol{\alpha}$ ( $\boldsymbol{\alpha}$  为某常数向量),完全不会影响模型的拟合。为此,通常将某方案(比如,方案 1)作为“参照方案”(base category),然后令其相应系数  $\boldsymbol{\beta}_1 = \mathbf{0}$ 。由此,个体  $i$  选择方案  $j$  的概率为

$$P(y_i=j | \mathbf{x}_i) = \begin{cases} \frac{1}{1 + \sum_{k=2}^J \exp(\mathbf{x}'_i \boldsymbol{\beta}_k)} & (j=1) \\ \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_j)}{1 + \sum_{k=2}^J \exp(\mathbf{x}'_i \boldsymbol{\beta}_k)} & (j=2, \dots, J) \end{cases} \quad (12.4)$$

其中,“ $j=1$ ”所对应的方案为参照方案。此模型称为“多项 Logit”(Multinomial Logit),可用 MLE 进行估计。个体  $i$  的似然函数为

$$L_i(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J) = \prod_{j=1}^J [P(y_i=j | \mathbf{x}_i)]^{1(y_i=j)} \quad (12.5)$$

其对数似然函数为

$$\ln L_i(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J) = \sum_{j=1}^J \mathbf{1}(y_i=j) \cdot \ln P(y_i=j | \mathbf{x}_i) \quad (12.6)$$

其中,  $\mathbf{1}(\cdot)$  为示性函数(indicator function), 即如果括号中的表达式成立, 则取值为 1; 反之, 取值为 0。将所有个体的对数似然函数加总, 即得到整个样本的对数似然函数, 将其最大化则得到系数估计值  $\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_J$ 。

另外, 在方程(12.2)中, 如果假设  $\{\varepsilon_{i1}, \dots, \varepsilon_{iJ}\}$  服从  $J$  维正态分布, 则可得到“多项 Probit”(Multinomial Probit)模型, 但该模型的选择概率涉及高维积分, 不易计算。在扰动项  $\{\varepsilon_{i1}, \dots, \varepsilon_{iJ}\}$  相互独立的假定下, Stata 首先把此高维积分转化为一维积分, 然后进行数值计算(但此一维积分也没有解析解)。

## 12.2 条件 Logit 模型

多项 Logit 模型仅考虑不随方案而变的解释变量(比如, 个体收入), 但有些解释变量可能既随个体而变, 也随方案而变。比如, 在选择交通工具时, 乘车时间既因个体而异(不同个体的行车路线可能不同), 也因交通工具而异(坐公交的时间与乘出租车的时间不同)。这种解释变量称为“随方案而变”(alternative-specific), 既包括同时随方案与个体而变的变量, 也包括随方案而变但不随个体而变的变量(比如, 选择加入某协会时, 所有人交的会费都一样)。

此时, 个体  $i$  选择方案  $j$  所带来的随机效用为

$$U_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \varepsilon_{ij} \quad (i=1, \dots, n; j=1, \dots, J) \quad (12.7)$$

其中, 解释变量  $\mathbf{x}_{ij}$  表明, 它既随个体  $i$  而变, 也随方案  $j$  而变。系数  $\boldsymbol{\beta}$  表明,  $\mathbf{x}_{ij}$  对随机效用  $U_{ij}$  的作用不依赖于方案  $j$ , 比如乘车时间依个体与方案而变, 但乘车时间太长所带来的负效用是一致的。根据与多项 Logit 类似的推导可知, 个体  $i$  选择方案  $j$  的概率为

$$P(y_i=j | \mathbf{x}_{ij}) = \frac{\exp(\mathbf{x}'_{ij} \boldsymbol{\beta})}{\sum_{k=1}^J \exp(\mathbf{x}'_{ik} \boldsymbol{\beta})} \quad (12.8)$$

此模型称为“条件 Logit”(Conditional Logit), 也称为“McFadden 选择模型”(McFadden's Choice Model)(McFadden, 1974)<sup>①</sup>。条件 Logit 模型的估计方法与多项 Logit 模型类似, 即通过 MLE 估计以得到系数估计值  $\hat{\boldsymbol{\beta}}$ 。与多项 Logit 不同的是, 在条件 Logit 模型中, 由于系数  $\boldsymbol{\beta}$  不依赖于方案, 故不需要选择参照方案, 也不需要将  $\boldsymbol{\beta}$  的某部分标准化为 0。

## 12.3 混合 Logit 模型

上面分别考虑了解释变量不随方案而变的多项 Logit 模型, 以及解释变量随方案而变的条件 Logit 模型。很自然地, 可考虑这两种情况同时发生的混合情形。假设个体  $i$  选择方案  $j$  所能带来的随机效用为

$$U_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_i \boldsymbol{\gamma}_j + \varepsilon_{ij} \quad (i=1, \dots, n; j=1, \dots, J) \quad (12.9)$$

<sup>①</sup> Daniel McFadden 由于在离散选择领域的开创性贡献而与 James Heckman 分享 2000 年诺贝尔经济学奖。

其中,解释变量  $x_{ij}$  既随个体  $i$  而变,也随方案  $j$  而变;而解释变量  $z_i$  只随个体  $i$  而变。经过类似的推导可知,个体  $i$  选择方案  $j$  的概率为

$$P(y_i=j|x_{ij}) = \frac{\exp(x'_{ij}\beta + z'_i\beta_j)}{\sum_{k=1}^J \exp(x'_{ik}\beta + z'_i\beta_k)} \quad (12.10)$$

此模型在文献中称为“混合 Logit”(Mixed Logit)<sup>①</sup>,但 Stata 仍称其为条件 Logit。为了识别该模型,在方程(12.10)中,也需要选择一个参照方案(比如方案 1),然后令  $\beta_1 = 0$ 。

对于以上三种模型,当各方案本身的特征不重要,或缺乏有关方案特征的数据时,常使用多项 Logit 模型。如果需要考虑不同方案的特征,则应使用条件 Logit 模型或混合 Logit 模型。另外,在这些多值选择模型中,由于被解释变量的分布必然为“多项分布”(multinomial distribution),故一般不必使用稳健标准误,使用普通标准误即可;这一点类似于二值选择模型。然而,如果数据为聚类样本,则仍应使用聚类稳健的标准误(使用 Stata 命令选择项 vce(cluster clustvar))。

需要注意的是,在多项 Logit 与混合 Logit 模型中,对参数估计值  $\hat{\beta}_j$  的解释是以参照方案(base category)为转移的(可根据理论或方便来选择参照方案)。以多项 Logit 为例,假设“方案 1”或“方案  $j$ ”( $j \neq 1$ )必然发生(二者必居其一),则在此条件下,“方案  $j$ ”发生的条件概率为

$$P(y=j|y=1 \text{ or } j) = \frac{P(y=j)}{P(y=1) + P(y=j)} = \frac{\exp(x'_i\beta_j)}{1 + \exp(x'_i\beta_j)} \quad (12.11)$$

上式与二值选择的 Logit 模型具有完全相同的形式。而“几率比”(odds ratio)或“相对风险”(relative risk)为

$$\frac{P(y=j)}{P(y=1)} = \exp(x'_i\beta_j) \quad (12.12)$$

故对数几率比(log-odds ratio)为

$$\ln\left[\frac{P(y=j)}{P(y=1)}\right] = x'_i\beta_j \quad (12.13)$$

从条件概率  $P(y=j|y=1 \text{ or } j)$  的表达式可以看出,该条件概率并不依赖于任何其他方案。换言之,如果将多值选择模型中的任何两个方案单独挑出来,都是二值 Logit 模型。此假定称为“无关方案的独立性”(Independence of Irrelevant Alternatives,简记 IIA)。根据类似的推导可知,条件 Logit 模型也服从 IIA 假定。然而,在实践中,如果不同方案之间很类似,则 IIA 假设不一定满足。这是多项 Logit、条件 Logit 与混合 Logit 模型的共同缺点。

比如,假设共有三个备选的交通方式,即自驾车、乘红色公共汽车(red bus)、乘蓝色公共汽车(blue bus)。根据“无关选择的独立性”假定,在给定选择“自驾车”或“乘红色公共汽车”条件下,选择“自驾车”的条件概率与是否存在“乘蓝色公共汽车”的选择无关。然而,由于“蓝色公共汽车”与“红色公共汽车”仅仅颜色不同,故在加上“乘蓝色公共汽车”的选择之后,对选择“自驾车”的概率没有什么影响,但将使选择“乘红色公共汽车”的概率降低一半。因此,引入“蓝色公共汽车”后,上述条件概率将增大,与“无关选择的独立性”假定相矛盾。这个问题也被称为“红车蓝车问题”(red bus-blue bus problem)。

对于 IIA 假定,检验方法之一为豪斯曼检验。其基本思想是,如果 IIA 假定成立,则去掉某个方

<sup>①</sup> “Mixed logit”也被用来指另一完全不同的模型,即“随机系数 Logit”(random parameters logit),参见 Cameron and Trivedi (2005),在此不作介绍。

案不影响对其他方案参数的一致估计,只是降低了效率。这意味着,在 IIA 原假设成立的情况下,去掉某个方案后子样本的系数估计值(记为  $\hat{\beta}_R$ )与全样本的系数估计值(记为  $\hat{\beta}_F$ ,不含被去掉方案的对应系数)没有系统差别。为此,Hausman and McFadden (1984)提出以下统计量:

$$(\hat{\beta}_R - \hat{\beta}_F)' [\widehat{\text{Var}}(\hat{\beta}_R) - \widehat{\text{Var}}(\hat{\beta}_F)]^{-1} (\hat{\beta}_R - \hat{\beta}_F) \xrightarrow{d} \chi^2(m) \quad (12.14)$$

其中, $m$  等于  $\hat{\beta}_R$  的维度。检验 IIA 假定的方法之二由 Small and Hsiao (1985) 提出。然而,Cheng and Long (2007) 通过蒙特卡罗方法发现,这两个检验的小样本性质都不好(即有限样本下统计量的真实分布偏离方程(12.14)的渐近分布较远),故其结论只具有参考价值(两者的结论可能矛盾)。

在 Stata 中,多项选择模型所使用的数据格式依赖于解释变量的类型。对于所有解释变量都只随个体而变(case-specific)的多项 Logit 或多项 Probit 模型,应使用“宽形格式”(wide form),即同一个个体的所有数据均出现在 Stata 数据表格的同一行。反之,在条件 Logit 模型与混合 Logit 模型中,由于存在随方案而变(alternative-specific)的解释变量,故应使用“长形格式”(long form),即同一个体的数据出现在 Stata 数据表格的几行,对应于备选的几个方案(一个方案占一行)。在 Stata 中,可通过命令 reshape 在宽形与长形格式之间转换,参见第 23 章。

多项 Logit 模型的 Stata 命令为

```
mlogit y x1 x2 x3 ,rrr base (#)
```

其中,选择项“base (#)”用于指定参照组(base category);如果不指定,则默认使用观测值最多的方案为参照方案。选择项“rrr”表示汇报“相对风险比率”(Relative Risk Ratio,简记 RRR),即汇报  $\exp(\beta_j)$ ,而非  $\beta_j$ 。

在估计多项 Logit 模型后,可通过非官方命令 mlogtest 对 IIA 假定进行检验:

```
. net install sg155 (下载安装命令 mlogtest)①  
. mlogtest,hausman base  
. mlogtest,smhsiao base
```

其中,选择项“hausman”表示进行豪斯曼检验,选择项“smhsiao”表示进行 Small-Hsiao 检验,而选择项“base”表示在检验中包括“去掉参照方案,而以剩余方案中观测值最多的方案为参照方案”的检验。

多项 Probit 模型的 Stata 命令为

```
mprobit y x1 x2 x3 ,rrr base (#)
```

条件 Logit 模型的 Stata 命令为

```
clogit y x1 x2 x3 ,group(varname) or
```

其中,必选项“group(varname)”用来指定由归属同一个个体的观测值所构成的一组(因为数据按长形排列);选择项“or”表示汇报“几率比”(odds ratio)。命令 clogit 不能直接处理只随个体而变(case-specific)的解释变量。

混合 Logit 模型的 Stata 命令为

```
asclogit y x1 x2 x3 ,case (varname) alternatives (varname) casevars  
(varname) base (#) or
```

其中,命令“asclogit”表示“alternative-specific conditional logit”,必选项“case(varname)”用

① 或输入命令“findit sg155”寻找下载地址。

来指定个体(case),必选项“alternatives(varname)”用来指定方案(数据也按长形排列),选择项“casevars(varname)”表示只随个体而变的解释变量,选择项“base(#)”用来指定参照方案,选择项“or”表示汇报“几率比”(odds ratio)。

下面以数据集 nomocc2.dta 为例<sup>①</sup>,进行多项 Logit 与多项 Probit 估计。此数据集来自 1982 年美国 General Social Survey,该问卷调查将受访者职业分为五类(occ),即服务人员(menial),蓝领(blue collar),工匠(craft),白领:white collar 以及专业人士(professional)。解释变量包括:是否白人(white),受教育年限(ed)以及工龄(exper)。显然,这三个解释变量都只依赖于个体,而不依赖于方案,故使用多项 Logit 或多项 Probit 回归。

```
. use nomocc2.dta, clear
```

先看一下数据的基本特征。

```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
occ	337	3.397626	1.367913	1	5
white	337	.9169139	.2764227	0	1
ed	337	13.09496	2.946427	3	20
exper	337	20.50148	13.95936	2	66

其次,通过前 6 个数据来看数据格式。

```
. list in 1/6
```

	occ	white	ed	exper
1.	Menial	1	11	3
2.	Menial	1	12	14
3.	Menial	1	12	44
4.	Menial	1	12	18
5.	Menial	0	14	24
6.	Menial	1	13	38

由于数据按职业排序,故前 6 名都从事服务业。其中,5 名为白人,受教育年限从 11~14 年不等,而工龄从 3~44 年不等。由于数据表中每行对应于一名个体,故数据格式为宽形(wide form)。

下面通过列表粗略地考察受教育年限(ed)与职业(occ)的关系:

```
. table occ, contents(N ed mean ed sd ed)
```

其中,命令“table”表示将变量的统计特征列表(tables of summary statistics),而选择项“contents()”则用来指定最多 5 个统计量,“N ed mean ed sd ed”表示罗列变量 ed 的样本容量、均值与标准差(按变量 occ 分为 5 个子样本)。

Occupation	N(ed)	mean(ed)	sd(ed)
Menial	31	11.774194	2.186469
BlueCol	69	11.217391	2.571733
Craft	84	11.964286	2.119596
WhiteCol	41	13.170732	2.096455
Prof	112	15.4375	2.608998

<sup>①</sup> 此例来自 Long and Feese (2006)。

从上表可知,在样本中专业人士最多,且平均受教育年限最高;而蓝领的受教育程度最低。下面进行多项 Logit 回归。

. mlogit occ white ed exper,nolog

Multinomial logistic regression						Number of obs = 337
						LR chi2(12) = 166.09
						Prob > chi2 = 0.0000
						Pseudo R2 = 0.1629
occ		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
<i>Menial</i>						
white		-1.774306	.7550543	-2.35	0.019	-3.254186 -.2944273
ed		-.7788519	.1146293	-6.79	0.000	-1.003521 -.5541826
exper		-.0356509	.018037	-1.98	0.048	-.0710028 -.000299
_cons		11.51833	1.849356	6.23	0.000	7.893659 15.143
<i>BlueCol</i>						
white		-.5378027	.7996033	-0.67	0.501	-2.104996 1.029391
ed		-.8782767	.1005446	-8.74	0.000	-1.07534 -.6812128
exper		-.0309296	.0144086	-2.15	0.032	-.05917 -.0026893
_cons		12.25956	1.668144	7.35	0.000	8.990061 15.52907
<i>Craft</i>						
white		-1.301963	.647416	-2.01	0.044	-2.570875 -.0330509
ed		-.6850365	.0892996	-7.67	0.000	-.8600605 -.5100126
exper		-.0079671	.0127055	-0.63	0.531	-.0328693 .0169351
_cons		10.42698	1.517943	6.87	0.000	7.451864 13.40209
<i>WhiteCol</i>						
white		-.2029212	.8693072	-0.23	0.815	-1.906732 1.50089
ed		-.4256943	.0922192	-4.62	0.000	-.6064407 -.2449479
exper		-.001055	.0143582	-0.07	0.941	-.0291967 .0270866
_cons		5.279722	1.684006	3.14	0.002	1.979132 8.580313
Prof	(base outcome)					

由于没有指定参照方案(base outcome),故命令 mlogit 自动选择观测值最多的方案(即专业人士)为参照方案。上表显示,在 5% 的显著性水平上,给定其他变量,白人(white)更不可能选择服务业或工匠;但是否白人对于选择蓝领或白领没有显著影响。受教育程度(ed)越高,越不可能选择除专业人士以外的职业。工龄越长(exper),越不可能选择服务业或蓝领;工龄对于选择工匠或白领无显著影响。

由于 IIA 假定是多项 Logit 模型的前提,下面检验 IIA 假定是否满足。

. mlogtest hausman base

**** Hausman tests of IIA assumption				
Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.				
You used the old syntax of hausman. Click <a href="#">here</a> to learn about the new syntax.				
(storing estimation results as _HAUSMAN)				
Omitted	chi2	df	P>chi2	evidence
Menial	7.324	12	0.835	for Ho
BlueCol	0.320	12	1.000	for Ho
Craft	-14.436	12	1.000	for Ho
WhiteCol	-5.541	11	1.000	for Ho
Prof	15.566	12	0.212	for Ho

上表的前四行豪斯曼检验结果显示,去掉四个非参照方案(nonbase alternatives)中的任何一个方案,都不会拒绝 IIA 的原假设。由于使用了选择项“base”,故上表第五行计算去掉参照方案(Prof),而以剩余方案中观测值最多的方案作为参照方案的检验结果,同样也不拒绝 IIA 假设。但由于某种原因,却无法进行 Small - Hsiao 检验。

```
. mlogtest, smhsiao base
```

```
**** Small-Hsiao tests of IIA assumption
Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.
equation 5 not found
r(303);
```

上文提及,这两个检验的小样本性质均不好,只具有参考价值;但至少没有发现违背 IIA 假定的迹象(可能这五类职业太不相同了)。

如要显示相对风险比率,可输入命令:

```
. mlogit occ white ed exper,nolog rrr
```

Multinomial logistic regression						Number of obs = 337
						LR chi2(12) = 166.09
						Prob > chi2 = 0.0000
Log likelihood = -426.80048						Pseudo R2 = 0.1629
		RRR	Std. Err.	z	P> z	[95% Conf. Interval]
Menial						
white		.169601	.128058	-2.35	0.019	.0386123 .7449581
ed		.4589326	.0526071	-6.79	0.000	.3665863 .5745417
exper		.9649771	.0174053	-1.98	0.048	.9314593 .999701
_cons		100542	185937.9	6.23	0.000	2680.234 3771571
BlueCol						
white		.5840301	.4669924	-0.67	0.501	.1218461 2.79936
ed		.4154983	.0417761	-8.74	0.000	.3411816 .5060029
exper		.9695438	.0139698	-2.15	0.032	.9425465 .9973143
_cons		210989.6	351961.2	7.35	0.000	8022.948 5548662
Craft						
white		.2719974	.1760954	-2.01	0.044	.0764686 .9674893
ed		.5040718	.0450134	-7.67	0.000	.4231365 .600488
exper		.9920646	.0126046	-0.63	0.531	.967665 1.017079
_cons		33758.18	51243	6.87	0.000	1723.071 661385.6
WhiteCol						
white		.8163426	.7096525	-0.23	0.815	.1485651 4.485678
ed		.653316	.0602483	-4.62	0.000	.5452883 .7827453
exper		.9989455	.0143431	-0.07	0.941	.9712254 1.027457
_cons		196.3154	330.5962	3.14	0.002	7.236457 5325.773
Prof	(base outcome)					

下面,根据模型预测个体选择各种职业的可能性,分别记为 `occ1,occ2,occ3,occ4,occ5`,并显示对前 5 个观测值的预测结果。

```
. predict occ1 occ2 occ3 occ4 occ5
(option pr assumed; predicted probabilities)
.list occ1 -occ5 in 1/5
```

	occ1	occ2	occ3	occ4	occ5
1.	.1681295	.4128002	.2760952	.085288	.0576871
2.	.1257816	.2945018	.3076293	.1328948	.1391926
3.	.0644456	.1738508	.3616529	.1922331	.2078175
4.	.1161744	.2771936	.3174044	.1409616	.148266
5.	.1691383	.0988214	.410424	.1063373	.215279

也可以选择其他职业作为参照方案,比如服务业。

```
. mlogit occ white ed exper,base(1) nolog
```

Multinomial logistic regression		Number of obs = 337				
		LR chi2(12) = 166.09				
		Prob > chi2 = 0.0000				
Log likelihood = -426.80048		Pseudo R2 = 0.1629				
occ	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Menial	(base outcome)					
BlueCol						
white	1.236504	.7244352	1.71	0.088	-.1833631	2.656371
ed	-.0994247	.1022812	-0.97	0.331	-.2998922	.1010428
exper	.0047212	.0173984	0.27	0.786	-.0293789	.0388214
_cons	.7412336	1.51954	0.49	0.626	-2.23701	3.719477
Craft						
white	.4723436	.6043097	0.78	0.434	-.7120817	1.656769
ed	.0938154	.097555	0.96	0.336	-.0973888	.2850197
exper	.0276838	.0166737	1.66	0.097	-.004996	.0603636
_cons	-1.091353	1.450218	-0.75	0.452	-3.933728	1.751022
WhiteCol						
white	1.571385	.9027216	1.74	0.082	-.1979166	3.340687
ed	.3531577	.1172786	3.01	0.003	.1232959	.5830194
exper	.0345959	.0188294	1.84	0.066	-.002309	.0715007
_cons	-6.238608	1.899094	-3.29	0.001	-9.960764	-2.516453
Prof						
white	1.774306	.7550543	2.35	0.019	.2944273	3.254186
ed	.7788519	.1146293	6.79	0.000	.5541826	1.003521
exper	.0356509	.018037	1.98	0.048	.000299	.0710028
_cons	-11.51833	1.849356	-6.23	0.000	-15.143	-7.893659

从上表可知,系数估计值随参照方案的不同而变化。下面,使用 multinomial probit 来估计此模型。

```
. mprobit occ white ed exper,nolog
```

Multinomial probit regression		Number of obs = 337				
		Wald chi2(12) = 105.61				
		Prob > chi2 = 0.0000				
occ	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Menial						
white	-1.144907	.5027501	-2.28	0.023	-2.130279	-.1595352
ed	-.5094985	.0698816	-7.29	0.000	-.6464639	-.3725331
exper	-.0234636	.0109546	-2.14	0.032	-.0449343	-.0019929
_cons	7.46242	1.145854	6.51	0.000	5.216587	9.708253
BlueCol						
white	-.392222	.5182974	-0.76	0.449	-1.408066	.6236222
ed	-.5845723	.063011	-9.28	0.000	-.7080715	-.461073
exper	-.0225903	.00984	-2.30	0.022	-.0418764	-.0033042
_cons	8.188586	1.069264	7.66	0.000	6.092867	10.28431
Craft						
white	-.8903573	.457069	-1.95	0.051	-1.786196	.0054814
ed	-.4718874	.0579237	-8.15	0.000	-.5854157	-.3583591
exper	-.0077824	.0090923	-0.86	0.392	-.025603	.0100382
_cons	7.140264	.9896954	7.21	0.000	5.200496	9.080031
WhiteCol						
white	-.1434167	.5530156	-0.26	0.795	-1.227307	.9404739
ed	-.3038566	.0576254	-5.27	0.000	-.4168003	-.1909129
exper	-.0039043	.0095574	-0.41	0.683	-.0226365	.0148279
_cons	3.76544	1.036649	3.63	0.000	1.733645	5.797234
Prof	(base outcome)					

从上表可知,多项 Probit 的系数估计值与多项 Logit 并不相同,但二者的系数不具可比性;具有可比性的是两个模型的预测概率。为此,计算多项 Probit 模型所预测的各种职业的选择概率,分别记为 occ1p,occ2p,occ3p,occ4p,occ5p。

```
. predict occ1p occ2p occ3p occ4p occ5p  
(option pr assumed; predicted probabilities)
```

下面计算两个模型职业预测概率的相关性。

```
. corr occ1 occ1p
```

	occ1	occ1p
occ1	1.0000	
occ1p	0.9979	1.0000

```
. corr occ2 occ2p
```

	occ2	occ2p
occ2	1.0000	
occ2p	0.9985	1.0000

```
. corr occ3 occ3p
```

	occ3	occ3p
occ3	1.0000	
occ3p	0.9935	1.0000

```
. corr occ4 occ4p
```

	occ4	occ4p
occ4	1.0000	
occ4p	0.9929	1.0000

```
. corr occ5 occ5p
```

	occ5	occ5p
occ5	1.0000	
occ5p	0.9989	1.0000

由以上结果可知,两个模型所预测的职业选择概率高度一致,相关系数均在 99% 以上。这意味着,使用多项 Logit 或多项 Probit 在实际上并无多少区别;只是多项 Probit 的计算时间更长,且无法从几率比角度解释系数估计值,故实践中常使用多项 Logit。

下面以数据集 travel2.dta 为例<sup>①</sup>,进行条件 Logit 与混合 Logit 的估计。该数据集的观测单位为 152 组人群(每组可视为一个小型旅行团),每个人群从以下三种度假旅行方式中选择一种,即火车,长途大巴,或自驾车。随方案而变的解释变量包括 time(总旅行时间)与 invc(乘车成本,in-vehicle cost)。不随方案而变的解释变量包括 hinc(家庭收入,household income)与 psize(旅行团人数,party size)。

首先,通过前 6 个观测值来考察数据格式。

```
. use travel2.dta,clear
.list id mode train bus time invc choice hinc psize in 1/6,sepby(id)
```

	id	mode	train	bus	time	invc	choice	hinc	psize
1.	1	Train	1	0	406	31	0	35	1
2.	1	Bus	0	1	452	25	0	35	1
3.	1	Car	0	0	180	10	1	35	1
4.	2	Train	1	0	398	31	0	30	2
5.	2	Bus	0	1	452	25	0	30	2
6.	2	Car	0	0	255	11	1	30	2

其中,选择项“sepby(id)”表示根据变量 id 的取值来画上表中的横线(默认每隔 5 个观测值画一条横线)。上表显示,每个旅行团(由变量 id 指定)对应于 3 行数据,每行对应于一种旅行方式(以变量 mode 表示)。虚拟变量 Train = 1 表示,该行数据对应于乘火车的方案;虚拟变量 Bus = 1 表示,该行数据对应于乘长途大巴的方案;如果 Train = 0 且 Bus = 0,则该行数据对应于自驾车的方案。即使对于同一旅行团,每个方案的总旅行时间、乘车成本也是不同的。另一方面,对于同一旅行团,其家庭收入与旅行团规模则不变。被解释变量 choice 为虚拟变量,表示选择哪一种方案(比如,第 1 个旅行团选择自驾车,故 car 所对应的那一行 choice = 1,而其他两行 choice = 0)。

尽管数据集中只有 152 个旅行团,但由于每个旅行团占据 3 行数据,故实际的样本容量为 456(即  $152 \times 3$ )。下面看一下数据的统计特征。

```
. sum id mode train bus time invc choice hinc psize
```

<sup>①</sup> 此数据集取自 Long and Freese (2006),最早由 Greene and Hensher (1995) 所使用。

Variable	Obs	Mean	Std. Dev.	Min	Max
id	456	103.8882	61.03044	1	210
mode	456	2	.8173933	1	3
train	456	.3333333	.4719223	0	1
bus	456	.3333333	.4719223	0	1
time	456	632.1096	270.2547	180	1440
invc	456	33.95175	21.795	2	109
choice	456	.3333333	.4719223	0	1
hinc	456	31.80921	19.25813	2	72
psize	456	1.809211	1.069457	1	6

上表显示,变量 id 的最大值为 210,但 id 的编码并不连续,实际的旅行团数为 152。

下面,进行条件 Logit 估计。由于命令 clogit 只接受随方案而变的解释变量,故仅使用总旅行时间(单位为分钟)、乘车成本、虚拟变量 train 以及虚拟变量 bus 为解释变量(以旅行方式 car 为参照方案)。

```
. clogit choice train bus time invc,group(id) nolog
```

Conditional (fixed-effects) logistic regression	Number of obs	=	456
	LR chi2(4)	=	172.06
	Prob > chi2	=	0.0000
Log likelihood = -80.961135	Pseudo R2	=	0.5152
choice	Coef.	Std. Err.	z P> z  [95% Conf. Interval]
train	2.671238	.4531611	5.89 0.000 1.783058 3.559417
bus	1.472335	.4007152	3.67 0.000 .6869474 2.257722
time	-.0191453	.0024509	-7.81 0.000 -.0239489 -.0143417
invc	-.0481658	.0119516	-4.03 0.000 -.0715905 -.0247411

上表显示,如果其他解释变量(time,invc)的取值相同,则旅行团最有可能选择火车,其次为长途大巴。另外,一个方案的总旅行时间越长,乘车成本越高,则选择该方案的概率越低。然而,由于这是非线性模型,故不易通过系数估计值来评价边际效应。为此,在上述命令中加上选择项“or”来计算风险比率。

```
. clogit choice train bus time invc,group(id) nolog or
```

Conditional (fixed-effects) logistic regression	Number of obs	=	456
	LR chi2(4)	=	172.06
	Prob > chi2	=	0.0000
Log likelihood = -80.961135	Pseudo R2	=	0.5152
choice	Odds Ratio	Std. Err.	z P> z  [95% Conf. Interval]
train	14.45786	6.551738	5.89 0.000 5.94802 35.14272
bus	4.359401	1.746879	3.67 0.000 1.987639 9.561286
time	.9810368	.0024044	-7.81 0.000 .9763356 .9857607
invc	.9529758	.0113896	-4.03 0.000 .930912 .9755624

从上表可知,变量 time 的风险比率为 0.98,这意味着在给定其他变量的情况下,一个方案的总旅行时间每增加 1 分钟,则选择此方案的概率将乘以 0.98,即下降 2%。变量 invc 的风险比率可类似地解释。另一方面,虚拟变量 bus 的风险比率为 4.36,这意味着,如果各方案的时间与成本均相等,则旅行团选择长途大巴的概率是选择自驾车概率的 4.36 倍;虚拟变量 train 的风险比

率也可类似地解释。

下面计算条件 Logit 模型的预测概率。

```
. predict prob
(option pcl assumed; probability of success given one success within
group)
```

看一下第 1 个旅行团的预测概率。

```
. list id mode prob choice time invc in 1 / 3
```

	id	mode	prob	choice	time	invc
1.	1	Train	.0642477	0	406	31
2.	1	Bus	.0107205	0	452	25
3.	1	Car	.9250318	1	180	10

上表显示,第 1 个旅行团实际选择自驾车(Car),而模型预测选择自驾车的概率高达 0.925,而且正好是旅行时间与成本最低的方案。

对于条件 Logit 模型,也可用命令 asclogit 来估计。

```
. asclogit choice time invc, case(id) alternatives(mode) base(3) nolog
```

Alternative-specific conditional logit		Number of obs = 456	
Case variable: id		Number of cases = 152	
Alternative variable: mode		Alts per case: min = 3	
		avg = 3.0	
		max = 3	
		Wald chi2(2) = 70.53	
Log likelihood = -80.961135		Prob > chi2 = 0.0000	
choice	Coef.	Std. Err.	z P> z  [95% Conf. Interval]
mode			
time	-.0191453	.0024509	-7.81 0.000 -.0239489 -.0143417
invc	-.0481658	.0119516	-4.03 0.000 -.0715905 -.0247411
Train			
_cons	2.671238	.4531611	5.89 0.000 1.783058 3.559417
Bus			
_cons	1.472335	.4007152	3.67 0.000 .6869474 2.257722
Car	(base alternative)		

对比上表与命令 clogit 的估计结果可知,二者的系数估计值与标准误均完全相同。使用命令 asclogit 的主要好处在于,它可以通过选择项“casevars(varname)”将只随个体而变的解释变量也包括在模型中。下面将解释变量“家庭收入”(hinc)与“旅行团人数”(psize)也包括进来,进行混合 Logit 模型估计。

```
. asclogit choice time invc, case(id) alternatives(mode) base(3)
casevars(hinc psize) nolog
```

Alternative-specific conditional logit Case variable: id	Number of obs = 456 Number of cases = 152
Alternative variable: mode	Alts per case: min = 3 avg = 3.0 max = 3
	Wald chi2(6) = 69.09 Prob > chi2 = 0.0000
Log likelihood = -77.504846	
choice	Coef. Std. Err. z P> z  [95% Conf. Interval]
mode	
time	-.0185035 .0025035 -7.39 0.000 -.0234103 -.0135966
invc	-.0402791 .0134851 -2.99 0.003 -.0667095 -.0138488
Train	
hinc	-.0342841 .0158471 -2.16 0.031 -.0653438 -.0032243
psize	-.0038421 .3098075 -0.01 0.990 -.6110537 .6033695
_cons	3.499641 .7579665 4.62 0.000 2.014054 4.985228
Bus	
hinc	-.0080174 .0200322 -0.40 0.689 -.0472798 .031245
psize	-.5141037 .4007015 -1.28 0.199 -.1299464 .2712569
_cons	2.486465 .8803649 2.82 0.005 .7609815 4.211949
Car	(base alternative)

上表显示,家庭收入(hinc)越高,越不倾向于选择火车;但对于选择长途大巴则无显著影响。旅行团规模(psize)没有显著影响。旅行时间(time)与乘车成本(invc)的估计系数依然显著为负,且与前面条件 Logit 模型的估计系数很接近。上表并不汇报准  $R^2$ ,但可以很容易地手工计算。从上表可知,该模型的对数似然函数为 -77.504846,下面估计一个只含常数项的模型。

```
. asclogit choice, case(id) alternatives(mode) base(3) nolog
```

Alternative-specific conditional logit Case variable: id	Number of obs = 456 Number of cases = 152
Alternative variable: mode	Alts per case: min = 3 avg = 3.0 max = 3
	Wald chi2(0) = Prob > chi2 =
Log likelihood = -160.00172	
choice	Coef. Std. Err. z P> z  [95% Conf. Interval]
Train	
_cons	.0655973 .1811689 0.36 0.717 -.2894872 .4206818
Bus	
_cons	-.67634 .2242376 -3.02 0.003 -1.115838 -.2368425
Car	(base alternative)

上表显示,只含常数项模型的对数似然函数为 -160.00172,故可计算准  $R^2$  如下。

```
. dis (160.00172 - 77.504846) / 160.00172  
.51559992
```

## 12.4 嵌套 Logit

上文提及,多项 Logit、条件 Logit 与混合 Logit 都须满足“无关方案的独立性”假定(IIA)。但在实践中,如果方案之间比较类似,则 IIA 假设可能不满足;限制了这些模型的应用。解决方法之一是,把相似的方案归入一组,允许同组内的方案相关,但不同组的方案相互独立。比如,在交通工具的选择上,可把“公交车、地铁”归入“公共交通组”,而把“自驾车、出租车”归入“私人交通组”,形成嵌套式(nested)的树形结构,参见图 12.1。

图 12.1 中的“交通工具”可视为此树形结构的“根部”(root),第一层(level 1)的“公共交通、私人交通”为“树干”(limb),而第二层(level 2)的“公交、地铁、自驾车、出租车”为“树枝”(branch)。类似的例子包括消费者对居住地(residential location)的选择,即首先选择住在哪个社区,然后再选择住在该社区的哪套房子。

也可将此树形结构视为“决策树”(decision tree),但关键在于它的统计性质,即在第二层的每组内部允许扰动项相关,比如 $(\varepsilon_{i, \text{bus}}, \varepsilon_{i, \text{subway}})$ 相关, $(\varepsilon_{i, \text{car}}, \varepsilon_{i, \text{taxi}})$ 相关;但这两组之间不相关,即 $(\varepsilon_{i, \text{bus}}, \varepsilon_{i, \text{subway}})$ 与 $(\varepsilon_{i, \text{car}}, \varepsilon_{i, \text{taxi}})$ 不相关。如果所有方案的扰动项都互不相关,则回到多项 Logit 或条件 Logit 的情形。

更一般地,假设在第一层共有  $J$  个树干备选,而树干  $j$  包含  $K_j$  个树枝(不同树干包含的树枝数可以不同),分别以  $j_1, \dots, j_{K_j}$  来指代。个体  $i$  选择树干  $j$  树枝  $k$  方案所能带来的效用为(为了简洁,略去个体  $i$  的下标)

$$U_{jk} = \mathbf{x}'_{jk} \boldsymbol{\beta} + \mathbf{z}'_j \boldsymbol{\gamma}_j + \varepsilon_{jk} \quad (j=1, \dots, J; k=1, \dots, K) \quad (12.15)$$

其中,解释变量  $\mathbf{z}_j$  只随树干而不随树枝方案而变(故其系数  $\boldsymbol{\gamma}_j$  可随  $j$  而变),而  $\mathbf{x}_{jk}$  同时随树干与树枝方案而变(故其系数  $\boldsymbol{\beta}$  没有下标,不随  $j, k$  而变)。McFadden (1978) 假定扰动项  $\varepsilon_{jk}$  服从“广义极值分布”(Generalized Extreme Value distribution,简记 GEV),其累积分布函数为

$$F(\varepsilon) = \exp[-G(e^{-\varepsilon_{11}}, \dots, e^{-\varepsilon_{1K_1}}; \dots; e^{-\varepsilon_{J1}}, \dots, e^{-\varepsilon_{JK_J}})] \quad (12.16)$$

其中,函数  $G(\cdot)$  的形式为

$$G(w) = G(w_{11}, \dots, w_{1K_1}; \dots; w_{J1}, \dots, w_{JK_J}) = \sum_{j=1}^J \left( \sum_{k=1}^{K_j} w_{jk}^{1/\tau_j} \right)^{\tau_j} \quad (12.17)$$

上式右边的每一项 $(\sum_{k=1}^{K_j} w_{jk}^{1/\tau_j})^{\tau_j}$ ,在形式上为微观经济学中的不变替代弹性函数(Constant Elasticity of Substitution,简记 CES),其参数 $\tau_j = \sqrt{1 - \text{Corr}(\varepsilon_{jk}, \varepsilon_{jl})}$ ,与 $(\varepsilon_{jk}, \varepsilon_{jl})$ 的相关系数呈反向变动的关系,故称为“不相似参数”(dissimilarity parameter),在 Stata 中记为 tau。显然, $0 \leq \tau_j \leq 1$ 。如果 $\tau_j = 1$  ( $j = 1, \dots, J$ ),则所有方案的扰动项均不相关,满足 IIA 假定,回到多项 Logit、条件 Logit 或混合 Logit 的情形。基于以上假设,可以证明,选择树干  $j$  树枝  $k$  方案的概率为

$$p_{jk} = p_j \times p_{kj} = \frac{\exp(\mathbf{z}'_j \boldsymbol{\gamma}_j + \tau_j I_j)}{\sum_{m=1}^J \exp(\mathbf{z}'_m \boldsymbol{\gamma}_m + \tau_m I_m)} \times \frac{\exp(\mathbf{x}'_{jk} \boldsymbol{\beta}_j / \tau_j)}{\sum_{l=1}^{K_j} \exp(\mathbf{x}'_{jl} \boldsymbol{\beta}_j / \tau_j)} \quad (12.18)$$

其中, $p_j$  为选择树干  $j$  的概率, $p_{kj}$  为在选择树干  $j$  的情况下,选择树枝  $k$  的条件概率;而  $I_j$  的定义为

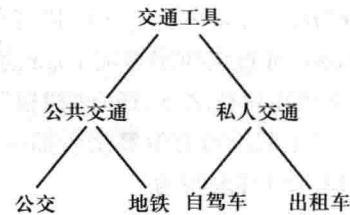


图 12.1 嵌套 Logit 的树形结构

$$I_j \equiv \ln \left[ \sum_{l=1}^{K_j} \exp(\mathbf{x}'_j \boldsymbol{\beta}_j / \tau_j) \right] \quad (12.19)$$

其中,  $I_j$  称为“包含价值”(inclusive value)或“对数和”(log-sum)。根据公式(12.18), 可写出整个样本的对数似然函数, 称为“全信息最大似然估计”(Full Information Maximum Likelihood, 简记 FIML)。另一方法为分步对  $p_j$  与  $p_{kj}$  进行 MLE 估计, 称为“有限信息最大似然估计”(Limited Information Maximum Likelihood, 简记 LIML), 但不如 FIML 有效率。进行 FIML 估计后, 可对联合假设“ $H_0: \tau_1 = \dots = \tau_J = 1$ ”进行似然比检验。如果接受原假设, 则 IIA 假定成立, 不需要使用嵌套 Logit, 可直接使用多项 Logit、条件 Logit 或混合 Logit 模型。另外, 也可以考虑更多层的嵌套结构, 比如在树枝之下, 还有“树权”(twig), 估计方法类似。

以上假设所有解释变量都随方案而变。如果解释变量包括不随方案而变的个体特征, 则方程(12.15)可改写为

$$U_{jk} = \mathbf{x}'_{jk} \boldsymbol{\beta} + \mathbf{z}'_j \boldsymbol{\gamma}_j + \mathbf{w}' \boldsymbol{\delta}_{jk} + \varepsilon_{jk} \quad (j = 1, \dots, J; k = 1, \dots, K) \quad (12.20)$$

在 Stata 中, 如果系数只随  $j$  而变(比如  $\boldsymbol{\gamma}_j$ ), 则将其归属于“level 1 equation”, 简记 lev1; 如果系数既随  $j$  也随  $k$  而变(比如  $\boldsymbol{\delta}_{jk}$ ), 则将其归属于“level 2 equation”, 简记 lev2; 既不随  $j$  也不随  $k$  而变的系数(比如  $\boldsymbol{\beta}$ ), 则直接归属于模型本身(需要注意的是, 以上只是 Stata 的称呼, 整个模型是一起估计的, 实际上只有一个方程)。

在 Stata 中, 估计嵌套 Logit 之前, 首先要定义树形的嵌套结构, 其命令为

```
. nlogitgen newaltvar = altvar (branchlist)
```

其中, “altvar”为原先定义方案的变量, “newaltvar”为新定义的方案分组变量, 而“branchlist”定义树枝的分组归类。

进行嵌套 Logit 估计的命令为

```
. nlogit y x1 x2 x3 || lev1_equation || lev2_equation, case(varname) base (#) notree
```

其中, “x1 x2 x3”为随方案而变(alternative-specific)的随机变量(其系数相当于  $\boldsymbol{\beta}$ , 不随  $j, k$  而变), “||”为分隔符(delimiter), “lev1\_equation”用来指定系数只随树干(level 1)而不随树枝(level 2)而变的解释变量(其系数相当于  $\boldsymbol{\gamma}_j$ ), “lev2\_equation”用来指定系数既随树干(level 1)又随树枝(level 2)而变的解释变量(其系数相当于  $\boldsymbol{\delta}_{jk}$ ), 必选项“case(varname)”用来指定个体, 选择项“base(#)”用来指定参照方案, 选择项“notree”表示不显示嵌套的树形结构(默认显示树形结构)。

继续以数据集 travel2.dta 为例。假设 Train 与 Bus 可归入一组(称为 public), 而 Car 自成一组(称为 private), 将定义嵌套结构的变量记为 type, 可输入以下命令来定义树形结构:

```
. nlogitgen type = mode(public: Train | Bus, private: Car)
```

```
new variable type is generated with 2 groups
label list lb_type
lb_type:
    1 public
    2 private
```

如果要显示此树形结构, 可使用以下命令:

```
. nlogittree mode type
```

```
tree structure specified for the nested logit model
type      N      mode      N
public    304   Train 152
           Bus   152
private   152   Car   152
total     456
N = number of observations at each level
```

更直观地,通过前 6 个观测值来看新生成的变量 type:

```
. list id mode train bus time invc choice hinc type in 1/6, sepby(id)
```

	id	mode	train	bus	time	invc	choice	hinc	type
1.	1	Train	1	0	406	31	0	35	public
2.	1	Bus	0	1	452	25	0	35	public
3.	1	Car	0	0	180	10	1	35	private
4.	2	Train	1	0	398	31	0	30	public
5.	2	Bus	0	1	452	25	0	30	public
6.	2	Car	0	0	255	11	1	30	private

从上表可知,对于方案 Train 与 Bus,变量 Type 的取值为 public;而对于方案 Car,变量 Type 的取值为 private。在数据集中,有两个不随方案而变的解释变量,即 hinc 与 psize。为了演示的目的,假设 psize 的系数只随树干方案而变(仅取决于是否公共交通还是私人交通)<sup>①</sup>,而 hinc 的系数可随树枝方案而变(对于 Train, Bus, Car 各不相同)。

```
. nlogit choice time invc || type: psize, base(private) || mode: hinc,
base(3) case(id) nolog notree
```

RUM-consistent nested logit regression		Number of obs = 456					
Case variable: id		Number of cases = 152					
Alternative variable: mode		Alts per case: min = 3					
		avg = 3.0					
		max = 3					
		Wald chi2(5) = 40.43					
		Prob > chi2 = 0.0000					
Log likelihood = -71.596852							
choice		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
mode							
time		-.0131104	.0025116	-5.22	0.000	-.0180331	-.0081877
invc		-.023953	.0092628	-2.59	0.010	-.0421077	-.0057983
type equations							
public							
psize		-.2313463	.2475569	-0.93	0.350	-.7165488	.2538563
private							
psize		0	(base)				
mode equations							
Train							
hinc		-.0357403	.0138418	-2.58	0.010	-.0628696	-.0086109
_cons		3.301306	.6493802	5.08	0.000	2.028544	4.574068
Bus							
hinc		-.0192732	.0150936	-1.28	0.202	-.0488562	.0103097
_cons		2.431383	.6584803	3.69	0.000	1.140785	3.721981
Car							
hinc		0	(base)				
_cons		0	(base)				
dissimilarity parameters							
type							
/public_tau		.3271257	.1057812			.1197983	.5344531
/private_tau		1	595784.1			-1167714	1167716
LR test for IIA (tau = 1):				chi2(2) = 13.95	Prob > chi2 = 0.0009		

<sup>①</sup> 也可指定 psize 的系数随树枝方案而变。

上表最后一行的似然比检验强烈拒绝 IIA 假定, 故应使用嵌套 Logit 模型。另外, 旅行团规模 (psize) 对于选择公共交通 (public) 或私人交通 (private) 没有显著影响。家庭收入 (hinc) 越高, 越不倾向于选择火车, 但对选择长途大巴无显著影响。变量 time 与 invc 的系数依然显著地为负, 但估计值与前面的条件 Logit 或混合 Logit 模型的估计值有一定差距(由于后者没有考虑 Train 与 Bus 两个公共交通方案间的相关性, 故可能高估了变量 time 与 invc 的作用)。最后, 上表下部还显示, public 组的不相似参数为 0.327; 而对于 private 组, 由于只有一个方案, 故不相似参数标准化为 1。

需要注意的是, 使用嵌套 Logit 的前提是, 各备选方案之间应有一个清晰的嵌套结构; 而有些多值选择模型不存在清晰的嵌套结构。如果 IIA 假定不成立, 但又没有清晰的嵌套结构, 可考虑随机系数 Logit 模型(参见 Cameron and Trivedi, 2005, 在此从略)。

## 习 题

**12.1** (此例来自 Cameron and Trivedi, 2005) 考虑潜变量模型  $y_i^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon_i$ , 其中  $\varepsilon_i \sim N(0, 1)$ 。假设可观测变量  $y_i = 2$ , 如果  $y_i^* < \alpha$ ;  $y_i = 1$ , 如果  $\alpha \leq y_i^* < U_i$ ;  $y_i = 0$ , 如果  $y_i^* \geq U_i$ 。其中,  $U_i$  已知但可随个体  $i$  而变, 而  $\alpha$  为未知。

(1) 给定  $\mathbf{x}_i$ , 计算  $y_i = 0$ ,  $y_i = 1$ , 及  $y_i = 2$  的条件概率。

(2) 具体说明, 如何一致地估计  $\alpha$  与  $\boldsymbol{\beta}$ 。

**12.2** 数据集 mlogit.dta<sup>①</sup> 包含 200 名高中生的如下变量: 定性变量 ice\_cream(取值为 chocolate, vanilla 与 strawberry), female(是否女性), video(电子游戏成绩), puzzle(猜谜成绩)。

(1) 估计一个多项 Logit 模型, 以变量 female, video 与 puzzle 解释这些学生对冰淇淋口味的偏好。

(2) 解释此模型估计结果的经济意义。

(3) 对 IIA 假定, 进行豪斯曼检验。

(4) 根据多项 Logit 模型, 预测每位学生选择冰淇淋口味的可能性。

(5) 汇报此模型的相对风险比率(RRR)。

(6) 使用同样的变量, 估计一个多项 Probit 模型。

(7) 根据多项 Probit 模型, 预测每位学生选择冰淇淋口味的可能性。

(8) 比较这两种模型预测结果的相关性。

**12.3** 数据集 travel.dta<sup>②</sup> 包含有关交通工具选择的以下变量: subject(个体 ID), mode(1 = Air, 2 = Train, 3 = Bus, 4 = Car), choice(1 = 选中的交通工具), time(等待时间), cost(成本), income(家庭收入), air(1 = 乘飞机), train(坐火车), bus(1 = 坐公交车), car(1 = 自驾车), air\_inc(air 与 income 的互动项)

(1) 使用解释变量 time, cost, air\_inc, 估计条件 Logit 模型。

(2) 将交通工具分为两组, 即空中交通(fly)与地面交通(ground), 进行嵌套 Logit 估计。

① 此数据集的下载地址为 <http://www.ats.ucla.edu/stat/stata/output/mlogit.dta>。

② 此数据集来自 <http://pages.stern.nyu.edu/~wgreen/Text/tables/tablelist5.htm>。

# 第13章 排序与计数模型

## 13.1 排序模型

有时,离散数据有着天然的排序。比如,公司债券的评级(AAA,AA,A,B,C级),对“春节联欢晚会”的满意度(很满意、满意、不满意、很不满意)。又比如,Li and Zhou(2005)研究经济增长绩效对地方官员仕途的影响,以0表示“卸任”,1表示“留任或平级调动”,2表示“提拔”。这种数据被称为“排序数据”(ordered data)。如果使用 multinomial logit,将无视数据内在的排序,而 OLS 又把排序视为基数来处理。对于排序数据,仍然可以使用潜变量法来推导出 MLE 估计量。

假设  $y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$  ( $y^*$  不可观测),而选择规则为

$$y = \begin{cases} 0, & \text{若 } y^* \leq r_0 \\ 1, & \text{若 } r_0 < y^* \leq r_1 \\ 2, & \text{若 } r_1 < y^* \leq r_2 \\ \dots\dots \\ J, & \text{若 } r_{J-1} \leq y^* \end{cases} \quad (13.1)$$

其中, $r_0 < r_1 < r_2 < \dots < r_{J-1}$  为待估参数,称为“切点”(cutoff points,在 Stata 输出结果中简称“cut”),参见图 13.1。

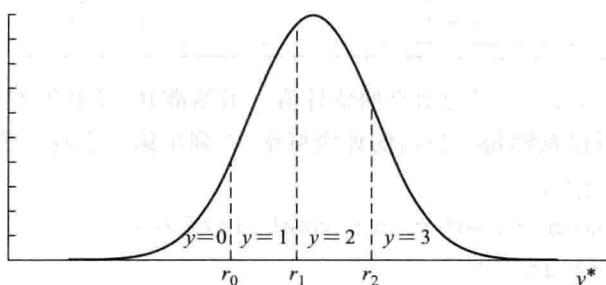


图 13.1 ordered logit 示意图

假设  $\varepsilon \sim N(0, 1)$ (将扰动项  $\varepsilon$  的方差标准化为 1),则

$$P(y=0 | \mathbf{x}) = P(y^* \leq r_0 | \mathbf{x}) = P(\mathbf{x}'\boldsymbol{\beta} + \varepsilon \leq r_0 | \mathbf{x}) \quad (13.2)$$

$$= P(\varepsilon \leq r_0 - \mathbf{x}'\boldsymbol{\beta} | \mathbf{x}) = \Phi(r_0 - \mathbf{x}'\boldsymbol{\beta})$$

$$P(y=1 | \mathbf{x}) = P(r_0 < y^* \leq r_1 | \mathbf{x})$$

$$= P(y^* \leq r_1 | \mathbf{x}) - P(y^* \leq r_0 | \mathbf{x})$$

$$= P(\mathbf{x}'\boldsymbol{\beta} + \varepsilon \leq r_1 | \mathbf{x}) - \Phi(r_0 - \mathbf{x}'\boldsymbol{\beta})$$

$$= P(\varepsilon \leq r_1 - \mathbf{x}'\boldsymbol{\beta} | \mathbf{x}) - \Phi(r_0 - \mathbf{x}'\boldsymbol{\beta}) \quad (13.3)$$

$$= \Phi(r_1 - \mathbf{x}'\boldsymbol{\beta}) - \Phi(r_0 - \mathbf{x}'\boldsymbol{\beta})$$

$$P(y = 2 | \mathbf{x}) = \Phi(r_2 - \mathbf{x}'\boldsymbol{\beta}) - \Phi(r_1 - \mathbf{x}'\boldsymbol{\beta})$$

(13.4)

$$P(y = J | \mathbf{x}) = 1 - \Phi(r_{J-1} - \mathbf{x}'\boldsymbol{\beta})$$

这样,可写出样本似然函数,并得到 MLE 估计量,即 ordered probit 模型。如果假设扰动项服从逻辑分布,则可得到 ordered logit 模型。

排序模型的 Stata 命令为

oprobit y x1 x2 x3 (ordered probit 模型)

ologit y x1 x2 x3 (ordered logit 模型)

下面以数据集“panel84extract.dta”为例<sup>①</sup>,估计决定公司债券评级的排序数据模型。被解释变量为 rating83c(1983 年公司债券评级,取值 2~5,其中 5 为最优评级),而解释变量为 ia83(1983 年 income-to-asset 比率),dia(ia 比率从 1982 年到 1983 年的变化,即 ia83 减 ia82)。

首先进行 ordered probit 估计:

. use panel84extract.dta,clear

. oprobit rating83c ia83 dia,nolog

Ordered probit regression						Number of obs	=	98
						LR chi2(2)	=	10.33
						Prob > chi2	=	0.0057
						Pseudo R2	=	0.0388
Log likelihood = -127.87756								
rating83c	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]			
ia83	.0512509	.0167257	3.06	0.002	.018469	.0840327		
dia	-.0496082	.0254406	-1.95	0.051	-.0994709	.0002545		
/cut1	-.1877442	.2048422			-.5892276	.2137392		
/cut2	.6344613	.2156938			.2117092	1.057213		
/cut3	1.064523	.2231881			.6270819	1.501963		

上表中,“cut1,cut2,cut3”为切点的估计值。虽然准  $R^2$  仅为 0.04,但两个解释变量都比较显著。根据该模型,可以预测每个公司的评级概率,并列出第 1 个观测值的预测结果:

```
. predict p2 p3 p4 p5
(option pr assumed; predicted probabilities)
.list p2 p3 p4 p5 in 1/1
```

	p2	p3	p4	p5
1.	.2530799	.309446	.1590192	.2784549

下面进行 ordered logit 估计,然后预测每个公司的债券评级概率,并列出第 1 个观测值的预测结果:

```
. ologit rating83c ia83 dia,nolog
```

<sup>①</sup> 此例来自 Baum (2006)。

Ordered logistic regression						Number of obs = 98
						LR chi2(2) = 11.54
						Prob > chi2 = 0.0031
Log likelihood = -127.27146						Pseudo R2 = 0.0434
rating83c	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ia83	.0939166	.0296196	3.17	0.002	.0358633	.1519699
dia	-.0866925	.0449789	-1.93	0.054	-.1748496	.0014646
/cut1	-.1853053	.3571432			-.8852932	.5146826
/cut2	1.185726	.3882099			.4248488	1.946604
/cut3	1.908412	.4164896			1.092108	2.724717

```
. predict r2 r3 r4 r5
(option pr assumed; predicted probabilities)
.list r2 r3 r4 r5 in 1/1
```

	r2	r3	r4	r5
1.	.2532052	.3186541	.1615769	.2665637

根据上述结果可知,虽然 ordered probit 与 ordered logit 模型的系数与切点估计有所差别,但预测债券评级概率的结果却很接近。

## 13.2 泊松回归

有些被解释变量只能取非负整数,即  $0, 1, 2, \dots$ ,比如,专利个数、奥运金牌个数、子女人数、看病次数。对于这一类计数数据,常使用“泊松回归”(Poisson regression)。首先回顾泊松分布。假设在一次实验中某事件的发生概率为  $p$ ,共进行了  $n$  次相互独立的随机实验,记该事件的发生次数为  $Y$ ,则  $Y=y$  的概率为

$$P(Y=y) = C_n^y p^y (1-p)^{n-y} \quad (y=0, 1, \dots, n) \quad (13.5)$$

可以证明(参见标准的概率统计教材),当  $p \rightarrow 0, n \rightarrow \infty$ ,而  $np = \lambda > 0$  时,此概率的极限为泊松分布:

$$\lim_{n \rightarrow \infty} P(Y=y) = \lim_{n \rightarrow \infty} C_n^y p^y (1-p)^{n-y} = \frac{e^{-\lambda} \lambda^y}{y!} \quad (y=0, 1, \dots, n) \quad (13.6)$$

由于此分布在  $p \rightarrow 0$  时成立,故泊松分布也称为“稀有事件定律”(Law of rare events),即如果在一次实验中事件发生概率很小( $p$  很小),并进行了大量的实验( $n$  很大),则事件发生次数大致服从泊松分布。回到计数数据的情形,对于个体  $i$ ,记被解释变量为  $Y_i$ ,假设  $Y_i = y_i$  的概率由参数为  $\lambda_i$  的泊松分布决定:

$$P(Y_i = y_i | \mathbf{x}_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (y_i = 0, 1, 2, \dots) \quad (13.7)$$

其中,  $\lambda_i > 0$  为“泊松到达率”(Poisson arrival rate),表示事件发生的平均次数,由解释变量  $\mathbf{x}_i$  所决定。泊松分布的期望与方差都等于泊松到达率,即  $E(Y_i | \mathbf{x}_i) = \text{Var}(Y_i | \mathbf{x}_i) = \lambda_i$ 。为了保证  $\lambda_i$  非负,假设  $Y_i$  的“条件期望函数”(conditional mean function)为

$$E(Y_i | \mathbf{x}_i) = \lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta}) \quad (13.8)$$

因此,  $\ln \lambda_i = \mathbf{x}'_i \boldsymbol{\beta}$ , 为对数线性模型。假定样本独立同分布, 则样本的似然函数为

$$L(\boldsymbol{\beta}) = \frac{\exp\left(-\sum_{i=1}^n \lambda_i\right) \cdot \prod_{i=1}^n \lambda_i^{y_i}}{\prod_{i=1}^n y_i!} \quad (13.9)$$

其对数似然函数为

$$\begin{aligned} \ln L(\boldsymbol{\beta}) &= \sum_{i=1}^n [-\lambda_i + y_i \ln \lambda_i - \ln(y_i!)] \\ &= \sum_{i=1}^n [-\exp(\mathbf{x}'_i \boldsymbol{\beta}) + y_i \mathbf{x}'_i \boldsymbol{\beta} - \ln(y_i!)] \end{aligned} \quad (13.10)$$

最大化的一阶条件为

$$\sum_{i=1}^n [y_i - \exp(\mathbf{x}'_i \boldsymbol{\beta})] \mathbf{x}_i = \mathbf{0} \quad (13.11)$$

通过数值计算可得  $\hat{\boldsymbol{\beta}}_{MLE}$ 。根据 MLE 理论, 如果似然函数正确, 则  $\hat{\boldsymbol{\beta}}_{MLE}$  为一致估计量。事实上, 即使似然函数不正确, 由于泊松分布属于线性指数分布族, 故只要条件期望函数  $\lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$  正确, 则准最大似然估计 (QMLE)  $\hat{\boldsymbol{\beta}}_{QMLE}$  就是一致的。直观来看, 如果  $E(Y_i | \mathbf{x}_i) = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ , 则方程 (13.11) 左边的条件期望就为  $\mathbf{0}$ :

$$E\left\{\sum_{i=1}^n [y_i - \exp(\mathbf{x}'_i \boldsymbol{\beta})] \mathbf{x}_i | \mathbf{x}_i\right\} = \sum_{i=1}^n [E(y_i | \mathbf{x}_i) - \exp(\mathbf{x}'_i \boldsymbol{\beta})] \mathbf{x}_i = \mathbf{0} \quad (13.12)$$

然而, 如果似然函数不正确, 则常规的标准误 (比如, OIM 或 BHHH 法) 就不是真实标准误的一致估计量, 基于常规标准误的统计推断 (比如,  $t$  检验或  $F$  检验) 均失效。此时, 必须使用在 QMLE 基础上计算的稳健标准误 (参见第 6 章), 它对于似然函数是否正确比较稳健<sup>①</sup>。

显然,  $\hat{\boldsymbol{\beta}}_{MLE}$  并不表示边际效应。由于  $\ln \lambda_i = \mathbf{x}'_i \boldsymbol{\beta}$ , 故  $\frac{\partial \ln \lambda_i}{\partial x_k} = \beta_k$ 。因此, 可以将  $\beta_k$  解释为“半弹性” (semi-elasticity), 即当解释变量  $x_k$  增加微小量时, 事件的平均发生次数将增加多少百分点。另外, 由于泊松到达率  $\lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ , 故也可计算  $\exp(\beta_k)$ , 称为“发生率比” (Incidence Rate Ratio, 简记 IRR), 表示当  $x_k$  增加一单位时 (从  $x_k$  增加到  $x_k + 1$ ), 事件的平均发生次数将是原来的多少倍, 因为  $\exp[(x_k + 1)\beta_k]/\exp(x_k\beta_k) = \exp(\beta_k)$ 。对模型拟合优度的衡量可以通过准  $R^2$  来进行。

泊松分布描述的是, 在给定时间内某观测单位的事件发生次数。因此, 如果观测时间变长, 或观测单位的空间规模变大, 则事件发生的平均次数也应同比例增多。记个体  $i$  在单位时间内事件发生的平均次数为  $\phi_i$ 。如果不同个体的时间或空间规模不同, 记为  $T_i$ , 称为“暴露期” (exposure), 则该事件发生的平均次数也须相应调整为  $\phi_i T_i$ 。比如, 考察某疾病在不同城市的发病人数, 而各城市的人口基数不同。此时, 可将方程 (13.7) 改写为

$$P(Y_i = y_i | \mathbf{x}_i, T_i) = \frac{e^{-\phi_i T_i} (\phi_i T_i)^{y_i}}{y_i!} \quad (y_i = 0, 1, 2, \dots) \quad (13.13)$$

依然假设  $\phi_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ 。在上式中, 令  $\lambda_i \equiv \phi_i T_i$ , 则  $\ln \lambda_i = \mathbf{x}'_i \boldsymbol{\beta} + \ln T_i$ 。因此, 如果暴露期  $T_i$  随  $i$  而变, 则应把  $\ln T_i$  作为解释变量放入泊松回归, 并且令其系数为 1。如果暴露期不随个体而变, 则自动归入常数项, 无须作此特别处理。另外, 关于“零断尾泊松回归”, 参见第 14 章。

<sup>①</sup> 由于  $\text{Var}(y_i | \mathbf{x}_i) = \lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ , 故此模型本来就允许异方差, 因此其协方差矩阵是关于异方差稳健的。

泊松回归的 Stata 命令为

```
poisson y x1 x2 x3, r irr  
poisson y x1 x2 x3, r exposure(x1)  
poisson y x1 x2 x3, r offset(x1)
```

其中,选择项“r”表示使用稳健标准误;选择项“irr”表示显示发生率比;选择项“exposure(x1)”表示把  $\ln(x1)$  作为解释变量,并令其系数为 1;选择项“offset(x1)”表示把  $x1$  作为解释变量,并令其系数为 1。

### 13.3 负二项回归

泊松回归的局限是泊松分布的期望与方差一定相等,被称为“均等分散”(equidispersion);但这个特征常与实际数据不符。如果被解释变量的方差明显大于期望,即存在“过度分散”(overdispersion)。通常的处理方法是,在条件期望函数的对数表达式中加入一项:

$$\ln \lambda_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i \quad (13.14)$$

其中,随机变量  $\varepsilon_i$  表示条件期望函数中的不可观测部分或个体的异质性。由方程(13.14)可得:

$$\lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta}) \cdot \exp(\varepsilon_i) \equiv u_i v_i \quad (13.15)$$

其中,  $u_i \equiv \exp(\mathbf{x}'_i \boldsymbol{\beta})$  为  $x_i$  的确定性函数,而  $v_i \equiv \exp(\varepsilon_i) > 0$  仍为随机变量。给定  $\mathbf{x}_i$  与  $v_i$ ,则  $y_i$  依然服从泊松分布:

$$P(Y_i = y_i | \mathbf{x}_i, v_i) = \frac{e^{-u_i v_i} (u_i v_i)^{y_i}}{y_i!} \quad (y_i = 0, 1, 2, \dots) \quad (13.16)$$

但由于  $v_i$  不可观测,故无法对方程(13.16)进行估计。为此,记  $v_i$  的概率密度函数为  $g(v_i)$ ,则可以将  $v_i$  积分掉,以计算  $y_i$  的边缘密度:

$$P(Y_i = y_i | \mathbf{x}_i) = \int_0^\infty \frac{e^{-u_i v_i} (u_i v_i)^{y_i}}{y_i!} g(v_i) dv_i \quad (13.17)$$

由于  $v_i > 0$ ,故通常选择  $v_i$  服从 Gamma 分布(指数分布与卡方分布都是 Gamma 分布的特例)。特别地,假设  $v_i \sim \text{Gamma}(1/\alpha, \alpha)$ ,其中  $\alpha > 0$ 。由数理统计知识可知,对于  $\text{Gamma}(a, b)$ ,其期望为  $ab$ ,而方差为  $ab^2$ 。故  $E(v_i) = 1$ ,而  $\text{Var}(v_i) = \alpha$ (参数  $\alpha$  即  $v_i$  的方差)。将  $\text{Gamma}(1/\alpha, \alpha)$  的概率密度代入方程(13.17),可得到负二项分布的概率密度(证明参见 Cameron and Trivedi, 2005, p. 675),由此可以写出样本数据的似然函数,然后进行 MLE 估计。这称为“负二项回归”(negative binomial regression),因为它事实上假设样本来自“负二项分布”(negative binomial distribution)。

有关负二项分布,简单介绍如下。假设某事件在一次实验中成功的概率为  $\theta$  ( $0 < \theta < 1$ )。记  $Y$  为在第  $J$  次成功前失败的总次数<sup>①</sup>,则离散随机变量  $Y$  的分布律为

$$P(Y = y | \theta, J) = C_{y+J-1}^{J-1} \theta^J (1-\theta)^y \quad (y = 0, 1, 2, \dots) \quad (13.18)$$

由于第  $(y+J)$  次一定为成功,故只要在前面的  $(y+J-1)$  次中找出成功的  $(J-1)$  次的组合次数即可。如果  $J=1$ ,则称为“几何分布”(geometric distribution)。

可以证明,负二项回归模型的条件期望仍为  $E(Y_i | \mathbf{x}_i) = u_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ (负二项回归不影响

<sup>①</sup> 大概因为计算的是失败次数,故称为“负二项分布”。

条件期望),而条件方差为

$$\text{Var}(Y_i | \mathbf{x}_i) = u_i + \alpha u_i^2 > u_i = E(Y_i | \mathbf{x}_i) \quad (13.19)$$

这表明,在负二项回归中,条件方差大于条件期望(在泊松回归中,二者相等)。由于条件方差的表达式包含条件期望  $u_i$  的平方项,故也称为“NB2 模型”。从方程(13.19)可以看出,条件方差是  $\alpha$  的增函数,故  $\alpha$  称为“过度分散参数”(overdispersion parameter);特别地,当  $\alpha \rightarrow 0$  时,泊松回归实际上是负二项回归的特例。因此,在进行负二项回归后,只要对原假设 “ $H_0: \alpha = 0$ ” 进行检验,即可确定应使用负二项回归还是泊松回归。为此,Stata 命令提供一个 LR 检验,其原假设为“不存在过度分散,应使用泊松回归”。在完成负二项回归估计后,Stata 将自动输出检验结果“LR test of alpha = 0”,其中过度分散参数“alpha = 0”对应于“泊松回归”;即如果原假设“alpha = 0”成立,则可使用泊松回归。如果使用选择项“r”(稳健标准误),则 Stata 将只输出 alpha 的 95% 置信区间。根据这个置信区间,同样可以检验原假设“alpha = 0”。

另外,如果将  $\text{Gamma}(1/\alpha, \alpha)$  中的  $\alpha$  换为  $\delta/u_i$ ,即假设  $v_i \sim \text{Gamma}(u_i/\delta, \delta/u_i)$ ,其中  $\delta > 0$ ,则条件方差函数(13.19)变为

$$\text{Var}(Y_i | \mathbf{x}_i) = u_i + (\delta/u_i) u_i^2 = u_i + \delta u_i > u_i \quad (13.20)$$

此时,条件方差变为条件期望  $u_i$  的一次函数,故称为“NB1 模型”,是负二项回归的另一形式。如果  $\delta \rightarrow 0$ ,则回到泊松回归的特例。此时,Stata 会汇报对原假设“delta = 0”的检验结果。在实践中,常使用 NB2 模型。在多数情况下,NB2 模型的二次条件方差函数更符合数据的特点。使用 NB2 模型的另一好处是,即使似然函数不正确,只要条件期望函数正确,则  $\hat{\beta}_{\text{QMLE}}$  依然是一致估计(但  $\hat{\alpha}_{\text{QMLE}}$  不是一致估计),此时应使用稳健标准误;而 NB1 模型则无此优点,参见 Cameron and Trivedi (1998)。

对于负二项回归,如果个体的暴露期  $T_i$  不同,同样应将  $\ln T_i$  作为解释变量,并令其系数为 1。另外,关于“零断尾负二项回归”,参见第 14 章。

在 NB2 负二项回归中,条件方差函数主要由参数  $\alpha$  来刻画,不随个体  $i$  而变。作为一个推广,也可让此参数依个体  $i$  而变,记为  $\alpha_i$ ,并让  $\ln \alpha_i$  依赖于变量  $\mathbf{z}_i$ ( $\mathbf{z}_i$  可以与解释变量  $\mathbf{x}_i$  有重叠),正如  $\ln \lambda_i$  依赖于解释变量  $\mathbf{x}_i$ ,然后使用 MLE 对条件均值方程与条件方差方程同时进行估计,这称为“广义负二项回归”(generalized negative binomial regression);但在实践中较少使用。

负二项回归的 Stata 命令为

```
nbreg y x1 x2 x3, r exposure(x1)
```

```
nbreg y x1 x2 x3, r dispersion(constant) offset(x1)
```

其中,选择项“r”表示使用稳健标准误;选择项“dispersion(constant)”表示使用 NB1 模型(默认为 NB2 模型);选择项“exposure(x1)”表示把  $\ln x_1$  作为解释变量,并令其系数为 1;选择项“offset(x1)”表示把  $x_1$  作为解释变量,并令其系数为 1。

广义负二项回归的 Stata 命令为

```
gnbreg y x1 x2 x3, r lnalpha(z1 z2)
```

其中,选择项“r”表示使用稳健标准误;选择项“lnalpha(z1 z2)”表示  $\ln \alpha_i = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2}$ 。如果不使用选择项“lnalpha(varlist)”,则默认为  $\ln \alpha_i = \gamma_0$ ,即  $\alpha$  为常数,将得到与负二项回归 nbreg 同样的结果。

究竟何时使用泊松回归或负二项回归?在某种意义上,泊松回归与负二项回归的关系有如线性模型中 OLS 与 WLS 的关系。即使数据中存在过度分散,“泊松回归 + 稳健标准误”依

然提供了对参数及标准误的一致估计;这类似于在异方差的情况下使用“OLS + 稳健标准误”。另一方面,如果比较了解条件方差函数,则“负二项回归 + 稳健标准误”可提供更有效率的估计;正如在异方差的情况下使用 WLS 可提高估计效率。因此,如果研究者只关心参数  $\beta$  的估计值,或许泊松回归就足够了;但如果希望预测 “ $Y = y$ ” 的发生概率,则可考虑负二项回归。另外,对 “ $H_0 : \alpha = 0$ ” 的 LR 检验结果也提供了在泊松回归与负二项回归之间选择的参考依据。

## 13.4 零膨胀泊松回归与负二项回归

如果计数数据中含有大量的“0”值(zero outcome),则可考虑使用“零膨胀泊松回归”(Zero-inflated Poisson Regression, 简记 ZIP)或“零膨胀负二项回归”(Zero-inflated Negative Binomial Regression, 简记 ZINB)。从理论上来说,决策可能分两阶段进行。首先,决定“取零”(无)或“取正整数”(有),这相当于二值选择。其次,如果决定“取正整数”,则进一步确定具体选择哪个正整数。为此,假定被解释变量  $y_i$  服从以下“混合分布”(mixed distribution):

$$\begin{cases} P(y_i = 0 | \mathbf{x}_i) = \theta \\ P(y_i = j | \mathbf{x}_i) = \frac{(1 - \theta)e^{-\lambda_i} \lambda_i^j}{j! (1 - e^{-\lambda_i})} \quad (j = 1, 2, \dots) \end{cases} \quad (13.21)$$

其中,  $\lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ , 而  $\theta > 0$  与  $\boldsymbol{\beta}$  为待估参数。可以证明,  $\sum_{j=0}^{\infty} P(y_i = j | \mathbf{x}_i) = 1$ (参见习题)。因此,这是一个离散随机变量的分布律。进一步,可以让  $\theta$  依赖于解释变量  $\mathbf{z}_i$ ( $\mathbf{z}_i$  可以等于  $\mathbf{x}_i$ , 或与  $\mathbf{x}_i$  有重叠部分),并用 Logit 模型来估计此二值选择问题,即  $y_i = 0$  或  $y_i > 0$ 。使用 MLE 估计以上模型,即得到“零膨胀泊松回归”。类似地,可以定义“零膨胀负二项回归”。

究竟应该使用标准的泊松回归(standard Poisson)还是零膨胀泊松回归(ZIP)? Stata 提供了一个“Vuong 统计量”(Vuong, 1989),其渐近分布为标准正态。如果 Vuong 统计量很大(为正数),则应选择零膨胀泊松回归(或零膨胀负二项回归);反之,如果 Vuong 统计量很小(为负数),则应选择标准的泊松回归(或“标准的负二项回归”)。

零膨胀回归的 Stata 命令为

`zip y x1 x2 x3, inflate(varlist) vuong` (零膨胀泊松回归)

`zinb y x1 x2 x3, inflate(varlist) vuong` (零膨胀负二项回归)

其中,选择项“`inflate(varlist)`”不可缺少,它列出  $\mathbf{z}_i$  中的所有变量。如果  $\mathbf{z}_i$  中只有常数项,则使用选择项“`inflate(_cons)`”。选择项“`vuong`”表示显示 Vuong 统计量。

第 17 章将介绍计数数据的面板模型。

## 13.5 计数模型的 Stata 实例

下面以有关犯罪率的数据集“CRIME1.dta”为例<sup>①</sup>,使用 Stata 来处理计数数据。该数据集包含以下变量:narr86 (# of times arrested in 1986, 1986 年被捕次数), penv (proportion of prior

<sup>①</sup> 此例来自 Wooldridge (2003)。

convictions, 有前科的比例), avgsen( average sentence length in months, 平均判刑月数), tottime (total time in prison since 18 in months, 18岁以来入狱月数), ptime86 (months in prison during 1986, 1986 年入狱月数), qemp86 (# of quarters employed in 1986, 1986 年就业季度数), inc86 (legal income in 1986 in \$ 100s, 1986 年合法收入), black (是否黑人), hispan (是否拉丁裔), born60 (是否生于 1960 年)。考虑以下模型:

$$\begin{aligned} \text{narr86}_i = & \beta_0 + \beta_1 \text{pcnv}_i + \beta_2 \text{avgse}n_i + \beta_3 \text{tottime}_i + \beta_4 \text{ptime86}_i \\ & + \beta_5 \text{qemp86}_i + \beta_6 \text{inc86}_i + \beta_7 \text{black}_i + \beta_8 \text{hispan}_i + \beta_9 \text{born60}_i + \varepsilon_i \end{aligned} \quad (13.22)$$

首先看一下被解释变量 narr86 (1986 年被逮捕的次数) 的分布:

```
. use CRIME1.dta, clear
.tab narr86
```

# times arrested, 1986	Freq.	Percent	Cum.
0	1,970	72.29	72.29
1	559	20.51	92.81
2	121	4.44	97.25
3	42	1.54	98.79
4	12	0.44	99.23
5	13	0.48	99.71
6	4	0.15	99.85
7	1	0.04	99.89
9	1	0.04	99.93
10	1	0.04	99.96
12	1	0.04	100.00
Total	2,725	100.00	

从上表可知, 在样本容量 2 725 人中, 有 1 970 人在 1986 年从未被逮捕。最高被逮捕纪录为 12 次。更直观地, 可以看直方图。

```
. histogram narr86, discrete frequency
```

其中, 选择项“discrete”表示 narr86 为离散变量, 而“frequency”表示纵坐标为频数, 结果参见图 13.2。

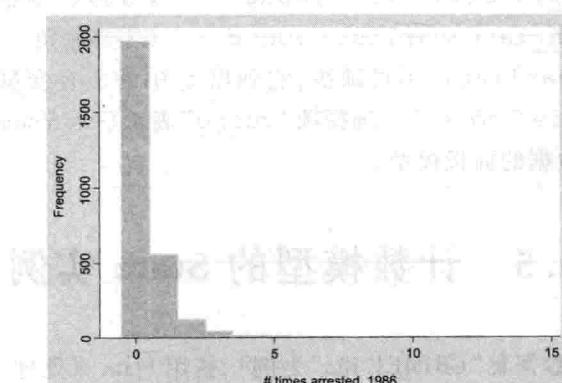


图 13.2 被捕次数 narr86 的直方图

显然,被解释变量为计数数据。尽管如此,作为对照,首先进行 OLS 回归<sup>①</sup>:

```
. reg narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan  
born60, r
```

Linear regression		Number of obs = 2725				
		F( 9, 2715) = 25.93				
		Prob > F = 0.0000				
		R-squared = 0.0725				
		Root MSE = .82873				
narr86		Robust				
	Coef.	Std. Err.				
pcnv	-.131886	.0335876	-3.93	0.000	-.1977458	-.0660262
avgse	-.0113316	.0141409	-0.80	0.423	-.0390595	.0163963
tottime	.0120693	.0131776	0.92	0.360	-.0137699	.0379084
ptime86	-.0408735	.0067985	-6.01	0.000	-.0542043	-.0275426
qemp86	-.0513099	.014205	-3.61	0.000	-.0791636	-.0234562
inc86	-.0014617	.0002289	-6.38	0.000	-.0019106	-.0010128
black	.3270097	.0584381	5.60	0.000	.2124221	.4415973
hispan	.1938094	.0401625	4.83	0.000	.1150572	.2725616
born60	-.022465	.032094	-0.70	0.484	-.0853961	.0404661
_cons	.576566	.0426021	13.53	0.000	.4930302	.6601019

虽然  $R^2$  仅为 0.07,但多数解释变量都很显著。下面进行泊松回归,并使用稳健标准误:

```
. poisson narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan  
born60, r nolog
```

Poisson regression		Number of obs = 2725				
		Wald chi2(9) = 246.22				
		Prob > chi2 = 0.0000				
		Pseudo R2 = 0.0791				
narr86		Robust				
	Coef.	Std. Err.				
pcnv	-.4015713	.1011619	-3.97	0.000	-.5998449	-.2032976
avgse	-.0237723	.0236078	-1.01	0.314	-.0700427	.0224981
tottime	.0244904	.0205023	1.19	0.232	-.0156934	.0646741
ptime86	-.0985584	.0223035	-4.42	0.000	-.1422724	-.0548445
qemp86	-.0380187	.0341509	-1.11	0.266	-.1049532	.0289158
inc86	-.0080807	.0012276	-6.58	0.000	-.0104867	-.0056747
black	.6608376	.0994572	6.64	0.000	.4659051	.85577
hispan	.4998133	.0923874	5.41	0.000	.3187374	.6808892
born60	-.0510286	.0811403	-0.63	0.529	-.2100606	.1080034
_cons	-.5995888	.0893463	-6.71	0.000	-.7747044	-.4244732

尽管泊松回归的估计系数与 OLS 差别很大,但二者并不具可比性。为便于比较,计算泊松回归的平均边际效应:

<sup>①</sup> 对计数数据进行 OLS 回归时,常把被解释变量取对数,即把  $\ln y$  对  $x$  进行 OLS 回归。但由于此处 narr86 可以等于 0,故无法直接取对数。解决方法包括把  $\ln(y+1)$  对  $x$  进行 OLS 回归,或以  $\ln 0.5$  来替代  $\ln 0$ 。一般来说,用 OLS 处理计数数据的主要目的为“探索性数据分析”(exploratory data analysis);比如,计数模型的某个解释变量为内生变量,而一般计量软件尚无法直接处理计数模型中的内生变量问题。

```
. margins, dydx(*)
```

Average marginal effects		Number of obs = 2725				
Model VCE : Robust						
Expression : Predicted number of events, predict()						
dy/dx w.r.t. : pcnv avgse tottime ptime86 qemp86 inc86 black hispan born60						
<hr/>						
		Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
pcnv	-.1623969	.0404199	-4.02	0.000	-.2416184	-.0831754
avgse	-.0096136	.0095756	-1.00	0.315	-.0283815	.0091543
tottime	.009904	.0083122	1.19	0.233	-.0063877	.0261957
ptime86	-.0398574	.0091649	-4.35	0.000	-.0578202	-.0218946
qemp86	-.0153749	.0139299	-1.10	0.270	-.0426771	.0119272
inc86	-.0032679	.0005148	-6.35	0.000	-.0042769	-.0022588
black	.2672451	.0420561	6.35	0.000	.1848168	.3496735
hispan	.2021263	.0382924	5.28	0.000	.1270745	.2771781
born60	-.0206361	.0328797	-0.63	0.530	-.0850793	.043807

从上表可知,泊松模型的平均边际效应与 OLS 的回归系数很接近。为便于解释系数,下面计算发生率比。

```
. poisson narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan  
born60,r irr nolog
```

Poisson regression		Number of obs = 2725				
		Wald chi2(9) = 246.22				
		Prob > chi2 = 0.0000				
Log pseudolikelihood = -2248.7611		Pseudo R2 = 0.0791				
<hr/>						
narr86		Robust				
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
pcnv	.6692676	.0677044	-3.97	0.000	.5488968	.8160353
avgse	.976508	.0230532	-1.01	0.314	.932354	1.022753
tottime	1.024793	.0210106	1.19	0.232	.9844291	1.066811
ptime86	.9061427	.0202101	-4.42	0.000	.8673849	.9466324
qemp86	.9626949	.0328769	-1.11	0.266	.9003667	1.029338
inc86	.9919519	.0012177	-6.58	0.000	.9895681	.9943414
black	1.936414	.1925902	6.64	0.000	1.593456	2.353186
hispan	1.648413	.1522926	5.41	0.000	1.37539	1.975634
born60	.9502515	.0771037	-0.63	0.529	.8105351	1.114052
_cons	.5490374	.0490545	-6.71	0.000	.46084	.6541143

上表显示,给定其他变量,黑人被逮捕的平均次数比白人多 93.6%;其他变量的效应也可类似地解释。使用泊松回归的前提之一是,被解释变量的期望与方差相等。为此,考察 narr86 的统计特征。

```
. sum narr86,detail
```

# times arrested, 1986			
Percentiles		Smallest	
1%	0	0	
5%	0	0	
10%	0	0	Obs 2725
25%	0	0	Sum of Wgt. 2725
50%	0		Mean .4044037
		Largest	Std. Dev. .8590768
75%	1	7	
90%	1	9	Variance .7380129
95%	2	10	Skewness 4.111134
99%	4	12	Kurtosis 32.66587

结果显示,样本方差(Variance)几乎是样本均值(Mean)的两倍。为放松此假定,下面进行负二项回归(NB2):

```
. nbreg narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan  
born60 ,r nolog
```

Negative binomial regression		Number of obs = 2725			
Dispersion	= mean	Wald chi2(9) = 246.07			
Log pseudolikelihood = -2157.628		Prob > chi2 = 0.0000			
<b>Robust</b>					
narr86	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
pcnv	-.4770963	.1013889	-4.71	0.000	-.6758148 -.2783778
avgse	-.0173385	.0243654	-0.71	0.477	-.0650939 .0304168
tottime	.0197394	.0201359	0.98	0.327	-.0197263 .0592051
ptime86	-.1073997	.0230838	-4.65	0.000	-.1526432 -.0621562
qemp86	-.0504884	.0355401	-1.42	0.155	-.1201457 .0191689
inc86	-.0077126	.0012147	-6.35	0.000	-.0100934 -.0053318
black	.6560406	.0980021	6.69	0.000	.4639601 .8481211
hispan	.5048465	.0923508	5.47	0.000	.3238422 .6858508
born60	-.046412	.0798343	-0.58	0.561	-.2028843 .1100603
_cons	-.5637368	.0929054	-6.07	0.000	-.7458281 -.3816455
/lnalpha	-.0738912	.1459399			-.3599282 .2121458
alpha	.9287728	.135545			.6977264 1.236328

上表显示, alpha 的 95% 置信区间为(0.70, 1.24), 故可在 5% 的显著性水平上拒绝过度分散参数“alpha = 0”的原假设(对应于泊松回归), 即认为应使用负二项回归。为演示目的, 下面估计 NB1 负二项回归(尽管 NB1 不如 NB2 稳健)。

```
. nbreg narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan  
born60 ,r d(c) nolog
```

Negative binomial regression		Number of obs = 2725			
Dispersion = constant		Wald chi2(9) = 252.36			
Log pseudolikelihood = -2163.7034		Prob > chi2 = 0.0000			
<hr/>					
narr86		Robust			
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
pcnv	-.5581407	.0938184	-5.95	0.000	-.7420214 -.3742601
avgse	-.0134707	.0271602	-0.50	0.620	-.0667038 .0397624
tottime	.0173537	.0245949	0.71	0.480	-.0308514 .0655588
ptime86	-.1034884	.0213284	-4.85	0.000	-.1452912 -.0616856
qemp86	-.0030067	.0315804	-0.10	0.924	-.0649031 .0588898
inc86	-.0075688	.0011525	-6.57	0.000	-.0098276 -.00531
black	.623694	.0859821	7.25	0.000	.4551722 .7922159
hispan	.4491803	.0851075	5.28	0.000	.2823727 .6159879
born60	-.0155624	.0726358	-0.21	0.830	-.157926 .1268012
_cons	-.6198049	.0794918	-7.80	0.000	-.7756059 -.464004
/lndelta	-.7792728	.1637161			-1.10015 -.4583952
delta	.4587395	.075103			.332821 .6322976

上表显示, delta 的 95% 置信区间为 (0.33, 0.63), 故可在 5% 的显著性水平上拒绝 “delta = 0” 的原假设(对应于泊松回归), 认为应使用负二项回归。假设条件方差参数  $\alpha$  依赖于 pcnv 与 ptime86, 可进行以下广义负二项回归。

```
. gnbreg narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan
born60, r lna(pcnv ptime86) nolog
```

Generalized negative binomial regression		Number of obs = 2725			
		Wald chi2(9) = 201.09			
		Prob > chi2 = 0.0000			
Log pseudolikelihood = -2130.1596		Pseudo R2 = 0.0505			
<hr/>					
narr86		Robust			
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
pcnv	-.3950848	.1103196	-3.58	0.000	-.6113072 -.1788624
avgse	-.0144596	.0271741	-0.53	0.595	-.06772 .0388007
tottime	.0175512	.0211157	0.83	0.406	-.0238348 .0589371
ptime86	.1705355	.0731268	2.33	0.020	.0272095 .3138615
qemp86	-.0330864	.0342524	-0.97	0.334	-.1002198 .034047
inc86	-.0074203	.0011977	-6.20	0.000	-.0097677 -.0050729
black	.6395796	.0982466	6.51	0.000	.4470198 .8321395
hispan	.4099333	.0908442	4.51	0.000	.231882 .5879846
born60	-.0945714	.0797743	-1.19	0.236	-.2509261 .0617833
_cons	-.6190885	.091377	-6.78	0.000	-.7981842 -.4399928
lnalpha					
pcnv	2.057551	.4733131	4.35	0.000	1.129874 2.985228
ptime86	.3532886	.0428277	8.25	0.000	.2693479 .4372293
_cons	-.9914996	.3291731	-3.01	0.003	-1.636667 -.3463322

上表下部显示, 条件方差方程 (lnalpha) 的两个变量均高度显著。而且, 广义负二项回归的对数似然函数 (-2130.1596) 也大于相应的负二项回归 (-2163.7034), 表明模型的拟合优度提高了。由于被解释变量的大多数取值为 0, 下面考虑零膨胀泊松回归:

```
. zip narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan
born60, inf(_cons) vuong nolog
```

Zero-inflated Poisson regression						
Number of obs						= 2725
Nonzero obs						= 755
Zero obs						= 1970
Inflation model = logit						
Log likelihood = -2204.81						
LR chi2(9)						= 290.06
Prob > chi2						= 0.0000
narr86	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
narr86						
pcnv	-.3599746	.0967869	-3.72	0.000	-.5496734	-.1702758
avgse	-.031854	.0241106	-1.32	0.186	-.07911	.015402
tottime	.0323113	.0180048	1.79	0.073	-.0029774	.0676001
ptime86	-.1030596	.0235883	-4.37	0.000	-.1492918	-.0568274
qemp86	-.0698728	.032354	-2.16	0.031	-.1332855	-.0064601
inc86	-.0074046	.0010912	-6.79	0.000	-.0095433	-.005266
black	.6028718	.0834132	7.23	0.000	.4393849	.7663588
hispan	.4899381	.0827269	5.92	0.000	.3277963	.6520798
born60	-.0607905	.0709329	-0.86	0.391	-.1998165	.0782356
_cons	-.1201744	.087349	-1.38	0.169	-.2913753	.0510264
inflate						
_cons	-.5970509	.1258435	-4.74	0.000	-.8436996	-.3504022
Vuong test of zip vs. standard Poisson:						
z =						3.23 Pr>z = 0.0006

上表显示,Vuong 统计量为 3.23,远远大于 1.96(比该统计量更大的概率仅为 0.0006),故拒绝“标准泊松回归”,认为应使用“零膨胀泊松回归”。

下面进行零膨胀负二项回归:

```
. zinb narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan
born, inf(_cons) vuong nocon nolog
```

Zero-inflated negative binomial regression						
Number of obs						= 2725
Nonzero obs						= 755
Zero obs						= 1970
Inflation model = logit						
Log likelihood = -2174.656						
Wald chi2(9)						= 296.49
Prob > chi2						= 0.0000
narr86	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
narr86						
pcnv	-.6181529	.1047333	-5.90	0.000	-.8234264	-.4128794
avgse	-.022769	.0268294	-0.85	0.396	-.0753536	.0298157
tottime	.0231646	.0198022	1.17	0.242	-.0156469	.0619761
ptime86	-.1210453	.0253992	-4.77	0.000	-.1708269	-.0712638
qemp86	-.1349078	.032077	-4.21	0.000	-.1977776	-.0720381
inc86	-.0070007	.0011356	-6.16	0.000	-.0092264	-.0047749
black	.4710273	.0861432	5.47	0.000	.3021898	.6398648
hispan	.3884124	.0879129	4.42	0.000	.2161063	.5607185
born60	-.1547468	.0758095	-2.04	0.041	-.3033307	-.006163
inflate						
_cons	-1.442806	.3069071	-4.70	0.000	-2.044333	-.8412795
/lnalpha	-.4957576	.2157331	-2.30	0.022	-.9185868	-.0729284
alpha	.6091093	.1314051			.3990826	.9296674
Vuong test of zinb vs. standard negative binomial: z =						
z =						-3.66 Pr>z = 0.9999

其中,选择项“nocon”表示不带常数项<sup>①</sup>。上表显示, alpha 的 95% 置信区间为 (0.40, 0.93), 故可在 5% 的显著性水平上拒绝 “alpha = 0”的原假设(对应于泊松回归), 即认为应使用负二项回归。另一方面, Vuong 统计量为 -3.66, 远远小于 -1.96 (比该统计量更小的概率仅为 0.0001), 故拒绝“零膨胀负二项回归”, 认为应使用“标准负二项回归”。总之, 在以上各种模型设定中, 更倾向于选择负二项回归。

## 习 题

**13.1** 对于零膨胀泊松回归, 证明  $\sum_{j=0}^{\infty} P(y_i=j | \mathbf{x}_i) = 1$ 。

**13.2** 使用排序数据集“mus18data1.dta”<sup>②</sup>, 估计以下模型,

$$\text{health}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{linc}_i + \beta_3 \text{ndisease}_i + \varepsilon_i \quad (13.23)$$

其中, health 为健康状况 (1 = 差, 2 = 好, 3 = 非常好), age 为年龄, linc 为收入的对数, ndisease 为患慢性疾病的个数。

(1) 进行 OLS 回归;

(2) 进行 ordered probit 回归;

(3) 进行 ordered logit 回归。

**13.3** 使用数据集“poissonreg.dta”, 估计决定“初中生旷课天数”(daysabs)的计数模型, 解释变量为“语言艺术课成绩”(langarts)与“是否男性”(male)。

(1) 估计泊松回归模型;

(2) 估计负二项回归模型;

(3) 计算被解释变量的均值与方差, 是否方差比均值大很多?

(4) 根据 LR 检验, 应该使用泊松回归, 还是负二项回归?

(5) 估计零膨胀泊松回归模型;

(6) 估计零膨胀负二项回归模型;

(7) 应该使用零膨胀模型吗?

<sup>①</sup> 这样做的原因有二: 其一, 在前面的零膨胀泊松回归中, 常数项并不显著; 其二, 如果包括常数项, 在进行零膨胀负二项回归的数值计算时, 很难收敛。

<sup>②</sup> 此数据集截取自 Cameron and Trivedi (2010)。

# 第 14 章 受限被解释变量

在某些情况下, 被解释变量的取值范围可能受限制, 称为“受限被解释变量”( limited dependent variable)。本章考虑断尾回归( truncated regression)、归并回归( censored regression) 与样本选择模型( sample selection)。

## 14.1 断尾回归

对于线性模型  $y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i$  ( $i = 1, 2, \dots, n$ ) , 假设由于某种原因, 只有满足  $y_i \geq c$  ( $c$  为常数) 的数据才能观测到。因此, 当  $y_i < c$  时, 没有任何有关  $\{y_i, \mathbf{x}_i\}$  的数据。例如,  $y_i$  的总体为某地区所有企业的年销售收入, 而统计局只收集规模以上企业的数据, 比如  $y_i \geq 100\,000$ 。这样, 被解释变量在 100 000 处就存在“左边断尾”(left truncation)。

### 断尾随机变量的概率分布

一个随机变量  $y$  断尾后, 其概率密度函数也随之发生变化。记  $y$  原来的概率密度为  $f(y)$ , 则在  $c$  处左边断尾后的条件密度函数为

$$f(y | y > c) = \begin{cases} \frac{f(y)}{P(y > c)}, & \text{若 } y > c \\ 0, & \text{若 } y \leq c \end{cases} \quad (14.1)$$

由于概率密度函数曲线下的面积必须为 1, 故“断尾随机变量”(truncated random variable) 的密度函数均乘以因子  $\frac{1}{P(y > c)}$ 。断尾的效果参见图 14.1。

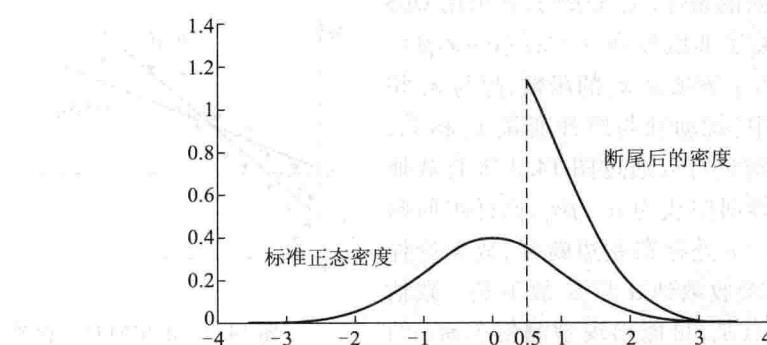


图 14.1 断尾的效果

显然, 断尾分布的期望也发生了变化。由于回归的实质就是估计条件期望函数, 下面以左边

断尾(left truncation)<sup>①</sup>为例来计算断尾分布的条件期望。首先,对于最简单的情形, $y \sim N(0, 1)$ ,可以证明(参见本章附录)

$$E(y | y > c) = \frac{\phi(c)}{1 - \Phi(c)} \quad (14.2)$$

对于任意实数  $c$ , 定义“反米尔斯比率”(Inverse Mill's Ratio, 简记 IMR)<sup>②</sup>为  $\lambda(c) = \frac{\phi(c)}{1 - \Phi(c)}$ , 则  $E(y | y > c) = \lambda(c)$ 。反米尔斯比率在  $c$  点的取值, 等于标准正态的密度函数  $\phi(c)$  除以密度函数曲线下比  $c$  大的阴影部分面积  $[1 - \Phi(c)]$ , 参见图 14.2。

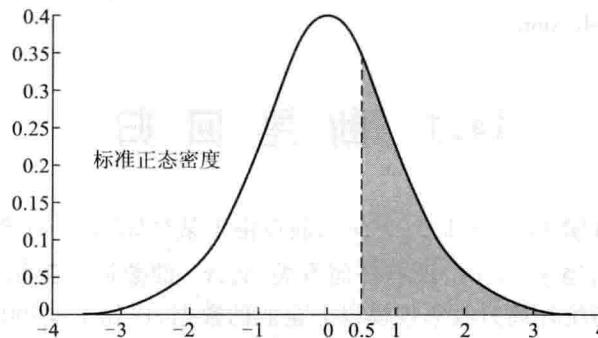


图 14.2 反米尔斯比率

其次, 对于一般的正态分布  $y \sim N(\mu, \sigma^2)$ , 定义  $z \equiv \frac{y - \mu}{\sigma} \sim N(0, 1)$ , 则  $y = \mu + \sigma z$ 。故

$$\begin{aligned} E(y | y > c) &= E(\mu + \sigma z | \mu + \sigma z > c) = E[\mu + \sigma z | z > (c - \mu)/\sigma] \\ &= \mu + \sigma E[z | z > (c - \mu)/\sigma] = \mu + \sigma \cdot \lambda[(c - \mu)/\sigma] \end{aligned} \quad (14.3)$$

其中, 最后一个等号使用了标准正态的断尾期望公式(14.2)。对于回归模型  $y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i$ , 假设  $\varepsilon_i | \mathbf{x}_i \sim N(0, \sigma^2)$ 。因此,  $y_i | \mathbf{x}_i \sim N(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2)$ 。套用方程(14.3)可得

$$E(y_i | y_i > c) = \mathbf{x}'_i \boldsymbol{\beta} + \sigma \cdot \lambda[(c - \mathbf{x}'_i \boldsymbol{\beta})/\sigma] \quad (14.4)$$

由于“ $y_i > c$ ”是样本可观测的条件, 上式表明, 如果用 OLS 估计  $y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i$ , 则遗漏了非线性项  $\sigma \cdot \lambda[(c - \mathbf{x}'_i \boldsymbol{\beta})/\sigma]$ , 被纳入扰动项中。由于该项是  $\mathbf{x}_i$  的函数, 故与  $\mathbf{x}_i$  相关。因此, 在 OLS 回归中, 扰动项与解释变量  $\mathbf{x}_i$  相关, 导致不一致的估计。此结论可以通过图 14.3 来直观地展示。在图 14.3 中, 总体回归线为  $\alpha + \beta x_i$ , 而样本回归线为  $\hat{\alpha} + \hat{\beta} x_i$ 。由于在  $y_i = c$  处存在左边断尾, 故无论样本容量多大,  $\hat{\alpha}$  与  $\hat{\beta}$  都不会收敛到  $\alpha$  与  $\beta$ , 故不是一致估计。OLS 回归的另一缺点是, 可能出现预测值  $\hat{y}_i \leq c$  的不可能情形。

使用 MLE 可以得到一致的估计。断尾前  $y_i$  的概率密度函数为

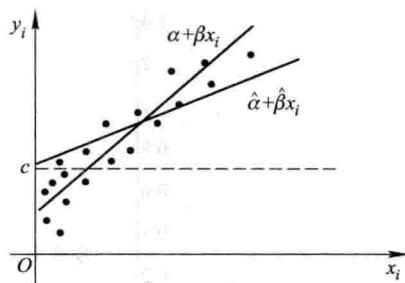


图 14.3 断尾回归示意图

<sup>①</sup> 右边断尾(right truncation)的情形可类似地计算。

<sup>②</sup> 也称为“风险函数”(hazard function)。

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma}\right)^2\right\} = \frac{1}{\sigma} \phi\left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma}\right) \quad (14.5)$$

样本被观测到的概率为

$$\begin{aligned} P(y_i > c | \mathbf{x}_i) &= 1 - P(y_i \leq c | \mathbf{x}_i) \\ &= 1 - P\left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \leq \frac{c - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \mid \mathbf{x}_i\right) \\ &= 1 - P\left(\frac{\varepsilon_i}{\sigma} \leq \frac{c - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \mid \mathbf{x}_i\right) \\ &= 1 - \Phi\left(\frac{c - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma}\right) \end{aligned} \quad (14.6)$$

其中,  $\varepsilon_i/\sigma$  服从标准正态分布。因此, 根据公式(14.1), 断尾后的条件密度为

$$f(y_i \mid y_i > c, \mathbf{x}_i) = \frac{\frac{1}{\sigma} \phi((y_i - \mathbf{x}'_i \boldsymbol{\beta})/\sigma)}{1 - \Phi((c - \mathbf{x}'_i \boldsymbol{\beta})/\sigma)} \quad (14.7)$$

由此, 可以计算出整个样本的似然函数, 然后使用 MLE。

断尾回归的 Stata 命令为

```
truncreg y x1 x2 x3 ,ll (#)ul (#)
```

其中, 选择项“ll (#)”表示 lower limit, 即左边断尾; 选择项“ul (#)”表示 upper limit, 即右边断尾; 如果同时使用这两个选择项, 则表示双边断尾。

以数据集 laborsub.dta 为例<sup>①</sup>, 估计一个决定妇女劳动时间的模型。该数据集包含以下变量: whrs(妻子的劳动时间, wife's hours of work), kl6(6 岁以下儿童数, # of kids less than 6), k618(6~18 岁少年数, # of kids between 6 and 18), wa(妻子年龄, wife's age), we(妻子教育年限, wife's years of education), 以及 lfp(是否就业, labor force participation)。

显然, 如果 lfp = 1(有工作), 则 whrs > 0。为此, 先看一下 lfp 的分布:

```
. use laborsub.dta, clear
. tab lfp
```

1 if woman worked in 1975	Freq.	Percent	Cum.
0	100	40.00	40.00
1	150	60.00	100.00
Total	250	100.00	

上表显示, 在样本容量 250 人中, 有 100 人不工作, 即工作时间为 0。作为对照, 先对有工作的 150 人的子样本进行 OLS 回归:

```
. reg whrs kl6 k618 wa we if whrs > 0
```

<sup>①</sup> 此例来自 Baum(2006)。

Source	SS	df	MS	Number of obs = 150		
Model	7326995.15	4	1831748.79	F( 4, 145) = 2.80		
Residual	94793104.2	145	653745.546	Prob > F = 0.0281		
Total	102120099	149	685369.794	R-squared = 0.0717		
				Adj R-squared = 0.0461		
				Root MSE = 808.55		

whrs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
k16	-421.4822	167.9734	-2.51	0.013	-753.4748 -89.48953
k618	-104.4571	54.18616	-1.93	0.056	-211.5538 2.639668
wa	-4.784917	9.690502	-0.49	0.622	-23.9378 14.36797
we	9.353195	31.23793	0.30	0.765	-52.38731 71.0937
_cons	1629.817	615.1301	2.65	0.009	414.0371 2845.597

下面进行断尾回归,假设在“whrs = 0”处存在左边断尾:

. truncreg whrs k16 k618 wa we, ll(0) nolog

Truncated regression					
Limit: lower = 0				Number of obs = 150	
upper = +inf				Wald chi2(4) = 10.05	
Log likelihood = -1200.9157				Prob > chi2 = 0.0395	
whrs	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
k16	-803.0042	321.3614	-2.50	0.012	-1432.861 -173.1474
k618	-172.875	88.72898	-1.95	0.051	-346.7806 1.030578
wa	-8.821123	14.36848	-0.61	0.539	-36.98283 19.34059
we	16.52873	46.50375	0.36	0.722	-74.61695 107.6744
_cons	1586.26	912.355	1.74	0.082	-201.9233 3374.442
/sigma	983.7262	94.44303	10.42	0.000	798.6213 1168.831

比较以上两表可知,断尾回归的结果与对子样本进行 OLS 回归有较大不同。

## 14.2 零断尾泊松回归与负二项回归

对于离散数据,也可能出现断尾的情形。特别地,对于计数数据(参见第13章),有时样本中仅包括正整数,而不包括取值为0的观测值。比如,在商场发放问卷调查,研究消费者每周去商场的次数;在公交车上发放问卷调查,研究乘车者每周坐公交车的次数。在这两个例子中,虽然被解释变量为计数数据,但取值都不为0,故在0处存在左边断尾,称为“零断尾”(zero-truncated)。

对于零断尾的计数模型,无论使用泊松回归还是负二项回归,如果不对似然函数进行适当调整,将得不到一致估计。记 $f(y)$ 为离散被解释变量 $y$ 的概率函数,而 $F(y) \equiv P(Y \leq y)$ 为其累积分布函数。如果存在零断尾,则断尾后的概率函数为

$$f(y | y \geq 1) = \frac{f(y)}{1 - F(0)}, \quad y = 1, 2, \dots \quad (14.8)$$

如果 $y$ 服从泊松分布,则

$$f(y|y \geq 1) = \frac{e^{-\lambda} \lambda^y}{y! (1 - e^{-\lambda})}, \quad y = 1, 2, \dots \quad (14.9)$$

由此可写出样本数据的似然函数, 进行 MLE 估计, 即得到“零断尾泊松回归”(zero-truncated Poisson regression)。类似地, 如果  $y$  服从负二项分布(可使用 NB1 或 NB2 模型), 可进行“零断尾负二项回归”(zero-truncated negative binomial regression)。相应的 Stata 命令为:

`ztp y x1 x2 x3, r` (零断尾泊松回归)

`ztnb y x1 x2 x3, r` (零断尾负二项回归, 默认为 NB2 模型)

`ztnb y x1 x2 x3, r dispersion(constant)` (零断尾负二项回归, NB1 模型)

其中, 选择项“r”表示使用稳健标准误。

下面以第 13 章有关犯罪率的数据集 CRIME1.dta 为例。为了使用零断尾回归, 首先去掉样本中没有犯罪记录的观测值。

. use CRIME1.dta, clear

. drop if narr86 == 0

(1970 observations deleted)

首先, 进行零断尾泊松回归。

. ztp narr86 pcnv avgse tottime ptime86 qemp86 inc86 black hispan born60, r nolog

Zero-truncated Poisson regression							
					Number of obs	=	755
					Wald chi2(9)	=	145.98
					Prob > chi2	=	0.0000
					Pseudo R2	=	0.1942
Log pseudolikelihood = -667.14707							
narr86	Robust	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
pcnv	.6609906	.1852481	3.57	0.000	.297911	1.02407	
avgse	-.0680206	.0271185	-2.51	0.012	-.1211718	-.0148694	
tottime	.0539391	.0143497	3.76	0.000	.0258143	.0820639	
ptime86	.0322599	.0475728	0.68	0.498	-.060981	.1255009	
qemp86	-.1939824	.0706526	-2.75	0.006	-.3324589	-.0555059	
inc86	-.0070621	.003688	-1.91	0.056	-.0142905	.0001663	
black	.385323	.1837578	2.10	0.036	.0251644	.7454817	
hispan	.3182917	.1715818	1.86	0.064	-.0180025	.6545858	
born60	-.2538259	.1493695	-1.70	0.089	-.5465847	.0389328	
_cons	-.1183123	.1836823	-0.64	0.520	-.478323	.2416983	

然而, 如果直接用所有解释变量进行零断尾负二项回归, 无论 NB1 还是 NB2, 都不会收敛。为了演示目的, 在解释变量中去掉在断尾回归中最不显著的两个变量 ptime86 与 born60, 再进行 NB2 回归。

. ztnb narr86 pcnv avgse tottime qemp86 inc86 black hispan, r nolog

Zero-truncated negative binomial regression		Number of obs = 755			
Dispersion	= mean	Wald chi2(7)		= 131.10	
Log likelihood	= -624.91294	Prob > chi2		= 0.0000	
<hr/>					
narr86	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
pcnv	.868443	.242886	3.58	0.000	.3923953 1.344491
avgse	-.0754736	.0378035	-2.00	0.046	-.1495671 -.0013801
tottime	.0623184	.0217107	2.87	0.004	.0197663 .1048706
qemp86	-.2788742	.0862416	-3.23	0.001	-.4479047 -.1098436
inc86	-.0069477	.0038483	-1.81	0.071	-.0144903 .0005949
black	.53225	.2279557	2.33	0.020	.085465 .9790349
hispan	.5326132	.2184151	2.44	0.015	.1045274 .9606989
_cons	-15.45257	4.655327	-3.32	0.001	-24.57684 -6.328298
/lnalpha	15.50324	4.785254		6.124316	24.88217
alpha	5407201	2.59e+07		456.8322	6.40e+10

### 14.3 随机前沿模型(选读)

在经济学中,常需估计生产函数或成本函数。教科书对生产函数  $f(\mathbf{x})$  的标准定义为在给定投入  $\mathbf{x}$  情况下的最大产出。但现实中的厂商可能达不到此最大产出的前沿(frontier)。为此,假设厂商  $i$  的产量为

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) \xi_i \quad (14.10)$$

其中,  $\boldsymbol{\beta}$  为待估参数;  $\xi_i$  为厂商  $i$  的效率水平,满足  $0 < \xi_i \leq 1$ 。如果  $\xi_i = 1$ , 则厂商  $i$  正好位于效率前沿。考虑到生产函数还会受到随机冲击,故将方程(14.10)改写为

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) \xi_i e^{v_i} \quad (14.11)$$

其中,  $e^{v_i} > 0$  为随机冲击。方程(14.11)意味着生产函数的前沿  $f(\mathbf{x}_i, \boldsymbol{\beta}) e^{v_i}$  是随机的,故此类模型称为“随机前沿模型”(stochastic frontier model),最早由 Aigner, Lovell and Schmidt(1977)提出。随机前沿模型在实证研究中应用广泛,Kumbhakar and Lovell(2000)为此领域的专著。

假设  $f(\mathbf{x}_i, \boldsymbol{\beta}) = e^{\beta_0} x_{i1}^{\beta_1} \cdots x_{iK}^{\beta_K}$  (Cobb-Douglas 生产函数,且共有  $K$  个投入品),则对方程(14.11)两边取对数可得

$$\ln y_i = \beta_0 + \sum_{k=1}^K \beta_k \ln x_{ki} + \ln \xi_i + v_i \quad (14.12)$$

由于  $0 < \xi_i \leq 1$ , 故  $\ln \xi_i \leq 0$ 。定义  $u_i \equiv -\ln \xi_i \geq 0$ , 则方程(14.12)可写为

$$\ln y_i = \beta_0 + \sum_{k=1}^K \beta_k \ln x_{ki} + \underbrace{v_i - u_i}_{\varepsilon_i}, \quad u_i \geq 0 \quad (14.13)$$

其中,  $u_i \geq 0$  为“无效率项”(inefficiency term),反映厂商  $i$  离效率前沿的距离;而  $v_i$  称为“idiosyncratic error”。假设  $v_i$  为 iid, 对称分布且独立于  $u_i$ , 则方程(14.13)的复合扰动项  $\varepsilon_i \equiv v_i - u_i$  为非对称分布(因为  $u_i \geq 0$ )。如果假设  $u_i, v_i$  都与  $\mathbf{x}_i$  不相关,则使用 OLS 可以一致地估计  $\beta_1, \dots, \beta_K$ ;但 OLS 无法一致地估计  $\beta_0$ , 因为复合扰动项的期望值  $E(\varepsilon_i) = E(v_i - u_i) = E(-u_i) \leq 0$ 。除了参数估计外,使用随机前沿模型的另一目的在于估计每个厂商的生产效率  $u_i$ ;但 OLS 无

法提供此信息。为了估计  $u_i$ , 必须对  $u_i, v_i$  的分布作假设; 在此基础上, 则可进行更有效率的 MLE 估计。首先, 假设  $u_i, v_i$  均为 iid 且相互独立, 而且独立于解释变量  $x_i$ 。其次, 假设  $v_i \sim N(0, \sigma_v^2)$ , 故其密度函数为  $\frac{1}{\sqrt{2\pi}\sigma_v} \exp(-v_i^2/2\sigma_v^2)$ 。最后, 有关  $u_i$  的分布, 常假定  $u_i \sim N^+(0, \sigma_u^2)$ , 即期望为 0 的正态分布, 但在原点左边断尾, 称为“半正态分布”(half-normal distribution); 此模型也称为“正态 - 半正态模型”(normal-half normal model)(因为  $v_i$  服从正态分布), 简称“半正态模型”。根据断尾分布的密度公式(14.1)可知,  $u_i$  的密度函数为  $\frac{2}{\sqrt{2\pi}\sigma_u} \exp(-u_i^2/2\sigma_u^2)$ 。由于  $u_i, v_i$  相互独立, 故其联合分布密度为二者边缘密度的乘积:

$$f(u_i, v_i) = \frac{1}{\pi\sigma_v\sigma_u} \exp\left\{-\frac{u_i^2}{2\sigma_u^2} - \frac{v_i^2}{2\sigma_v^2}\right\} \quad (14.14)$$

由此, 可证明  $\varepsilon_i \equiv v_i - u_i$  的密度函数为(Kumbhakar and Lovell, 2000, p. 75)

$$f(\varepsilon_i) = \frac{2}{\sigma} \phi(\varepsilon_i/\sigma) \Phi(-\varepsilon_i\lambda/\sigma) \quad (14.15)$$

其中,  $\sigma \equiv \sqrt{\sigma_u^2 + \sigma_v^2}$ ,  $\lambda \equiv \sigma_u/\sigma_v$ , 而  $\phi(\cdot)$ ,  $\Phi(\cdot)$  分别为标准正态的密度函数与累积分布函数。 $\lambda$  度量的是  $u$  与  $v$  对  $\varepsilon$  的相对贡献。由于  $\phi(\varepsilon_i/\sigma) = \frac{1}{\sqrt{2\pi}} \exp[-(\varepsilon_i/\sigma)^2/2]$ , 故厂商  $i$  的对数似然函数为(忽略常数项)

$$\ln f(\varepsilon_i) = -\ln \sigma - \frac{1}{2} (\varepsilon_i/\sigma)^2 + \ln \Phi(-\varepsilon_i\lambda/\sigma) \quad (14.16)$$

假设样本为 iid, 则样本容量为  $n$  的样本似然函数为

$$\ln L = -n \ln \sigma - \frac{1}{2} \sum_{i=1}^n (\varepsilon_i/\sigma)^2 + \sum_{i=1}^n \ln \Phi(-\varepsilon_i\lambda/\sigma) \quad (14.17)$$

最大化表达式(14.17)即得到对半正态模型的 MLE 估计。

由于半正态分布的众数(mode)为 0, 故半正态模型意味着位于效率前沿的厂商多于位于任何其他位置的厂商。此假定在现实中未必成立。一种可能的情形是, 只有少数厂商达到效率前沿或处于最无效率的状态, 而多数厂商处于中间状态。为此, Stevenson(1980)引入如下“正态 - 断尾正态模型”(normal-truncated normal model), 简称“断尾正态模型”, 即假设  $u_i \sim N^+(\mu, \sigma_u^2)$ , 即期望为  $\mu$  的正态分布, 但在原点左边断尾。显然, 半正态模型是断尾正态模型的特例, 即  $\mu = 0$  的情形。由于断尾正态模型比半正态模型多了一个待估计参数  $\mu$ (Stata 记为 mu), 故可更灵活地拟合数据。可以证明, 断尾正态模型的样本对数似然函数为

$$\ln L = -n \ln \sigma - n \ln \Phi(-\mu/\sigma_u) - \frac{1}{2} \sum_{i=1}^n \left(\frac{\varepsilon_i + \mu}{\sigma}\right)^2 + \sum_{i=1}^n \ln \Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon_i\lambda}{\sigma}\right) \quad (14.18)$$

显然, 在方程(14.18)中, 如果令  $\mu = 0$ , 则简化为半正态模型的方程(14.17)。另一常用模型假设  $u_i$  服从指数分布, 其密度函数为  $f(u_i) = \frac{1}{\sigma_u} \exp(-u_i/\sigma_u)$ ,  $u_i \geq 0$ , 称为“正态 - 指数模型”(normal-exponential model), 简称“指数模型”<sup>①</sup>。类似地, 可以证明, 指数模型的样本对数似然函

<sup>①</sup> 指数模型还可推广为 Gamma 模型, 因为单一参数的指数分布是双参数 Gamma 分布的特例, 参见 Kumbhakar and Lovell (2000) 以及 Greene(2012)。但 Stata 不提供对 Gamma 模型的估计。

数为(Kumbhakar and Lovell, 2000, p. 81)

$$\ln L = -n \ln \sigma_u - n \left( \frac{\sigma_v^2}{2\sigma_u^2} \right) + \sum_{i=1}^n \ln \Phi \left( -\frac{\varepsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) + \sum_{i=1}^n \left( \frac{\varepsilon_i}{\sigma_u} \right)^2 \quad (14.19)$$

对以上模型进行 MLE 估计后, 可计算残差  $\hat{\varepsilon}_i \equiv \ln y_i - \hat{\beta}_0 - \sum_{k=1}^K \hat{\beta}_k \ln x_{ki}$ , 但这依然不是对  $u_i$  的估计。由于  $\varepsilon_i = v_i - u_i$  包含了  $u_i$  的信息<sup>①</sup>, 故可通过条件分布  $f(u_i | \varepsilon_i)$  来对  $u_i$  进行点估计。最常用的方法是以条件期望  $E(u_i | \varepsilon_i)$  作为  $u_i$  的估计, 有时也以条件众数  $M(u_i | \varepsilon_i)$  作为  $u_i$  的估计(即条件分布  $f(u_i | \varepsilon_i)$  的众数)。

在以上推导中均假设样本为 iid, 排除了异方差的可能性。事实上, 由于厂商规模不同,  $u_i$  与  $v_i$  都可能存在异方差。对于半正态或指数模型, 可假设  $\sigma_i^2$  依赖于变量  $z = (z_1, \dots, z_m)$ <sup>②</sup>:

$$\sigma_i^2 = \exp(z'_i \boldsymbol{\delta}) \quad (14.20)$$

其中,  $\sigma_i^2$  可以是  $\text{Var}(u_i)$  或  $\text{Var}(v_i)$ , 而  $z_i$  为条件方差方程(14.20)的解释变量(可与  $x_i$  重叠)。将方程(14.20)代入样本对数似然函数, 即可进行 MLE 估计。通过检验系数  $\boldsymbol{\delta}$  的联合显著性即可判断是否存在异方差。

对于断尾正态模型, Stata 可以估计条件期望模型 (conditional mean model), 即让期望  $\mu_i$  依赖于变量  $w$ :

$$\mu_i = w'_i \gamma \quad (14.21)$$

其中,  $w_i$  为条件期望方程(14.21)的解释变量(可与  $x_i$  重叠)。

随机前沿模型可以很容易地用于估计成本函数。根据微观经济学的知识, 给定成本情况下的产出最大化问题, 其对偶问题就是给定产出情况下的成本最小化问题。对于 Cobb-Douglas 生产函数, 容易证明其成本函数在形式上也是 Cobb-Douglas 函数。经过与生产函数的随机前沿模型类似的推导可得

$$\ln c_i = \beta_0 + \beta_y \ln y_i + \sum_{k=1}^K \beta_k \ln p_{ki} + \underbrace{v_i + u_i}_{\varepsilon_i}, \quad u_i \geq 0 \quad (14.22)$$

其中,  $c_i$  为厂商  $i$  的成本,  $y_i$  为产出,  $p_{ki}$  为要素  $k$  的价格,  $u_i$  为无效率项 (inefficiency term), 而  $v_i$  为对成本函数的随机冲击。对于成本函数而言, 如果  $u_i = 0$ , 则达到最低成本的效率前沿; 反之, 如果  $u_i > 0$ , 则厂商需付出更高成本。由方程(14.22)可知, 对于成本函数的随机前沿模型, 可类似地进行 MLE 估计(只需略微调整对数似然函数即可)。

使用随机前沿模型的前提是无效率项  $u_i$  的存在。此假定可通过检验 “ $H_0 : \sigma_u^2 = 0 \text{ vs } H_1 : \sigma_u^2 > 0$ ” 来判断是否成立。如果接受原假设, 则无须使用随机前沿模型。但困难在于, 原假设  $\sigma_u^2 = 0$  正好落在  $\sigma_u^2$  的参数空间 ( $\sigma_u^2 \geq 0$ ) 的边界上, 导致标准的似然比检验不成立; 此时, 应使用单边的广义似然比检验 (one-sided generalized likelihood ratio test), 参见 Gutierrez et al (2001)。

检验无效率项  $u_i$  是否存在的另一方法通过检验 OLS 残差的偏度来进行, 其基本思想如下。如果不存在无效率项  $u_i$ , 则复合扰动项  $\varepsilon_i$  简化为  $v_i$ , 为对称分布, 偏度为 0; 反之, 如果存在无效率项  $u_i$ , 则复合扰动项  $\varepsilon_i$  的偏度将不为零。比如, 对于生产函数模型,  $\varepsilon_i = v_i - u_i$ ; 由于  $u_i$  与  $v_i$  相互独立, 故  $E(\varepsilon_i^3) \equiv E(v_i^3) - E(u_i^3) = -E(u_i^3) \leq 0$ (由于  $v_i$  为对称分布, 故  $E(v_i^3) = 0$ ),  $\varepsilon_i$  的偏度为

<sup>①</sup> 比如, 如果  $\varepsilon_i$  为正, 则可大致推断  $u_i$  较小; 反之, 如果  $\varepsilon_i$  为负, 则可大致推断  $u_i$  较大。

<sup>②</sup> 此处理方法类似于含异方差的 Probit 模型, 参见第 11 章。

负数,即  $\varepsilon_i$  更多地分布于原点左边。因此,可通过 OLS 残差来检验复合扰动项的分布是否有偏,其原假设为“ $H_0 : E(\varepsilon_i^3) = 0$ ”,而替代假设为“ $H_0 : E(\varepsilon_i^3) < 0$ ”,故为单边左侧检验;反之,对于成本函数模型, $\varepsilon_i \equiv v_i + u_i$ ;故  $\varepsilon_i$  的偏度为正数,即  $\varepsilon_i$  更多地分布于原点右边;因此替代假设为“ $H_0 : E(\varepsilon_i^3) > 0$ ”,为单边右侧检验。在原假设成立的条件下,Coelli(1995)证明检验统计量

$$\frac{m_3}{(6m_2^3/n)^{1/2}} \xrightarrow{d} N(0,1) \quad (14.23)$$

其中, $m_2$  与  $m_3$  分别为 OLS 残差的样本二阶矩与三阶矩<sup>①</sup>。根据此统计量进行上述的单边检验。如果接受原假设,则认为不存在无效率项  $u_i$ ,不必使用随机前沿模型;反之,则认为存在无效率项  $u_i$ 。此检验方法的优点在于它是基于 OLS 残差的,计算十分方便。

使用横截面数据的随机前沿模型一般假设  $u_i, v_i$  与  $x_i$  不相关。然而,由于厂商很可能知道自己的  $u_i$ ,并据此调整最优投入品  $x_i$ ,故此假定可能不成立,导致不一致的估计。但这是横截面数据所无法克服的困难。一般来说,需要使用面板数据,才能在  $u_i$  与  $x_i$  相关的情况下得到一致估计。有关面板随机前沿模型,参见第 17 章。

随机前沿模型的 Stata 命令格式为

```
frontier lny lnx1 lnx2 lnx3, cost distribution (exponential)
distribution(tnormal) uhet(varlist) vhet(varlist) cm(varlist)
```

其中,“lny”为对数形式的被解释变量,“lnx1 lnx2 lnx3”为对数形式的解释变量。选择项“cost”表示估计成本函数,默认估计生产函数。选择项“distribution(exponential)”与“distribution(tnormal)”分别表示估计指数模型与断尾正态模型,默认估计半正态模型,即“distribution(hnornormal)”。选择项“uhet(varlist)”用于指定决定异方差  $Var(u_i)$  的解释变量,选择项“vhet(varlist)”用于指定影响异方差  $Var(v_i)$  的解释变量;这两个选择项仅适用于半正态模型与指数模型(不适用于断尾正态模型)。选择项“cm(varlist)”用于指定影响条件期望  $\mu_i$  的解释变量(“cm”表示 conditional mean),此选择项仅适用于断尾正态模型。

下面以数据集 green9.dta 为例。该数据集来自 Greene(2008),最早为 Zellner and Revankar(1969)所使用。它涉及美国 25 个州的运输设备制造行业,主要变量包括:lnv(增加值对数)、lnk(资本对数)、lnl(劳动力对数)、state(州),以及 nfirm(该州的企业个数)。作为对照,首先以 OLS 估计生产函数。

```
. use green9.dta, clear
. reg lny lnk lnl, r
```

Linear regression						Number of obs = 25
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	F( 2, 22) = 625.05
lnv	.2454281	.0877269	2.80	0.010	.0634935 .4273626	Prob > F = 0.0000
lnk	.805183	.1230483	6.54	0.000	.5499965 1.060369	R-squared = 0.9731
lnl						Root MSE = .23571
_cons	1.844416	.2829646	6.52	0.000	1.257583 2.431248	

<sup>①</sup> 由于 OLS 残差的均值为 0,故残差的中心矩与原点矩相等。

其次,估计随机前沿的半正态模型。

```
. frontier lnv lnk lnl,nolog
```

Stoc. frontier normal/half-normal model						Number of obs = 25
						Wald chi2(2) = 743.71
						Prob > chi2 = 0.0000
Log likelihood = 2.4695222						
lnv	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnk	.2585478	.098764	2.62	0.009	.0649738	.4521218
lnl	.7802451	.1199399	6.51	0.000	.5451672	1.015323
_cons	2.081135	.281641	7.39	0.000	1.529128	2.633141
/lnsig2v	-3.48401	.6195353	-5.62	0.000	-4.698277	-2.269743
/lnsig2u	-3.014599	1.11694	-2.70	0.007	-5.203761	-.8254368
sigma_v	.1751688	.0542616			.0954514	.3214633
sigma_u	.2215073	.1237052			.074134	.6618486
sigma2	.0797496	.0426989			-.0039388	.163438
lambda	1.264536	.1678684			.9355204	1.593552

Likelihood-ratio test of sigma\_u=0: chibar2(01) = 0.43 Prob>=chibar2 = 0.256

上表中的  $\sigma_v$ ,  $\sigma_u$ ,  $\sigma^2 \equiv \sigma_u^2 + \sigma_v^2$  以及  $\lambda \equiv \sigma_u / \sigma_v$ 。由于  $\hat{\lambda} = 1.26$ , 故无效率项  $u$  在复合扰动项  $\varepsilon \equiv v - u$  占据了主导地位。然而, 上表最后一行的似然比检验却接受了原假设 “ $H_0 : \sigma_u^2 = 0$ ”, 即认为不存在无效率项, 不必使用随机前沿模型; 这很可能是由于样本容量过小所致(仅为 25)。下面, 估计每个州的无效率项  $u_i$ , 并将此估计值记为  $u_h$ 。

```
. predict u_h,u
```

其中, 选择项 “u” 表示使用  $E(u_i | \varepsilon_i)$  预测  $u_i$ , 而默认预测  $\ln \hat{y}_i$ 。

再次, 估计随机前沿的指数模型。

```
. frontier lnv lnk lnl,d(e) nolog
```

Stoc. frontier normal/exponential model						Number of obs = 25
						Wald chi2(2) = 845.68
						Prob > chi2 = 0.0000
Log likelihood = 2.8604897						
lnv	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnk	.2624859	.0919988	2.85	0.004	.0821717	.4428002
lnl	.7703795	.1109569	6.94	0.000	.5529079	.9878511
_cons	2.069242	.2356159	8.78	0.000	1.607444	2.531041
/lnsig2v	-3.527598	.4486176	-7.86	0.000	-4.406873	-2.648324
/lnsig2u	-4.002457	.9274575	-4.32	0.000	-5.820241	-2.184674
sigma_v	.1713925	.0384448			.1104231	.2660258
sigma_u	.1351691	.0626818			.0544692	.3354317
sigma2	.0476461	.0157921			.016694	.0785981
lambda	.7886525	.087684			.616795	.9605101

Likelihood-ratio test of sigma\_u=0: chibar2(01) = 1.21 Prob>=chibar2 = 0.135

上表最后一行依然接受了原假设 “ $H_0 : \sigma_u^2 = 0$ ”, 尽管检验的  $p$  值缩小为 0.135。下面, 估计每个州的无效率项  $u_i$ , 并将此估计值记为  $u_e$ 。

```
. predict u_e,u
```

由于样本容量过小,如果估计断尾正态模型,将无法收敛。下面罗列半正态模型与指数模型对无效率项的预测。

```
. list state u_h u_e
```

	state	u_h	u_e
1.	Alabama	.2011338	.1459287
2.	California	.1448097	.0972165
3.	Connecticut	.1903485	.134788
4.	Florida	.5175314	.5903303
5.	Georgia	.1039791	.0714099
6.	Illinois	.121267	.0830415
7.	Indiana	.2112821	.1545066
8.	Iowa	.2493315	.2007308
9.	Kansas	.1009952	.0685763
10.	Kentucky	.0562692	.0415244
11.	Louisiana	.2033273	.150664
12.	Maine	.2226316	.1724579
13.	Maryland	.1353406	.092455
14.	Massachusetts	.15637	.1093292
15.	Michigan	.1580957	.1075692
16.	Missouri	.1028805	.0704146
17.	New Jersey	.0958434	.0658799
18.	New York	.2778779	.2224942
19.	Ohio	.2291423	.1698186
20.	Pennsylvania	.1500667	.103029
21.	Texas	.2029787	.1455222
22.	Virginia	.1400013	.0967608
23.	Washington	.1104758	.0753325
24.	West Virginia	.1556139	.1123615
25.	Wisconsin	.1406707	.0970861

然后计算这两个模型预测结果的相关系数。

```
. corr u_h u_e
```

	u_h	u_e
u_h	1.0000	
u_e	0.9773	1.0000

由上表可知,半正态模型与指数模型对于  $u_i$  预测值的相关系数高达 0.98。这表明,从预测无效率项  $u_i$  的角度来看,无论使用何种模型,结果几乎相同,并不依赖于对  $u_i$  具体分布的假定。

虽然在样本容量较小的情况下,一般不宜估计含异方差的随机前沿模型<sup>①</sup>;但为了演示目的,下面估计允许  $u_i$  与  $v_i$  都有异方差的半正态模型,并假定异方差取决于变量 nfirm(企业个数)。

```
. frontier lnv lnk lnl,uhet(nfirm) vhet(nfirm) nolog
```

① 否则,可能导致对条件方差方程的参数估计不准确,参见 Stata 手册有关命令 frontier 的说明。

Stoc. frontier normal/half-normal model		Number of obs = 25					
		Wald chi2(2) = 2135.66		Prob > chi2 = 0.0000			
Log likelihood = 8.9688909							
lnv	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
lnv							
lnk	.2429007	.0540886	4.49	0.000	.136889 .3489125		
lnl	.9309447	.0667146	13.95	0.000	.8001864 1.061703		
_cons	1.537479	.1417207	10.85	0.000	1.259711 1.815246		
lnsig2v							
nfirm	-.1017466	.0362518	-2.81	0.005	-.1727988 -.0306944		
_cons	1.62307	1.891768	0.86	0.391	-2.084727 5.330867		
lnsig2u							
nfirm	.0055647	.0028622	1.94	0.052	-.0000451 .0111745		
_cons	-4.001382	.8248283	-4.85	0.000	-5.618016 -2.384748		

上表中部的结果表明,变量 nfir 对  $v_i$  的异方差有显著的负影响,  $p$  值为 0.005; 而上表底部的结果表明,变量 nfir 对  $u_i$  的异方差的影响也几乎在 5% 水平上显著( $p$  值为 0.052)。当然,由于样本容量太小,这些结果未必稳健或可信。

下面使用 Stata 提供的数据集 frontier2.dta 来演示断尾正态模型。该数据用于估计 1231 家企业的成本函数,主要变量包括 lncost(成本对数)、lnout(产出对数)、lnp\_k(资本价格对数)、lnp\_l(劳动价格对数)、quality(企业管理质量评级,取值为 1,2,3,4,5)。

```
. use frontier2,clear
. frontier lncost lnp_k lnp_l lnout, cost d(t) nolog
```

Stoc. frontier normal/truncated-normal model		Number of obs = 1231					
		Wald chi2(3) = 8.82		Prob > chi2 = 0.0318			
Log likelihood = -2386.2493							
lncost	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
lncost							
lnp_k	.3410717	.2363861	1.44	0.149	-.1222366 .80438		
lnp_l	.6608628	.4951499	1.33	0.182	-.3096131 1.631339		
lnout	.7528653	.3468968	2.17	0.030	.0729601 1.432771		
_cons	2.602609	1.083004	2.40	0.016	.4799595 4.725259		
/mu	1.095705	.881517	1.24	0.214	-.632037 2.823446		
/lnsigma2	1.5534	.1873464	8.29	0.000	1.186208 1.920592		
/ilgtgamma	1.257862	.2589522	4.86	0.000	.7503255 1.765399		
sigma2	4.727518	.8856833			3.274641 6.825001		
gamma	.7786579	.0446303			.6792496 .8538846		
sigma_u2	3.681119	.7503408			2.210478 5.15176		
sigma_v2	1.046399	.2660035			.5250413 1.567756		
H0: No inefficiency component:		z = 5.595		Prob>z = 0.000			

上表最后一行提供了 Coelli(1995) 基于 OLS 残差偏度的检验。检验结果表明,可以在 1% 水平上拒绝“没有无效率项”的原假设,即认为存在无效率项,应使用随机前沿模型。

对于断尾正态模型,也可通过选择项“cm(varlist)”来指定影响条件期望  $\mu_i$  的解释变量。假设  $\mu_i$  取决于企业  $i$  的管理质量 (quality)。由于变量 quality 分为五级,故可通过“i. quality”来

表示由 quality 所生产的四个虚拟变量<sup>①</sup>。相应的命令为

```
. frontier lncost lnp_k lnp_l lnout, cost d(t) cm (i.quality) nolog
```

Stoc. frontier normal/truncated-normal model				Number of obs = 1231	
				Wald chi2(3) = 9.31	Prob > chi2 = 0.0254
Log likelihood = -2382.3233					
lncost	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lncost					
lnp_k	.3611204	.2359749	1.53	0.126	-.1013819 .8236227
lnp_l	.680446	.4934935	1.38	0.168	-.2867835 1.647675
lnout	.7605533	.3466102	2.19	0.028	.0812098 1.439897
_cons	2.550769	1.078911	2.36	0.018	.4361417 4.665396
mu					
quality					
2	.5056067	.3382907	1.49	0.135	-.1574309 1.168644
3	.783223	.376807	2.08	0.038	.0446947 1.521751
4	.5577511	.3355061	1.66	0.096	-.0998288 1.215331
5	.6792882	.3428073	1.98	0.048	.0073981 1.351178
_cons	.6014025	.990167	0.61	0.544	-.1339289 2.542094
/lnsigma2	1.541784	.1790926	8.61	0.000	1.190769 1.892799
/ilgtgamma	1.242302	.2588968	4.80	0.000	.734874 1.749731
sigma2	4.67292	.8368852			3.289611 6.637923
gamma	.7759645	.0450075			.6758739 .8519189
sigma_u2	3.62602	.7139576			2.226689 5.025351
sigma_v2	1.0469	.2583469			.5405491 1.553251

上表中部的 mu 表示条件期望  $\mu$ , 其中“quality = 3”与“quality = 5”对应的虚拟变量均在 5% 水平上对条件期望有显著影响。

## 14.4 偶然断尾与样本选择

有时, 被解释变量  $y_i$  的断尾与另一变量  $z_i$  有关。这被称为“偶然断尾”( incidental truncation)或“样本选择”(sample selection)<sup>②</sup>, 称  $z_i$  为选择变量。

**例** 在美国的亚裔移民给人的整体印象是聪明能干。但在美国的亚裔并非亚洲人口的代表性样本。通常, 只有受过高等教育或具有吃苦冒险精神的亚裔才会“自我选择”(self selection)移民。此时, 决定移民与否的变量便对被解释变量产生了断尾作用。因此, “样本选择”将导致“选择性偏差”(selection bias)。

**例** 考察妇女的劳动力供给模型:

劳动时间方程  $hours = \alpha_0 + \alpha_1 wage + \alpha_2 children + \alpha_3 marriage + u$

工资方程  $w^o - w' = \beta_0 + \beta_1 age + \beta_2 education + \beta_3 children + \beta_4 location + v$

其中,  $w^o$  表示市场提供的工资 (offered wage),  $w'$  表示保留工资 (reservation wage)。如果  $w^o -$

<sup>①</sup> 这是 Stata 11 新增的功能。在 Stata 10 或更早的版本, 需手工生成这些虚拟变量。

<sup>②</sup> 也称为“Tobit II”模型。

$w^r < 0$ , 则该妇女将选择不工作, 也就无法观测到她的劳动时间(hours), 造成劳动时间方程的偶然断尾与样本选择问题。

为了纠正选择性偏差(selection bias), 首先必须搞清样本选择的具体形式, 即它究竟是怎么发生的, 然后再进行针对性的校正。

考虑二维正态随机向量 $(y, z)$ , 记期望为 $(\mu_y, \mu_z)$ , 标准差为 $(\sigma_y, \sigma_z)$ , 相关系数为 $\rho$ , 联合密度函数为 $f(y, z)$ 。假设决定个体是否有可能进入样本的“选择机制”(selection mechanism)为“选择变量 $z$ 大于某常数 $c$ ”。比如, 在妇女劳动力供给的例子中,  $z = w^o - w^r$ , 而 $c = 0$ 。显然, 断尾后的联合分布为

$$f(y, z | z > c) = \frac{f(y, z)}{\text{P}(z > c)} \quad (14.24)$$

将上式中的 $z$ 积分掉, 可以证明偶然断尾 $y$ 的条件期望为

$$\text{E}(y | z > c) = \mu_y + \rho \sigma_y \lambda[(c - \mu_z) / \sigma_z] \quad (14.25)$$

这个公式类似于普通断尾的公式(14.3), 其中 $\lambda(\cdot)$ 为反米尔斯比率(IMR)函数。显然, 如果 $\rho = 0$ ( $y$ 与 $z$ 相互独立), 则 $z$ 的选择过程并不对 $y$ 产生影响。如果 $\rho > 0$ (即 $y$ 与 $z$ 正相关), 则“ $z > c$ ”偶然断尾的结果是把 $y$ 的整个分布推向右边(因为 $\lambda(\cdot) > 0$ ), 从而使得条件期望 $\text{E}(y | z > c)$ 大于无条件期望 $\text{E}(y)$ 。类似地, 在“ $z < c$ ”条件下, 偶然断尾 $y$ 的条件期望为

$$\text{E}(y | z < c) = \mu_y - \rho \sigma_y \lambda[(\mu_z - c) / \sigma_z] \quad (14.26)$$

假设回归模型为 $y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i$  ( $i = 1, 2, \dots, n$ )。其中, 被解释变量 $y_i$ 是否可以观测取决于二值选择变量 $z_i$ (取值为0或1)

$$y_i = \begin{cases} \text{可观测, 若} & z_i = 1 \\ \text{不可观测, 若} & z_i = 0 \end{cases} \quad (14.27)$$

而决定二值变量 $z_i$ 的方程为

$$z_i = \begin{cases} 1, & \text{若 } z_i^* > 0 \\ 0, & \text{若 } z_i^* \leq 0 \end{cases} \quad (14.28)$$

$$z_i^* = \mathbf{w}'_i \boldsymbol{\gamma} + u_i \quad (14.29)$$

其中,  $z_i^*$ 为不可观测的潜变量。假设 $u_i$ 服从正态分布, 则 $z_i$ 为Probit模型, 故 $\text{P}(z_i = 1 | \mathbf{w}_i) = \Phi(\mathbf{w}'_i \boldsymbol{\gamma})$ 。可观测样本的条件期望为

$$\begin{aligned} \text{E}(y_i | y_i \text{可观测}) &= \text{E}(y_i | z_i^* > 0) = \text{E}(\mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i | \mathbf{w}'_i \boldsymbol{\gamma} + u_i > 0) \\ &= \text{E}(\mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i | u_i > -\mathbf{w}'_i \boldsymbol{\gamma}) = \mathbf{x}'_i \boldsymbol{\beta} + \text{E}(\varepsilon_i | u_i > -\mathbf{w}'_i \boldsymbol{\gamma}) \\ &= \mathbf{x}'_i \boldsymbol{\beta} + \rho \sigma_\varepsilon \lambda(-\mathbf{w}'_i \boldsymbol{\gamma}) \quad (\text{使用偶然断尾的期望公式}) \end{aligned}$$

其中,  $\text{E}(\varepsilon_i) = \text{E}(u_i) = 0$ , 并将Probit扰动项的标准差 $\sigma_u$ 标准化为1。显然, 如果直接用OLS来估计样本数据, 将遗漏非线性项 $\rho \sigma_\varepsilon \lambda(-\mathbf{w}'_i \boldsymbol{\gamma})$ 。一般来说,  $\mathbf{w}_i$ 与 $\mathbf{x}_i$ 相关(比如, 包含某个相同的变量), 故OLS将导致不一致的估计, 除非“ $\rho = 0$ ”(即 $y$ 与 $z$ 不相关)。考察解释变量 $x_{ik}$ 变动的边际效应可知

$$\frac{\partial \text{E}(y_i | z_i^* > 0)}{\partial x_{ik}} = \beta_k + \rho \sigma_\varepsilon \frac{\partial \lambda(-\mathbf{w}'_i \boldsymbol{\gamma})}{\partial x_{ik}} \quad (14.30)$$

其中, 右边第一项为 $x_{ik}$ 对 $y_i$ 的直接影响, 而第二项则是通过改变个体被选入样本的可能性而产生的间接影响(即选择性偏差)。

如果知道 $\boldsymbol{\gamma}$ , 就知道 $\lambda(-\mathbf{w}'_i \boldsymbol{\gamma})$ , 从而可以把它作为解释变量引入回归方程中。基于此,

Heckman(1979)提出“两步估计法”(two-step estimation),也称为“Heckit”<sup>①</sup>。

第一步:用 Probit 估计方程  $P(z_i = 1 | \mathbf{w}) = \Phi(\mathbf{w}' \boldsymbol{\gamma})$ , 得到估计值  $\hat{\boldsymbol{\gamma}}$ , 计算  $\hat{\lambda}(-\mathbf{w}' \hat{\boldsymbol{\gamma}})$ 。

第二步:用 OLS 回归  $y_i \xrightarrow{\text{OLS}} \mathbf{x}_i, \hat{\lambda}_i$ , 得到估计值  $\hat{\beta}, \hat{\rho}, \hat{\sigma}_e$ 。

另外一种更有效率的方法则是用 MLE 来估计该模型。在 Heckit 的两步估计法中,由于第一步的误差被带入第二步,故其效率不如 MLE 的整体估计。两步法的优点在于,操作简便;对于分布的假设也更弱(即使不假设二维正态分布,也可能成立)。

为了检验是否存在样本选择偏差(sample selection bias),可以使用似然比检验来检验原假设“ $H_0 : \rho = 0$ ”。如果使用 Heckit,则无法进行此似然比检验。另外,样本选择模型也被称为“Type II Tobit”。

样本选择模型的 Stata 命令为

heckman y x1 x2 x3, select(z1 z2)(默认使用 MLE, 选择方程的被解释变量为 y)

heckman y x1 x2 x3, select(z1 z2) twostep(两步法, 选择方程的被解释变量为 y)

heckman y x1 x2 x3, select(w = z1 z2)(默认使用 MLE, 选择方程的被解释变量为 w)

需要注意的是,对于没有观测值的  $y_i$ ,在 Stata 中应将其设为“缺失”(missing)。Stata 将缺失的观测值记为“.”。执行以上命令后,Stata 将汇报原回归方程与样本选择方程的系数估计值,并检验原假设“ $H_0 : \rho = 0$ ”。

下面再次以数据集 womenwk.dta 为例,但定义被解释变量为 lw(如果工作,则等于工资的对数;如果不工作,则为缺失值),进行如下“样本选择”MLE 回归:

```
. use womenwk.dta, clear
. heckman lw education age children, select (age married children
education) nolog
```

Heckman selection model (regression model with sample selection)		Number of obs		=	2000
		Censored obs		=	657
		Uncensored obs		=	1343
		Wald chi2(3)		=	454.78
Log likelihood = -1052.857		Prob > chi2		=	0.0000
lw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
education	.0397189	.0024525	16.20	0.000	.0349121 .0445256
age	.0075872	.0009748	7.78	0.000	.0056767 .0094977
children	-.0180477	.0064544	-2.80	0.005	-.0306981 -.0053973
_cons	2.305499	.0653024	35.30	0.000	2.177509 2.43349
select					
age	.0350233	.0042344	8.27	0.000	.0267241 .0433225
married	.4547724	.0735876	6.18	0.000	.3105434 .5990014
children	.4538372	.0288398	15.74	0.000	.3973122 .5103621
education	.0565136	.0110025	5.14	0.000	.0349492 .0780781
_cons	-2.478055	.1927823	-12.85	0.000	-2.855901 -2.100208
/athrho	.3377674	.1152251	2.93	0.003	.1119304 .5636045
/lnsigma	-1.375543	.0246873	-55.72	0.000	-1.423929 -1.327156
rho	.3254828	.1030183			.1114653 .5106469
sigma	.2527024	.0062385			.2407662 .2652304
lambda	.0822503	.0273475			.0286501 .1358505
LR test of indep. eqns. (rho = 0): chi2(1) = 5.53 Prob > chi2 = 0.0187					

<sup>①</sup> Heckman 因此于 2000 年获诺贝尔经济学奖。

上表底部的似然比检验显示,可以拒绝原假设“ $H_0 : \rho = 0$ ”,即应该使用样本选择模型。表中的“/athrho”指的是 $\rho$ 的双曲反正切(hyperbolic arctangent)函数,即 $\text{arctanh}(\rho) = \frac{1}{2}\ln[(1+\rho)/(1-\rho)]$ 。由于相关系数 $\rho$ 必须满足参数约束“ $|\rho| < 1$ ”,这给MLE的最大化问题带来不便。对其进行参数变换后(re-parameterization),新参数 $\text{arctanh}(\rho)$ 不再受任何约束。因此,可以先估计 $\text{arctanh}(\rho)$ ,然后根据MLE的不变性,反推出 $\hat{\rho}$ 。同理,“/lnsigma”指的是 $\ln\sigma_\epsilon$ 。由于标准差 $\sigma_\epsilon$ 必须满足 $\sigma_\epsilon > 0$ 的参数约束,故在参数变换后,首先估计 $\ln\sigma_\epsilon$ ,再反推出 $\hat{\sigma}_\epsilon$ 。

也可以使用Heckit的两步估计法:

```
. heckman lw education age children, select (age married children education) twostep nolog
```

		Heckman selection model -- two-step estimates		Number of obs		= 2000
		(regression model with sample selection)		Censored obs	=	657
				Uncensored obs	=	1343
				Wald chi2(3)		= 405.68
				Prob > chi2	=	0.0000
lw		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lw						
education		.0427067	.003106	13.75	0.000	.0366191 .0487944
age		.009322	.0014343	6.50	0.000	.0065108 .0121333
children		-.0019549	.0115202	-0.17	0.865	-.0245341 .0206242
_cons		2.124787	.1249789	17.00	0.000	1.879833 2.369741
select						
age		.0347211	.0042293	8.21	0.000	.0264318 .0430105
married		.4308575	.074208	5.81	0.000	.2854125 .5763025
children		.4473249	.0287417	15.56	0.000	.3909922 .5036576
education		.0583645	.0109742	5.32	0.000	.0368555 .0798735
_cons		-2.467365	.1925635	-12.81	0.000	-2.844782 -2.089948
mills						
lambda		.1822815	.0638285	2.86	0.004	.05718 .307383
rho		0.66698				
sigma		.27329216				

对比以上两表可知,这两种方法的估计结果很接近。

## 14.5 归并回归

受限被解释变量的另一种情形是,对于线性模型 $y_i = \mathbf{x}'\boldsymbol{\beta} + \varepsilon_i$ ,可能当 $y_i \geq c$ (或 $y_i \leq c$ )时,所有 $y_i$ 都被归并为 $c$ 。这种数据被称为“归并数据”(censored data)。

例(上不封顶的数据,top coding)在问卷调查中,常有诸如“收入在¥50 000及以上”这样的选项。对于这部分数据,我们只知道 $y_i \geq 50 000$ ,而不知道 $y_i$ 的具体取值。

例(边角解,corner solution)考虑买车的决定,并考察“买车开支”这个变量。如果不买车,则“买车开支”的最优解为边角解,即买车开支为0;反之,如果买车,则买车开支一定为正数。

例(边角解)考察“劳动时间”这个变量。对于失业或待业者而言,“劳动时间”的最优解

为边角解,即劳动时间为 0;而就业者的劳动时间一定为正数。

归并回归(censored regression)与断尾回归不同的是,虽然有全部的观测数据,但对于某些观测数据,被解释变量  $y_i$  被压缩在一个点上了。此时,  $y_i$  的概率分布就变成由一个离散点与一个连续分布所组成的混合分布(mixed distribution)。下面将说明,在这种情况下,如果用 OLS 来估计,无论使用的是整个样本,还是去掉离散点后的子样本,都不能得到一致的估计。

假设  $y_i^* = \mathbf{x}'\beta + \varepsilon_i$  ( $y_i^*$  不可观测), 扰动项  $\varepsilon_i | \mathbf{x}_i \sim N(0, \sigma^2)$ 。不失一般性, 为简便起见, 假定归并点为  $c = 0$ 。假设可以观测到  $y_i = \begin{cases} y_i^*, & \text{若 } y_i^* > 0, \\ 0, & \text{若 } y_i^* \leq 0 \end{cases}$ , 下面分别计算子样本的条件期望

$E(y_i | \mathbf{x}_i; y_i > 0)$ , 以及整个样本的条件期望  $E(y_i | \mathbf{x}_i)$ 。

对于满足条件“ $y_i > 0$ ”的子样本,

$$\begin{aligned} E(y_i | \mathbf{x}_i; y_i > 0) &= E(y_i^* | \mathbf{x}_i; y_i > 0) \quad (\text{给定 } y_i > 0, \text{ 必然 } y_i = y_i^*) \\ &= E(\mathbf{x}'\beta + \varepsilon_i | \mathbf{x}_i; y_i^* > 0) \\ &= \mathbf{x}'\beta + E(\varepsilon_i | \mathbf{x}_i; y_i^* > 0) \\ &= \mathbf{x}'\beta + E(\varepsilon_i | \mathbf{x}_i; \varepsilon_i > -\mathbf{x}'\beta) \\ &= \mathbf{x}'\beta + \sigma \cdot \lambda(-\mathbf{x}'\beta / \sigma) \quad (\text{使用正态分布的断尾公式, 且 } E(\varepsilon_i) = 0) \end{aligned}$$

因此, 在使用子样本进行回归时, 由于忽略了非线性项  $\sigma \cdot \lambda(-\mathbf{x}'\beta / \sigma)$ , 被纳入扰动项中, 导致扰动项与解释变量  $\mathbf{x}_i$  相关, 故 OLS 估计是不一致的。对于整个样本,

$$\begin{aligned} E(y_i | \mathbf{x}_i) &= 0 \cdot P(y_i = 0 | \mathbf{x}_i) + E(y_i | \mathbf{x}_i; y_i > 0) \cdot P(y_i > 0 | \mathbf{x}_i) \quad (14.31) \\ &= E(y_i^* | \mathbf{x}_i; y_i > 0) \cdot P(y_i > 0 | \mathbf{x}_i) \end{aligned}$$

其中,  $P(y_i > 0 | \mathbf{x}_i) = P(y_i^* > 0 | \mathbf{x}_i) = P(\mathbf{x}'\beta + \varepsilon_i > 0 | \mathbf{x}_i)$

$$\begin{aligned} &= P(\varepsilon_i > -\mathbf{x}'\beta | \mathbf{x}_i) = P\left(\frac{\varepsilon_i}{\sigma} > \frac{-\mathbf{x}'\beta}{\sigma} \middle| \mathbf{x}_i\right) \\ &= 1 - \Phi(-\mathbf{x}'\beta / \sigma) = \Phi(\mathbf{x}'\beta / \sigma) \quad (\text{标准正态以原点为对称}) \end{aligned}$$

因此,  $E(y_i | \mathbf{x}_i) = E(y_i^* | \mathbf{x}_i, y_i > 0) \cdot P(y_i > 0 | \mathbf{x}_i) = \Phi(\mathbf{x}'\beta / \sigma) [\mathbf{x}'\beta + \sigma \cdot \lambda(-\mathbf{x}'\beta / \sigma)]$  是解释变量  $\mathbf{x}_i$  的非线性函数。如果使用 OLS 对整个样本进行线性回归, 其非线性项将被纳入扰动项中, 导致不一致的估计。

Tobin(1958)提出用 MLE 估计这个模型, 该方法因此被称为“Tobit”, 也称为“Type I Tobit”或“归并回归”(censored regression)。在归并数据的情况下,  $y_i > 0$  时的概率密度依然不变, 仍为  $\frac{1}{\sigma} \phi((y_i - \mathbf{x}'\beta) / \sigma)$ ,  $\forall y_i > 0$ 。而  $y_i \leq 0$  时的分布却被挤到一个点“ $y_i = 0$ ”上了, 即  $P(y_i = 0 | \mathbf{x}) = 1 - P(y_i > 0 | \mathbf{x}) = 1 - \Phi(\mathbf{x}'\beta / \sigma)$ 。因此, 该混合分布的概率密度函数可以写为

$$f(y_i | \mathbf{x}) = [1 - \Phi(\mathbf{x}'\beta / \sigma)]^{1(y_i=0)} \left[ \frac{1}{\sigma} \phi((y_i - \mathbf{x}'\beta) / \sigma) \right]^{1(y_i>0)} \quad (14.32)$$

其中,  $\mathbf{1}(\cdot)$  为示性函数(indicator function), 即如果括号里的表达式为真, 取值为 1; 反之, 取值为 0。因此, 可以写出整个样本的似然函数, 然后使用 MLE 来估计。

归并回归的 Stata 命令为

`tobit y x1 x2 x3, ll(#) ul(#)`

其中, 选择项“ll(#)”表示 lower limit, 即左归并; 选择项“ul(#)”表示 upper limit, 即右归并; 如果同时使用这两个选择项, 则表示双边归并。如果把选择项“ll(#)”写为“ll”, 则 Stata 将把样本中  $y$  的最小值作为左归并点。类似地, 如果把选择项“ul(#)”写为“ul”, 则 Stata 将把样本中

$y$  的最大值作为右归并点。

Tobit 模型的一个缺陷是对分布的依赖性很强, 不够稳健。如果似然函数不正确, 比如扰动项不服从正态分布或存在异方差, 则 QMLE 估计就不一致。直观来看, 这是因为, 从方程(14.32)导出的 MLE 一阶条件是复杂的非线性函数。为此, 在使用 Tobit 模型时, 需要检验其正态性与同方差性。然而, 由于这是一个非线性模型, 故不能直接使用第 6 章的 JB 检验或第 7 章的异方差检验来检验正态性与同方差性。

为了检验正态性, 可将 Tobit 模型的 MLE 一阶条件视为某种矩条件, 并基于此进行“条件矩检验”(conditional moment test)(Skeels and Vella, 1999)。但 Drukker(2002)发现, 条件矩统计量的真实分布与渐近分布有相当差距, 存在较严重的显著性水平扭曲(size distortion), 故使用“参数自助法”(parametric bootstrap)来获得校正的临界值(有关显著性水平扭曲与自助法, 参见第 19 章)。条件矩检验可通过非官方命令 `tobcm` 来实现; 但该命令仅适用于左归并(left-censored), 且归并点为 0 的情形(这也是最常见的情形)。

```
net install tobcm.pkg① (下载安装命令 tobcm)
```

```
tobcm,pbs (对正态性假设进行条件矩检验)
```

其中, 选择项“`pbs`”表示使用参数自助法获得临界值。

为了检验同方差的原假设“ $H_0 : \sigma_i^2 = \sigma^2$ ”, 考虑替代假设“ $H_1 : \sigma_i^2 = \exp(z'_i\boldsymbol{\alpha})$ ”, 其中  $z_i$  通常等于解释变量  $x_i$ (但也可以不同)。然后, 可以通过辅助回归构建 LM 统计量来检验  $\boldsymbol{\alpha} = \mathbf{0}$ (类似于第 6 章第 7 节介绍的方法)。具体步骤及 Stata 操作, 参见 Cameron and Trivedi(2010, p. 550)。

如果发现扰动项并不服从正态分布或存在异方差, 解决方法之一为使用更稳健的“归并最小绝对离差法”(Censored Least Absolute Deviations, 简记 CLAD)(Powell, 1984)。其他解决方法包括上节的“样本选择模型”, 以及下节的“两部分模型”。

CLAD 法仅要求扰动项为 iid, 即使在非正态与异方差的情况下也能得到一致估计。而且, 在一定的正则条件下, 估计量服从渐近正态分布。首先, 将归并数据模型简洁地写为

$$y_i = \max(0, x'_i \boldsymbol{\beta} + \varepsilon_i) \quad (14.33)$$

即如果  $x'_i \boldsymbol{\beta} + \varepsilon_i \geq 0$ , 则  $y_i = x'_i \boldsymbol{\beta} + \varepsilon_i$ ; 反之,  $y_i = 0$ 。CLAD 法的目标函数为离差绝对值之和:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n |y_i - \max(0, x'_i \boldsymbol{\beta})| \quad (14.34)$$

选择  $\boldsymbol{\beta}$  使得离差绝对值之和最小化, 即可得到 CLAD 估计量  $\hat{\boldsymbol{\beta}}_{CLAD}$ , 可通过非官方 Stata 命令 `clad` 来实现。这实际上是中位数回归, 为分位数回归的特例(参见第 26 章)。 $\hat{\boldsymbol{\beta}}_{CLAD}$  的协方差矩阵可通过自助法得到(参见第 19 章)。

```
net install sg153.pkg② (下载安装命令 clad)
```

```
clad y x1 x2 x3, ll(#) ul(#) (进行 CLAD 估计)
```

其中, 选择项“`ll(#)`”与“`ul(#)`”分别表示左归并点与右归并点。

下面以美国妇女就业数据集 `womenwk.dta` 为例<sup>③</sup>, 估计一个工资决定模型。被解释变量为 `lwf`(equals log of wage if working and 0 if not working, 如果工作, 则等于工资的对数; 如果不工作, 则等于 0)。解释变量为 `age`(年龄), `married`(婚否), `children`(子女数), 与 `education`(教育年限)。

作为对照, 首先对整个样本进行 OLS 回归:

① 或输入命令“`findit tobcm`”寻找下载地址。

② 或输入命令“`findit clad`”寻找下载地址。

③ 此数据集曾用于第 11 章二值模型的估计。

```
. use womenwk.dta,clear  
. reg lwf age married children education,r  
. estimates store ols
```

Linear regression		Number of obs = 2000				
		F( 4, 1995) = 197.56				
		Prob > F = 0.0000				
		R-squared = 0.2120				
		Root MSE = 1.3218				
		<i>Robust</i>				
lwf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0363624	.0038582	9.42	0.000	.0287959	.0439288
married	.3188214	.0710831	4.49	0.000	.1794164	.4582263
children	.3305009	.0181054	18.25	0.000	.2949935	.3660084
education	.0843345	.010691	7.89	0.000	.0633677	.1053012
_cons	-1.077738	.1659689	-6.49	0.000	-1.403228	-.7522473

其次，画被解释变量 lwf 的直方图，结果如图 14.4。

. hist lwf

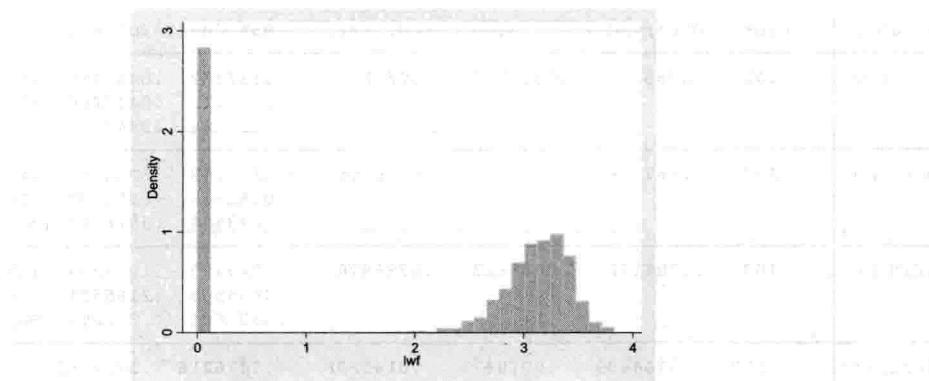


图 14.4 被解释变量 lwf 的直方图

图 14.4 显示，在相当多的观测点上，被解释变量 lwf 的取值为 0。为此，进行归并回归，假设左归并点为 0：

```
. tobit lwf age married children education, ll(0)  
. estimates store tobit
```

Tobit regression		Number of obs	=	2000
		LR chi2(4)	=	461.85
		Prob > chi2	=	0.0000
Log likelihood = -3349.9685		Pseudo R2	=	0.0645
lwf	Coef.	Std. Err.	t	P> t
age	.052157	.0057457	9.08	0.000
married	.4841801	.1035188	4.68	0.000
children	.4860021	.0317054	15.33	0.000
education	.1149492	.0150913	7.62	0.000
_cons	-2.807696	.2632565	-10.67	0.000
/sigma	1.872811	.040014		1.794337
Obs. summary:	657	left-censored observations at lwf<=0		
	1343	uncensored observations		
	0	right-censored observations		

下面,使用命令 `tobcm` 检验扰动项是否服从正态分布。

```
. tobcm,p
```

Conditional moment test against the null of normal errors			
critical values			
CM	%10	%5	%1
1370.2	5.16182	7.4033384	16.871058

结果显示,条件矩统计量高达 1 370.2,故强烈拒绝“扰动项服从正态分布”的原假设。为此,考虑在非正态情况下也成立的 CLAD 估计法。

```
. clad lwf age married children education, ll(0)
. estimates store clad
```

Initial sample size = 2000						
Final sample size = 2000						
Pseudo R2 = .06322683						
Bootstrap statistics						
Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]	
age	100	.0268424	.0012875	.0066561	.0136352 .017413 .016788	.0400496 .0448288 .0444136 (N) (P) (BC)
married	100	.1567166	.0158871	.0876595	-.0172189 .0442563 .043674	.3306521 .355791 .3516315 (N) (P) (BC)
children	100	.1206152	.0045623	.0388678	.043493 .0668565 .0631661	.1977374 .2165551 .215001 (N) (P) (BC)
education	100	.0764499	.0018672	.0145288	.0476216 .0554966 .0605122	.1052782 .1169837 .1173066 (N) (P) (BC)
const	100	.4908203	-.1189834	.5931536	-.6861252 -1.09028 -.978726	1.667766 1.235906 1.336596 (N) (P) (BC)
N = normal, P = percentile, BC = bias-corrected						

下面将 OLS,Tobit 与 CLAD 的估计结果列表比较。

```
. esttab ols tobit clad, mtitles se star(*0.1 **0.05 ***0.01)
```

	(1) ols	(2) tobit	(3) clad
main age	0.0364*** (0.00386)	0.0522*** (0.00575)	0.0356*** (0.00364)
married	0.319*** (0.0711)	0.484*** (0.104)	0.175*** (0.0650)
children	0.331*** (0.0181)	0.486*** (0.0317)	0.139*** (0.0199)
education	0.0843*** (0.0107)	0.115*** (0.0151)	0.0879*** (0.00959)
_cons	-1.078*** (0.166)	-2.808*** (0.263)	-0.122 (0.160)
sigma _cons		1.873*** (0.0400)	
N	2000	2000	2000
Standard errors in parentheses			
* p<0.1, ** p<0.05, *** p<0.01			

从上表可知,对于变量 married 与 children 而言,CLAD 估计值与 Tobit 估计值相差较大,尽管这两种估计法在变量显著性上没有质的不同。Wooldridge(2010,p.689)指出,如果 Tobit 模型的设定正确(比如,满足正态性与同方差性),则 CLAD 与 Tobit 的估计结果应相差不多。从这个角度来看,也可将 CLAD 的估计结果大致视为对 Tobit 模型的设定检验(specification test)。在本例中,由于 CLAD 估计值与 Tobit 相差较大,故可认为 Tobit 模型的设定有误,并倾向于使用 CLAD 模型。

## 14.6 归并数据的两部分模型

上节提到,如果归并数据的扰动项不服从正态分布或存在异方差,则 Tobit MLE 估计是不一致的。Tobit 模型的特征是,被解释变量为 0 或正数的产生机制是一样的;但此假定未必成立。正如第 13 章零膨胀泊松回归的情形,有时可将决策分为两个阶段,并允许两个阶段拥有不同的决定机制。比如,在第一阶段决定是否参加工作(或购买耐用消费品),称为“参与决策”(participation decision);而在第二阶段决定工作时间(或购买耐用消费品的开支),称为“数量决策”(amount decision)。

为此,作为 Tobit 模型的推广,Cragg(1971)提出了以下“两部分模型”(two-part model)。首先,定义虚拟变量  $d$ ,将样本中的活动参与者(比如,工作者或购买者)记为  $d = 1$ ,非参与者记为  $d = 0$ 。对于参与者,可以观测到  $y > 0$ ;而对于非参与者,只能观测到  $y = 0$ 。对于非参与者而言,只能估计  $P(d = 0)$ ,即“ $d = 0$ ”的概率。对于参与者而言,给定  $d = 1$  的条件密度可写为  $f(y|d=1) = P(d=1|x)f(y|d=1,x)$ 。因此,关于  $y$  的两部分模型为:

$$f(y|x) = \begin{cases} P(d=0|x) & \text{若 } y=0 \\ P(d=1|x)f(y|d=1,x) & \text{若 } y>0 \end{cases} \quad (14.35)$$

其中,为了保证参与者的 $y$ 为正数, $f(y|d=1,x)$ 应为某取值为正的随机变量的密度函数,比如对数正态分布(log-normal)。为此,在实践中常把 $\ln y$ 作为第二部分的被解释变量。显然,对于第一阶段的参与决策 $d$ ,可以用Probit或Logit来建模。从潜变量的角度来看,如果 $x'_i\beta + \varepsilon_i$ 超过了0,则 $d=1$ 。因此,这个模型也称为“跨栏模型”(hurdle model),因为只有跨过这个栏(hurdle)或门槛,才能参与活动。

对于两部分模型,通常假设其两个部分相互独立,故可分别进行估计(如果不相互独立,则应使用本章第4节的样本选择模型)。对于第一部分的二值选择模型,可以使用全样本进行Probit或Logit估计;对于第二部分的模型,则可使用由参与者组成的子样本进行OLS估计。通常,第一部分与第二部分模型所使用的解释变量完全相同;但也可以不同(这正是两部分模型的灵活性)。

继续以上节的美国妇女就业数据集为例。

```
. use womenwk.dta, clear
```

对于第一部分模型,首先生成一个被解释的虚拟变量,然后使用全样本数据进行Probit回归。

```
. gen lwd = (lwf > 0)
. probit lwd age married children education, nolog
```

Probit regression						Number of obs	=	2000
						LR chi2(4)	=	478.32
						Prob > chi2	=	0.0000
						Pseudo R2	=	0.1889
Log likelihood = -1027.0616								
lwd	Coef.	Std. Err.	z	P> z		[95% Conf. Interval]		
age	.0347211	.0042293	8.21	0.000		.0264318	.0430105	
married	.4308575	.074208	5.81	0.000		.2854125	.5763025	
children	.4473249	.0287417	15.56	0.000		.3909922	.5036576	
education	.0583645	.0109742	5.32	0.000		.0368555	.0798735	
_cons	-2.467365	.1925635	-12.81	0.000		-2.844782	-2.089948	

上表显示,各解释变量对于参加工作的概率都有显著的正影响。对于第二部分模型,则仅用参加工作者的子样本进行OLS线性回归。

```
. reg lwf age married children education if lwd = 1, r
```

Linear regression						Number of obs	=	1343
						F( 4, 1338)	=	123.87
						Prob > F	=	0.0000
						R-squared	=	0.2570
						Root MSE	=	.24733
lwf	Coef.	Std. Err.	t	P> t		[95% Conf. Interval]		
age	.0066617	.000873	7.63	0.000		.0049492	.0083743	
married	-.0298994	.016201	-1.85	0.065		-.0616815	.0018826	
children	-.0321762	.0048209	-6.67	0.000		-.0416336	-.0227188	
education	.0380189	.0021444	17.73	0.000		.0338121	.0422257	
_cons	2.448392	.0426429	57.42	0.000		2.364738	2.532046	

上表显示,两部分模型与Tobit模型估计结果的最大不同是,变量married与children的系数

显著地为负(前者在 10% 水平上显著)。这表明,虽然已婚与孩子数量能提高进入就业市场的概率,但却会降低进入就业市场后的工资水平。在两部分模型中,是否依然存在非正态扰动项或异方差?由于这是线性回归模型,故可进行常规的异方差检验。

```
. qui reg lwf age married children education if lwd = 1
. estat imtest,white
```

```
White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity
```

```
chi2(13)      =      45.78
Prob > chi2   =     0.0000
```

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	45.78	13	0.0000
Skewness	37.56	4	0.0000
Kurtosis	6.75	1	0.0094
Total	90.09	18	0.0000

上表显示,BP 检验的结果强烈拒绝同方差的原假设,即存在异方差。为了检验扰动项的正态性,首先生成残差。

```
. predict lwf_resid,resid
. sktest lwf_resid
```

Variable	Obs	Skewness/Kurtosis tests for Normality			Prob>chi2
		Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	
lwf_resid	2.0e+03	0.0000	.	.	.

结果也强烈地拒绝了正态性的原假设。更直观地,可以画残差的核密度图,结果见图 14.5。

```
. kdensity lwf_resid
```

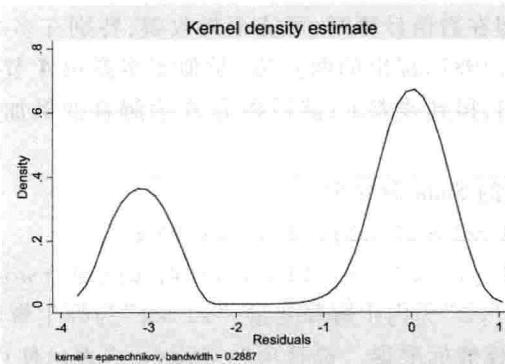


图 14.5 两部分模型残差的核密度图

从图 14.5 可以看出,扰动项的分布明显地存在双峰,与正态分布相去甚远。然而,对于线性回归模型,非正态性与异方差并不影响估计量的一致性与渐近正态性,只要使用异方差稳健的标

准误即可。

## 14.7 含内生解释变量的 Tobit 模型(选读)

对于归并数据使用 Tobit 模型,有时会遇到解释变量为内生变量的情形。不失一般性,假设在 0 处存在左边归并。考虑以下模型:

$$y_{1i}^* = \mathbf{x}'_i \boldsymbol{\delta} + \beta y_{2i} + u_i \quad (14.36)$$

$$y_{2i} = \mathbf{x}'_i \boldsymbol{\gamma}_1 + \mathbf{z}'_i \boldsymbol{\gamma}_2 + v_i \quad (14.37)$$

$$y_{1i} = \begin{cases} y_{1i}^* & \text{若 } y_{1i}^* > 0 \\ 0 & \text{若 } y_{1i}^* = 0 \end{cases} \quad (14.38)$$

其中, $y_{1i}$ 为可观测的被解释变量, $y_{1i}^*$ 为不可观测的潜变量, $y_{2i}$ 是模型中唯一的内生解释变量(以下方法可推广到多个内生解释变量的情形)。含内生变量 Tobit 模型的估计方法类似于含内生变量的 Probit 模型(IV Probit,参见第 11 章)。假设扰动项( $u_i, v_i$ )服从期望为 0 的二维正态分布:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \alpha \sigma_v \\ \alpha \sigma_v & \sigma_v^2 \end{pmatrix} \right] \quad (14.39)$$

其中, $u_i$ 的方差被标准化为 1,而  $\alpha$  为  $(u_i, v_i)$  的相关系数。显然,由于  $v_i$  服从正态分布,故  $y_{2i}$ 也服从正态分布,因此  $y_{2i}$ 必须是一个连续变量。进一步,假设  $(u_i, v_i)$  独立于  $\mathbf{x}_i$  与  $\mathbf{z}_i$ ,故在方程(14.36)中, $\mathbf{x}_i$  为外生解释变量。而且, $\mathbf{z}_i$  可以作为方程(14.36)中内生变量  $y_{2i}$  的工具变量,因为  $\mathbf{z}_i$  与内生变量  $y_{2i}$  相关(参见方程(14.37)),并且  $\mathbf{z}_i$  与  $u_i$  无关。在此模型中, $y_{2i}$  的内生性完全来自于  $u_i$  与  $v_i$  的相关性;如果二者的相关系数  $\alpha=0$ (Stata 记为 alpha),则  $y_{2i}$  为外生变量。因此,对于  $y_{2i}$  内生性的检验可通过检验“ $H_0 : \alpha=0$ ”来进行。

对于由方程(14.36)~(14.39)构成的模型,在给定  $\mathbf{x}_i$  与  $\mathbf{z}_i$  的情况下,( $y_{1i}, y_{2i}$ )的条件概率分布已完全确定。将联合概率密度  $f(y_{1i}, y_{2i} | \mathbf{x}_i, \mathbf{z}_i)$  分解为  $f(y_{1i} | y_{2i}, \mathbf{x}_i, \mathbf{z}_i)f(y_{2i} | \mathbf{x}_i, \mathbf{z}_i)$ ,则可写出样本数据( $y_{1i}, y_{2i}$ )的似然函数,进行最有效率的 MLE 估计。这种方法称为“工具变量 Tobit”(Instrumental Variable Tobit,简记 IV Tobit)。

尽管 MLE 最有效率,但在数值计算时,可能不易收敛,特别在多个内生解释变量的情形下。为此,可以考虑使用 Newey(1987)提出的两步法(类似于本章第 4 节介绍的两步法),即首先对方程(14.37)进行 OLS 回归,得到残差  $\hat{v}_i$ ;然后将  $\hat{v}_i$  作为解释变量加入到方程(14.36)中,进行 Tobit 估计。

含内生变量 Tobit 模型的 Stata 命令为

```
. ivtobit y1 x1 x2 (y2 = z1 z2), ll(#) ul(#) r  
. ivtobit y1 x1 x2 (y2 = z1 z2), ll(#) ul(#) first twostep
```

其中,“y1”为被解释变量,“y2”为内生解释变量,“x1 x2”为外生解释变量,而“z1 z2”为工具变量;选择项“r”表示使用稳健标准误。选择项“twostep”表示使用两步法,而默认进行 MLE 估计;选择项“first”表示显示第一步回归的结果。如果把选择项“ll(#)”写为“ll”,则 Stata 将把样本中“y1”的最小值作为左归并点。类似地,如果把选择项“ul(#)”写为“ul”,则 Stata 将把样本中“y1”的最大值作为右归并点。

下面,以 Stata 提供的数据集 laborsup.dta 为例。被解释变量为 fem\_inc(妇女收入),解释变

量包括 `fem_educ`(妇女教育水平), `kids`(孩子数量), 以及 `other_inc`(其他家庭收入)。假设如果妇女不工作, 则可领取 \$10 000 福利与儿童赡养费。因此, 妇女收入肯定大于或等于 \$10 000, 故在 \$10 000 处存在左边归并。另一方面, `other_inc`(其他家庭收入)可能为内生变量, 我们使用 `male_inc`(丈夫收入)作为工具变量。

```
. use laborsup.dta
```

首先, 进行 IV Tobit 估计。

```
. ivtobit fem_inc fem_educ kids (other_inc = male_educ), ll nolog
```

		Number of obs = 500			
		Wald chi2(3) = 117.42			
		Prob > chi2 = 0.0000			
Log likelihood = -3226.0845					
		Coef.	Std. Err.	z	P> z  [95% Conf. Interval]
other_inc		-0.9045399	.1329762	-6.80	0.000 -1.165168 -.6439114
fem_educ		3.272391	.3968708	8.25	0.000 2.494538 4.050243
kids		-3.312357	.7218628	-4.59	0.000 -4.727182 -1.897532
_cons		19.24735	7.372391	2.61	0.009 4.797725 33.69697
/alpha		.2907654	.1379965	2.11	0.035 .0202972 .5612336
/lns		2.874031	.0506672	56.72	0.000 2.774725 2.973337
/lnv		2.813383	.0316228	88.97	0.000 2.751404 2.875363
s		17.70826	.897228		16.03422 19.55707
v		16.66621	.5270318		15.66461 17.73186
Instrumented: other_inc					
Instruments: fem_educ kids male_educ					
Wald test of exogeneity (/alpha = 0): chi2(1) = 4.44 Prob > chi2 = 0.0351					
Obs. summary: 272 left-censored observations at fem_inc<=10 228 uncensored observations 0 right-censored observations					

上表底部的沃尔德检验结果表明, 可以拒绝 “ $\alpha=0$ ” 的外生性原假设, 即认为存在内生变量。其次, 进行两步法估计, 并显示第一步回归的结果。

```
. ivtobit fem_inc fem_educ kids (other_inc = male_educ), ll  
first twostep
```

First-stage regression					
Source	SS	df	MS	Number of obs = 500	
Model	28864.2732	3	9621.42439	F( 3, 496) = 34.36	
Residual	138881.269	496	280.002558	Prob > F = 0.0000	
Total	167745.542	499	336.163411	R-squared = 0.1721	
				Adj R-squared = 0.1671	
				Root MSE = 16.733	
other_inc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
male_educ	2.845253	.2838838	10.02	0.000	2.28749 3.403016
fem_educ	.3351866	.2837344	1.18	0.238	-.2222829 .8926562
kids	.8329056	.5497701	1.52	0.130	-.2472597 1.913071
_cons	9.872562	5.049432	1.96	0.051	-.0483506 19.79347

在第一步回归中,变量 male\_educ 的系数显著为正,整个方程的  $F$  值为 34.36,故不存在弱工具变量。

Two-step tobit with endogenous regressors						Number of obs = 500
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
other_inc	-0.9045397	.1330015	-6.80	0.000	-1.165218	-.6438616
fem_educ	3.27239	.3969399	8.24	0.000	2.494402	4.050378
kids	-3.312356	.7220066	-4.59	0.000	-4.727463	-1.897249
_cons	19.24735	7.37392	2.61	0.009	4.794728	33.69997

Instrumented:	other_inc
Instruments:	fem_educ kids male_educ
<hr/>	
Wald test of exogeneity:	chi2(1) = 4.64
Prob > chi2 = 0.0312	
<hr/>	
Obs. summary:	272 left-censored observations at fem_inc<=10
	228 uncensored observations
	0 right-censored observations

上表底部的 Wald 检验结果再次拒绝了外生性的原假设。经目测可知,两步法的估计值与 IV Tobit 十分接近。

## 习 题

**14.1** 数据集 tobacco.dta<sup>①</sup> 包含了 1995 年 2 724 个比利时家庭的以下变量:share2(budget share of tobacco, 烟草占预算比重),age(年龄),lnx(家庭开支的对数),nadults(number of adults in household,家庭中成年人数),nkids(number of kids > 2 year old,两岁以上儿童数),nkids2(number of kids≤2 year old,两岁及以下儿童数)。考虑以下恩格尔函数(用于烟草的开支比重随着总开支而变化):

$$\text{share2}_i = \beta_0 + \beta_1 \ln x_i + \beta_2 \text{age}_i + \beta_3 \text{nadults}_i + \beta_4 \text{nkids}_i + \beta_5 \text{nkids2}_i + \varepsilon_i$$

- (1) 进行 OLS 回归;
- (2) 假设在被解释变量“share2 = 0”处存在左边断尾,进行断尾回归;
- (3) 假设在被解释变量“share2 = 0”处存在左边归并,进行归并回归(Tobit);
- (4) 将被解释变量“share2 = 0”的取值定义为“缺失”,然后用 MLE 进行样本选择回归,选择方程的解释变量为“age”,“nadults”,“nkids”与“nkids2”(选择是否抽烟,可能不取决于收入,而取决于对自己及家人健康的考虑)。
- (5) 恩格尔定律适用于烟草吗(即用于烟草的开支比重随着总开支的增加而下降)?

**14.2** 重复上题,但把被解释变量改为“share1”(budget share of alcohol, 酒精占预算比重)。

## 附 录

标准正态的断尾期望公式

证明:假设  $y \sim N(0,1)$ ,  $c$  为任意常数,则

<sup>①</sup> 此数据集来自 Verbeek(2004)。

$$\begin{aligned}
 E(y|y>c) &= \int_c^\infty y \frac{\phi(y)}{1-\Phi(c)} dy \\
 &= \frac{1}{1-\Phi(c)} \int_c^\infty y \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
 &= \frac{1}{1-\Phi(c)} \frac{1}{\sqrt{2\pi}} \int_c^\infty e^{-y^2/2} d(y^2/2) \\
 &= \frac{1}{1-\Phi(c)} \frac{1}{\sqrt{2\pi}} \int_{c^2/2}^\infty e^{-z} dz \\
 &= \frac{1}{1-\Phi(c)} \left[ \frac{1}{\sqrt{2\pi}} e^{-c^2/2} \right] \\
 &= \frac{\phi(c)}{1-\Phi(c)}
 \end{aligned}$$

在上面的推导中, 使用了积分变换  $z=y^2/2$ 。

通过这个简单的推导, 我们可以知道分布参数对分布的影响。对于正态分布, 其均值和方差都是重要的参数。

### 从概率论到数理统计学 (上)

	$\mu = 0$	$\mu = 1$	$\mu = -1$	$\mu = 2$	$\mu = -2$
$\sigma = 1$					
$\sigma = 2$					
$\sigma = 3$					

通过观察我们可以看到, 正态分布的形状由两个参数决定, 即均值  $\mu$  和标准差  $\sigma$ 。

均值  $\mu$  表示分布的中心位置, 而标准差  $\sigma$  表示分布的离散程度。

当  $\mu=0$  时, 分布称为零均值正态分布 (mean-zero normal distribution)。

当  $\sigma=1$  时, 分布称为单位正态分布 (unit normal distribution)。

当  $\mu=1$  时, 分布称为右偏态分布 (right-skewed distribution)。

当  $\mu=-1$  时, 分布称为左偏态分布 (left-skewed distribution)。

当  $\mu=2$  时, 分布称为高均值分布 (high-mean distribution)。

当  $\mu=-2$  时, 分布称为低均值分布 (low-mean distribution)。

当  $\sigma=2$  时, 分布称为宽分布 (wide distribution)。

当  $\sigma=3$  时, 分布称为非常宽分布 (extremely wide distribution)。

因此, 正态分布的形状完全由两个参数决定, 即均值  $\mu$  和标准差  $\sigma$ 。

当  $\mu=0$  且  $\sigma=1$  时, 分布称为标准正态分布 (standard normal distribution)。

当  $\mu=1$  且  $\sigma=2$  时, 分布称为右偏态分布 (right-skewed distribution)。

当  $\mu=-1$  且  $\sigma=2$  时, 分布称为左偏态分布 (left-skewed distribution)。

当  $\mu=2$  且  $\sigma=3$  时, 分布称为高均值分布 (high-mean distribution)。

当  $\mu=-2$  且  $\sigma=3$  时, 分布称为低均值分布 (low-mean distribution)。

因此, 正态分布的形状完全由两个参数决定, 即均值  $\mu$  和标准差  $\sigma$ 。

因此, 正态分布的形状完全由两个参数决定, 即均值  $\mu$  和标准差  $\sigma$ 。

# 第 15 章 短 面 板

## 15.1 面板数据的特点

面板数据 (panel data 或 longitudinal data, 也译为“平行数据”), 指的是在一段时间内跟踪同一组个体 (individual) 的数据。它既有横截面的维度 ( $n$  位个体), 又有时间维度 ( $T$  个时期)。一个  $T = 3$  的面板数据结构如表 15.1<sup>①</sup>。

表 15.1 面板数据的结构(长形排列)

	$y$	$x_1$	$x_2$	$x_3$
Individual 1: $t = 1$				
Individual 1: $t = 2$				
Individual 1: $t = 3$				
.....				
Individual $n$ : $t = 1$				
Individual $n$ : $t = 2$				
Individual $n$ : $t = 3$				

通常的面板数据  $T$  较小, 而  $n$  较大, 在使用大样本理论时让  $n$  趋于无穷大。这种面板数据被称为“短面板”(short panel)。反之, 如果  $T$  较大, 而  $n$  较小, 则被称为“长面板”(long panel)。

在面板模型中, 如果解释变量包含被解释变量的滞后值, 则称为“动态面板”(dynamic panel); 反之, 则称为“静态面板”(static panel)。本章关注静态的短面板, 第 16 章讨论长面板与动态面板, 第 17 章考虑非线性面板。

如果在面板数据中, 每个时期在样本中的个体完全一样, 则称为“平衡面板数据”(balanced panel); 反之, 则称为“非平衡面板数据”(unbalanced panel)。我们主要关注平衡面板, 但将在本章第 10 节讨论非平衡面板。

面板数据的主要优点如下。

(1) 可以解决遗漏变量问题: 遗漏变量偏差是一个普遍存在的问题。虽然可以用工具变量法解决, 但有效的工具变量常常很难找。遗漏变量常常是由于不可观测的个体差异或“异质性”(heterogeneity)造成的, 如果这种个体差异“不随时间而改变”(time invariant), 则面板数据提供了解决遗漏变量问题的又一利器。

(2) 提供更多个体动态行为的信息: 由于面板数据同时有横截面与时间两个维度, 有时它可以解决单独的截面数据或时间序列数据所不能解决的问题。比如, 考虑如何区分规模效应与

<sup>①</sup> 也可将数据排成“宽形”(wide form), 参见第 23 章。一般以“长形”(long form) 排列最为常见。

技术进步对企业生产效率的影响。对于截面数据来说,由于没有时间维度,故无法观测到技术进步。然而,对于单个企业的时间序列数据来说,我们无法区分其生产效率的提高究竟有多少是由于规模扩大,有多少是由于技术进步。又比如,对于失业问题,截面数据能告诉我们在某个时点上哪些人失业,而时间序列数据能告诉我们某个人就业与失业的历史,但这两种数据均无法告诉我们是否失业的总是同一批人(意味着低流转率,low turnover rate),还是失业的人群总在变动(意味着高流转率,high turnover rate)。如果有面板数据,就可能解决上述问题。

(3) 样本容量较大:由于同时有截面维度与时间维度,通常面板数据的样本容量更大,从而可以提高估计的精确度。

当然,面板数据也会带来一些问题,比如,样本数据通常不满足独立同分布的假定,因为同一个体在不同期的扰动项一般存在自相关。另外,面板数据的收集成本通常较高,不易获得。

## 15.2 面板数据的估计策略

估计面板数据的一个极端策略是将其看成截面数据而进行混合回归(pooled regression),即要求样本中每个个体都拥有完全相同的回归方程。另一极端策略则是,为每个个体估计一个单独的回归方程。前者忽略了个体间不可观测或被遗漏的异质性,而该异质性可能与解释变量相关从而导致估计不一致。后者则忽略了个体间的共性,也可能没有足够大的样本容量。因此,在实践中常采用折中的估计策略,即假定个体的回归方程拥有相同的斜率,但可以有不同的截距项,以此来捕捉异质性,参见图 15.1。这种模型被称为“个体效应模型”(individual-specific effects model),即

$$y_{it} = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \boldsymbol{\delta} + u_i + \varepsilon_{it} \quad (i=1, \dots, n; t=1, \dots, T) \quad (15.1)$$

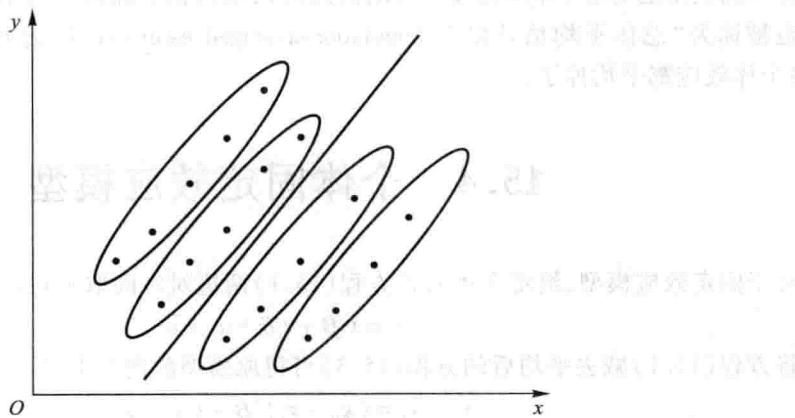


图 15.1 面板数据中不同个体的截距项可以不同

其中, $\mathbf{z}_i$  为不随时间而变(time invariant)的个体特征(即  $\mathbf{z}_{it} = \mathbf{z}_i, \forall t$ ), 比如性别; 而  $\mathbf{x}_i$  可以随个体及时间而变(time-varying)。扰动项由  $(u_i + \varepsilon_{it})$  两部分构成, 称为“复合扰动项”(composite error term), 而方程(15.1)也称为“复合扰动项模型”(error components model)。其中, 不可观测的随机变量  $u_i$  是代表个体异质性的截距项, 故方程(15.1)也称为“不可观测效应模型”(unobserved effects model)。在较早的文献中有时将  $u_i$  视为常数, 但这也只是随机变量的特例, 即退化的随机

变量。 $\varepsilon_{it}$  为随个体与时间而改变的扰动项。假设  $\{\varepsilon_{it}\}$  为独立同分布 (在下一章的长面板中将放松此假定), 且与  $u_i$  不相关。

如果  $u_i$  与某个解释变量相关, 则进一步称之为“固定效应模型”( Fixed Effects Model, 简记 FE)<sup>①</sup>。在这种情况下, OLS 是不一致的。解决的方法是将模型转换, 消去  $u_i$  后获得一致估计量。

如果  $u_i$  与所有解释变量 ( $x_{it}, z_i$ ) 均不相关, 则进一步称之为“随机效应模型”( Random Effects Model, 简记 RE)。从经济理论的角度来看, 随机效应模型比较少见<sup>②</sup>, 但仍须通过数据来检验究竟该用随机效应还是固定效应模型。

显然, 与截面数据相比, 面板数据提供了更为丰富的模型与估计方法。

## 15.3 混合回归

如果所有个体都拥有完全一样的回归方程, 则方程 (15.1) 可以写为

$$y_{it} = \alpha + x_{it}' \beta + z_i' \delta + \varepsilon_{it} \quad (15.2)$$

其中,  $x_{it}$  不包括常数项。这样, 就可以把所有的数据放在一起, 像对待横截面数据那样进行 OLS 回归, 故被称为“混合回归”( pooled regression)。由于面板数据的特点, 虽然通常可以假设不同个体之间的扰动项相互独立, 但同一个体在不同时期的扰动项之间往往存在自相关。此时, 对标准误的估计应该使用聚类稳健的标准误 (cluster-robust standard error), 而所谓聚类 (cluster), 就是由每个个体不同时期的所有观测值所组成。同一聚类 (个体) 的观测值允许存在相关性, 而不同聚类 (个体) 的观测值则不相关。

混合回归的基本假设是不存在个体效应。对于这个假设必须进行统计检验。由于个体效应以两种不同的形态存在 (即固定效应与随机效应), 故将在下面两节分别介绍其检验方法。混合回归也被称为“总体平均估计量”( Population-averaged estimator, 简记 PA), 因为可以把它理解为, 将个体效应都平均掉了。

## 15.4 个体固定效应模型

对于固定效应模型, 给定个体  $i$ , 将方程 (15.1) 两边对时间取平均可得

$$\bar{y}_i = \bar{x}_i' \beta + z_i' \delta + u_i + \bar{\varepsilon}_i \quad (15.3)$$

将方程 (15.1) 减去平均后的方程 (15.3) 可得原模型的离差形式<sup>③</sup>

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (15.4)$$

定义  $\tilde{y}_{it} \equiv y_{it} - \bar{y}_i$ ,  $\tilde{x}_{it} \equiv x_{it} - \bar{x}_i$ ,  $\tilde{\varepsilon}_{it} \equiv \varepsilon_{it} - \bar{\varepsilon}_i$ , 则

$$\tilde{y}_{it} = \tilde{x}_{it}' \beta + \tilde{\varepsilon}_{it} \quad (15.5)$$

<sup>①</sup> 固定效应 (fixed effects) 这个名词有时易引起误解。即使在固定效应模型中, 个体效应  $u_i$  也是随机的 (尽管其取值不随时间而变), 而非固定 (fixed) 的常数。但这已是大家都接受的传统称呼。

<sup>②</sup> 一般来说, 不可观测的异质性通常会对解释变量有影响。

<sup>③</sup> 这种变换被称为“mean-differencing”或“time demeaning”。

由于上式中已将  $u_i$  消去, 故只要  $\hat{\varepsilon}_i$  与  $\hat{x}_i$  不相关, 则可以用 OLS 一致地估计  $\beta$ , 称为“固定效应估计量”(Fixed Effects Estimator), 记为  $\hat{\beta}_{FE}$ 。由于  $\hat{\beta}_{FE}$  主要使用了每位个体的组内离差信息, 故也称为“组内估计量”(within estimator)。即使个体特征  $u_i$  与解释变量  $x_i$  相关, 只要使用组内估计量, 就可以得到一致估计, 这是面板数据的一大优势。然而, 在作离差转换的过程中,  $z'_i \delta$  也被消掉了, 故无法估计  $\delta$ 。也就是说,  $\hat{\beta}_{FE}$  无法估计不随时间而变的变量之影响, 这是 FE 的一大缺点。另外, 为了保证  $(\varepsilon_i - \bar{\varepsilon}_i)$  与  $(x_i - \bar{x}_i)$  不相关, 则要求第  $i$  个观测值满足严格外生性, 即  $E(\varepsilon_i | x_{i1}, \dots, x_{iT}) = 0$ , 因为  $\bar{x}_i$  中包含了所有  $(x_{i1}, \dots, x_{iT})$  的信息。换言之, 扰动项必须与各期的解释变量均不相关(而不仅仅是当期的解释变量), 这是一个比较强的假定。

如果在原方程中引入  $(n-1)$  个虚拟变量(如果没有截距项, 则引入  $n$  个虚拟变量)来代表不同的个体, 则可以得到与上述离差模型同样的结果。因此, FE 也被称为“最小二乘虚拟变量模型”(Least Square Dummy Variable Model, 简记 LSDV)。为什么二者的结果相同? 直观来说, 这正如线性回归与离差形式的回归在某种意义上是等价的<sup>①</sup>。比如,

$$y_i = \alpha + \beta x_i + \varepsilon_i \Leftrightarrow y_i - \bar{y} = \beta(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon}) \quad (15.6)$$

但是, 如果作完 LSDV 后发现某些个体的虚拟变量不显著而将其删去, 那么 LSDV 的结果就不会与 FE 相同。使用 LSDV 的好处是可以得到对个体异质性  $u_i$  的估计。LSDV 法的缺点是, 如果  $n$  很大, 则须在回归方程中引入很多虚拟变量, 可能超出计量软件(比如 Stata)所允许的解释变量个数。

## 15.5 时间固定效应

以上个体固定效应模型解决了不随时间而变(time invariant)但随个体而异的遗漏变量问题。类似地, 引入时间固定效应, 则可以解决不随个体而变(individual invariant)但随时间而变(time varying)的遗漏变量问题。假设模型为

$$y_{it} = x'_{it} \beta + z'_i \delta + \gamma S_t + u_i + \varepsilon_{it} \quad (15.7)$$

其中,  $S_t$  不可观测。定义  $\lambda_t \equiv \gamma S_t$ , 则上式可以写为

$$y_{it} = x'_{it} \beta + z'_i \delta + \lambda_t + u_i + \varepsilon_{it} \quad (15.8)$$

在上式中, 可将  $\lambda_t$  视为第  $t$  期独有的截距项, 并将其解释为“第  $t$  期”对被解释变量  $y$  的效应。因此, 这些  $\lambda_1, \dots, \lambda_T$  称为“时间固定效应”(time fixed effects)。显然, 对于方程(15.8), 可以使用 LSDV 法来估计, 即对每个时期定义一个虚拟变量, 然后把  $(T-1)$  个时间虚拟变量包括在回归方程中(未包括的时间虚拟变量即为基期), 比如

$$y_{it} = x'_{it} \beta + z'_i \delta + \gamma_2 D_{2,t} + \dots + \gamma_T D_{T,t} + u_i + \varepsilon_{it} \quad (15.9)$$

其中, 时间虚拟变量  $D_{2,t} = 1$ , 如果  $t=2$ ;  $D_{2,t} = 0$ , 如果  $t \neq 2$ ; 以此类推。由于方程(15.9)既考虑了个体固定效应, 又考虑了时间固定效应, 故称为“双向固定效应”(Two-way FE)。相应地, 如果仅考虑个体固定效应(即上节的模型), 则称为“单向固定效应”(One-way FE)。

在有些情况下, 为了节省参数, 可以引入一个时间趋势项, 以替代上述  $(T-1)$  个时间虚拟变量:

$$y_{it} = x'_{it} \beta + z'_i \delta + \gamma t + u_i + \varepsilon_{it} \quad (15.10)$$

<sup>①</sup> 严格证明参见 Cameron and Trivedi(2005), 第 732 页。

显然,上式隐含着一个较强的假定,即每个时期的时间效应相等,每期均增加  $\gamma$ 。如果此假定不太可能成立,则应使用方程(15.9)的时间虚拟变量法(可通过该法的估计结果判断每期的时间效应是否大致相等)。

## 15.6 一阶差分法

对于固定效应模型,可对方程(15.1)两边进行一阶差分,以消去个体效应  $u_i$ (但同时也把  $z'_i\delta$  消掉了),

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}) \quad (15.11)$$

对此差分形式的方程使用 OLS 即得到“一阶差分估计量”(First Differencing Estimator),记为  $\hat{\boldsymbol{\beta}}_{FD}$ 。由于  $u_i$  不再出现在差分方程中,只要扰动项的一阶差分  $(\varepsilon_{it} - \varepsilon_{i,t-1})$  与解释变量的一阶差分  $(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})$  不相关,则  $\hat{\boldsymbol{\beta}}_{FD}$  是一致的。此一致性条件比保证  $\hat{\boldsymbol{\beta}}_{FE}$  一致的严格外生性假定更弱,这是  $\hat{\boldsymbol{\beta}}_{FD}$  的主要优点。可以证明(参见习题),如果  $T=2$ ,则  $\hat{\boldsymbol{\beta}}_{FD} = \hat{\boldsymbol{\beta}}_{FE}$ 。但对于  $T>2$ ,如果  $\{\varepsilon_{it}\}$  为独立同分布的,则组内估计量  $\hat{\boldsymbol{\beta}}_{FE}$  比一阶差分估计量  $\hat{\boldsymbol{\beta}}_{FD}$  更有效率。因此,在实践上,主要使用  $\hat{\boldsymbol{\beta}}_{FE}$ ,而较少用  $\hat{\boldsymbol{\beta}}_{FD}$ 。但对于动态面板(参见第 16 章),严格外生性假定无法满足,故转而使用差分法。

## 15.7 随机效应模型

对于回归方程  $y_{it} = \mathbf{x}'_i \boldsymbol{\beta} + z'_i \delta + u_i + \varepsilon_{it}$ ,随机效应模型假设  $u_i$  与解释变量  $\{\mathbf{x}_i, z_i\}$  均不相关,故 OLS 是一致的。然而,由于扰动项由  $(u_i + \varepsilon_{it})$  组成,不是球型扰动项,因此 OLS 不是最有效的。假设不同个体之间的扰动项互不相关。由于  $u_i$  的存在,同一个体不同时期的扰动项之间存在自相关,

$$\text{Cov}(u_i + \varepsilon_{it}, u_s + \varepsilon_{is}) = \begin{cases} \sigma_u^2, & \text{若 } t \neq s \\ \sigma_u^2 + \sigma_\varepsilon^2, & \text{若 } t = s \end{cases} \quad (15.12)$$

其中,  $\sigma_u^2$  为  $u_i$  的方差(不随  $i$  变化),而  $\sigma_\varepsilon^2$  为  $\varepsilon_{it}$  的方差(不随  $i, t$  变化)。当  $t \neq s$  时,其自相关系数为

$$\rho \equiv \text{Corr}(u_i + \varepsilon_{it}, u_s + \varepsilon_{is}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} \quad (15.13)$$

显然,同一个体不同时期的扰动项之间的自相关系数  $\rho$  不随时间距离( $t-s$ )而改变,故随机效应模型也被称为“等相关模型”(equicorrelated model)或“可交换扰动项模型”(exchangeable errors model)<sup>①</sup>。 $\rho$  越大,则复合扰动项  $(u_i + \varepsilon_{it})$  中个体效应的部分( $u_i$ )越重要。Stata 记  $\rho$  为“rho”。

同一个体扰动项的协方差阵为

<sup>①</sup> 对于自回归模型,一般来说,扰动项之间的自相关系数  $\rho$ (或其绝对值)随时间而递减。

$$\Sigma = \begin{pmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_e^2 \end{pmatrix}_{T \times T} \quad (15.14)$$

由表达式(15.14)可知,同一个体的扰动项具有相同的方差,但存在组内自相关。整个样本扰动项的协方差阵为块对角矩阵(block diagonal matrix),

$$\Omega = \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix}_{nT \times nT} \quad (15.15)$$

由于 OLS 是一致的,且其扰动项为( $u_i + \varepsilon_{iu}$ ),故可以用 OLS 的残差来估计( $\sigma_u^2 + \sigma_e^2$ )。另一方面,FE 也是一致的,且其扰动项为( $\varepsilon_{iu} - \bar{\varepsilon}_i$ ),故可以用 FE 的残差来估计  $\sigma_e^2$ (Stata 记  $\hat{\sigma}_u$  为“sigma\_u”,记  $\hat{\sigma}_e$  为“sigma\_e”)。然后,就可以使用可行广义最小二乘法(FGLS)来估计原模型,得到“随机效应估计量”(Random Effects Estimator),记为  $\hat{\beta}_{RE}$ 。具体来说,用 OLS 来估计以下“广义离差”(quasi-demeaned)模型,

$$y_{it} - \hat{\theta}\bar{y}_i = (\mathbf{x}_{it} - \hat{\theta}\bar{\mathbf{x}}_i)' \boldsymbol{\beta} + (1 - \hat{\theta})\mathbf{z}'_i \boldsymbol{\delta} + \underbrace{[(1 - \hat{\theta})u_i + (\varepsilon_{it} - \hat{\theta}\bar{\varepsilon}_i)]}_{\text{误差项}} \quad (15.16)$$

其中, $\hat{\theta}$ 是 $\theta = 1 - \frac{\sigma_e}{(T\sigma_u^2 + \sigma_e^2)^{1/2}}$ 的一致估计量(Stata 记  $\hat{\theta}$  为“theta”)。可以证明(参见习题),方程(15.16)的扰动项不再有自相关。显然, $0 \leq \theta \leq 1$ 。进一步,如果  $\hat{\theta} = 0$ ,则为混合回归。而如果  $\hat{\theta} = 1$ ,则为组内估计量。

对于随机效应模型,如果进一步假设扰动项服从正态分布,则可写出样本的对数似然函数,然后进行最大似然估计(MLE)。

## 15.8 组间估计量

对于随机效应模型,还可以使用“组间估计量”。如果每位个体的时间序列数据较不准确或噪音较大,可对每位个体取时间平均值,然后用平均值来回归。

$$\bar{y}_i = \bar{\mathbf{x}}'_i \boldsymbol{\beta} + \mathbf{z}'_i \boldsymbol{\delta} + u_i + \bar{\varepsilon}_i \quad (i = 1, \dots, n) \quad (15.17)$$

对上式使用 OLS,就可以得到“组间估计量”(Between Estimator),记为  $\hat{\beta}_{BE}$ 。由于  $\{\bar{\mathbf{x}}_i, \bar{\mathbf{z}}_i\}$  中包含了  $\{\mathbf{x}_i, \mathbf{z}_i\}$  的信息,如果  $u_i$  与解释变量  $\{\mathbf{x}_i, \mathbf{z}_i\}$  相关,则  $\hat{\beta}_{BE}$  是不一致的。因此,不能在固定效应模型下使用组间估计法。即使在随机效应模型下,由于面板数据被压缩为截面数据,损失了较多信息量,故组间估计法也不常用。

## 15.9 拟合优度的度量

对于线性回归模型,一般用  $R^2$  衡量拟合优度。在有常数项的情况下,线性回归模型的  $R^2$  就等于被解释变量  $y$  与 OLS 预测值  $\hat{y}$  之间相关系数的平方,即  $R^2 = [\text{corr}(y, \hat{y})]^2$ (参见第 3 章习

题)。对于面板模型,如果使用混合回归,则可直接用混合回归的  $R^2$  衡量拟合优度。但如果使用固定效应、随机效应或组间回归,拟合优度的度量略为复杂。任意给定估计量  $(\hat{\beta}, \hat{\delta})$ , Stata 提供了以下三种  $R^2$ (不一定具有线性回归  $R^2$  的全部性质)。

首先,对应于原模型(15.1),称  $[\text{Corr}(y_{it}, \mathbf{x}'_i \hat{\beta} + z'_i \hat{\delta})]^2$  为“整体  $R^2$ ”( $R^2$  overall),衡量估计量  $(\hat{\beta}, \hat{\delta})$  对原模型的拟合优度。

其次,对应于组内模型(15.5),称  $[\text{Corr}(\bar{y}_{it}, \bar{\mathbf{x}}'_i \hat{\beta})]^2$  为“组内  $R^2$ ”( $R^2$  within),衡量估计量  $(\hat{\beta}, \hat{\delta})$  对组内模型的拟合优度。

再次,对应于组间模型(15.17),称  $[\text{Corr}(\bar{y}_i, \bar{\mathbf{x}}'_i \hat{\beta} + z'_i \hat{\delta})]^2$  为“组间  $R^2$ ”( $R^2$  between),衡量估计量  $(\hat{\beta}, \hat{\delta})$  对组间模型的拟合优度。

显然,无论固定效应、随机效应还是组间回归,都可以计算以上三种  $R^2$ 。对于固定效应模型,建议使用组内  $R^2$ ,也就是组内方程(15.5)的 OLS  $R^2$ ;而整体  $R^2$  与组间  $R^2$  只是相应的相关系数平方而已。对于组间回归模型,建议使用组间  $R^2$ ,也就是组间方程(15.17)的 OLS  $R^2$ ;而整体  $R^2$  与组内  $R^2$  只是相应的相关系数平方而已。然而,对于随机效应模型,这三种  $R^2$  都只是相应的相关系数平方,而非方程(15.16)的 OLS  $R^2$ 。

## 15.10 非平衡面板

在面板数据中,如果每个时期在样本中的个体完全一样,则称为“平衡面板数据”(balanced panel)。然而,有时某些个体的数据可能缺失(比如,个体死亡、企业倒闭或被兼并、个体不再参与调查),或者新的个体在后来才加入到调查中来。在这种情况下,每个时期观测到的个体不完全相同,被称为“非平衡面板数据”(unbalanced panel)或“不完全面板”(incomplete panel)。

显然,非平衡面板数据并不影响计算离差形式的组内估计量(within estimator),因此,固定效应模型的估计可以照样进行。对于随机效应模型而言,非平衡面板数据也没有实质性影响,只要在做广义离差变换时让  $\theta_i = 1 - \frac{\sigma_\epsilon}{(T_i \sigma_u^2 + \sigma_\epsilon^2)^{1/2}}$  ( $T_i$  为第  $i$  位个体的时间维度),就可以照常进行 FGLS 估计。当然,非平衡面板数据使得估计量及其协方差矩阵的数学表达式更加复杂,但这些都由 Stata 在幕后进行。

非平衡面板可能出现的最大问题是,那些原来在样本中但后来丢掉的个体,如果其“丢掉”的原因是内生的(即与扰动项相关),则会导致样本不具有代表性(不再是随机样本),从而导致估计量不一致。比如,低收入的人群更容易从面板数据中丢掉。此时,需要用“样本选择模型”(sample selection model)来进行处理,参见第 14 章。

如果从非平衡面板数据中提取一个平衡的面板数据子集<sup>①</sup>,然后进行数据处理,则必然会损失样本容量,降低估计效率。更进一步,如果人为“丢掉”的个体并非完全随机,则同样会破坏样本的随机性。

<sup>①</sup> 可通过连玉君老师提供的 Stata 命令“xtbalance”来实现,下载方法为“ssc install xtbalance”,使用方法详见“help xtbalance”。

## 15.11 究竟该用固定效应还是随机效应模型

在处理面板数据时,究竟应该使用固定效应模型还是随机效应模型是一个根本问题。为此,希望检验原假设“ $H_0: u_i$  与  $\mathbf{x}_{it}, \mathbf{z}_i$  不相关”(即随机效应模型为正确模型)。无论原假设成立与否,FE 都是一致的。然而,如果原假设成立,则 RE 比 FE 更有效。但如果原假设不成立,则 RE 不一致。因此,如果  $H_0$  成立,则 FE 与 RE 估计量将共同收敛于真实的参数值,故  $(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \xrightarrow{P} \mathbf{0}$ 。反之,如果二者的差距过大,则倾向于拒绝原假设。豪斯曼检验(Hausman, 1978)的统计量为

$$(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})' \left[ \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{FE}) - \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{RE}) \right]^{-1} (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \xrightarrow{d} \chi^2(K) \quad (15.18)$$

其中, $K$  为  $\hat{\boldsymbol{\beta}}_{FE}$  的维度,即  $\mathbf{x}_{it}$  中所包含的随时间而变的解释变量个数(因为  $\hat{\boldsymbol{\beta}}_{FE}$  无法估计不随时间而变的解释变量系数)。如果该统计量大于临界值,则拒绝  $H_0$ 。

上述检验的缺点是,它假设在  $H_0$  成立的情况下,  $\hat{\boldsymbol{\beta}}_{RE}$  是最有效率的(fully efficient)。然而,如果扰动项存在异方差,则  $\hat{\boldsymbol{\beta}}_{RE}$  并非最有效率的估计量。因此,上面的检验不适用于异方差的情形。

解决方法之一,通过自助法(bootstrap),即计算机模拟再抽样(resampling)的方法来计算  $\text{Var}(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})$ ,参见第 19 章。

解决方法之二,进行以下辅助回归(Wooldridge, 2010),

$$\mathbf{y}_{it} - \hat{\theta}\bar{\mathbf{y}}_i = (\mathbf{x}_{it} - \hat{\theta}\bar{\mathbf{x}}_i)' \boldsymbol{\beta} + (1 - \hat{\theta})\mathbf{z}'_i \boldsymbol{\delta} + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \boldsymbol{\gamma} + [(1 - \hat{\theta})u_i + (\varepsilon_{it} - \hat{\theta}\bar{\varepsilon}_i)] \quad (15.19)$$

然后,使用聚类稳健标准误来检验原假设“ $H_0: \boldsymbol{\gamma} = \mathbf{0}$ ”。这个检验方法在异方差的情况下也适用。由于总可以把原模型变换为方程(15.16),故在上式中,  $\boldsymbol{\gamma} = \mathbf{0}$ 。如果随机效应模型成立,则 OLS 是一致的,故  $\underset{n \rightarrow \infty}{\text{plim}} \hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} = \mathbf{0}$ 。反之,如果固定效应模型成立,扰动项  $[(1 - \hat{\theta})u_i + (\varepsilon_{it} - \hat{\theta}\bar{\varepsilon}_i)]$  与  $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$  相关(因为  $u_i$  与  $\mathbf{x}_{it}$  相关),故 OLS 是不一致的,即  $\underset{n \rightarrow \infty}{\text{plim}} \hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^* \neq \boldsymbol{\gamma} = \mathbf{0}$ 。因此,拒绝 “ $H_0: \boldsymbol{\gamma} = \mathbf{0}$ ”,则意味着拒绝随机效应,接受固定效应。对于非平衡面板,则以  $\hat{\theta}_i$  替代方程(15.19)中的  $\hat{\theta}$  即可。

## 15.12 个体时间趋势

迄今为止,本章所考虑的个体效应都体现为个体截距项  $u_i$ 。事实上,个体异质性还可能表现为个体的不同时间趋势。比如,在一个跨国面板中,各国的经济增长率可能不同。为此,考虑以下模型:

$$\mathbf{y}_{it} = \mathbf{x}'_{it} \boldsymbol{\beta} + \mathbf{z}'_i \boldsymbol{\delta} + \gamma_i t + u_i + \varepsilon_{it} \quad (15.20)$$

其中,  $\gamma_i t$  为个体时间趋势。类似于  $u_i$  的性质,一般将  $\gamma_i$  视为来自某分布的随机变量(但从该分布随机抽出一个观测值后,就不再随时间而变)。因此,此模型称为“随机趋势模型”(random trend model)。另外,如果  $y_{it}$  取对数形式(比如  $\ln GDP_{it}$ ),则  $\gamma_i$  可解释为在给定  $(\mathbf{x}_{it}, \mathbf{z}_i)$  条件下的

平均增长率(即 $\partial E(\ln GDP_u)/\partial t$ ),故此模型也称为“随机增长模型”(random growth model)。

为了估计此模型,首先对方程(15.20)两边做差分,以去掉 $u_i$ :

$$\Delta y_{it} = \Delta x'_{it} \beta + \gamma_i + \Delta \varepsilon_{it} \quad (15.21)$$

其中, $\gamma_i t - \gamma_i(t-1) = \gamma_i$ 。在形式上,方程(15.21)与标准的个体效应模型(15.1)一模一样。如果 $\gamma_i$ 与解释变量 $\Delta x_{it}$ 不相关,则可用随机效应模型(RE)估计方程(15.21)。更一般地,如果 $\gamma_i$ 与解释变量 $\Delta x_{it}$ 相关,则可用固定效应模型(FE)或一阶差分法(FD)估计方程(15.21)。

## 15.13 短面板的 Stata 命令及实例

### 1. 面板数据的设定

设定面板数据的 Stata 命令为

```
xtset panelvar timevar
```

命令“`xtset`”告诉 Stata 你的数据为面板数据,其中面板(个体)变量“`panelvar`”的取值必须为整数且不重复,相当于把样本中每个个体进行编号;而“`timevar`”为时间变量。假如“`panelvar`”本来是字符串(比如,国家名字 `country`),则可以使用以下命令将其转换为数字型变量:

```
encode country, gen(cntry)
```

这样,变量 `cntry` 就以“1,2,3,...”来指代不同的国家。

显示面板数据统计特性的 Stata 命令为

```
xtdes                   (显示面板数据的结构,是否为平衡面板)
```

```
xtsum                   (显示组内、组间与整体的统计指标)
```

```
xttab varname         (显示组内、组间与整体的分布频率,tab 指的是 tabulate)
```

`xtline varname` (对每位个体分别显示该变量的时间序列图;如果希望将所有个体的时间序列图叠放在一起,可加上选择项 `overlay`)

以数据集“`traffic.dta`”为例<sup>①</sup>。该面板数据集包含了美国 48 个州 1982—1988 年“交通死亡率”(traffic fatality rates)的相关变量:`fatal`(交通死亡率),`beertax`(啤酒税),`spircons`(酒精消费量),`unrate`(失业率),`perinck`(人均收入,以千元记)<sup>②</sup>,`state`(州),`year`(年)。

首先,设定 `state` 与 `year` 为面板(个体)变量及时间变量:

```
. use traffic.dta, clear  
. xtset state year
```

```
panel variable: state (strongly balanced)  
time variable: year, 1982 to 1988  
delta: 1 unit
```

上表显示,这是一个平衡的面板数据。其次,显示数据集的结构:

```
. xtdes
```

① 此例来自 Baum(2006)。

② Per capita income in thousands.

state: 1, 4, ..., 56	n =	48					
year: 1982, 1983, ..., 1988	T =	7					
Delta(year) = 1 unit							
Span(year) = 7 periods							
(state*year uniquely identifies each observation)							
Distribution of T_i:	min	5%	25%	50%	75%	95%	max
	7	7	7	7	7	7	7
Freq.	Percent	Cum.	Pattern				
48	100.00	100.00	1111111				
48	100.00		XXXXXXXX				

上表清晰地显示,  $n=48$ , 而  $T=7$ 。由于  $n$  大而  $T$  小, 故这是一个短面板。

再次, 显示数据集中以上变量的统计特征:

```
. xtsum fatal beertax spircons unrata perinck state year
```

Variable		Mean	Std. Dev.	Min	Max	Observations
fatal	overall	2.040444	.5701938	.82121	4.21784	N = 336
	between		.5461407	1.110077	3.653197	n = 48
	within		.1794253	1.45556	2.962664	T = 7
beertax	overall	.513256	.4778442	.0433109	2.720764	N = 336
	between		.4789513	.0481679	2.440507	n = 48
	within		.0552203	.1415352	.7935126	T = 7
spircons	overall	1.75369	.6835745	.79	4.9	N = 336
	between		.6734649	.8614286	4.388572	n = 48
	within		.147792	1.255119	2.265119	T = 7
unrate	overall	7.346726	2.533405	2.4	18	N = 336
	between		1.953377	4.1	13.2	n = 48
	within		1.634257	4.046726	12.14673	T = 7
perinck	overall	13.88018	2.253046	9.513762	22.19345	N = 336
	between		2.122712	9.95087	19.51582	n = 48
	within		.8068546	11.43261	16.55782	T = 7
state	overall	30.1875	15.30985	1	56	N = 336
	between		15.44883	1	56	n = 48
	within	0	30.1875	30.1875	30.1875	T = 7
year	overall	1985	2.002983	1982	1988	N = 336
	between		0	1985	1985	n = 48
	within		2.002983	1982	1988	T = 7

上表显示, 变量 state 的组内(within) 标准差为 0, 因为分在同一组的数据属于同一个州。另一方面, 变量 year 的组间(between) 标准差为 0, 因为不同组的这一变量取值完全相同。

下面, 看一下被解释变量 fatal 在 48 个州的时间趋势图, 结果如图 15.2。

```
. xtline fatal
```

从上图可以看出, 不同州的交通死亡率(fatal)的时间趋势不尽相同, 有些州很平稳(比如, 佛罗里达州, FL), 有些州呈下降趋势(比如, 俄克拉荷马州, OK), 而有些州则为上升趋势(比如, 南卡罗来纳州, SC)。在一定程度上, 交通死亡率的州际差异有助于估计决定交通死亡率的因素。

## 2. 混合回归

作为参照系, 首先进行混合回归。其 Stata 命令的基本格式为

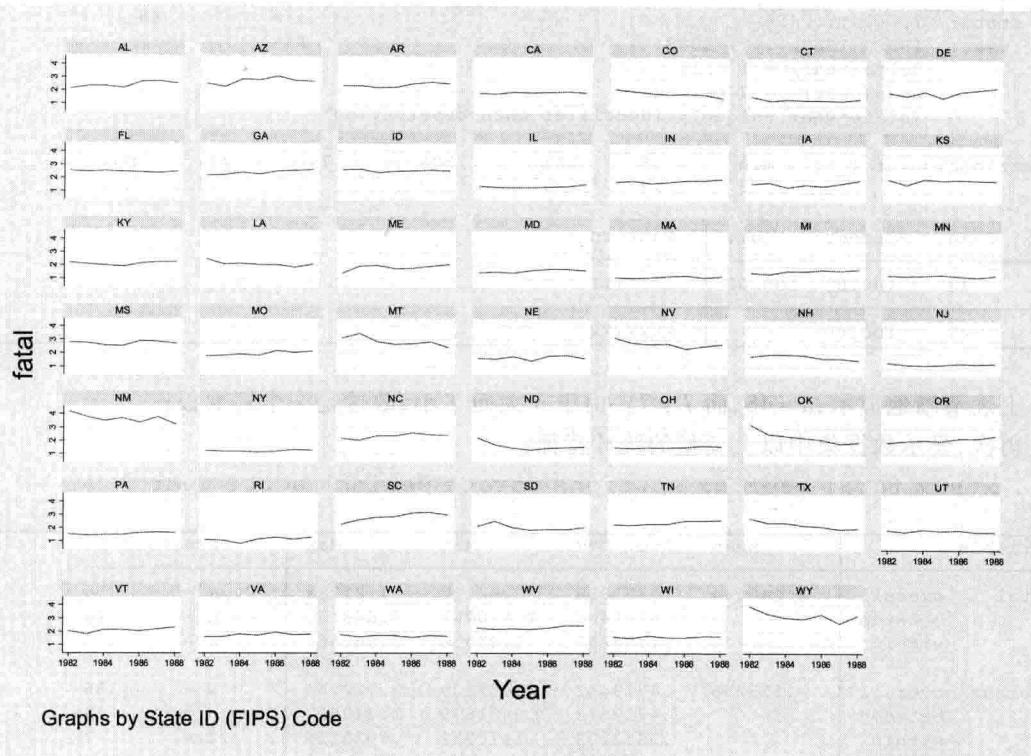


图 15.2 48个州的时间趋势图

```
reg y x1 x2 x3, vce(cluster id)
```

其中,“id”指用来确定每位个体的变量。

```
. reg fatal beertax spircons unrate perinck, vce(cluster state)
. estimates store OLS
```

Linear regression						Number of obs = 336
						F( 4, 47) = 17.43
						Prob > F = 0.0000
						R-squared = 0.3019
						Root MSE = .47929
(Std. Err. adjusted for 48 clusters in state)						
fatal	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
beertax	.0971997	.116884	0.83	0.410	-.1379406	.33234
spircons	.1623471	.1070988	1.52	0.136	-.053108	.3778021
unrate	-.0291014	.0208687	-1.39	0.170	-.0710838	.012881
perinck	-.1584291	.0371404	-4.27	0.000	-.233146	-.0837122
_cons	4.118674	.6765746	6.09	0.000	2.757581	5.479766

其中,选择项“vce(cluster state)”表示,使用以“state”为聚类变量的聚类稳健标准误。如果使用普通标准误,则为

```
. reg fatal beertax spircons unrate perinck
```

Source	SS	df	MS	Number of obs = 336		
Model	32.878445	4	8.21961124	F( 4, 331) =	35.78	
Residual	76.037065	331	.22971923	Prob > F =	0.0000	
Total	108.91551	335	.325120925	R-squared =	0.3019	
				Adj R-squared =	0.2934	
				Root MSE =	.47929	

fatal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
beertax	.0971997	.0615518	1.58	0.115	-.0238823 .2182817
spircons	.1623471	.0432466	3.75	0.000	.0772743 .2474198
unrate	-.0291014	.0127156	-2.29	0.023	-.0541151 -.0040877
perinck	-.1584291	.0169869	-9.33	0.000	-.1918449 -.1250132
_cons	4.118674	.296995	13.87	0.000	3.534438 4.702909

对比普通标准误与聚类稳健标准误可知,前者大约仅是后者的一半。由于同一个州不同期之间的扰动项一般存在自相关,而默认的普通标准误计算方法假设扰动项为独立同分布的,故普通标准误的估计并不准确。

### 3. 固定效应

由于每个州的“州情”不同,可能存在不随时间而变的遗漏变量,故考虑使用固定效应模型(FE)。固定效应模型(组内估计量)的 Stata 命令基本格式为

```
xtreg y x1 x2 x3, fe r
```

LSDV 法的 Stata 命令为

```
reg y x1 x2 x3 i.id,r
```

其中,选择项“r”表示使用聚类稳健标准误;如果使用选择项“vce(cluster id)”也能达到完全相同的效果<sup>①</sup>。“id”表示用来确定个体的变量,“i.id”则表示根据变量“id”而生成的虚拟变量。

首先使用组内估计量:

```
. xtreg fatal beertax spircons unrate perinck,fe r  
. estimates store FE_robust
```

Fixed-effects (within) regression		Number of obs = 336			
Group variable: state		Number of groups = 48			
R-sq:	within = 0.3526	Obs per group: min = 7			
	between = 0.1146	avg = 7.0			
	overall = 0.0863	max = 7			
		F(4,47) = 21.27			
		Prob > F = 0.0000			
(Std. Err. adjusted for 48 clusters in state)					
fatal	Coef.	Robust	t	P> t	[95% Conf. Interval]
beertax	-.4840728	.2218754	-2.18	0.034	-.9304285 -.037717
spircons	.8169652	.1272627	6.42	0.000	.5609456 1.072985
unrate	-.0290499	.0094581	-3.07	0.004	-.0480772 -.0100227
perinck	.1047103	.0341455	3.07	0.004	.0360184 .1734022
_cons	-.383783	.7091738	-0.54	0.591	-1.810457 1.042891
sigma_u	1.1181913				
sigma_e	.15678965				
rho	.98071823	(fraction of variance due to u_i)			

① 在使用命令 xtreg 时,Stata 已经知道这是面板数据,故使用选择项“r”或“vce(robust)”,也能得到完全相同的聚类稳健标准误。

上表的 Stata 输出结果中包括一个常数项 (\_cons), 这是所有个体效应  $u_i$  的平均值。上表最后一行显示, “rho = 0.98”, 故复合扰动项 ( $u_i + \varepsilon_u$ ) 的方差主要来自个体效应  $u_i$  的变动。

究竟应该使用混合回归还是固定效应模型呢?

在使用命令“`xtreg, fe`”时,如果不加选择项“`r`”,则输出结果还包含一个  $F$  检验,其原假设为“ $H_0 : \text{all } u_i = 0$ ”,即混合回归是可以接受的:

```
. xtreg fatal beertax spircons unrate perinck, fe
. estimates store FE
```

Fixed-effects (within) regression		Number of obs = 336	
Group variable: state		Number of groups = 48	
R-sq:	within = 0.3526	Obs per group: min = 7	
	between = 0.1146	avg = 7.0	
	overall = 0.0863	max = 7	
corr(u_i, Xb)	= -0.8804	F(4, 284)	= 38.68
		Prob > F	= 0.0000
fatal	Coef.	Std. Err.	t P> t  [95% Conf. Interval]
beertax	- .4840728	.1625106	-2.98 0.003 -.8039508 -.1641948
spircons	.8169652	.0792118	10.31 0.000 .6610484 .9728819
unrate	-.0290499	.0090274	-3.22 0.001 -.0468191 -.0112808
perinck	.1047103	.0205986	5.08 0.000 .064165 .1452555
_cons	-.383783	.4201781	-0.91 0.362 -1.210841 .4432754
sigma_u	1.1181913		
sigma_e	.15678965		
rho	.98071823		(fraction of variance due to u_i)
F test that all u_i=0:		F(47, 284) = 59.77	Prob > F = 0.0000

对于原假设“ $H_0 : \text{all } u_i = 0$ ”,由于上表最后一行  $F$  检验的  $p$  值为 0.000 0,故强烈拒绝原假设,即认为 FE 明显优于混合回归,应该允许每位个体拥有自己的截距项。然而,由于未使用聚类稳健标准误,故这个  $F$  检验并不有效,因为普通标准误大约只是聚类稳健标准误的一半<sup>①</sup>。为此,进一步通过 LSDV 法来考察:

```
. reg fatal beertax spircons unrate perinck i.state, vce(cluster state)
. estimates store LSDV
```

① 尽管如此,由于  $p$  值很小,为 0.000 0,故即使按聚类标准误来计算  $F$  值,大致也能拒绝原假设。

Linear regression		Number of obs = 336				
		F( 3, 47) = .	Prob > F = .	R-squared = 0.9359	Root MSE = .15679	
(Std. Err. adjusted for 48 clusters in state)						
		Robust				
fatal	Coef.	Std. Err.	t	P> t	[ 95% Conf. Interval]	
beertax	-.4840728	.2395323	-2.02	0.049	-.9659495	-.002196
spircons	.8169652	.1373903	5.95	0.000	.5405716	1.093359
unrate	-.0290499	.0102108	-2.85	0.007	-.0495914	-.0085085
perinck	.1047103	.0368628	2.84	0.007	.0305519	.1788687
state						
4	-1.146089	.3464123	-3.31	0.002	-1.842981	-.4491975
5	-.3422802	.2516739	-1.36	0.180	-.8485829	.1640224
6	-2.478906	.4335366	-5.72	0.000	-3.35107	-1.606743
8	-2.359016	.3971192	-5.94	0.000	-3.157917	-1.560115
9	-3.517465	.4888933	-7.19	0.000	-4.500992	-2.533939
10	-2.621819	.4507483	-5.82	0.000	-3.528608	-1.715031
12	-1.484055	.2576804	-5.76	0.000	-2.002441	-.9656691
13	-.4193311	.222352	-1.89	0.065	-.8666457	.0279835
16	-.475246	.3043054	-1.56	0.125	-1.08743	.1369375
17	-2.66262	.3935873	-6.77	0.000	-3.454415	-1.870824
18	-1.551228	.3208617	-4.83	0.000	-2.196718	-.9057372
19	-1.512083	.2850596	-5.30	0.000	-2.085549	-.9386167
20	-1.425295	.2782851	-5.12	0.000	-1.985132	-.8654573
21	-1.008736	.3429057	-2.94	0.005	-1.698573	-.3188985
22	-.9904429	.2155297	-4.60	0.000	-1.424033	-.556853
23	-1.728957	.2458348	-7.03	0.000	-2.223513	-1.234401
24	-2.971806	.4333382	-6.86	0.000	-3.84357	-2.100042
25	-3.542156	.4444227	-7.97	0.000	-4.43622	-2.648093
26	-2.039022	.3147281	-6.48	0.000	-2.672174	-1.405871
27	-2.726702	.3617594	-7.54	0.000	-3.454468	-1.998936
28	.1623338	.156378	1.04	0.305	-.152258	.4769257
29	-1.473051	.312394	-4.72	0.000	-2.101507	-.8445953
30	-.6003982	.327793	-1.83	0.073	-1.259833	.0590363
31	-1.812884	.2951366	-6.14	0.000	-2.406622	-1.219146
32	-3.433805	.6642254	-5.17	0.000	-4.770054	-2.097556
33	-4.226736	.5945373	-7.11	0.000	-5.422791	-3.030682
34	-3.425003	.4771024	-7.18	0.000	-4.384809	-2.465196
35	.4537402	.3027855	1.50	0.141	-.1553857	1.062866
36	-3.146312	.4211584	-7.47	0.000	-3.993574	-2.299051
37	-.6387593	.1008379	-6.33	0.000	-.8416191	-.4358996
38	-2.124115	.3257623	-6.52	0.000	-2.779464	-1.468765
39	-1.689753	.2868615	-5.89	0.000	-2.266843	-1.112662
40	-.6842853	.1722774	-3.97	0.000	-1.030863	-.3377078
41	-1.388066	.347058	-4.00	0.000	-2.086257	-.6898754
42	-1.92044	.323107	-5.94	0.000	-2.570447	-1.270432
44	-3.097237	.3984525	-7.77	0.000	-3.89882	-2.295654
45	-.0251951	.0789182	-0.32	0.751	-.1839581	.1335679
46	-1.345773	.2595865	-5.18	0.000	-1.867994	-.8235526
47	-.780764	.3155413	-2.47	0.017	-1.415551	-.1459769
48	-1.117104	.2845582	-3.93	0.000	-1.689561	-.544647
49	-.7663286	.2268098	-3.38	0.001	-1.222611	-.3100462
50	-1.844306	.29671	-6.22	0.000	-2.441209	-1.247402
51	-1.891904	.2537621	-7.46	0.000	-2.402408	-1.381401
53	-2.176212	.3644988	-5.97	0.000	-2.909489	-1.442935
54	-.2142207	.2815892	-0.76	0.451	-.780705	.3522637
55	-2.41626	.3894621	-6.20	0.000	-3.199757	-1.632763
56	-.7183534	.402788	-1.78	0.081	-1.528658	.0919516
_cons	1.269226	.7728348	1.64	0.107	-.2855167	2.823969

从上表可知,大多数个体虚拟变量均很显著( $p$ 值为0.000),故可放心地拒绝“所有个体虚拟变量都为0”的原假设,即认为存在个体效应,不应使用混合回归。

对于固定效应模型,也可使用一阶差分法(FD)。Stata没有专门执行一阶差分法的命令,但在使用命令“`xtserial, output`”对组内自相关进行检验时(参见第16章),可以附带提供一阶差分法的结果:

```
. xtserial fatal beertax spircons unrate perinck, output
. estimates store FD
```

Linear regression		Number of obs = 288 F( 4, 47) = 3.66 Prob > F = 0.0113 R-squared = 0.0814 Root MSE = .19047			
		(Std. Err. adjusted for 48 clusters in state)			
D.fatal		Robust Coef.	Std. Err.	t	P> t  [95% Conf. Interval]
beertax	D1.	.1187701	.2276875	0.52	0.604 -.339278 .5768183
spircons	D1.	.523584	.1633135	3.21	0.002 .1950396 .8521285
unrate	D1.	.003399	.0118903	0.29	0.776 -.0205212 .0273192
perinck	D1.	.1417981	.0468699	3.03	0.004 .047508 .2360882

Wooldridge test for autocorrelation in panel data  
H0: no first-order autocorrelation  
F( 1, 47) = 6.246  
Prob > F = 0.0160

从上表可以看出,一阶差分估计量(FD)的估计系数与组内估计量(FE)有较大差别。比如,如果使用FD估计法,啤酒税(beertax)对交通死亡率(fatal)的影响系数竟然为正(0.12),尽管在统计上不显著( $p$ 值为0.60)。一般认为,FE比FD更有效率,故较少使用FD。

在使用命令`xtserial`计算一阶差分估计量时,同时得到了对“面板一阶自相关”的检验。该检验的 $p$ 值为0.016,即在5%的显著性水平上存在面板自相关。但这是一个只有7期的短面板( $T=7$ ),我们可以不考虑“面板自相关”的问题。

也可以在固定效应模型中考虑时间效应,即双向固定效应(Two-way FE)。为此,定义年度虚拟变量:

```
. tab year, gen(year)
```

Year	Freq.	Percent	Cum.
1982	48	14.29	14.29
1983	48	14.29	28.57
1984	48	14.29	42.86
1985	48	14.29	57.14
1986	48	14.29	71.43
1987	48	14.29	85.71
1988	48	14.29	100.00
Total	336	100.00	

此命令将在 Stata 的变量窗口生成时间虚拟变量 year1, year2, ..., year7。

```
. xtreg fatal beertax spircons unrate perinck year2 -year7, fe r
. estimates store FE_TW
```

Fixed-effects (within) regression						Number of obs = 336				
Group variable: state						Number of groups = 48				
R-sq: within = 0.4528						Obs per group: min = 7				
between = 0.1090						avg = 7.0				
overall = 0.0770						max = 7				
corr(u_i, Xb) = -0.8728						F(10, 47) = 14.13				
						Prob > F = 0.0000				
(Std. Err. adjusted for 48 clusters in state)										
fatal	Robust									
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]					
beertax	-.4347195	.2442775	-1.78	0.082	-.9261425	.0567036				
spircons	.805857	.1161087	6.94	0.000	.5722764	1.039438				
unrate	-.0549084	.011763	-4.67	0.000	-.0785725	-.0312443				
perinck	.0882636	.0322971	2.73	0.009	.0232901	.153237				
year2	-.0533713	.0312438	-1.71	0.094	-.1162256	.0094831				
year3	-.1649828	.0439375	-3.75	0.000	-.2533737	-.076592				
year4	-.1997376	.0496167	-4.03	0.000	-.2995535	-.0999218				
year5	-.0508034	.0661756	-0.77	0.447	-.1839315	.0823248				
year6	-.1000728	.0756768	-1.32	0.192	-.2523149	.0521693				
year7	-.134057	.0863572	-1.55	0.127	-.3077854	.0396714				
_cons	.1290568	.6238303	0.21	0.837	-1.125928	1.384041				
sigma_u	1.0987683									
sigma_e	.14570531									
rho	.98271904	(fraction of variance due to u_i)								

其中, year1(即 1982 年)被作为基期(对应于常数项\_cons),而不包括在上述回归命令中<sup>①</sup>。时间效应的符号均为负,但有些年度虚拟变量显著,而有些则不显著。下面检验所有年度虚拟变量的联合显著性:

```
. test year2 year3 year4 year5 year6 year7
```

( 1) year2 = 0
( 2) year3 = 0
( 3) year4 = 0
( 4) year5 = 0
( 5) year6 = 0
( 6) year7 = 0
F( 6, 47) = 8.21
Prob > F = 0.0000

结果强烈拒绝“无时间效应”的原假设,认为应在模型中包括时间效应。在 Stata 12 中,还可以直接用以下命令来估计双向固定效应模型(不必生成时间虚拟变量):

```
. xtreg fatal beertax spircons unrate perinck i.year, fe r
```

① 否则,将出现虚拟变量陷阱,导致完全多重共线性。

Fixed-effects (within) regression		Number of obs = 336	
Group variable: state		Number of groups = 48	
R-sq:	within = 0.4528	Obs per group:	min = 7
	between = 0.1090		avg = 7.0
	overall = 0.0770		max = 7
corr(u_i, Xb)	= -0.8728	F(10, 47)	= 14.13
		Prob > F	= 0.0000
(Std. Err. adjusted for 48 clusters in state)			
fatal	Coef.	Robust	
		Std. Err.	t P> t  [95% Conf. Interval]
beertax	-.4347195	.2442775	-1.78 0.082 -.9261425 .0567036
spircons	.805857	.1161087	6.94 0.000 .5722764 1.039438
unrate	-.0549084	.011763	-4.67 0.000 -.0785725 -.0312443
perinck	.0882636	.0322971	2.73 0.009 .0232901 .153237
year			
1983	-.0533713	.0312438	-1.71 0.094 -.1162256 .0094831
1984	-.1649828	.0439375	-3.75 0.000 -.2533737 -.076592
1985	-.1997376	.0496167	-4.03 0.000 -.2995535 -.0999218
1986	-.0508034	.0661756	-0.77 0.447 -.1839315 .0823248
1987	-.1000728	.0756768	-1.32 0.192 -.2523149 .0521693
1988	-.134057	.0863572	-1.55 0.127 -.3077854 .0396714
_cons	.1290568	.6238303	0.21 0.837 -1.125928 1.384041
sigma_u	1.0987683		
sigma_e	.14570531		
rho	.98271904	(fraction of variance due to u_i)	

#### 4. 随机效应

以上结果已基本确认了个体效应的存在,但个体效应仍可能以随机效应(RE)的形式存在。随机效应的 Stata 命令为

xtreg y x1 x2 x3,re r theta (随机效应 FGLS)

xtreg y x1 x2 x3,mle (随机效应 MLE)

其中,选择项“r”表示使用聚类稳健标准误;如果使用选择项“vce(cluster id)”也能达到完全相同的效果。选择项“theta”表示显示用于随机效应估计的  $\theta$  值。Breusch and Pagan(1980)提供了一个检验个体效应的 LM 检验(LM test for individual-specific effects),其原假设为“ $H_0 : \sigma_u^2 = 0$ ”,而备择假设为“ $H_1 : \sigma_u^2 \neq 0$ ”。如果拒绝  $H_0$ ,则说明原模型中应该有一个反映个体特性的随机扰动项  $u_i$ ,而不应该使用混合回归。该 LM 检验的 Stata 命令为“xttest0”(在执行命令“xtreg,re”后才能进行)。

```
. xtreg fatal beertax spircons unrate perinck,re r theta
. estimates store RE
```

Random-effects GLS regression	Number of obs	=	336		
Group variable: state	Number of groups	=	48		
R-sq: within = 0.2263	Obs per group: min =	7			
between = 0.0123	avg =	7.0			
overall = 0.0042	max =	7			
corr(u_i, X) = 0 (assumed)	Wald chi2(4)	=	42.30		
theta = .85922063	Prob > chi2	=	0.0000		
(Std. Err. adjusted for 48 clusters in state)					
	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
beertax	.0442768	.1283189	0.35	0.730	-.2072237 .2957772
spircons	.3024711	.1080475	2.80	0.005	.0907019 .5142404
unrate	-.0491381	.009052	-5.43	0.000	-.0668798 -.0313964
perinck	-.0110727	.024949	-0.44	0.657	-.0599718 .0378263
_cons	2.001973	.5226507	3.83	0.000	.977597 3.02635
sigma_u	.41675665				
sigma_e	.15678965				
rho	.87601197	(fraction of variance due to u_i)			

上表最后三行显示,  $\sigma_u = 0.41675665$ ,  $\sigma_e = 0.41675665$ , 而  $\rho \equiv \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = 0.87601197$ 。

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects												
fatal[state,t] = Xb + u[state] + e[state,t]												
Estimated results:												
Var sd = sqrt(Var)												
<table border="1"> <thead> <tr> <th></th><th>Var</th><th>sd = sqrt(Var)</th></tr> </thead> <tbody> <tr> <td>fatal</td><td>.3251209</td><td>.5701938</td></tr> <tr> <td>e</td><td>.024583</td><td>.1567897</td></tr> <tr> <td>u</td><td>.1736861</td><td>.4167567</td></tr> </tbody> </table>		Var	sd = sqrt(Var)	fatal	.3251209	.5701938	e	.024583	.1567897	u	.1736861	.4167567
	Var	sd = sqrt(Var)										
fatal	.3251209	.5701938										
e	.024583	.1567897										
u	.1736861	.4167567										
Test: Var(u) = 0												
chibar2(01) = 550.22												
Prob > chibar2 = 0.0000												

上表显示, LM 检验强烈拒绝“不存在个体随机效应”的原假设, 即认为在“随机效应”与“混合回归”二者之间, 应该选择“随机效应”。

作为对照, 也可以对随机效应模型进行 MLE 估计:

. xtreg fatal beertax spircons unrate perinck, mle nolog  
. estimates store MLE

Random-effects ML regression		Number of obs	=	336
Group variable: state		Number of groups	=	48
Random effects u_i ~ Gaussian		Obs per group: min	=	7
		avg	=	7.0
		max	=	7
		LR chi2(4)	=	67.39
Log likelihood = 12.784035		Prob > chi2	=	0.0000
fatal	Coef.	Std. Err.	z	P> z  [95% Conf. Interval]
beertax	-.1982558	.1505869	-1.32	0.188 -.4934007 .0968891
spircons	.5815301	.0871644	6.67	0.000 .4106909 .7523692
unrate	-.0381467	.0091776	-4.16	0.000 -.0561344 -.0201589
perinck	.0543694	.0223621	2.43	0.015 .0105405 .0981984
_cons	.6479714	.4682047	1.38	0.166 -.2696929 1.565636
/sigma_u	.8559797	.1147673		.6581681 1.113243
/sigma_e	.159379	.0070215		.1461946 .1737525
rho	.9664931	.0096949		.9425673 .9815318

Likelihood-ratio test of sigma\_u=0: chibar2(01)= 479.84 Prob>chibar2 = 0.000

上表显示,随机效应 MLE 的系数估计值与随机效应 FGLS 有所不同,但在性质上依然类似。另外,上表最后一行强烈拒绝原假设“ $H_0 : \sigma_u = 0$ ”,即认为存在个体随机效应,不应进行混合回归。

## 5. 组间估计量

组间估计量(BE)的 Stata 命令为

xtreg y x1 x2 x3, be

如果数据质量不好,可考虑使用组间估计量,但会损失较多的信息量。在此,也把它作为一种对照的估计方法:

```
. xtreg fatal beertax spircons unrate perinck, be  
. estimates store BE
```

Between regression (regression on group means)	Number of obs	=	336
Group variable: state	Number of groups	=	48
R-sq: within = 0.0479	Obs per group: min	=	7
between = 0.4565	avg	=	7.0
overall = 0.2583	max	=	7
	F(4,43)	=	9.03
sd(u_i + avg(e_i)) = .4209489	Prob > F	=	0.0000
fatal	Coef.	Std. Err.	t P> t  [95% Conf. Interval]
beertax	.0740362	.1456333	0.51 0.614 -.2196614 .3677338
spircons	.2997517	.1128135	2.66 0.011 .0722417 .5272618
unrate	.0322333	.038005	0.85 0.401 -.0444111 .1088776
perinck	-.1841747	.0422241	-4.36 0.000 -.2693277 -.0990218
_cons	3.796343	.7502025	5.06 0.000 2.283415 5.309271

## 6. 固定效应还是随机效应:豪斯曼检验

在处理面板数据时,究竟使用固定效应还是随机效应模型,这是一个基本问题。为此,进行豪斯曼检验。豪斯曼检验的 Stata 命令如下:

xtreg y x1 x2 x3, fe (固定效应估计)

```

estimates store FE (存储结果)
xtreg y x1 x2 x3,re (随机效应估计)
estimates store RE (存储结果)
hausman FE RE,constant sigmamore (豪斯曼检验)

```

其中,选择项“constant”表示在比较系数估计值时包括常数项(默认设置不包括常数项),而选择项“sigmamore”表示统一使用更有效率的那个估计量(即随机效应估计量)的方差估计。由于在前面我们已经存储了相关的结果,故可以直接进行豪斯曼检验:

```
. hausman FE RE,constant sigmamore
```

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) FE	(B) RE		
beertax	-.4840728	.0442768	-.5283495	.1462601
spircons	.8169652	.3024711	.514494	.0663028
unrate	-.0290499	-.0491381	.0200882	.0037898
perinck	.1047103	-.0110727	.115783	.0140556
_cons	-.383783	2.001973	-2.385756	.3078253

b = consistent under  $H_0$  and  $H_a$ ; obtained from xtreg  
 B = inconsistent under  $H_a$ , efficient under  $H_0$ ; obtained from xtreg

Test:  $H_0$ : difference in coefficients not systematic

$$\begin{aligned} \text{chi2}(5) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 90.46 \\ \text{Prob>chi2} &= 0.0000 \\ (V_b-V_B &\text{ is not positive definite}) \end{aligned}$$

由于  $p$  值为 0.0000, 故强烈拒绝原假设“ $H_0: u_i$  与  $\mathbf{x}_u, \mathbf{z}_u$  不相关”, 认为应该使用固定效应模型, 而非随机效应模型。

然而, 传统的豪斯曼检验假定, 在  $H_0$  成立的情况下, 随机效应模型是最有效率的(fully efficient)。这意味着,  $u_i$  与  $\varepsilon_u$  都必须是独立同分布的<sup>①</sup>。因此, 如果聚类稳健标准误与普通标准误相差较大(在本例中, 二者约相差一倍), 则传统的豪斯曼检验不适用。事实上, 在 Stata 中进行以上豪斯曼检验时, 如果使用聚类稳健标准误, 比如“`xtreg y x1 x2 x3, fe vce(cluster id)`”, 则 Stata 将无法执行“`hausman FE RE`”命令。

解决方法之一是使用自助法(参见第 19 章)。解决方法之二是进行以下辅助回归:

$$y_{iu} - \hat{\theta}\bar{y}_i = (\mathbf{x}_{iu} - \hat{\theta}\bar{\mathbf{x}}_i)' \boldsymbol{\beta} + (1 - \hat{\theta})\mathbf{z}'_i \boldsymbol{\delta} + (\mathbf{x}_{iu} - \bar{\mathbf{x}}_i)' \boldsymbol{\gamma} + [(1 - \hat{\theta})u_i + (\varepsilon_{iu} - \hat{\theta}\bar{\varepsilon}_i)] \quad (15.22)$$

然后, 使用聚类稳健标准误来检验原假设“ $H_0: \boldsymbol{\gamma} = \mathbf{0}$ ”。为了演示的目的, 首先介绍如何在 Stata 中手工进行此辅助回归(也可直接使用非官方命令 `xtoverid`, 参见下文)<sup>②</sup>。首先, 进行 RE 估计, 得到  $\hat{\theta}$ (使用命令“`xtreg, re`”时, 加上选择项“`theta`”, 即会显示  $\hat{\theta}$  的估计值; 或者使用命令“`scalar theta = e(theta)`”从 RE 回归结果中直接获得  $\hat{\theta}$  的取值); 然后用  $\hat{\theta}$  对模型进行如上变换(生成相应的广义离差变量), 最后进行辅助回归与检验:

```
. quietly xtreg fatal beertax spircons unrate perinck,re
```

① 否则, 即使 FGLS 也不是最有效率的, 因为 FGLS 必须首先估计一些参数, 再进行 GLS 估计。

② 此程序仅适用于平衡面板, 因为非平衡面板会有很多  $\hat{\theta}_i$ , 程序将更为复杂。对于非平衡面板, 可直接使用下文介绍的非官方命令 `xtoverid`。

```

. scalar theta = e(theta)
. global yandxforhausman fatal beertax spircons unrate perinck
. sort state
. foreach x of varlist $yandxforhausman{
2. by state:egen mean`x'=mean(`x')
3. gen md`x'=`x'-mean`x'
4. gen red`x'=`x'-theta * mean`x'
5. }
. quietly reg redfatal redbeertax redspircons redunrate redperinck
mdbeertax mdspircons mdunrate mdperinck, vce(cluster state)
. test mdbeertax mdspircons mdunrate mdperinck

```

```

( 1)  mdbeertax = 0
( 2)  mdspircons = 0
( 3)  mdunrate = 0
( 4)  mdperinck = 0

F(  4,      47) =    15.86
Prob > F =      0.0000

```

其中,“global”为“全局宏”(global macro),即用“\$yandxforhausman”来表示“fatal beertax spircons unrate perinck”。如此定义之后,但凡看见“\$yandxforhausman”,就视同“fatal beertax spircons unrate perinck”。“foreach”为循环语句,表示对变量名单“varlist”上的每个变量“x”都进行同样的操作。“egen”为命令“gen”的扩展(extension),使用“egen”可以调用更多的函数。新变量名的前缀“md”表示原变量的“mean-differences”,即 $(\bar{x}_u - \bar{x}_i)$ ;而新变量名的前缀“red”表示原变量的“RE-differences”,即 $(x_u - \hat{\theta} \bar{x}_i)$ 。由于该检验的p值为0.0000,故强烈拒绝“ $H_0 : \gamma = 0$ ”,即拒绝随机效应,认为应该使用固定效应模型。

除了使用以上手程序,更直接的方法是下载非官方命令xtoverid来进行此辅助回归。此处“overid”指的是过度识别检验(overidentification test),因为随机效应模型与固定效应模型相比,前者多了“个体异质性 $u_i$ 与解释变量不相关”的约束条件,也可视为过度识别条件。

```
. ssc install xtoverid (下载安装命令 xtoverid)
```

在使用命令xtoverid之前,须先以稳健标准误来执行命令“xtreg,re”。

```
. quietly xtreg fatal beertax spircons unrate perinck, re r
. xtoverid
```

```
Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re robust cluster(state)
Sargan-Hansen statistic 63.427 Chi-sq(4) P-value = 0.0000
```

命令xtoverid汇报的 $\chi^2(4)$ 统计量为63.427,p值为0.0000。根据第5章附录,F分布与 $\chi^2$ 分布在大样本下是等价的。根据上文的手程序,F统计量为 $F(4,47) = 15.86$ ,则相应的 $\chi^2$ 统计量为 $\chi^2(4) = 4 \times F(4,47) = 4 \times 15.86 = 63.44$ 。因此,使用命令xtoverid的结果与手程序完全一致(只是小数点的精确位数不同)。使用命令xtoverid的好处在于,它也适用于非平衡面板。

为了便于比较,将以上各主要方法的系数估计值及标准误列表:

```
. estimates table OLS FE_robust FE_TW RE BE,b se
```

Variable	OLS	FE_robust	FE_TW	RE	BE
beertax	.0971997 .11688401	-.48407277 .22187542	-.43471948 .24427755	.04427676 .12046126	.07403619 .14563334
spircons	.16234707 .10709883	.81696515 .12726271	.805857 .11610873	.30247114 .06429535	.29975174 .11281346
unrate	-.02910139 .02086869	-.02904993 .00945812	-.05490839 .01176299	-.0491381 .00981968	.03223326 .03800499
perinck	-.15842906 .03714042	.10471027 .0341455	.08826358 .03229713	-.01107273 .01947461	-.18417475 .04222407
year2			-.05337126 .03124376		
year3			-.16498284 .04393752		
year4			-.19973763 .04961666		
year5			-.05080338 .0661756		
year6			-.10007279 .0756768		
year7			-.13405703 .08635725		
_cons	4.1186735 .67657457	-.38378301 .70917383	.12905677 .62383029	2.0019735 .38112466	3.796343 .75020251

legend: b/se

从上表可知,不同方法的系数估计值差别较大。其中,固定效应(FE\_robust, FE\_TW)的系数估计值与其他估计法(OLS, RE, BE)的差别最大。另一方面,在固定效应模型中加入时间效应(FE\_TW),对估计系数的影响不大。

## 习 题

15.1 证明当  $T = 2$  时,  $\hat{\beta}_{FD} = \hat{\beta}_{FE}$ 。

15.2 证明经过广义差分变换后的随机效应模型,其扰动项不再存在自相关,即

$$\text{Cov}[(1-\theta)u_i + (\varepsilon_{it} - \theta\bar{\varepsilon}_i), (1-\theta)u_j + (\varepsilon_{jt} - \theta\bar{\varepsilon}_j)] = 0, \quad t \neq s \quad (15.23)$$

其中,  $\theta = 1 - \frac{\sigma_\varepsilon}{(T\sigma_u^2 + \sigma_\varepsilon^2)^{1/2}}$ 。

15.3 面板数据集 mus08psidextract.dta 包含 595 名美国工人 1976—1982 年的以下变量:lwage(工资对数), ed(教育年限), exp(工龄), exp2(exp 的平方), wks(工作周数, weeks worked)。考虑以下模型:

$$\text{lwage}_i = \alpha + \beta_1 \exp_i + \beta_2 \exp2_i + \beta_3 \text{wks}_i + \beta_4 \text{ed}_i + u_i + \varepsilon_i \quad (15.24)$$

- (1) 进行混合回归;
- (2) 对随机效应模型进行 FGLS 估计,检验是否存在个体效应;
- (3) 对随机效应模型进行 MLE 估计;
- (4) 对固定效应模型计算组内估计量;
- (5) 对固定效应模型进行 LSDV 估计,检验是否存在个体效应;
- (6) 估计带时间效应的固定效应模型;
- (7) 计算一阶差分估计量;
- (8) 计算组间估计量;
- (9) 进行传统的豪斯曼检验;
- (10) 进行稳健的豪斯曼检验。

# 第 16 章 长面板与动态面板

## 16.1 长面板的估计策略

上一章关注短面板。对于短面板,由于时间维度  $T$  较小,每位个体的信息较少,无法探讨扰动项  $\{\varepsilon_{it}\}$  是否存在自相关,故一般假设  $\{\varepsilon_{it}\}$  独立同分布。对于长面板(long panel),由于  $T$  较大,信息较多,故可以放松这个假定,考虑  $\{\varepsilon_{it}\}$  可能存在的异方差与自相关。在长面板中,由于  $n$  相对于  $T$  较小,对于可能存在的固定效应,只要加入个体虚拟变量即可(即 LSDV 法)。对于时间效应,可以通过加上时间趋势项或其平方项来控制(由于  $T$  较大,如果加上时间虚拟变量,则可能损失较多的自由度)。

为此,考虑以下模型:

$$y_{it} = \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it} \quad (16.1)$$

其中,  $\mathbf{x}_{it}$  可以包括常数项、时间趋势项(或其平方项)、个体虚拟变量以及不随时间变化的解释变量  $z_i$ 。下面考虑扰动项  $\{\varepsilon_{it}\}$  存在异方差或自相关的几种情形。

(1) 记个体  $i$  的扰动项方差为  $\sigma_i^2 \equiv \text{Var}(\varepsilon_{it})$ 。如果存在  $\sigma_i^2 \neq \sigma_j^2 (i \neq j)$ , 则称扰动项  $\{\varepsilon_{it}\}$  存在“组间异方差”(groupwise heteroskedasticity)。

(2) 如果存在  $\text{Cov}(\varepsilon_{it}, \varepsilon_{is}) \neq 0 (t \neq s, \forall i)$ , 则称扰动项  $\{\varepsilon_{it}\}$  存在“组内自相关”(autocorrelation within panel)。

(3) 如果存在  $\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) \neq 0 (i \neq j, \forall t)$ , 则称扰动项  $\{\varepsilon_{it}\}$  存在“组间同期相关”(contemporaneous correlation)或“截面相关”(cross-sectional correlation)。比如,对于省际面板数据(provincial panel data),相邻省份之间的同期经济活动可能通过贸易或投资相互影响。这种相关也称为“空间相关”(spatial correlation)。

对于扰动项  $\{\varepsilon_{it}\}$  可能存在的组间异方差、组内自相关或组间同期相关,主要有两类处理方法。方法一,继续使用 OLS(即 LSDV)来估计系数,只对标准误差进行校正(即下文的面板校正标准误差)。方法二,对异方差或自相关的具体形式进行假设,然后使用可行广义最小二乘法(FGLS)进行估计。

## 16.2 面板校正标准误

对于模型(16.1),即使扰动项  $\{\varepsilon_{it}\}$  存在组间异方差或组间同期相关,OLS(即 LSDV)也依然是一致的。此时,只要使用“组间异方差、组间同期相关”稳健的标准误差即可,即“面板校正标准误差”(Panel-Corrected Standard Error,简记 PCSE)。相应的 Stata 命令为

`xtpcse y x1 x2 x3, hetonly`

其中,选择项“`hetonly`”表示存在组间异方差,但不存在组间同期相关;默认既存在组间异方差,也存在组间同期相关。

下面以数据集 `mus08cigar.dta` 为例<sup>①</sup>。该面板数据集包含了美国 10 个州 1963—1992 年有关香烟消费量的相关变量:`lnc`(人均香烟销售量的对数),`lnp`(实际香烟价格的对数)<sup>②</sup>,`lnpmin`(相邻州最低香烟价格的对数),`lny`(人均可支配收入的对数),`state`(州),`year`(年)。由于  $n = 10$ ,而  $T = 30$ ,故这是一个长面板。

考虑估计如下香烟需求函数:

$$\ln c_i = \beta_0 + \beta_1 \ln p_i + \beta_2 \ln p_{\min i} + \beta_3 \ln y_i + u_i + \varepsilon_i \quad (16.2)$$

其中,变量 `lnp` 代表价格效应,`lnpmin` 代表替代效应(可以购买价格更低的邻州香烟),而 `lny` 则代表收入效应。

```
.use mus08cigar.dta,clear
```

为了考虑时间效应,生成时间趋势变量 `t`<sup>③</sup>,然后用 LSDV 法估计双向固定效应模型(作为对比,先不考虑组间异方差或同期相关):

```
.gen t=year-62
.reg lnc lnp lnpmin lny i.state t,vce(cluster state)
.estimates store OLS
```

Linear regression						Number of obs = 300
						F( 3, 9) = .
						Prob > F = .
						R-squared = 0.7203
						Root MSE = .11203
(Std. Err. adjusted for 10 clusters in state)						
lnc	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lnp	-1.027181	.4412156	-2.33	0.045	-2.02528	-.0290819
lnpmin	.5100582	.2627025	1.94	0.084	-.0842161	1.104332
lny	.4975365	.4139086	1.20	0.260	-.4387899	1.433863
state						
2	-.0773908	.0594636	-1.30	0.225	-.2119069	.0571252
3	.088557	.0090317	9.81	0.000	.0681259	.1089881
4	-.1809375	.1695839	-1.07	0.314	-.5645629	.202688
5	-.1066138	.2409767	-0.44	0.669	-.651741	.4385135
6	.2177434	.1063333	2.05	0.071	-.0227993	.458286
7	.115543	.1959351	0.59	0.570	-.327693	.5587791
8	.1068277	.124702	0.86	0.414	-.1752679	.3889233
9	.0433207	.04618	0.94	0.373	-.0611458	.1477872
10	-.133583	.0333562	-4.00	0.003	-.2090399	-.0581261
t	-.0429824	.0305275	-1.41	0.193	-.1120405	.0260757
_cons	3.488747	4.03674	0.86	0.410	-5.642993	12.62049

上表显示,有些州虚拟变量很显著,即存在固定效应,应允许每个州拥有自己的截距项;而时间效应则不太显著,  $p$  值高达 0.19。然而,上表中的稳健标准误未考虑可能存在的组间异方差与

① 此例来自 Cameron and Trivedi (2009)。

② 不同州的香烟税不同,故香烟的实际价格不同。

③ 为了减少样本容量的损失,此处不用年度虚拟变量。

组间同期相关。为此,使用面板校正标准误进行估计。

```
. xtpcse lnc lnp lnpmin lny i.state t
. estimates store PCSE
```

Linear regression, correlated panels corrected standard errors (PCSEs)						
Group variable:	state	Number of obs	=	300		
Time variable:	year	Number of groups	=	10		
Panels:	correlated (balanced)	Obs per group:	min =	30		
Autocorrelation:	no autocorrelation		avg =	30		
			max =	30		
Estimated covariances	= 55	R-squared	= 0.7203			
Estimated autocorrelations	= 0	Wald chi2(13)	= 2147.36			
Estimated coefficients	= 14	Prob > chi2	= 0.0000			
		Panel-corrected				
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnc						
lnp	-1.027181	.1332425	-7.71	0.000	-1.288332	-.7660305
lnpmin	.5100582	.1393134	3.66	0.000	.237009	.7831074
lny	.4975365	.1804528	2.76	0.006	.1438554	.8512176
state						
2	-.0773908	.0389125	-1.99	0.047	-.153658	-.0011237
3	.088557	.011143	7.95	0.000	.0667171	.1103968
4	-.1809375	.0829463	-2.18	0.029	-.3435093	-.0183657
5	-.1066138	.1018461	-1.05	0.295	-.3062284	.0930008
6	.2177434	.049543	4.40	0.000	.1206408	.3148459
7	.115543	.0980092	1.18	0.238	-.0765516	.3076376
8	.1068277	.0483694	2.21	0.027	.0120255	.20163
9	.0433207	.0202198	2.14	0.032	.0036907	.0829508
10	-.133583	.0221487	-6.03	0.000	-.1769937	-.0901722
t	-.0429824	.014234	-3.02	0.003	-.0708806	-.0150843
_cons	3.488747	1.637622	2.13	0.033	.2790665	6.698428

从上表可知,命令 xtpcse 的估计系数与命令 reg 的估计系数完全一样,只是标准误不同。在本例中,面板校正标准误差似乎反而更小。

### 16.3 仅解决组内自相关的 FGLS

对于模型(16.1),假设扰动项  $\varepsilon_u$  服从 AR(1) 过程,即

$$\varepsilon_u = \rho_i \varepsilon_{i,t-1} + v_u \quad (16.3)$$

其中,  $|\rho_i| < 1$ ,  $\{v_u\}$  为独立同分布的,且期望为 0。如果  $\rho_i = \rho$  ( $i = 1, \dots, n$ ), 则所有个体的扰动项都服从自回归系数相同的 AR(1) 过程。使用 Prais-Winsten 估计法对原模型进行广义差分变换(参见第 8 章),即可得到 FGLS 估计量。相应的 Stata 命令为

```
xtpcse y x1 x2 x3, corr(ar1) corr(psar1)
```

其中,选择项“corr(ar1)”对应于  $\rho_i = \rho$  的情形;而选择项“corr(psar1)”则允许每个面板有自己的  $\rho_i$ , psar1 指 panel-specific ar1。如果  $T$  并不比  $n$  大很多,则建议使用选择项“corr(ar1)”,即约束每个面板(个体)的自回归系数  $\rho$  均相等,因为时间维度  $T$  可能无法提供足够的信息来分别估计每个面板自己的  $\rho_i$ 。

如果既不指定 corr(ar1),也不指定 corr(psar1),则进行 OLS 回归。无论使用 OLS 还是 Prais-Winsten 估计法,命令 xtpcse 均提供面板校正标准误(在组间异方差与同期相关的情况下依然成立)。

继续上面的实例。考虑存在组内自相关的情形,但要求各组的自回归系数相同:

```
. xtpcse lnc lnp lnpm lnny i.state t,corr(ar1)
. estimates store AR1
```

Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)						
Group variable:	state	Number of obs	=	300		
Time variable:	year	Number of groups	=	10		
Panels:	correlated (balanced)	Obs per group:	min =	30		
Autocorrelation:	common AR(1)		avg =	30		
			max =	30		
Estimated covariances	=	55	R-squared	=	0.9824	
Estimated autocorrelations	=	1	Wald chi2(13)	=	504.30	
Estimated coefficients	=	14	Prob > chi2	=	0.0000	
Panel-corrected						
lnc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnp	-.3440621	.0571589	-6.02	0.000	-.4560914	-.2320328
lnpm	.1000875	.0736082	1.36	0.174	-.044182	.2443569
lny	.5661073	.1566725	3.61	0.000	.2590348	.8731798
state						
2	-.0784486	.0519567	-1.51	0.131	-.1802818	.0233846
3	.0954506	.0261747	3.65	0.000	.0441491	.146752
4	-.207494	.0810219	-2.56	0.010	-.366294	-.048694
5	-.1811666	.0977461	-1.85	0.064	-.3727454	.0104123
6	.185644	.0554283	3.35	0.001	.0770065	.2942816
7	.0701358	.120881	0.58	0.562	-.1667866	.3070582
8	.0397148	.0507738	0.78	0.434	-.0598001	.1392297
9	.0231754	.025083	0.92	0.356	-.0259863	.0723371
10	-.1289763	.0387423	-3.33	0.001	-.2049099	-.0530428
t	-.0498643	.0121104	-4.12	0.000	-.0736003	-.0261283
_cons	1.746348	1.305801	1.34	0.181	-.8129742	4.305671
rho	.7936188					

从上表可知,是否考虑组内自相关,对估计系数有较大影响。下面考虑允许各组自回归系数不同的组内自相关情形:

```
. xtpcse lnc lnp lnpm lnny i.state t,corr(psar1)
. estimates store PSAR1
```

Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)						
Group variable: state		Number of obs = 300				
Time variable: year		Number of groups = 10				
Panels: correlated (balanced)		Obs per group: min = 30				
Autocorrelation: panel-specific AR(1)		avg = 30				
		max = 30				
Estimated covariances	= 55	R-squared	= 0.9954			
Estimated autocorrelations	= 10	Wald chi2(12)	= 855.24			
Estimated coefficients	= 13	Prob > chi2	= 0.0000			
<hr/>						
lnc	Panel-corrected					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnp	-.2960784	.0539584	-5.49	0.000	-.401835	-.1903218
lnpmin	.0507118	.0673426	0.75	0.451	-.0812772	.1827009
lny	.5326598	.1350574	3.94	0.000	.2679522	.7973674
state						
2	-.0639957	.0802776	-0.80	0.425	-.2213369	.0933456
3	0	(omitted)				
4	-.187228	.0981734	-1.91	0.057	-.3796444	.0051885
5	-.163645	.1128024	-1.45	0.147	-.3847336	.0574436
6	.1980726	.0809248	2.45	0.014	.0394629	.3566822
7	.0796166	.1709107	0.47	0.641	-.2553622	.4145954
8	.0492827	.0817016	0.60	0.546	-.1108495	.2094149
9	.0263335	.0630703	0.42	0.676	-.097282	.149949
10	-.1229701	.065134	-1.89	0.059	-.2506303	.0046901
t	-.0488194	.0103959	-4.70	0.000	-.0691949	-.0284439
_cons	2.015998	1.142949	1.76	0.078	-.2241421	4.256138
rhos	= .9414582	.66151	1	.7045066	.7351521	...
						.6905643

其中,第3个州的州虚拟变量系数估计值及其标准误,由于过于接近于0而被Stata自动删除(omitted)。为了便于比较,将以上各种方法的系数估计值及标准误列表:

```
.esttab OLS PCSE AR1 PSAR1,r2 se mtitles star(* 0.1 ** 0.05 *** 0.01)
```

	(1) OLS	(2) PCSE	(3) AR1	(4) PSAR1
lnp	-1.027** (0.441)	-1.027*** (0.133)	-0.344*** (0.0572)	-0.296*** (0.0540)
lnpmin	0.510* (0.263)	0.510*** (0.139)	0.100 (0.0736)	0.0507 (0.0673)
lny	0.498 (0.414)	0.498*** (0.180)	0.566*** (0.157)	0.533*** (0.135)
1b.state	0 (.)	0 (.)	0 (.)	0 (.)
2.state	-0.0774 (0.0595)	-0.0774** (0.0389)	-0.0784 (0.0520)	-0.0640 (0.0803)
3.state	0.0886*** (0.00903)	0.0886*** (0.0111)	0.0955*** (0.0262)	
4.state	-0.181 (0.170)	-0.181** (0.0829)	-0.207** (0.0810)	-0.187* (0.0982)
5.state	-0.107 (0.241)	-0.107 (0.102)	-0.181* (0.0977)	-0.164 (0.113)
6.state	0.218* (0.106)	0.218*** (0.0495)	0.186*** (0.0554)	0.198** (0.0809)
7.state	0.116 (0.196)	0.116 (0.0980)	0.0701 (0.121)	0.0796 (0.171)
8.state	0.107 (0.125)	0.107** (0.0484)	0.0397 (0.0508)	0.0493 (0.0817)
9.state	0.0433 (0.0462)	0.0433** (0.0202)	0.0232 (0.0251)	0.0263 (0.0631)
10.state	-0.134*** (0.0334)	-0.134*** (0.0221)	-0.129*** (0.0387)	-0.123* (0.0651)
t	-0.0430 (0.0305)	-0.0430*** (0.0142)	-0.0499*** (0.0121)	-0.0488*** (0.0104)
3o.state				0 (.)
_cons	3.489 (4.037)	3.489** (1.638)	1.746 (1.306)	2.016* (1.143)
N	300	300	300	300
R-sq	0.720	0.720	0.982	0.995

Standard errors in parentheses  
\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

从上表可知, AR1 与 PSAR1 的系数估计值比较接近, 但与 OLS 的结果相去甚远。究竟选择 AR1 还是 OLS, 需要对组内自相关进行检验, 参见下文。

## 16.4 全面 FGLS

虽然命令 `xtpcse` 提供了组间异方差与同期相关稳健的面板校正标准误差,但在进行 FGLS 估计时仅针对组内自相关,并未考虑组间异方差或同期相关。更为全面的 FGLS 估计则同时考虑这三个因素。为此,可先对方程(16.1)进行 OLS 估计,然后使用残差  $\{e_{it}\}$  来估计  $\sigma_{it}$  的协方差矩阵,以此进行 FGLS 估计。还可进行迭代 FGLS 估计,即使用 FGLS 的残差再进行 FGLS 估计,不断迭代,直至收敛。相应的 Stata 命令为

```
xtgls y x1 x2 x3, panels(option) corr(option) igls
```

其中,选择项“`panels(iid)`”假定不同个体的扰动项为独立同分布的,“`panels(het)`”假定不同个体的扰动项相互独立但可以有不同的方差,而“`panels(corr)`”假定不同个体的扰动项同期相关且有不同的方差。选择项“`corr(ar1)`”对应于“ $\rho_i = \rho$ ”的组内自相关情形,而“`corr(psar1)`”则允许每个面板有自己的自回归系数。选择项“`igls`”表示使用迭代式 FGLS。

继续以数据集 `mus08cigar.dta` 为例。首先,考虑同时存在组间异方差、同期相关以及组内自相关(自回归系数相同)的情形。

```
. xtgls lnc lnp lnpmin lny i.state t, panels(corr) corr(ar1)
```

Cross-sectional time-series FGLS regression						
Coefficients: generalized least squares						
Panels: heteroskedastic with cross-sectional correlation						
Correlation: common AR(1) coefficient for all panels (0.7967)						
Estimated covariances	=	55	Number of obs	=	300	
Estimated autocorrelations	=	1	Number of groups	=	10	
Estimated coefficients	=	14	Time periods	=	30	
			Wald chi2(13)	=	797.11	
			Prob > chi2	=	0.0000	
lnc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnp	-.3629823	.0239353	-15.17	0.000	-.4098947	-.3160699
lnpmin	.0258594	.0302743	0.85	0.393	-.033477	.0851959
lny	.5116673	.0733079	6.98	0.000	.3679865	.655348
state						
2	-.0719573	.0471313	-1.53	0.127	-.164333	.0204185
3	.0913241	.0260485	3.51	0.000	.04027	.1423782
4	-.1844595	.056133	-3.29	0.001	-.2944782	-.0744407
5	-.1466676	.0636223	-2.31	0.021	-.271365	-.0219701
6	.198174	.0424493	4.67	0.000	.114975	.2813731
7	.0861722	.1025173	0.84	0.401	-.1147579	.2871024
8	.0548621	.0386937	1.42	0.156	-.0209761	.1307003
9	.0227095	.0204346	1.11	0.266	-.0173416	.0627607
10	-.1279293	.0359818	-3.56	0.000	-.1984523	-.0574063
t	-.0448556	.0057821	-7.76	0.000	-.0561884	-.0335228
_cons	2.550302	.5964829	4.28	0.000	1.381217	3.719388

其次,考虑各组自回归系数不同的情形。

```
. xtgls lnc lnp lnpmin lny i.state t, panels(corr) corr(psar1)
```

## Cross-sectional time-series FGLS regression

Coefficients: generalized least squares  
 Panels: heteroskedastic with cross-sectional correlation  
 Correlation: panel-specific AR(1)

Estimated covariances	=	55	Number of obs	=	300
Estimated autocorrelations	=	10	Number of groups	=	10
Estimated coefficients	=	14	Time periods	=	30
			Wald chi2(13)	=	1246.78
			Prob > chi2	=	0.0000

lnc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnp	-.3541126	.024395	-14.52	0.000	-.4019259 -.3062994
lnpmin	.0185409	.030352	0.61	0.541	-.040948 .0780297
lny	.5459469	.0712941	7.66	0.000	.406213 .6856808
state					
2	-.1229944	.0660784	-1.86	0.063	-.2525056 .0065169
3	-.3649205	.2134793	-1.71	0.087	-.7833323 .0534913
4	-.2405925	.0717201	-3.35	0.001	-.3811613 -.1000236
5	-.2086117	.0793916	-2.63	0.009	-.3642163 -.0530071
6	.1417767	.0643671	2.20	0.028	.0156194 .2679339
7	-.0337763	.1435772	-0.24	0.814	-.3151824 .2476297
8	-.00008845	.0647305	-0.01	0.989	-.1277539 .1259849
9	-.0350394	.0544958	-0.64	0.520	-.1418491 .0717703
10	-.1888756	.0569821	-3.31	0.001	-.3005584 -.0771928
t					
cons	-.0487174	.0055724	-8.74	0.000	-.059639 -.0377957
	2.356635	.5905045	3.99	0.000	1.199267 3.514002

由以上两表可知,各种 FGLS 估计结果不尽相同。究竟使用何种 FGLS 估计,则取决于对组间异方差、组内自相关与组间同期相关的检验。总的来说,“OLS + 面板校正标准误差”最为稳健,全面的 FGLS 估计最有效率,而仅解决组内自相关的 FGLS 则介于二者之间。有关如何在稳健性与有效性之间进行权衡,参见第 7 章第 4 节。

## 16.5 组间异方差的检验

Greene (2000, p. 598) 提供了一个对组间异方差的沃尔德检验。考虑原假设“不同个体的扰动项方差均相等”,即  $H_0: \sigma_i^2 = \sigma^2 (i = 1, \dots, n)$ 。在原假设成立的前提下,

$$\frac{\hat{\sigma}_i^2 - \sigma^2}{\sqrt{\text{Var}(\hat{\sigma}_i^2)}} \xrightarrow{d} N(0, 1) \quad (16.4)$$

其中,  $\hat{\sigma}_i^2 = \sum_{t=1}^T e_{it}^2 / T$  为  $\sigma^2$  的一致估计量,  $e_{it}$  为对应于  $\varepsilon_{it}$  的残差。将上式平方可得

$$\frac{(\hat{\sigma}_i^2 - \sigma^2)^2}{\text{Var}(\hat{\sigma}_i^2)} \xrightarrow{d} \chi^2(1) \quad (16.5)$$

另一方面,  $\text{Var}(\hat{\sigma}_i^2)$  的一致估计量为

$$\widehat{\text{Var}(\hat{\sigma}_i^2)} = \frac{1}{T} \frac{1}{T-1} \sum_{t=1}^T (e_{it}^2 - \hat{\sigma}_i^2)^2 \quad (16.6)$$

进一步假设每位个体的扰动项相互独立,可得沃尔德统计量:

$$W \equiv \sum_{i=1}^n \frac{(\hat{\sigma}_i^2 - \sigma^2)^2}{\text{Var}(\hat{\sigma}_i^2)} \xrightarrow{d} \chi^2(n) \quad (16.7)$$

此沃尔德检验可通过非官方命令 `xttest3` 来实现,该命令只能在运行命令“`xtreg, fe`”或“`xtgls`”之后才能使用。

```
. ssc install xttest3      (下载安装命令 xttest3)
. quietly xtreg lnc lnp lnpmin lny i.state t,r fe
. xttest3
```

```
Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (10) =      378.90
Prob>chi2 =      0.0000
```

```
. quietly xtgls lnc lnp lnpmin lny i.state t
. xttest3
```

```
Modified Wald test for groupwise heteroskedasticity
in cross-sectional time-series FGLS regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (10) =      378.90
Prob>chi2 =      0.0000
```

由以上结果可知,无论在命令“`xtreg, fe`”或“`xtgls`”之后使用命令 `xttest3`,其统计量完全一样,均强烈拒绝同方差的原假设,认为存在组间异方差。

## 16.6 组内自相关的检验

Wooldridge (2002, p. 282 – 283) 提供了一个对组内自相关的沃尔德检验。考虑原假设“不存在组内自相关”,即  $H_0: \text{Cov}(\varepsilon_{it}, \varepsilon_{is}) = 0 (t \neq s, \forall i)$ 。给定个体  $i$ ,对方程(16.1)进行一阶差分可得

$$\Delta y_{it} = \Delta \mathbf{x}'_{it} \boldsymbol{\beta} + \Delta \varepsilon_{it} \quad (16.8)$$

在  $\{\varepsilon_{it}\}$  不存在组内自相关的原假设下,方程(16.8)扰动项  $\Delta \varepsilon_{it}$  的方差与自协方差(autocovariance)分别为

$$\text{Var}(\Delta \varepsilon_{it}) = \text{Var}(\varepsilon_{it} - \varepsilon_{i,t-1}) = \text{Var}(\varepsilon_{it}) + \text{Var}(\varepsilon_{i,t-1}) = 2\sigma_\varepsilon^2 \quad (16.9)$$

$$\begin{aligned} \text{Cov}(\Delta \varepsilon_{it}, \Delta \varepsilon_{i,t-1}) &= \text{Cov}(\varepsilon_{it} - \varepsilon_{i,t-1}, \varepsilon_{i,t-1} - \varepsilon_{i,t-2}) \\ &= -\text{Cov}(\varepsilon_{i,t-1}, \varepsilon_{i,t-1}) = -\text{Var}(\varepsilon_{i,t-1}) = -\sigma_\varepsilon^2 \end{aligned} \quad (16.10)$$

故自相关系数为

$$\text{Corr}(\Delta \varepsilon_{it}, \Delta \varepsilon_{i,t-1}) = \frac{\text{Cov}(\Delta \varepsilon_{it}, \Delta \varepsilon_{i,t-1})}{\text{Var}(\Delta \varepsilon_{it})} = \frac{-\sigma_\varepsilon^2}{2\sigma_\varepsilon^2} = -0.5 \quad (16.11)$$

记  $\Delta \varepsilon_{it}$  的样本值为  $e_{it}$ (即一阶差分回归的残差),对  $e_{it}$  进行一阶自回归:

$$e_{it} = \rho e_{i,t-1} + error_{it} \quad (i=1, \dots, n; t=3, \dots, T) \quad (16.12)$$

然后对原假设“ $H_0: \rho = -0.5$ ”进行沃尔德检验( $t$ 或 $F$ 检验)。在Stata中,可以通过下载命令xtserial来实现这个检验:

```
net install st0039      (下载安装命令 xtserial)
```

```
xtserial y x1 x2 x3, output
```

其中,选择项“output”表示显示一阶差分回归的结果。

继续上面的例子。由于该命令不接受以“i.state”表示州虚拟变量,故首先生成州虚拟变量,以state1-state10来表示。

```
. tab state, gen(state)
```

U.S. state	Freq.	Percent	Cum.
1	30	10.00	10.00
2	30	10.00	20.00
3	30	10.00	30.00
4	30	10.00	40.00
5	30	10.00	50.00
6	30	10.00	60.00
7	30	10.00	70.00
8	30	10.00	80.00
9	30	10.00	90.00
10	30	10.00	100.00
Total	300	100.00	

```
. xtserial lnc lnp lnpm lnny state2-state10 t
```

```
Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
F( 1,    9) =     89.304
Prob > F =    0.0000
```

结果强烈拒绝“不存在一阶组内自相关”的原假设。由于命令xtserial不允许使用“i.state”,故只能使用上文生成的州虚拟变量“state2-state10”。

## 16.7 组间同期相关的检验

考虑原假设“不存在组间同期相关”,即 $H_0: \text{Cov}(e_i, e_j) = 0 (i \neq j, \forall t)$ 。如果此原假设成立,则根据残差计算的不同个体扰动项的相关系数应接近于0。如果将这些相关系数排成一个矩阵,即“残差相关系数矩阵”(correlation matrix of residuals),则该矩阵非主对角线元素应离0不远。根据残差相关系数矩阵,可以设计以下几种检验。

Greene (2000, p. 601)提供了一个对组间同期相关的Breusch-Pagan LM检验,可由非官方命令xttest2来实现。此命令只能在运行命令“xtreg, fe”,“xtgls”或“ivreg2”之后使用。

```
. ssc install xttest2①      (下载安装命令 xttest2)
. quietly xtreg lnc lnp lnpmin lny t,fe
. xttest2
```

Correlation matrix of residuals:

	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10
e1	1.0000									
e2	-0.0937	1.0000								
e3	0.9592	-0.0621	1.0000							
e4	-0.4242	0.3875	-0.4670	1.0000						
e5	-0.5426	0.3441	-0.5872	0.5519	1.0000					
e6	0.0245	0.5696	-0.0405	0.5177	0.5805	1.0000				
e7	-0.7434	0.4153	-0.7509	0.5701	0.8446	0.4893	1.0000			
e8	0.5650	0.5380	0.5281	0.1007	-0.2150	0.4899	-0.3263	1.0000		
e9	0.8337	0.2859	0.8507	-0.2972	-0.3914	0.1548	-0.5800	0.7129	1.0000	
e10	0.7510	0.3314	0.7628	0.0002	-0.1575	0.2508	-0.4293	0.6318	0.8345	1.0000

Breusch-Pagan LM test of independence: chi2(45) = 376.963, Pr = 0.0000  
Based on 30 complete observations over panel units

上表显示,Breusch-Pagan LM 统计量的  $p$  值为 0.000 0,故强烈拒绝“无同期相关”的原假设,认为存在同期相关。

以上 LM 检验的缺陷是,它仅适用于长面板。而检验组间同期相关的另一非官方命令 xtcsd 则也适用于  $n$  大  $T$  小的短面板,其中“csd”代表“cross-sectional dependence”(截面相关)。其原假设也是“不存在组间同期相关”。它包括三种检验方法,分别由 Friedman (1937),Frees (1995,2004),以及 Pesaran (2004) 所提出;前两种为半参数检验(参见第 27 章),而后一种为通常的参数检验。命令 xtcsd 须在运行命令 xtreg 之后才能使用。

```
ssc install xtcsd          (下载安装命令 xtcsd)
xtcsd,pesaran abs show   (Pesaran (2004) 的检验,统计量服从标准正态分布)
xtcsd,friedman abs show (Friedman (1937) 的检验,统计量服从  $\chi^2$  分布)
xtcsd,frees abs show    (Frees (1995,2004) 的检验)
```

其中,选择项“show”表示显示残差的相关系数矩阵,而选择项“abs”表示显示该矩阵非主对角线元素的绝对值之平均。回到数据集 mus08cigar.dta 的例子:

```
. quietly xtreg lnc lnp lnpmin lny t,fe
. xtcsd,pes
```

Pesaran's test of cross sectional independence = 7.159, Pr = 0.0000

```
. xtcsd,fri
```

Friedman's test of cross sectional independence = 81.481, Pr = 0.0000

```
. xtcsd,fre abs show
```

<sup>①</sup> 如果以较早版本的命令 xttest2 来检验此数据集,可能会出现出错信息,诸如“Correlation matrix of residuals is singular,not possible with test”。此时,应使用命令“ssc install xttest2,replace”来下载最新版的命令 xttest2 (选择项“replace”表示替换旧版命令)。

Correlation matrix of residuals:										
	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
r1	1.0000									
r2	-0.0937	1.0000								
r3	0.9592	-0.0621	1.0000							
r4	-0.4242	0.3875	-0.4670	1.0000						
r5	-0.5426	0.3441	-0.5872	0.5519	1.0000					
r6	0.0245	0.5696	-0.0405	0.5177	0.5805	1.0000				
r7	-0.7434	0.4153	-0.7509	0.5701	0.8446	0.4893	1.0000			
r8	0.5650	0.5380	0.5281	0.1007	-0.2150	0.4899	-0.3263	1.0000		
r9	0.8337	0.2859	0.8507	-0.2972	-0.3914	0.1548	-0.5800	0.7129	1.0000	
r10	0.7510	0.3314	0.7628	0.0002	-0.1575	0.2508	-0.4293	0.6318	0.8345	1.0000

Frees' test of cross sectional independence =	2.306
-----	
Critical values from Frees' Q distribution	
alpha = 0.10 :	0.0861
alpha = 0.05 :	0.1119
alpha = 0.01 :	0.1598
Average absolute value of the off-diagonal elements =	
0.466	

以上三种检验的  $p$  值均小于 0.01, 故强烈拒绝“无组间同期相关”的原假设, 认为存在组间同期相关。上表最后一行显示, 非主对角线元素绝对值的平均值高达 0.466。命令“xtcsd, fre abs show”所显示的残差相关系数矩阵与命令“xttest2”所显示完全相同。另外, 还可以把面板数据排成“宽形”(wide form), 然后使用“似不相关回归”来检验组间截面相关(参见第 23 章), 但略为繁琐。

## 16.8 变系数模型

对于长面板数据, 由于样本容量大, 除了可以让每位个体拥有自己的截距项或时间趋势项外, 还可以允许每位个体的回归方程斜率也不同, 这被称为“变系数模型”。“变系数模型”可分为两大类, 取决于将“可变系数”视为常数还是随机变量。

### 1. 将可变系数视为常数

假设  $y_{it} = \mathbf{x}'_i \boldsymbol{\beta}_i + \varepsilon_{it}$ , 其中  $\boldsymbol{\beta}_i$  为个体  $i$  对应的系数。此时, 可以对每个个体方程进行“分别回归”(separate regressions)。但如果不同个体的扰动项相关, 则分别回归效率不高, 因为它忽略了不同方程扰动项相关性的可用信息。有效率的做法是, 把所有个体回归方程叠放(stack), 然后使用“似不相关回归”(SUR, 参见第 23 章)对整个方程系统进行“系统估计”(system estimation)。

使用这个方法的缺点是, 可能需要估计较多参数, 从而损失自由度。作为一种折中, 可以考虑“部分变系数模型”, 即允许  $\boldsymbol{\beta}_i$  中的部分系数(比如, 研究者感兴趣的系数)依个体而变, 而其余系数则不变。在这种情况下, 不再适用 SUR, 因为各个体方程除了扰动项相关外, 还拥有部分相同的系数(跨方程约束)。此时, 可以使用 LSDV 法, 即在回归方程中, 引入个体虚拟变量, 以及虚拟变量与  $\mathbf{x}_{it}$  中可变系数之解释变量的互动项。

继续前面的例子。假设人均收入(lny)的回归系数在每个州各不相同, 而其他解释变量的系数相同。

```
. reg lnc lnp lnpmin lny i.state i.state#c.lny t, vce(cluster state)
```

Linear regression							Number of obs = 300
					F( 2, 298) = .		
					Prob > F = .		
					R-squared = 0.9394		
					Root MSE = .05301		
(Std. Err. adjusted for 10 clusters in state)							
lnc	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]		
lnp	-.4153795	.1225736	-3.39	0.008	-.6926602	-.1380989	
lnpmin	.0765648	.1063314	0.72	0.490	-.1639735	.3171032	
lny	.6781285	.0888178	7.64	0.000	.4772086	.8790484	
state							
2	1.445143	.0460256	31.40	0.000	1.341026	1.54926	
3	-.109939	.0385508	-2.85	0.019	-.197147	-.0227311	
4	2.127437	.1178778	18.05	0.000	1.860779	2.394095	
5	1.894687	.0803023	23.59	0.000	1.71303	2.076343	
6	1.714489	.0646474	26.52	0.000	1.568247	1.860732	
7	4.817635	.1558296	30.92	0.000	4.465124	5.170146	
8	1.212741	.0803827	15.09	0.000	1.030903	1.394579	
9	.3182395	.0088129	36.11	0.000	.2983034	.3381756	
10	.5024975	.0815632	6.16	0.000	.3179888	.6870062	
state#c.lny							
2	-.1768319	.0043996	-40.19	0.000	-.1867846	-.1668792	
3	.0227827	.0045784	4.98	0.001	.0124256	.0331398	
4	-.2656664	.0102793	-25.84	0.000	-.2889197	-.242413	
5	-.2357472	.0084874	-27.78	0.000	-.254947	-.2165473	
6	-.1770753	.0054496	-32.49	0.000	-.1894032	-.1647474	
7	-.5347635	.0182173	-29.35	0.000	-.5759738	-.4935531	
8	-.1357293	.0098855	-13.73	0.000	-.1580919	-.1133667	
9	-.0355302	.0004363	-81.44	0.000	-.0365171	-.0345433	
10	-.0742162	.0084699	-8.76	0.000	-.0933764	-.0550559	
t	-.0463632	.007114	-6.52	0.000	-.0624563	-.0302702	
_cons	1.168383	.8387717	1.39	0.197	-.7290501	3.065817	

其中，“i.state”表示州虚拟变量，“i.state#c.lny”表示州虚拟变量与 lny 的互动项；“c”表示“continuous”，用以指明 lny 为连续型变量。这些变量在 Stata 中称为“因子变量”(factor variables)，是变量窗口所列变量的扩展，使用方法参见“`help fvvarlist`”。

上表结果显示，州虚拟变量与 lny 的互动项均十分显著，故应使用变系数模型。

## 2. 随机系数模型 (Random Coefficient Model)

在上一章考虑个体效应  $u_i$  或个体时间趋势  $\gamma_{it}$  时，曾分别将  $u_i$  或  $\gamma_i$  视为随机变量。类似地，可将系数(斜率) $\beta_i$  视为随机变量(仿佛从某个总体中获取随机样本，抽样后则取值不再随时间而变)，并假设

$$\beta_i = \beta + \nu_i \quad (16.13)$$

其中， $\beta$  为常数向量，而  $\nu_i$  为随机向量，且满足条件期望  $E(\nu_i | x_i) = \mathbf{0}$ (故影响斜率的随机因素  $\nu_i$  与解释变量  $x_i$  不相关)，条件协方差矩阵  $\text{Var}(\nu_i | x_i) = \Sigma$ (对角矩阵)。因此，

$$y_{it} = x'_{it}\beta_i + \varepsilon_{it} = x'_{it}(\beta + \nu_i) + \varepsilon_{it} = x'_{it}\beta + (x'_{it}\nu_i + \varepsilon_{it}) \quad (16.14)$$

由于  $E(\nu_i | x_i) = \mathbf{0}$ ，通过使用迭代期望定律，可以证明复合扰动项  $(x'_{it}\nu_i + \varepsilon_{it})$  与解释变量  $x_{it}$  不相关(参见习题)。因此，OLS 是一致的。然而，复合扰动项  $(x'_{it}\nu_i + \varepsilon_{it})$  的协方差矩阵为分块

对角矩阵,并非单位矩阵(类似于随机效应模型的情形,参见第15章)。Swamy(1970)提出用FGLS来估计此模型,即利用OLS残差来估计协方差矩阵中的参数,然后再使用GLS。

随机系数模型的Stata命令为

```
xtrc y x1 x2 x3 ,betas
```

其中,选择项“betas”表示,显示对每组系数的估计。该命令提供了一个检验参数稳定性(test of parameter constancy)的 $\chi^2$ 统计量及其p值,其原假设为“ $H_0: \beta_1 = \dots = \beta_n$ ”。如果拒绝此原假设,则认为应使用变系数模型。

继续前面的例子:

```
. xtrc lnc lnp lnpmin lny,beta
```

Random-coefficients regression		Number of obs		=	300
Group variable: state		Number of groups		=	10
		Obs per group:	min	=	30
			avg	=	30.0
			max	=	30
		Wald chi2(3)		=	99.29
		Prob > chi2		=	0.0000
lnc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnp	-.4523322	.1322364	-3.42	0.001	-.7115108 -.1931536
lnpmin	-.0824106	.1196998	-0.69	0.491	-.3170178 .1521967
lny	-.0696917	.0531879	-1.31	0.190	-.173938 .0345545
_cons	7.833393	.5180469	15.12	0.000	6.81804 8.848746
Test of parameter constancy:		chi2(36) =	3611.15	Prob > chi2 =	0.0000

上表给出了对于共同的 $\beta$ 系数向量估计值。最后一行的参数稳定性检验的p值为0.0000,故强烈拒绝“参数不变”的原假设,认为应使用变系数模型。对每组系数 $\beta_i$ 的估计结果如下。

Group-specific coefficients						
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Group 1						
lnp	-.6233233	.1659849	-3.76	0.000	-.9486476	-.2979989
lnpmin	.0943415	.1701196	0.55	0.579	-.2390868	.4277698
lny	.0980417	.0116712	8.40	0.000	.0751665	.1209168
_cons	6.242134	.2118045	29.47	0.000	5.827005	6.657264
Group 2						
lnp	-.2608889	.1671412	-1.56	0.119	-.5884796	.0667018
lnpmin	-.432141	.1654173	-2.61	0.009	-.7563529	-.1079291
lny	-.0678897	.0166591	-4.08	0.000	-.1005409	-.0352384
_cons	8.400657	.2507142	33.51	0.000	7.909266	8.892047
Group 3						
lnp	-.1361494	.1333002	-1.02	0.307	-.397413	.1251142
lnpmin	-.2787844	.1422594	-1.96	0.050	-.5576077	.0000388
lny	.1271675	.0093934	13.54	0.000	.1087568	.1455781
_cons	5.513523	.1886807	29.22	0.000	5.143716	5.883331
Group 4						
lnp	-.5621989	.1049334	-5.36	0.000	-.7678645	-.3565332
lnpmin	-.1648781	.1120782	-1.47	0.141	-.3845473	.0547912
lny	-.1911483	.0139909	-13.66	0.000	-.21857	-.1637266
_cons	9.707348	.2044615	47.48	0.000	9.306611	10.10809
Group 5						
lnp	-.7750139	.1123276	-6.90	0.000	-.9951719	-.5548559
lnpmin	.269899	.1201674	2.25	0.025	.0343751	.5054228
lny	-.1288682	.0146083	-8.82	0.000	-.1575	-.1002364
_cons	8.285053	.2041314	40.59	0.000	7.884963	8.685144
Group 6						
lnp	-.1837786	.0896636	-2.05	0.040	-.359516	-.0080413
lnpmin	-.1098901	.0926034	-1.19	0.235	-.2913895	.0716093
lny	-.1160046	.0100373	-11.56	0.000	-.1356772	-.0963319
_cons	7.343327	.1871318	39.24	0.000	6.976555	7.710098
Group 7						
lnp	-.5065479	.1197776	-4.23	0.000	-.7413077	-.2717881
lnpmin	.1068615	.1397386	0.76	0.444	-.1670212	.3807442
lny	-.4296878	.025671	-16.74	0.000	-.480002	-.3793736
_cons	10.69059	.3418134	31.28	0.000	10.02065	11.36054
Group 8						
lnp	-.2531896	.110291	-2.30	0.022	-.469356	-.0370233
lnpmin	-.2451675	.1257458	-1.95	0.051	-.4916248	.0012898
lny	-.0547215	.0133811	-4.09	0.000	-.080948	-.0284949
_cons	7.574233	.2314107	32.73	0.000	7.120677	8.02779
Group 9						
lnp	-.1528873	.1217504	-1.26	0.209	-.3915137	.0857392
lnpmin	-.3632993	.1235095	-2.94	0.003	-.6053735	-.121225
lny	.0726175	.0103202	7.04	0.000	.0523903	.0928447
_cons	6.425592	.2026637	31.71	0.000	6.028378	6.822805
Group 10						
lnp	-1.069344	.1829536	-5.84	0.000	-1.427927	-.7107616
lnpmin	.2989525	.1885074	1.59	0.113	-.0705152	.6684202
lny	-.006424	.0195543	-0.33	0.743	-.0447497	.0319016
_cons	8.151466	.2968129	27.46	0.000	7.569723	8.733208

## 16.9 面板工具变量法

虽然面板数据能在一定程度上解决遗漏变量(个体异质性)问题,但如果回归模型本身包含内生解释变量(endogenous regressors),则仍需使用工具变量法(参见第10章)。实际操作通常分为两步,即首先对模型进行变换以解决遗漏变量问题(比如,使用固定效应模型FE或一阶差分法FD),然后对变换后的模型使用二阶段最小二乘法(2SLS)。当然,如果认为不存在遗漏变量问题,则可省略第一步。考虑以下几种情形。

### 1. 对固定效应模型先进行离差变换,再使用工具变量法

对于固定效应模型,组内估计量把 $(y_{it} - \bar{y}_i)$ 对 $(x_{it} - \bar{x}_i)$ 进行OLS回归。现在假设 $x_{it}$ 包含内生解释变量,而 $z_{it}$ 为有效工具变量( $x_{it}$ 中的外生解释变量也被包括在 $z_{it}$ 中)。此时,可以使用工具变量 $(z_{it} - \bar{z}_i)$ ,把 $(y_{it} - \bar{y}_i)$ 对 $(x_{it} - \bar{x}_i)$ 进行2SLS回归。

对应的Stata命令为

```
xtivreg y [varlist1](varlist2 = varlist_iv),fe
```

其中,“varlist1”为外生解释变量,“varlist2”为内生解释变量,而“varlist\_iv”为工具变量,选择项“fe”表示组内估计法。

### 2. 对固定效应模型先进行一阶差分,再使用工具变量法

对于固定效应模型,也可以先进行一阶差分,然后使用工具变量 $(z_{it} - z_{i,t-1})$ ,把 $(y_{it} - y_{i,t-1})$ 对 $(x_{it} - x_{i,t-1})$ 进行2SLS回归。

对应的Stata命令为

```
xtivreg y [varlist1](varlist2 = varlist_iv),fd
```

其中,选择项“fd”表示一阶差分法。

### 3. 对随机效应模型使用工具变量法

此时,可以先对随机效应模型进行FGLS变换,然后对变换后的模型<sup>①</sup>进行2SLS回归。对应的Stata命令为

```
xtivreg y [varlist1](varlist2 = varlist_iv),re
```

其中,选择项“re”表示“随机效应”模型。

需要注意的是,Stata命令xtivreg并不提供选择项“vce(robust)”来保证使用聚类稳健标准误。但可以使用选择项“vce(bootstrap)”以得到稳健的“自助标准误”(参见第19章)。

当工具变量个数多于内生解释变量个数时,对面板数据进行GMM估计会更有效率,但需要下载非官方Stata命令xtivreg2来执行。该命令只能处理固定效应模型,即先对模型进行FE或FD变换后,再对变换后的模型使用GMM:

```
ssc install xtidivreg2 (下载安装命令xtidivreg2)
```

```
xtidivreg2 y [varlist1](varlist2 = varlist_iv),fe gmm (先进行FE变换)
```

```
xtidivreg2 y [varlist1](varlist2 = varlist_iv),fd gmm (先进行FD变换)
```

<sup>①</sup> 对于工具变量 $z_{it}$ 也要做相应的变换。

#### 4. 面板工具变量法的过度识别检验

对于面板工具变量法的过度识别检验(有关过度识别检验的概念,参见第10章),可通过非官方命令`xtoverid`来实现。该命令须在运行命令`xtreg`,`xtivreg`,`xtivreg2`或`xhtaylor`(参见下文)之后才能使用,下载安装方法为“`ssc install xtoverid`”,安装后可参考“`help xtoverid`”。

## 16.10 豪斯曼-泰勒估计量(选读)

固定效应模型的主要缺点是无法估计不随时间而变的变量系数。而这些不随时间变化的变量可能恰恰是我们感兴趣的,比如性别、受教育程度对工资的作用。如果有足够的工具变量,可以直接用工具变量法进行估计,即需要找到与内生解释变量相关,但与个体效应 $u_i$ 无关的有效工具变量。

在上一章介绍的固定效应或随机效应模型中,要么全部解释变量与 $u_i$ 相关(固定效应),要么全部解释变量都与 $u_i$ 不相关(随机效应)。介于二者之间的混合情形是,某些解释变量与 $u_i$ 相关,而其他解释变量与 $u_i$ 不相关。在这种情况下,有可能使用工具变量法得到对不随时间变化的变量系数的一致估计。考虑以下面板模型:

$$y_{it} = \mathbf{x}'_{1,it}\boldsymbol{\beta}_1 + \mathbf{x}'_{2,it}\boldsymbol{\beta}_2 + \mathbf{z}'_{1,i}\boldsymbol{\delta}_1 + \mathbf{z}'_{2,i}\boldsymbol{\delta}_2 + u_i + \varepsilon_{it} \quad (16.15)$$

其中,解释变量 $\mathbf{x}$ 随时间变化,而 $\mathbf{z}$ 不随时间变化。带下标1的变量(即 $\mathbf{x}_1$ 与 $\mathbf{z}_1$ )为外生变量(与 $u_i$ 不相关),而带下标2的变量(即 $\mathbf{x}_2$ 与 $\mathbf{z}_2$ )为内生变量(与 $u_i$ 相关)。所有解释变量均与 $\varepsilon_{it}$ 不相关。由于 $\mathbf{x}_{1,it}$ 与 $\mathbf{z}_{1,i}$ 外生,故可以用自己作为自己的工具变量。对于 $\mathbf{x}_{2,it}$ ,可以使用 $(\mathbf{x}_{2,it} - \bar{\mathbf{x}}_{2,i})$ 作为工具变量。显然, $\mathbf{x}_{2,it}$ 与 $(\mathbf{x}_{2,it} - \bar{\mathbf{x}}_{2,i})$ 相关;另一方面,根据迭代期望定律,

$$\begin{aligned} E[(\mathbf{x}_{2,it} - \bar{\mathbf{x}}_{2,i})u_i] &= E_{u_i}E[(\mathbf{x}_{2,it} - \bar{\mathbf{x}}_{2,i})|u_i] \\ &= E_{u_i}\{u_iE[(\mathbf{x}_{2,it} - \bar{\mathbf{x}}_{2,i})|u_i]\} \\ &= E_{u_i}\{u_i[E(\mathbf{x}_{2,it}|u_i) - E(\bar{\mathbf{x}}_{2,i}|u_i)]\} \\ &= E_{u_i}\{u_i \cdot \mathbf{0}\} = \mathbf{0} \end{aligned} \quad (16.16)$$

对于 $\mathbf{z}_{2,i}$ ,则可以使用 $\bar{\mathbf{x}}_1$ 作为工具变量,即用随时间变化的外生变量的平均值作为不随时间变化的内生变量的工具变量。一般来说, $\mathbf{z}_{2,i}$ 会与 $\bar{\mathbf{x}}_1$ 相关(二者同为 $y_i$ 的解释变量);另一方面,根据定义, $Cov(\bar{\mathbf{x}}_1, u_i) = 0$ 。因此, $\bar{\mathbf{x}}_1$ 为有效工具变量。

显然,为了使用工具变量法,必须要求 $\mathbf{x}_{1,it}$ 中所包含的外生变量个数比 $\mathbf{z}_{2,it}$ 中所包含的内生变量个数更多。使用以上工具变量进行2SLS估计,就得到“豪斯曼-泰勒估计量”(Hausman and Taylor, 1981)。Amemiya and MacCurdy(1986)则提出将 $(\mathbf{x}_{1,1t} - \bar{\mathbf{x}}_{1,1t}), (\mathbf{x}_{1,2t} - \bar{\mathbf{x}}_{1,2t}), \dots, (\mathbf{x}_{1,It} - \bar{\mathbf{x}}_{1,It})$ 也作为工具变量使用,以增加估计的效率。

使用豪斯曼-泰勒估计量的重要前提是,所有解释变量均与 $\varepsilon_{it}$ 不相关,而且部分解释变量与 $u_i$ 不相关。在实证研究中,需要说明这些条件为何能够满足;否则,将导致不一致的估计。

豪斯曼-泰勒估计量的Stata命令为

```
xhtaylor depvar indepvars, endog(varlist) amacurdy
```

其中,选择项“`endog(varlist)`”用来指定内生解释变量。如果加上选择项“`amacurdy`”,则使用Amemiya-MacCurdy估计量;否则,即为默认的豪斯曼-泰勒估计量。

## 16.11 动态面板

面板数据的一个优点是可以对个体的动态行为进行建模。有些经济理论认为,由于惯性或部分调整,个体的当前行为取决于过去行为,比如资本存量的调整。如果在面板模型中,解释变量包含了被解释变量的滞后值,则称之为“动态面板数据”(Dynamic Panel Data,简称DPD)。对于动态面板数据,即使组内估计量(FE)也是不一致的(Nickell,1981)。比如,假设

$$y_{it} = \alpha + \rho y_{i,t-1} + \beta x_{it} + u_i + \varepsilon_{it} \quad (t=2, \dots, T) \quad (16.17)$$

则其离差形式为

$$y_{it} - \bar{y}_i = \rho(y_{i,t-1} - \bar{y}_i) + \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (t=2, \dots, T) \quad (16.18)$$

其中,  $\bar{y}_i \equiv \frac{1}{T-1} \sum_{t=2}^T y_{it}$ ,  $\bar{L}y_i \equiv \frac{1}{T-1} \sum_{t=2}^T y_{i,t-1}$ ,  $\bar{x}_i \equiv \frac{1}{T-1} \sum_{t=2}^T x_{it}$ ,  $\bar{\varepsilon}_i \equiv \frac{1}{T-1} \sum_{t=2}^T \varepsilon_{it}$  为时间平均值。显然,由于  $\bar{L}y_i$  中包含  $\{y_{i1}, \dots, y_{i,T-1}\}$  的信息,而  $\{y_{i1}, \dots, y_{i,T-1}\}$  与  $(\varepsilon_{it} - \bar{\varepsilon}_i)$  相关,故  $\bar{L}y_i$  肯定与  $(\varepsilon_{it} - \bar{\varepsilon}_i)$  相关。因此,FE 是不一致的,称之为“动态面板偏差”(dynamic panel bias)。显然,对于长面板而言,  $n$  较小而  $T$  较大,故动态面板偏差较小(如果  $T \rightarrow \infty$ , 则  $\bar{L}y_i \xrightarrow{P} E(y_i)$ ,  $\bar{\varepsilon}_i \xrightarrow{P} 0$ ),可通过校正偏差的方法得到一致估计,参见本章第13节。本节主要针对  $n$  较大而  $T$  较小的短动态面板(short dynamic panel)<sup>①</sup>。

### 1. 差分 GMM

更一般地,考虑以下动态面板模型:

$$y_{it} = \alpha + \rho y_{i,t-1} + x'_{it} \beta + z'_it \delta + u_i + \varepsilon_{it} \quad (t=2, \dots, T) \quad (16.19)$$

先作一阶差分以消去个体效应  $u_i$ ,可得

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta x'_{it} \beta + \Delta \varepsilon_{it} \quad (t=2, \dots, T) \quad (16.20)$$

然而,  $\Delta y_{i,t-1} \equiv y_{i,t-1} - y_{i,t-2}$  依然与  $\Delta \varepsilon_{it} \equiv \varepsilon_{it} - \varepsilon_{i,t-1}$  相关,因为  $y_{i,t-1}$  与  $\varepsilon_{i,t-1}$  相关。因此,  $\Delta y_{i,t-1}$  为内生变量,需要寻找适当的工具变量才能得到一致估计。为此,Anderson and Hsiao(1981)提出使用  $y_{i,t-2}$  作为  $\Delta y_{i,t-1}$  的工具变量,然后进行 2SLS 估计,称为“Anderson-Hsiao 估计量”。显然,  $y_{i,t-2}$  与  $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$  相关。如果  $\{\varepsilon_{it}\}$  不存在自相关(对此假设须进行检验,参见下节),则  $y_{i,t-2}$  与  $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$  不相关(虽然  $y_{i,t-2}$  依赖于  $\varepsilon_{i,t-2}$ ,但  $\varepsilon_{it}, \varepsilon_{i,t-1}$  均与  $\varepsilon_{i,t-2}$  不相关)。因此,在  $\{\varepsilon_{it}\}$  不存在自相关的前提下,  $y_{i,t-2}$  是有效工具变量。然而,根据同样的逻辑,更高阶的滞后变量  $\{y_{i,t-3}, y_{i,t-4}, \dots\}$  也是有效工具变量,而 Anderson-Hsiao 估计量未加以利用,故不是最有效率的。Arellano and Bond(1991)使用所有可能的滞后变量作为工具变量(显然工具变量个数多于内生变量个数),进行 GMM 估计<sup>②</sup>。这就是“Arellano-Bond 估计量”,也被称为“差分 GMM”(Difference GMM)(因为是对差分后的方程进行 GMM 估计)。使用差分 GMM 的前提为扰动项  $\{\varepsilon_{it}\}$  不存在自相关,即  $Cov(\varepsilon_{it}, \varepsilon_{is}) = 0, t \neq s, \forall i$ 。

在使用高阶滞后变量进行 GMM 估计时,存在一个技术细节。记由工具变量组成的矩阵为  $Z$ ,则工具变量  $y_{i,t-2}$  将为  $Z$  贡献以下一个列向量:

<sup>①</sup> 本节的差分 GMM 与系统 GMM 方法仅适用于短动态面板,因为它们的大样本性质建立在于给定时间维度  $T$  的情况下,让横截面维度  $n$  趋于无穷。对于长动态面板,可考虑使用本章第13节的偏差校正 LSDV 法。

<sup>②</sup> 也可用 2SLS 来估计,但不如 GMM 有效率。

$$(\cdot \ y_{it} \cdots y_{i,T-2})' \quad (16.21)$$

其中,第一个点“·”表示缺失值,这意味着将丢失第一行观测数据,损失样本容量。类似地,使用工具变量  $y_{i,t-3}$  将损失前两行数据;以此类推,使用越高阶滞后作为工具变量,则损失的样本容量越多。为此,Holtz-Eakin et al (1988) 提出使用一系列的工具变量来表示  $y_{i,t-2}$ ,其中每个工具变量对应于一个时期,而将缺失值用 0 来代替:

$$\begin{pmatrix} 0 & 0 & \cdots & 0 \\ y_{it} & 0 & \cdots & 0 \\ 0 & y_{i2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & y_{i,T-2} \end{pmatrix} \quad (16.22)$$

这种形式的工具变量被称为“GMM 式”(GMM-type, GMM-style)或“展开式”(uncollapsed)工具变量<sup>①</sup>,形如(16.22)。传统形式的工具变量则称为“标准式”(standard)、“IV 式”(IV-style)或“collapsed”(折叠式)工具变量,形如(16.21)。在差分 GMM 以及下文的系统 GMM 估计中,为了尽量不损失样本容量,一般默认使用 GMM 式工具变量。但如果工具变量并非滞后变量(比如,额外的工具变量,或外生变量作为自己的工具变量),则仍可使用标准式工具变量。由表达式(16.22)可知,如果使用 GMM 式工具变量,则工具变量的总数是时间维度  $T$  的二次函数,可能导致很多工具变量<sup>②</sup>。

然而,差分 GMM 在作差分时也会带来以下四个问题。

(1) 如果  $x_{it}$  仅为前定变量(predetermined)而非“严格外生”(strictly exogenous),即虽然  $x_{it}$  与当期  $\varepsilon_{it}$  不相关,但与  $\varepsilon_{i,t-1}$  相关,则经过差分后,  $\Delta x_{it} = x_{it} - x_{i,t-1}$  就可能与  $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$  相关,致使  $\Delta x_{it}$  成为内生变量。此时,可以使用  $\{x_{i,t-1}, x_{i,t-2}, \dots\}$  作为  $\Delta x_{it}$  的工具变量。

(2) 如果  $T$  很大,则会有很多工具变量,容易出现弱工具变量(weak instruments)问题(通常滞后越期则相关性越弱),产生偏差。Windmeijer (2005) 使用差分 GMM 估计一个  $n=100, T=8$  的动态面板,经模拟发现,将工具变量个数由 28 减为 13 可使系数估计值的平均偏差(average bias)下降 40%。工具变量过多的另一后果是,它会弱化 Hansen 统计量(用于工具变量的过度识别检验),甚至出现  $p$  值等于 1 的不可信结果。解决方法之一是在使用 Stata 命令 xtabond 时,限制最多使用  $q$  阶滞后变量作为工具变量,文献中对于如何选择工具变量个数并无明确指南。为了稳健起见,可使用两种(或多种)不同的工具变量个数,并比较其结果;如果二者结果相差很大,则应选择工具变量较少的那个模型。解决方法之二为使用折叠的 IV 式工具变量,而不使用展开的 GMM 式工具变量。

(3) 不随时间变化的变量  $z_i$  被消掉了,故差分 GMM 无法估计  $z_i$  的系数。

(4) 如果序列  $\{y_{it}\}$  具有很强的持续性(highly persistent),即一阶自回归系数接近于 1,则  $y_{i,t-2}$  与  $\Delta y_{i,t-1} \equiv y_{i,t-1} - y_{i,t-2}$  的相关性可能很弱,导致弱工具变量问题。在极端情况下,如果  $\{y_{it}\}$  为随机游走(random walk),则  $\Delta y_{i,t-1}$  为白噪声,完全与  $y_{i,t-2}$  不相关,导致  $y_{i,t-2}$  不再是  $\Delta y_{i,t-1}$  的有效工具变量。因此,如果被解释变量的持续性很强,可能不适用差分 GMM。

<sup>①</sup> 这个名词可能引起误解,因为使用标准式工具变量也可以进行 GMM 估计。

<sup>②</sup> 由于工具变量个数约为时间维度  $T$  的二次函数,故矩条件协方差矩阵中的元素个数为  $T$  的四次函数,在使用有限样本估计如此大的矩阵时可能有困难,有时导致所估计矩条件协方差矩阵不可逆(singular),此时可使用广义逆矩阵(generalized inverse),不影响 GMM 的估计。另一方面,如果 Stata 出现“singular”的提示,则意味着可能工具变量太多。

## 2. 水平 GMM

为了解决上述问题(3)与(4), Arellano and Bover(1995)重新回到了差分之前的水平方程(level equation)(16.19), 并使用 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 作为 $y_{i,t-1}$ 的工具变量<sup>①</sup>。显然, 二者是相关的。另一方面, 如果 $\{\varepsilon_u\}$ 不存在自相关, 则 $E(\Delta y_{i,t-s}\varepsilon_u) = E(y_{i,t-s}\varepsilon_u) - E(y_{i,t-s-1}\varepsilon_u) = 0 - 0 = 0, s \geq 1$ ; 但必须假设 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 与个体效应 $u_i$ 不相关, 才能保证这些工具变量与水平方程的复合扰动项( $u_i + \varepsilon_u$ )不相关。假定 $|\rho| < 1$ , 则 $\{y_u\}$ 将趋于某均衡点 $y_i^*$ ( $y_i^*$ 取决于 $u_i$ ), 而“ $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 与 $u_i$ 不相关”的假设意味着趋于均衡点的速度 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 与固定效应 $u_i$ 无关。因此, 在整个样本期间,  $\{y_u\}$ 应该在均衡点 $y_i^*$ 附近, 才能保证这个假定成立<sup>②</sup>。在以上条件都满足的情况下, 可以使用 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 作为工具变量对水平方程进行 GMM 估计。这称为“水平 GMM”(Level GMM)。使用水平 GMM 的前提包括: 扰动项 $\{\varepsilon_u\}$ 不存在自相关; 而且 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 与个体效应 $u_i$ 不相关。

## 3. 系统 GMM

Blundell and Bond(1998)则将差分 GMM 与水平 GMM 结合在一起, 将差分方程与水平方程作为一个方程系统进行 GMM 估计<sup>③</sup>, 称为“系统 GMM”(System GMM)。与差分 GMM 相比, 系统 GMM 的优点是可以提高估计的效率, 并且可以估计不随时间变化的变量 $z_i$ 的系数(因为系统 GMM 包含对水平方程的估计)。其缺点是, 必须额外地假定 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 与 $u_i$ 无关; 如果这个条件无法满足, 则不能使用系统 GMM。使用系统 GMM 的前提包括(与水平 GMM 相同): 扰动项 $\{\varepsilon_u\}$ 不存在自相关; 而且 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 与个体效应 $u_i$ 不相关<sup>④</sup>。

在更一般的动态面板模型中, 解释变量可以包括被解释变量的多阶滞后值,

$$y_u = \alpha + \rho_1 y_{i,t-1} + \rho_2 y_{i,t-2} + \dots + \rho_p y_{i,t-p} + x'_u \beta + z'_i \delta + u_i + \varepsilon_u \quad (16.23)$$

对此, 可以基本沿用以上的方法。在以上的所有 GMM 估计中, 都可以指定额外的工具变量(在滞后变量及差分滞后变量以外的工具变量)。另外, 在水平方程中, 如果 $x_u$ 也包括内生变量, 则同样可以用其滞后值作为工具变量。

**例** Shen and Yao(2008)使用动态面板数据来分析村庄选举对该村居民收入分配的影响。

**例** Acemoglu et al(2008)使用差分 GMM 估计跨国面板数据, 发现人均收入(per capita income)对民主(democracy)的作用不显著。然而, Che et al(2013)使用系统 GMM 估计同样的数据集, 却发现人均收入对民主有显著的正作用。Che et al(2013)认为差分 GMM 存在弱工具变量问题, 而系统 GMM 比差分 GMM 更有效率。

# 16.12 动态面板的 Stata 命令及实例

差分 GMM 的 Stata 命令为

<sup>①</sup> 这种做法正好与差分 GMM 相反。差分 GMM 以水平变量 $\{y_{i,t-2}, y_{i,t-3}, \dots\}$ 作为差分变量 $\Delta y_{i,t-1}$ 的工具变量; 而水平 GMM 则以差分变量 $\{\Delta y_{i,t-1}, \Delta y_{i,t-2}, \dots\}$ 作为水平变量 $y_{i,t-1}$ 的工具变量。

<sup>②</sup> 经济理论(比如, Solow 增长模型)通常认为, 趋于均衡点(steady state)的速度取决于均衡点 $y_i^*$ 的位置, 而 $y_i^*$ 又依赖于 $u_i$ , 故收敛速度与个体特征 $u_i$ 相关。只有在均衡点 $y_i^*$ 附近, 收敛速度才与 $y_i^*$ 及 $u_i$ 无关。

<sup>③</sup> 有关方程系统, 参见第 23~24 章。

<sup>④</sup> 使用系统 GMM 的两个前提(尤其是第二个前提)常被遗忘, 故在此重复强调一下。然而, 第二个前提目前尚无法进行严格的统计检验, 只能根据经济常识判断经济变量是否在稳态(steady state)附近; 比如, 发达国家的经济体很可能在稳态附近, 而转型经济体或发展中国家则可能离稳态较远(由于频繁的制度变迁等原因)。

```
xtabond depvar indepvars, lags(p) maxldep(q) twostep vce(robust) pre(varlist) endogenous(varlist) inst(varlist)
```

其中,“*depvar*”为被解释变量,“*indepvars*”为满足严格外生性的解释变量。选择项“*lags(p)*”表示使用被解释变量的*p*阶滞后值作为解释变量,默认值为*p*=1;“*maxldep(q)*”表示最多使用*q*阶被解释变量的滞后值作为工具变量,默认值为使用所有可能的滞后值。选择项“*twostep*”表示使用GMM,默认方法为2SLS<sup>①</sup>;选择项“*pre(varlist)*”,“*endogenous(varlist)*”,“*inst(varlist)*”分别用来指定前定变量、内生变量与额外的工具变量。选择项“*vce(robust)*”表示使用稳健标准误,该稳健标准误允许 $\varepsilon_u$ 存在异方差,而且针对两步GMM的估计进行了调整,在Stata中称为“WC-Robust Standard Error”(Windmeijer, 2005;其中WC表示“Windmeijer bias-corrected”)。

系统GMM的Stata命令为<sup>②</sup>

```
xtdpdsys depvar [indepvars], lags(p) maxldep(q) twostep vce(robust) pre(varlist) endogenous(varlist) inst(varlist)
```

下面以数据集mus08psidextract.dta为例<sup>③</sup>。该面板数据集来自Panel Study of Income Dynamics(PSID)<sup>④</sup>,包含595名美国工人1976—1982年有关工资的以下变量( $n=595, T=7$ 的短面板):lwage(工资对数),ed(教育年限),exp(工龄),exp2(工龄平方),wks(weeks worked,工作周数),occ(是否蓝领工人),ind(是否在制造业工作),ms(婚否),south(是否在美国南方),smsa(是否住大城市),union(是否由工会合同确定工资),fem(是否女性),blk(是否黑人)。

首先以差分GMM估计以下动态面板模型:

$$\begin{aligned} \text{lwage}_{it} = & \alpha + \rho_1 \text{lwage}_{i,t-1} + \rho_2 \text{lwage}_{i,t-2} + \beta_1 \text{occ}_{it} + \beta_2 \text{south}_{it} + \beta_3 \text{smsa}_{it} + \\ & \beta_4 \text{ind}_{it} + \beta_5 \text{wks}_{it} + \beta_6 \text{wks}_{i,t-1} + \beta_7 \text{ms}_{it} + \beta_8 \text{union}_{it} + u_i + \varepsilon_{it} \end{aligned} \quad (16.24)$$

其中,occ,south,smsa,ind被视为外生解释变量,wks及其一阶滞后被视为前定解释变量,而ms,union被视为内生解释变量<sup>⑤</sup>。由于差分GMM首先对原方程进行差分,所有不随时间变化的变量都将无法估计(比如ed,fem,blk),故不包括在模型中。

Arellano-Bond估计量的Stata命令为

```
. use mus08psidextract.dta, clear
. xtabond lwage occ south smsa ind, lags(2) maxldep(3) pre(wks, lag(1,2))
  endogenous(ms, lag(0,2)) endogenous(union, lag(0,2)) twostep vce(robust)
. estimates store DiffGMM
```

<sup>①</sup> 可以将2SLS视为“one-step GMM”。

<sup>②</sup> 也可使用非官方命令xtabond2进行差分GMM或系统GMM估计,下载方法为“ssc install xtabond2”,参见下文。

<sup>③</sup> 此例来自Cameron and Trivedi(2009)。

<sup>④</sup> 该数据集最早为Baltagi and Khanti-Akom(1990)所使用。

<sup>⑤</sup> 为了演示的目的,解释变量中未包括exp,因为对于这个数据集而言,很难区分工龄与工资滞后值对工资的影响。如果在解释变量中同时包括工龄与工资滞后值,将导致估计不准确。

Arellano-Bond dynamic panel-data estimation		Number of obs	=	2380	
Group variable: id		Number of groups	=	595	
Time variable: t		Obs per group:	min =	4	
			avg =	4	
			max =	4	
Number of instruments =	40	Wald chi2(10)	=	1287.77	
		Prob > chi2	=	0.0000	
Two-step results		(Std. Err. adjusted for clustering on id)			
lwage	Coef.	WC-Robust Std. Err.	z	P> z	[95% Conf. Interval]
lwage					
L1.	.611753	.0373491	16.38	0.000	.5385501 .6849559
L2.	.2409058	.0319939	7.53	0.000	.1781989 .3036127
wks					
--	-.0159751	.0082523	-1.94	0.053	-.0321493 .000199
L1.	.0039944	.0027425	1.46	0.145	-.0013807 .0093695
ms	.1859324	.144458	1.29	0.198	-.0972 .4690649
union	-.1531329	.1677842	-0.91	0.361	-.4819839 .1757181
occ	-.0357509	.0347705	-1.03	0.304	-.1038999 .032398
south	-.0250368	.2150806	-0.12	0.907	-.446587 .3965134
smsa	-.0848223	.0525243	-1.61	0.106	-.187768 .0181235
ind	.0227008	.0424207	0.54	0.593	-.0604422 .1058437
_cons	1.639999	.4981019	3.29	0.001	.6637377 2.616261
Instruments for differenced equation					
GMM-type: L(2/4).lwage L(1/2).L.wks L(2/3).ms L(2/3).union					
Standard: D.occ D.south D.smsa D.ind					
Instruments for level equation					
Standard: _cons					

其中,选择项“lags(2)”表示解释变量中包含被解释变量的一阶与二阶滞后;选择项“maxldep(3)”表示最多使用被解释变量的三个滞后值作为工具变量<sup>①</sup>;选择项“pre(wks,lag(1,2))”<sup>②</sup>表示变量wks及其一阶滞后L.wks为前定解释变量,而使用其两个更高阶滞后值(即二阶与三阶)为工具变量;选择项“endogenous(ms,lag(0,2))”<sup>③</sup>指定变量ms为内生解释变量(其中的“0”表示内生解释变量不包含ms的任何滞后),而最多使用其两个更高阶滞后值(即二阶与三阶)为工具变量<sup>④</sup>。

上表的下部显示了所使用的全部工具变量,比如,“L(2/4).lwage”表示lwage的二至四阶滞后(如果存在),即lwage<sub>i,t-2</sub>,lwage<sub>i,t-3</sub>,lwage<sub>i,t-4</sub>,以此类推。此表的左上部显示,该差分GMM估计共使用了40个工具变量(矩条件)。

① 比如,当T=7时,使用lwage<sub>i,5</sub>,lwage<sub>i,4</sub>,lwage<sub>i,3</sub>为工具变量。

② 该选择项的一般格式为“pre(varlist [,lagstruct(prelags,premaxlags)])”,其中“prelags”指定把直至“prelags”阶的滞后值也作为解释变量(默认值prelags=0),而“premaxlags”则指定最多使用“premaxlags”个更高阶滞后值作为工具变量。

③ 该选择项的一般格式为“endogenous(varlist [,lagstruct(endlags,endmaxlags)])”,其中“endlags”指定把直至“endlags”阶的滞后值也作为解释变量(默认值endlags=0),而“endmaxlags”则指定最多使用“endmaxlags”个更高阶滞后值作为工具变量。

④ 虽然原模型仅包括ms的当期变量,但在差分方程中则包含ms的一阶滞后值,故两个更高阶滞后值为二阶与三阶滞后值。

表中结果显示,被解释变量的一阶与二阶滞后值都很显著,而其他解释变量则不太显著。

作为一致估计,差分GMM能够成立的前提是,扰动项 $\{\varepsilon_{it}\}$ 不存在自相关。对此要进行检验。即使原假设“扰动项 $\{\varepsilon_{it}\}$ 无自相关”成立,“扰动项的一阶差分”(first-differenced errors)仍将存在一阶自相关,因为

$$\text{Cov}(\Delta\varepsilon_{it}, \Delta\varepsilon_{i,t-1}) = \text{Cov}(\varepsilon_{it} - \varepsilon_{i,t-1}, \varepsilon_{i,t-1} - \varepsilon_{i,t-2}) = -\text{Cov}(\varepsilon_{i,t-1}, \varepsilon_{i,t-1}) = -\text{Var}(\varepsilon_{i,t-1}) \neq 0 \quad (16.25)$$

但扰动项的差分将不存在二阶或更高阶的自相关,即 $\text{Cov}(\Delta\varepsilon_{it}, \Delta\varepsilon_{i,t-k}) = 0, k \geq 2$ 。为此,可以通过检验扰动项的差分是否存在一阶与二阶自相关,来检验原假设。其Stata命令为

```
. estat abond
```

Arellano-Bond test for zero autocorrelation in first-differenced errors			
Order	z	Prob > z	
1	-4.5244	0.0000	
2	-1.6041	0.1087	
H0: no autocorrelation			

结果显示,扰动项的差分存在一阶自相关,但不存在二阶自相关,故接受原假设“扰动项 $\{\varepsilon_{it}\}$ 无自相关”,可以使用差分GMM。

也可以检验扰动项的差分是否存在更高阶的自相关:

```
. estat abond, artests(3)
```

Arellano-Bond test for zero autocorrelation in first-differenced errors			
Order	z	Prob > z	
1	-4.5244	0.0000	
2	-1.6041	0.1087	
3	.35729	0.7209	
H0: no autocorrelation			

其中,选择项“artests(3)”表示,检验扰动项的差分 $\Delta\varepsilon_{it}$ 是否存在一阶至三阶自相关;默认值为检验至二阶自相关,即“artests(2)”。结果显示,也不存在三阶自相关,故可以接受原假设。

其次,由于此差分GMM使用了40个工具变量,故需要进行过度识别检验。为此,我们重新运行以上“xtabond”命令,但略去选择项“vce(robust)”<sup>①</sup>。

```
. quietly xtabond lwage occ south smsa ind, lags(2) maxldep(3) pre(wks, lag(1,2)) endogenous(ms, lag(0,2)) endogenous(union, lag(0,2)) twoStep  
. estat sargan
```

Sargan test of overidentifying restrictions	
H0: overidentifying restrictions are valid	
chi2(29) =	39.87571
Prob > chi2 =	0.0860

<sup>①</sup> 如果使用选择项“vce(robust)”,则无法运用下面的“estat sargan”命令,因为该命令假定扰动项 $\varepsilon_{it}$ 为独立同分布(在独立同分布的情况下,不需要使用稳健标准误)。

上表显示，在5%的显著性水平上，无法拒绝“所有工具变量均有效”的原假设，因为  $p$  值 = 0.086 > 0.05。

下面使用系统 GMM 来估计以上模型：

```
. xtdpdsys lwage occ south smsa ind, lags(2) maxldep(3) pre(wks, lag(1, 2)) endogenous(ms, lag(0, 2)) endogenous(union, lag(0, 2)) twostep vce(robust)
```

```
. estimates store SysGMM
```

System dynamic panel-data estimation		Number of obs	=	2975
Group variable: id		Number of groups	=	595
Time variable: t		Obs per group:	min =	5
			avg =	5
			max =	5
Number of instruments =	60	Wald chi2(10)	=	2270.88
		Prob > chi2	=	0.0000
Two-step results				
lwage	Coef.	Std. Err.	z	P> z  [95% Conf. Interval]
lwage				
L1.	.6017533	.0291502	20.64	0.000 .5446199 .6588866
L2.	.2880537	.0285319	10.10	0.000 .2321322 .3439752
wks				
--.	-.0014979	.0056143	-0.27	0.790 -.0125017 .009506
L1.	.0006786	.0015694	0.43	0.665 -.0023973 .0037545
ms	.0395337	.0558543	0.71	0.479 -.0699386 .1490061
union	-.0422409	.0719919	-0.59	0.557 -.1833423 .0988606
occ	-.0508803	.0331149	-1.54	0.124 -.1157843 .0140237
south	-.1062817	.083753	-1.27	0.204 -.2704346 .0578713
smsa	-.0483567	.0479016	-1.01	0.313 -.1422422 .0455288
ind	.0144749	.031448	0.46	0.645 -.0471621 .0761118
_cons	.9584113	.3632287	2.64	0.008 .2464961 1.670327
Instruments for differenced equation				
GMM-type: L(2/4).lwage L(1/2).L.wks L(2/3).ms L(2/3).union				
Standard: D.occ D.south D.smsa D.ind				
Instruments for level equation				
GMM-type: LD.lwage LD.wks LD.ms LD.union				
Standard: _cons				

对比差分 GMM 与系统 GMM 的系数估计值与标准误：

```
. estimates table DiffGMM SysGMM, b se
```

Variable	DiffGMM	SysGMM
lwage		
L1.	.61175299	.60175325
	.03734912	.0291502
L2.	.24090583	.28805368
	.0319939	.02853189
wks		
--.	-.01597512	-.00149786
	.00825228	.00561431
L1.	.00399441	.0006786
	.00274247	.00156937
ms	.18593243	.03953373
	.14445799	.05585427
union	-.15313291	-.04224089
	.16778419	.07199186
occ	-.03575093	-.05088031
	.03477051	.03311489
south	-.02503681	-.10628169
	.2150806	.08375304
smsa	-.08482225	-.04835667
	.05252431	.04790164
ind	.02270076	.01447487
	.04242066	.031448
_cons	1.6399994	.95841132
	.49810188	.36322872

legend: b/se

从上表可知,系统 GMM 的系数估计值与差分 GMM 很接近,但前者的标准误比后者更小。也许因为使用了更多的工具变量(共 60 个),系统 GMM 估计得更准确些。下面对系统 GMM 扰动项的自相关性进行检验:

```
. estat abond
```

Arellano-Bond test for zero autocorrelation in first-differenced errors		
Order	z	Prob > z
1	-4.1072	0.0000
2	-2.149	0.0316
H0: no autocorrelation		

上表显示,可以在 5% 的显著性水平上拒绝“扰动项差分的二阶自相关系数为 0 的假设”,因为  $p$  值 = 0.03 < 0.05。因此,可以拒绝“扰动项  $\{\varepsilon_u\}$  无自相关”的原假设。这使得我们怀疑系统 GMM 是否适用。

继续进行过度识别检验:

```
. quietly xtdpdsys lwage occ south smsa ind, lags(2) maxldep(3) pre(wks,
lag(1,2)) endogenous(ms, lag(0,2)) endogenous(union, lag(0,2)) twoStep
. estat sargan
```

Sargan test of overidentifying restrictions	
H0: overidentifying restrictions are valid	
chi2(49)	= 84.02056
Prob > chi2	= 0.0014

结果强烈拒绝“所有工具变量都有效”的原假设,这意味着某些新增工具变量与扰动项相关,不是有效工具变量。

以上两个检验结果表明,对于这个数据集而言,虽然系统 GMM 可能更有效率,但其成立的前提似乎并不满足。然而,这也可能因为模型设定不当。为了解决扰动项自相关问题,我们在解释变量中引入被解释变量的三阶滞后<sup>①</sup>;为了改善工具变量的有效性,对于内生解释变量 ms, union 仅使用一个更高阶滞后值(即二阶滞后)为工具变量,但最多使用被解释变量的五个滞后值作为工具变量,重新进行系统 GMM 估计与检验:

```
. xtddpsys lwage occ south smsa ind, lags(3) maxldep(5) pre(wks, lag(1,2))
endogenous(ms, lag(0,1)) endogenous(union, lag(0,1)) twostep vce(robust)
```

System dynamic panel-data estimation		Number of obs		= 2380	
Group variable: id		Number of groups		= 595	
Time variable: t		Obs per group:		min =	4
				avg =	4
				max =	4
Number of instruments = 45		Wald chi2(11) = 2236.60		Prob > chi2 = 0.0000	
Two-step results					
lwage	Coef.	WC-Robust Std. Err.	z	P> z	[95% Conf. Interval]
L1.	.5851347	.044352	13.19	0.000	.4982063 .6720631
L2.	.2115442	.0347287	6.09	0.000	.1434771 .2796113
L3.	.094682	.0329862	2.87	0.004	.0300304 .1593337
wks					
--.	-.0034873	.0052114	-0.67	0.503	-.0137015 .006727
L1.	.0030127	.0017871	1.69	0.092	-.0004899 .0065154
ms	.0666652	.0658209	1.01	0.311	-.0623414 .1956718
union	.0167684	.0848163	0.20	0.843	-.1494686 .1830053
occ	-.0764385	.0334374	-2.29	0.022	-.1419746 -.0109024
south	-.1013125	.0765927	-1.32	0.186	-.2514313 .0488064
smsa	-.0706818	.0611874	-1.16	0.248	-.190607 .0492434
ind	.0439888	.0386347	1.14	0.255	-.0317339 .1197114
_cons	.9141559	.3503796	2.61	0.009	.2274245 1.600887
Instruments for differenced equation					
GMM-type: L(2/6).lwage L(1/2).L.wks L(2/2).ms L(2/2).union					
Standard: D.occ D.south D.smsa D.ind					
Instruments for level equation					
GMM-type: LD.lwage LD.wks LD.ms LD.union					
Standard: _cons					

从上表可知,被解释变量的三阶滞后均很显著。另外,工具变量的个数由 60 减少为 45。

```
. estat abond
```

<sup>①</sup> 解决“扰动项自相关”问题的另一方法是使用命令 xtddpd。该命令允许动态面板模型的扰动项  $\varepsilon_{it}$  服从较低阶的移动平均过程(a moving average process of low order),但句型更为复杂,参见“`help xtddpd`”。

Arellano-Bond test for zero autocorrelation in first-differenced errors

Order	z	Prob > z
1	-7.1406	0.0000
2	-1.0119	0.3116

H0: no autocorrelation

上表显示,可以接受扰动项无自相关的原假设。

```
. quietly xtddpdsys lwage occ south smsa ind, lags(3) maxldep(5) pre(wks,
lag(1,2)) endogenous(ms,lag(0,1)) endogenous(union,lag(0,1)) twostep
.estat sargan
```

Sargan test of overidentifying restrictions

H0: overidentifying restrictions are valid

```
chi2(33)      =  46.83873
Prob > chi2   =  0.0559
```

结果显示,可以在5%的显著性水平上接受“所有工具变量都有效”的原假设。这表明,对修正的模型设定,可以进行系统GMM估计。

由于官方命令xtabond与xtddpdsys均不提供异方差稳健的Hansen统计量(仅提供基于iid假设的Sargan统计量),故下面转而使用非官方命令xtabond2。该命令的格式与官方命令xtabond有所不同:

```
xtabond2 y l.y l2.y x1 x2 x3, gmmstyle(varlist) ivstyle(varlist)
nolevel twostep vce(robust)
```

其中,“y”为被解释变量,而“l.y l2.y x1 x2 x3”为解释变量(这里假设被解释变量的两阶滞后为解释变量)。选择项“gmmstyle(varlist)”指定GMM式工具变量,而选择项“ivstyle(varlist)”指定IV式工具变量。选择项“nolevel”表示不估计水平方程,即为差分GMM;默认认为系统GMM。选择项“twostep”与“vce(robust)”的含义与官方命令xtabond相同。更多说明,参见“help xtabond2”以及Roodman(2009)。

为了演示的目的,下面使用xtabond2来复制本节的差分GMM估计结果。

```
.ssc install xtabond2 (下载安装命令xtabond2)
.xtabond2 lwage L(1/2).lwage L(0/1).wks ms union occ south smsa ind,
gmm(lwage, lag(2 4)) gmm(wks ms union, lag(2 3)) iv(occ south smsa ind)
nolevel twostep robust
```

其中,选择项“gmm(lwage, lag(2 4))”表示使用变量lwage的2—4阶滞后作为GMM式工具变量,选择项“gmm(wks ms union, lag(2 3))”表示使用变量wks,ms与union的2—3阶滞后作为GMM式工具变量,而选择项“iv(occ south smsa ind)”表示使用变量occ,south,smsa与ind作为自身的IV式工具变量(假设它们为外生变量)。

Dynamic panel-data estimation, two-step difference GMM						
		Corrected				
	l wage	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lwage	L1.	.611753	.0373491	16.38	0.000	.5385501 .6849559
	L2.	.2409058	.0319939	7.53	0.000	.1781989 .3036127
wks	--.	-.0159751	.0082523	-1.94	0.053	-.0321493 .000199
	L1.	.0039944	.0027425	1.46	0.145	-.0013807 .0093695
ms	.1859324	.144458	1.29	0.198	-.0972	.4690649
union	-.1531329	.1677842	-0.91	0.361	-.4819839	.1757181
occ	-.0357509	.0347705	-1.03	0.304	-.1038999	.032398
south	-.0250368	.2150806	-0.12	0.907	-.446587	.3965134
smsa	-.0848223	.0525243	-1.61	0.106	-.187768	.0181235
ind	.0227008	.0424207	0.54	0.593	-.0604422	.1058437
Instruments for first differences equation						
Standard						
D.(occ south smsa ind)						
GMM-type (missing=0, separate instruments for each period unless collapsed)						
L(2/3).(wks ms union)						
L(2/4).l wage						
Arellano-Bond test for AR(1) in first differences: z = -4.52 Pr > z = 0.000						
Arellano-Bond test for AR(2) in first differences: z = -1.60 Pr > z = 0.109						
Sargan test of overid. restrictions: chi2(29) = 59.55 Prob > chi2 = 0.001 (Not robust, but not weakened by many instruments.)						
Hansen test of overid. restrictions: chi2(29) = 39.88 Prob > chi2 = 0.086 (Robust, but weakened by many instruments.)						
Difference-in-Hansen tests of exogeneity of instrument subsets:						
gmm(l wage, lag(2 4))						
Hansen test excluding group: chi2(18) = 23.59 Prob > chi2 = 0.169						
Difference (null H = exogenous): chi2(11) = 16.29 Prob > chi2 = 0.131						
gmm(wks ms union, lag(2 3))						
Hansen test excluding group: chi2(5) = 6.43 Prob > chi2 = 0.266						
Difference (null H = exogenous): chi2(24) = 33.44 Prob > chi2 = 0.095						
iv(occ south smsa ind)						
Hansen test excluding group: chi2(25) = 28.00 Prob > chi2 = 0.308						
Difference (null H = exogenous): chi2(4) = 11.87 Prob > chi2 = 0.018						

从上表可知, 使用 xtabond2 得到的系数与标准误估计值与上文使用 xtabond 的结果完全相同。然而, 上表中的 Sargan 统计量为 59.55, Hansen 统计量为 39.88; 却与 xtabond 的 Sargan 统计量相左(碰巧也是 39.88); 尽管不影响检验结果。一种可能性是, xtabond 汇报的 Sargan 统计量其实就是 Hansen 统计量。另外, 正如上表所示, Sargan 统计量与 Hansen 统计量各有千秋。尽管前者不稳健, 但不受工具变量过多的影响; 而后者虽然稳健, 却可能因工具变量过多而失效。xtabond2 的另一优势是, 它还可检验部分工具变量的外生性, 即上表中的“Difference-in-Hansen tests”(参见第 10 章)。

## 16.13 偏差校正 LSDV 法

上节介绍的差分 GMM 与系统 GMM 方法主要适用于短动态面板。这是因为,虽然基于 IV 或 GMM 的估计方法是一致估计量(即当  $n \rightarrow \infty$  时,没有偏差),但对于  $n$  较小而  $T$  较大的长面板则可能存在较严重的偏差。Nickell (1981) 证明,动态面板偏差在数量级上与  $T^{-1}$  相当(通常记为  $O(T^{-1})$ ),故当  $T \rightarrow \infty$  时,动态面板偏差趋向于 0。因此,对于长面板,可考虑使用以下“偏差校正 LSDV 法”(Biased - corrected LSDV, 简记 LSDVC)。蒙特卡罗模拟(参见第 19 章)结果显示,对于  $n$  较小的长面板,无论在偏差大小还是均方误差(RMSE)方面,LSDVC 法都明显优于差分 GMM 或系统 GMM (Kiviet, 1995; Judson and Owen, 1999)。

LSDVC 法的基本思想是,首先使用 LSDV 法估计动态面板模型,记估计系数为  $\hat{\beta}_{LSDV}$ ;其次,估计 LSDV 法的偏差,记为  $\widehat{Bias}$ ;最后,将 LSDV 系数估计值减去此偏差,即得到偏差校正后的一致估计:

$$\hat{\beta}_{LSDVC} = \hat{\beta}_{LSDV} - \widehat{Bias} \quad (16.26)$$

其中,估计量  $\hat{\beta}_{LSDVC}$  的标准误差可通过自助法得到(参见第 19 章)。LSDVC 法的局限是,它要求所有解释变量都是严格外生的,而差分 GMM 或系统 GMM 可通过引入工具变量来解决内生变量或前定变量的问题。

LSDVC 法可通过非官方命令 `xtlsdvc` 来实现(Bruno, 2005),其基本格式为

```
xtlsdvc y x1 x2 x3, initial(estimator) bias(#) vcov(#)
```

此命令仅允许以被解释变量的一阶滞后作为解释变量。其中,选择项“`initial(estimator)`”用来指定偏差校正的初始值;比如,“`i(ah)`”表示使用 Anderson-Hsiao 估计量作为初始值,“`i(ab)`”表示使用 Arellano-Bond 差分 GMM 估计量作为初始值,而“`i(bb)`”表示使用 Blundell-Bond 系统 GMM 估计量作为初始值。选择项“`bias(#)`”用来指定偏差校正的精确度;比如,“`bias(1)`”的精确度为  $O(T^{-1})$ ,“`bias(2)`”的精确度为  $O(n^{-1}T^{-1})$ ,而“`bias(3)`”的精确度为  $O(n^{-1}T^{-2})$ (这类似于在进行泰勒展开时,使用更高阶的近似)。选择项“`vcov(#)`”指定用来估计方差-协方差矩阵(variance-covariance matrix)的自助法重复抽样的次数,一般可设为“`vcov(50)`”。

下面以数据集 `mus08cigar.dta` 为例(纯粹为了演示的目的)。

```
. use mus08cigar.dta, clear
. gen t = year - 62
. ssc install xtsdvc (下载安装命令 xtsdvc)
. xtsdvc lnc lnp lnpm lnny t, initial(ab) vcov(50) bias(3)①
```

① 运行此命令可能费时较长。

LSDVC dynamic regression  
(bootstrapped SE)

	lnc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnc	1.058233	.0580338	18.23	0.000	.9444888	1.171977
L1.						
lnp	-.0299203	.050643	-0.59	0.555	-.1291788	.0693381
lnpmin	.0056437	.0522719	0.11	0.914	-.0968075	.1080948
lny	.0611414	.0988631	0.62	0.536	-.1326266	.2549095
t	-.0051682	.0076777	-0.67	0.501	-.0202161	.0098798

## 16.14 重复截面数据与组群分析

面板数据的最大优点是能够跟踪个体随时间的变化。但有时,虽然可以得到不同时点上的数据,但每个时点上样本所包含的个体却完全不同。比如,英国家庭支出调查(U. K. Family Expenditure Survey)每年均调查不同的家庭<sup>①</sup>。这类数据,虽然也有横截面与时间两个维度,但却不是真正的面板数据,被称为“重复截面”(repeated cross-sections)数据<sup>②</sup>。

对于随机效应模型,即个体效应不与解释变量相关的情形,可以直接用OLS混合回归估计重复截面数据。由于每位个体在样本中只出现一次,故可以认为样本为独立同分布的,即在截面与时间维度上都是独立同分布的。因此,并不需要如一般面板数据那样使用聚类稳健标准误,只需使用普通标准误即可(但仍可能需要使用异方差稳健的标准误)。

对于更常见的固定效应模型,即个体效应与解释变量相关的情形,混合回归将导致不一致的估计。另一方面,由于每位个体在样本中只出现一次,组内估计法(FE)与一阶差分法(FD)均不可行。为此,Browning, Deaton and Irish(1985)与 Deaton(1985)提出,将样本中的个体分成几个固定的“组群”(cohort),构建“准面板”(pseudo panel)或“合成面板”(synthetic panel),以控制固定效应,进行“组群分析”(cohort analysis)。

一个组群由拥有共同特征(比如性别、出生年份、教育水平)的一组个体所组成,其成员是固定不变的。比如,可以按照出生年份来区分组群,“80后组群”指的是在1980年至1989年出生的人,而“90后组群”指的是在1990年至1999年出生的人,等等。

考虑如下固定效应模型:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + u_i + \varepsilon_{it} \quad (t=1, \dots, T) \quad (16.27)$$

其中, $u_i$ 表示“个体固定效应”(individual fixed effect),即 $u_i$ 与 $\mathbf{x}_{it}$ 相关。由于每位个体仅出现一次,故无法使用组内估计法或一阶差分法消去 $u_i$ 。假设样本中所有个体可以按照某种特征分为C个组群,并记组群的下标为 $c=1, \dots, C$ 。设 $g$ 为决定个体的组群归属的变量,即第*i*个个体属于组群*c*,如果 $g_i \in I_c$ ,其中 $I_c$ 为一个集合。

给定组群的归属,对方程(16.27)两边求条件期望可得

$$E(y_{it} | g_i \in I_c) = E(\mathbf{x}'_{it} | g_i \in I_c)\boldsymbol{\beta} + E(u_i | g_i \in I_c) + E(\varepsilon_{it} | g_i \in I_c) \quad (16.28)$$

以“\*”来表示以上的条件期望,即“组群的总体平均值”(population cohort average):

① 如果每年都调查同样的家庭,则会对这些家庭造成较大负担,不利于对家庭真实信息的收集。

② 也称为“Time Series of Independent Cross-Sectional Data”(TSICS)。

$$y_{ct}^* = \mathbf{x}_{ct}'\boldsymbol{\beta} + u_c^* + \varepsilon_{ct}^* \quad (c=1, \dots, C; t=1, \dots, T) \quad (16.29)$$

其中,  $u_c^* \equiv E(u_i | g_i \in I_c)$  被称为“组群固定效应”(cohort fixed effects), 它衡量的是不同组群之间的异质性差异<sup>①</sup>。假设组群固定效应  $u_c^*$  不随时间而改变, 且  $\varepsilon_{ct}^*$  不与  $\mathbf{x}_{ct}$  相关。然而, 由于  $u_i$  与  $\mathbf{x}_i$  相关, 故  $u_c^*$  也与  $\mathbf{x}_{ct}^*$  相关, 因此仍然可以视为固定效应模型。

由于组群的总体平均值不可观测, 故使用“组群时点平均值”(cohort-time average)  $\bar{y}_{ct}$  与  $\bar{\mathbf{x}}_{ct}$  来替代, 即给定时间, 对组群中的个体进行平均。因此,

$$\bar{y}_{ct} = \bar{\mathbf{x}}_{ct}'\boldsymbol{\beta} + \bar{u}_c + \bar{\varepsilon}_{ct} \quad (c=1, \dots, C; t=1, \dots, T) \quad (16.30)$$

其中,  $\bar{y}_{ct} \equiv \frac{1}{n_{ct}} \sum_{g_{it} \in I_c} y_{it}$ ,  $n_{ct}$  为在第  $t$  期第  $c$  个组群所包含的个体数;  $\bar{\mathbf{x}}_{ct}', \bar{u}_c, \bar{\varepsilon}_{ct}$  的定义类似。显然, 在这个替代过程中引入了新的误差, 比如  $\bar{y}_{ct} = y_{ct}^* + \xi_{ct}$ ,  $\bar{\mathbf{x}}_{ct} = \mathbf{x}_{ct}^* + \mathbf{v}_{ct}$ 。如果每个组群在每个时点上的观测值足够多, 则可以忽略“测量误差”(measurement error)。在这种情况下, 对方程 (16.30) 使用组内估计法, 即可消去  $\bar{u}_c$ , 得到参数  $\boldsymbol{\beta}$  的一致估计。具体来说, 就是把  $(\bar{y}_{ct} - \bar{y}_c)$  对  $(\bar{\mathbf{x}}_{ct} - \bar{\mathbf{x}}_c)$  进行 OLS 回归, 其中  $\bar{y}_c \equiv \frac{1}{T} \sum_{t=1}^T \bar{y}_{ct}$ ,  $\bar{\mathbf{x}}_c \equiv \frac{1}{T} \sum_{t=1}^T \bar{\mathbf{x}}_{ct}$  为给定组群对时间进行平均。反之, 如果每个组群在每个时点上的观测值较少, 则需要对测量误差进行调整, 参见 Cameron and Trivedi (2005, p. 772)。

组群分析的主要缺点:(1)由于组群的界定有一定的任意性, 故对组群内个体的同质性假定(intracohort homogeneity)可能过强;(2)将个体数据加总平均为组群数据, 将导致信息的散失, 也可能出现异方差问题(因为组群的规模往往不同);(3)如果组群的个数  $C$  不大, 建立于  $C \rightarrow \infty$  基础上的大样本理论可能失效。

例 周绍杰、张俊森、李宏彬(2009)使用组群分析研究中国城市家庭的消费与储蓄行为。

## 习题

**16.1** 对于随机系数模型, 证明复合扰动项  $(\mathbf{x}'_i \mathbf{v}_i + \varepsilon_{ii})$  与解释变量  $\mathbf{x}_i$  不相关。

**16.2** 面板数据集 grunfeld.dta 包含了 10 个公司 1935—1954 年的以下变量: invest(投资额), mvalue(公司市场市值), 与 kstock(公司资本存量)。考虑估计以下投资函数:

$$\text{invest}_i = \beta_0 + \beta_1 \text{mvalue}_i + \beta_2 \text{kstock}_i + u_i + \varepsilon_i \quad (16.31)$$

参照本章实例, 将其视为“长面板数据”进行各种估计与检验。

**16.3** 尝试对数据集 traffic.dta 进行动态面板估计。动态面板模型适用于该数据集吗?

① 正如个体固定效应(individual fixed effects)衡量的是不同个体之间的差异。

# 第17章 非线性面板

第15、16章介绍了线性面板。对于面板数据，如果被解释变量为虚拟变量、计数变量、受限变量等，则为非线性面板。一般来说，非线性面板不再适用线性面板的方法。

## 17.1 面板二值选择模型

对于面板数据，如果被解释变量为虚拟变量，则称为“面板二值选择模型”(binary choice model for panel data)。对于二值选择行为，通常可通过一个“潜变量”(latent variable)来概括该行为的净收益(收益减去成本)。如果净收益大于0，则选择做；否则，选择不做。假设净收益为

$$y_{it}^* = \mathbf{x}'_{it}\boldsymbol{\beta} + u_i + \varepsilon_{it} \quad (i=1, \dots, n; t=1, \dots, T) \quad (17.1)$$

其中，净收益 $y_{it}^*$ 为不可观测的潜变量， $u_i$ 为个体效应(individual effects)，而解释变量 $\mathbf{x}_{it}$ 不含常数项。个体的选择规则为

$$y_{it} = \begin{cases} 1 & \text{若 } y_{it}^* > 0 \\ 0 & \text{若 } y_{it}^* \leq 0 \end{cases} \quad (17.2)$$

给定 $\mathbf{x}_{it}, \boldsymbol{\beta}, u_i$ ，则有

$$\begin{aligned} P(y_{it} = 1 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) &= P(y_{it}^* > 0 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\ &= P(\mathbf{x}'_{it}\boldsymbol{\beta} + u_i + \varepsilon_{it} > 0 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\ &= P(\varepsilon_{it} > -u_i - \mathbf{x}'_{it}\boldsymbol{\beta} | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\ &= P(\varepsilon_{it} < u_i + \mathbf{x}'_{it}\boldsymbol{\beta} | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\ &= F(u_i + \mathbf{x}'_{it}\boldsymbol{\beta}) \end{aligned} \quad (17.3)$$

其中， $F(\cdot)$ 为 $\varepsilon_{it}$ 的累积分布函数(cdf)，并假设 $\varepsilon_{it}$ 的密度函数关于原点对称。如果 $\varepsilon_{it} \sim N(0, 1)$ ( $\varepsilon_{it}$ 的方差被标准化为1)，则为Probit模型

$$P(y_{it} = 1 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \Phi(u_i + \mathbf{x}'_{it}\boldsymbol{\beta}) \quad (17.4)$$

如果 $\varepsilon_{it}$ 服从逻辑分布，则为Logit模型

$$P(y_{it} = 1 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \Lambda(u_i + \mathbf{x}'_{it}\boldsymbol{\beta}) = \frac{e^{u_i + \mathbf{x}'_{it}\boldsymbol{\beta}}}{1 + e^{u_i + \mathbf{x}'_{it}\boldsymbol{\beta}}} \quad (17.5)$$

面板二值选择模型的主要估计方法包括混合回归、随机效应估计与固定效应估计。在方程(17.1)中，如果 $u_1 = u_2 = \dots = u_n$ ，则为混合回归(pooled probit或pooled logit)，可将此面板数据作为截面数据处理(因为不存在个体效应)。此时，只需使用截面数据的相关Stata命令即可进行混合回归。由于同一个体不同时期的扰动项可能存在自相关，故应使用以面板为聚类的稳健标准误(cluster-robust standard error)。

混合回归的Stata命令为

`probit y x1 x2 x3, vce(cluster id)` (混合 probit 回归)

`logit y x1 x2 x3, vce(cluster id)` (混合 logit 回归)

其中,  $id$  为用来确定面板单位的变量。

## 17.2 面板二值选择模型的随机效应估计

更一般地, 我们允许个体效应的存在, 即不同的个体拥有不同的  $u_i$ 。如果  $u_i$  与所有解释变量  $\mathbf{x}_u$  均不相关, 则称为“随机效应模型”(Random Effects Model, 简记 RE)。如果  $u_i$  与某个解释变量相关, 则称为“固定效应模型”(Fixed Effects Model, 简记 FE)。

首先考虑随机效应模型。对于线性面板的随机效应模型, 一般使用广义最小二乘法(GLS)进行估计。但非线性面板不便使用 GLS, 故转而进行最大似然估计(MLE)。假设  $u_i \sim N(0, \sigma_u^2)$ , 记其密度函数为  $g(u_i)$ 。下面以 Logit 模型为例。给定  $u_i$ , 则个体  $i$  的条件分布为(推导过程可参考第 11 章的相关内容)

$$f(y_{i1}, y_{i2}, \dots, y_{iT} | \mathbf{x}_u, \boldsymbol{\beta}, u_i) = \prod_{t=1}^T [\Lambda(u_i + \mathbf{x}'_u \boldsymbol{\beta})]^{y_t} [1 - \Lambda(u_i + \mathbf{x}'_u \boldsymbol{\beta})]^{1-y_t} \quad (17.6)$$

但上式中的  $u_i$  不可观测。为此, 记  $(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$  的联合密度为  $f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$ , 并进行如下分解

$$f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) = f(y_{i1}, y_{i2}, \dots, y_{iT} | u_i) \cdot g(u_i) \quad (17.7)$$

在  $(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$  的联合密度中, 将  $u_i$  积分积掉, 即可得到  $(y_{i1}, y_{i2}, \dots, y_{iT})$  的边缘密度(这是根据联合密度求边缘密度的一般方法),

$$\begin{aligned} f(y_{i1}, y_{i2}, \dots, y_{iT}) &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) du_i \\ &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT} | u_i) \cdot g(u_i) du_i \\ &= \int_{-\infty}^{+\infty} \left\{ \prod_{t=1}^T [\Lambda(u_i + \mathbf{x}'_u \boldsymbol{\beta})]^{y_t} [1 - \Lambda(u_i + \mathbf{x}'_u \boldsymbol{\beta})]^{1-y_t} \right\} \cdot g(u_i) du_i \end{aligned} \quad (17.8)$$

上式的积分没有解析解, Butler and Moffitt (1982) 提出使用“Gauss-Hermite quadrature”方法进行数值积分(numerical integration)。数值积分的基本思想就是以一系列矩形或梯形的面积之和来近似曲线下的面积。数学家已对这种方法进行了不少改进以提高计算速度与精度。Stata 的默认方法为在 12 个点上进行“adaptive Gauss-Hermite quadrature”计算。由于此积分的精确度依赖于数值计算的点数, 故 Stata 提供命令 `quadchk` 来检验此数值积分的精确度, 即使用有别于 12 的计算点数, 以比较结果的稳定性。如果结果不稳健, 则应增加计算点数, 直至得到稳定结果。

假设不同个体的观测值相互独立, 则可以写出整个样本的似然函数。最大化此似然函数即得到对  $\boldsymbol{\beta}$  的“随机效应 Logit 估计量”(Random Effect Logit)。如果将方程(17.8)中的逻辑分布  $\Lambda(\cdot)$  改为正态分布  $\Phi(\cdot)$ , 则为“随机效应 Probit 估计量”(Random Effect Probit)。由于在进行随机效应估计时已将  $u_i$  积分积掉, 故得不到对个体效应  $u_i$  的估计, 也因此无法预测  $y_i$  的发生概率或计算解释变量对  $y_i$  的边际效应。解决方法之一是假设  $u_i = 0$ 。

由于不同个体的观测值相互独立,故不同个体之间的扰动项也互不相关。但由于  $u_i$  的存在,同一个体不同时期的扰动项之间仍存在自相关。

$$\text{Cov}(u_i + \varepsilon_{iu}, u_t + \varepsilon_{it}) = \begin{cases} \sigma_u^2 & \text{若 } t \neq s \\ \sigma_u^2 + \sigma_\varepsilon^2 & \text{若 } t = s \end{cases} \quad (17.9)$$

其中,  $\sigma_u^2$  为  $u_i$  的方差,而  $\sigma_\varepsilon^2$  为  $\varepsilon_{iu}$  的方差。当  $t \neq s$  时,其自相关系数为

$$\rho \equiv \text{Corr}(u_i + \varepsilon_{iu}, u_t + \varepsilon_{it}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} \quad (17.10)$$

$\rho$  越大,则复合扰动项( $u_i + \varepsilon_{iu}$ )中个体效应的部分( $u_i$ )越重要。特别地,如果  $\rho = 0$ ,则说明  $\sigma_u^2 = 0$ ,即不存在个体随机效应,应选择混合回归。

对于面板二值选择模型,其随机效应估计的 Stata 命令为

`xtprobit y x1 x2 x3` (默认为随机效应 probit)

`xtlogit y x1 x2 x3` (默认为随机效应 logit)

在 Stata 的输出结果中,包含了对原假设“ $H_0: \rho = 0$ ”的 LR 检验结果(Stata 将  $\rho$  记为“rho”)。如果拒绝“ $H_0: \rho = 0$ ”,则认为应采用随机效应模型;反之,则支持混合回归。

## 17.3 面板二值选择模型的固定效应估计

在面板二值选择模型中,如果个体效应  $u_i$  与解释变量  $x_{iu}$  相关,则为固定效应模型,

$$P(y_{iu} = 1 | x_{iu}, \beta, u_i) = F(u_i + x'_{iu}\beta) \quad (17.11)$$

其中,  $F(\cdot)$  为  $\Phi(\cdot)$  或  $\Lambda(\cdot)$ 。此时,如果使用随机效应模型或混合回归将得不到一致估计。对于线性面板数据,一般通过做组内变换( $y_{iu} - \bar{y}_i$ )或一阶差分( $y_{iu} - y_{i,t-1}$ )来消去固定效应  $u_i$ 。但对于非线性面板数据,这些变换一般不起作用,因为无法建立可观测的( $y_{iu} - \bar{y}_i$ )与不可观测的( $y_{iu}^* - \bar{y}_i^*$ ) (即潜变量  $y_{iu}^*$  的组内离差)之间的对应关系。

对于二值选择的固定效应模型,即使在模型中增加个体虚拟变量(相当于 LSDV 法),也仍然得不到一致估计(除非  $T \rightarrow \infty$ )。这是因为,当  $n \rightarrow \infty$  时,待估计的个体效应  $\{u_i\}_{i=1}^n$  的个数也随之增加,这些  $\{u_i\}_{i=1}^n$  被称为“伴生参数”( incidental parameters)<sup>①</sup>。另一方面,每个  $u_i$  只能由个体  $i$  的  $T$  个观测值来估计,而  $T$  并不增加。对于现实中的数据, $T$  为有限值,而且常常很小(短面板)。因此,当  $n \rightarrow \infty$  而  $T$  为有限值,对  $\{u_i\}_{i=1}^n$  的估计将是不一致的(不是因为  $\hat{u}_i$  收敛到不等于  $u_i$  的常数,而是  $\hat{u}_i$  根本就不收敛)。更为严重的问题是,在非线性面板中,对于  $u_i$  的不一致估计还会“污染” $\beta$  的估计,导致对  $\beta$  的估计也不一致,这被称为“伴生参数问题”( incidental parameters problem)。

在线性面板中,可以通过做组内变换或差分变换来解决伴生参数问题。对于固定效应的面板 Probit 模型,目前尚无法解决此伴生参数问题。对于固定效应的面板 Logit 模型,可以通过寻找  $u_i$  的“充分统计量”( sufficient statistic),然后在给定此充分统计量的条件下进行“条件最大似然估计”( conditional MLE)。

① 虽然  $u_i$  为随机变量,但抽样后即为确定值,对于  $y_{iu}$  而言可以视为参数。而且,我们主要对估计  $\beta$  感兴趣,故称  $u_i$  为“伴生参数”。有关“伴生参数问题”,参见 Neyman and Scott (1948) 以及 Lancaster (2000)。

充分统计量是统计学中的一个概念。考虑总体参数  $\theta$  与统计量  $W$ 。如果统计量  $W$  包含了样本中所有可用来估计  $\theta$  的信息，则称  $W$  为参数  $\theta$  的充分统计量。换言之，给定  $W$  之后，任何根据样本计算的其他统计量都不可能提供关于  $\theta$  的额外信息。因此，样本数据在给定充分统计量  $W$  后的条件分布将不再依赖于  $\theta$ （这正是充分统计量的严格定义）；否则，控制了  $W$  之后的条件分布仍将包含  $\theta$  的信息。

对于 Logit 模型，Chamberlain (1980) 提出使用  $n_i \equiv \sum_{t=1}^T y_{it}$  作为  $u_i$  的充分统计量，并计算在给定  $n_i$  情况下的条件似然函数（根据充分统计量的性质，此条件似然函数不再依赖于  $u_i$ ），然后进行条件最大似然估计。然而，对于 Probit 模型，却找不到  $u_i$  的充分统计量。Logit 模型之所以存在  $u_i$  的充分统计量，是由逻辑函数的特殊指数形式所决定的。

以最简单的两期模型为例进行说明，即  $T = 2$ 。此时，对于个体  $i$ ，只有以下三种可能情形，即  $n_i = y_{i1} + y_{i2} = 0, 1$  或  $2$ 。下面分别考虑这三种情形。

(1)  $n_i = y_{i1} + y_{i2} = 0$ 。此时，必然有  $y_{i1} = y_{i2} = 0$ ，即  $P(y_{i1} = 0, y_{i2} = 0 | n_i = 0) = 1$ ，其对数似然函数为  $\ln 1 = 0$ ，故对整个样本的对数似然函数没有贡献。直观来看，由于此条件似然函数的取值为常数（无论参数  $\beta$  如何取值），故此观测值不包含任何可用于估计  $\beta$  的信息。因此，在进行条件似然估计法时，是否包括这些观测值并不影响估计结果；事实上，等于损失了这些样本的观测值。

(2)  $n_i = y_{i1} + y_{i2} = 2$ 。此时，必然有  $y_{i1} = y_{i2} = 1$ ，即  $P(y_{i1} = 1, y_{i2} = 1 | n_i = 2) = 1$ 。同理，这些观测值也不包含任何有助于估计  $\beta$  的信息，也可以忽略。

(3)  $n_i = y_{i1} + y_{i2} = 1$ 。此时，或者  $(y_{i1} = 0, y_{i2} = 1)$ ，或者  $(y_{i1} = 1, y_{i2} = 0)$ 。分别计算其条件概率如下。

$$P(y_{i1} = 0, y_{i2} = 1 | n_i = 1) = \frac{P(y_{i1} = 0, y_{i2} = 1)}{P(y_{i1} = 0, y_{i2} = 1) + P(y_{i1} = 1, y_{i2} = 0)} \quad (17.12)$$

$$P(y_{i1} = 1, y_{i2} = 0 | n_i = 1) = \frac{P(y_{i1} = 1, y_{i2} = 0)}{P(y_{i1} = 0, y_{i2} = 1) + P(y_{i1} = 1, y_{i2} = 0)} \quad (17.13)$$

假设在给定  $u_i$  与  $x_i$  的条件下， $y_{i1}$  与  $y_{i2}$  相互独立，则

$$P(y_{i1} = 0, y_{i2} = 1) = \frac{1}{1 + e^{u_i + x_{i1}'\beta}} \cdot \frac{e^{u_i + x_{i2}'\beta}}{1 + e^{u_i + x_{i2}'\beta}} \quad (17.14)$$

$$P(y_{i1} = 1, y_{i2} = 0) = \frac{e^{u_i + x_{i1}'\beta}}{1 + e^{u_i + x_{i1}'\beta}} \cdot \frac{1}{1 + e^{u_i + x_{i2}'\beta}} \quad (17.15)$$

将表达式(17.14)与(17.15)代入方程(17.12)可得

$$\begin{aligned} P(y_{i1} = 0, y_{i2} = 1 | n_i = 1) &= \frac{e^{u_i + x_{i2}'\beta}}{e^{u_i + x_{i1}'\beta} + e^{u_i + x_{i2}'\beta}} \\ &= \frac{e^{x_{i2}'\beta}}{e^{x_{i1}'\beta} + e^{x_{i2}'\beta}} = \frac{e^{(x_{i2} - x_{i1})'\beta}}{1 + e^{(x_{i2} - x_{i1})'\beta}} = \Lambda[(x_{i2} - x_{i1})'\beta] \end{aligned} \quad (17.16)$$

注意到在上式中，分子与分母中的  $e^{u_i}$  被同时消去了，这正是逻辑分布的方便之处。类似地，将表达式(17.14)与(17.15)代入方程(17.13)可得

$$P(y_{i1} = 1, y_{i2} = 0 | n_i = 1) = \Lambda[-(x_{i2} - x_{i1})'\beta] = 1 - \Lambda[(x_{i2} - x_{i1})'\beta] \quad (17.17)$$

定义虚拟变量  $d_i = 1$ ，如果  $(y_{i1} = 0, y_{i2} = 1)$ ； $d_i = 0$ ，如果  $(y_{i1} = 1, y_{i2} = 0)$ 。综合方程(17.16)与(17.17)，可将个体  $i$  的条件对数似然函数写为

$$\ln L_i(\beta) = \{d_i \ln \Lambda[(x_{i2} - x_{i1})'\beta] + (1 - d_i) \ln \{1 - \Lambda[(x_{i2} - x_{i1})'\beta]\}\} \cdot \mathbf{1}(n_i = 1) \quad (17.18)$$

其中,  $\mathbf{1}(\cdot)$  为示性函数,  $\mathbf{1}(n_i = 1)$  表示仅考虑  $n_i = 1$  的观测值。将上式对  $i$  加总, 即可得到整个样本的条件对数似然函数。从方程(17.18)可以看出:

- (1) 给定  $n_i$  的条件似然函数不再依赖于  $u_i$  (已在前面的推导中消去);
- (2) 此条件似然函数仍为 Logit 模型, 只是解释变量变为  $(x_{i2} - x_{i1})$ ;
- (3) 不随时间变化的变量将无法识别其系数, 因为  $(x_{i2} - x_{i1}) = 0$ 。另外, 与随机效应模型相比, 固定效应的似然函数(17.18)并不包含积分, 故没有必要进行数值积分。

更一般地, 对于  $T > 2$ , 可以计算给定  $n_i = 1, n_i = 2, \dots$ , 或  $n_i = T - 1$  的条件似然函数。固定效应模型的缺点是, 将损失所有  $n_i = 0$  或  $n_i = T$  的观测值, 导致样本容量减少。另外, 由于条件似然函数估计法消去了  $u_i$ , 故无法估计个体效应  $u_i$ , 也因此无法预测  $y_i$  的发生概率或解释变量对  $y_i$  的边际效应 (它们的表达式都包含  $u_i$ )。解决方法之一是假设  $u_i = 0$ 。

如何在固定效应模型与混合回归之间进行选择? 可以使用以下的豪斯曼检验。原假设为不存在个体效应, 即“ $H_0: u_i = u$ ”。如果原假设成立, 则固定效应模型与混合回归都是一致的, 但混合回归更有效率 (固定效应模型未利用“ $H_0: u_i = u$ ”的信息, 而且可能损失样本容量)。反之, 如果原假设不成立, 则固定效应模型是一致的, 而混合回归不一致。如果二者的系数估计值相差较大 (以二次型来衡量), 则倾向于拒绝原假设, 接受存在个体效应的替代假设。更进一步, 对于固定效应模型与随机效应模型之间的选择, 也可以进行豪斯曼检验。

固定效应面板 logit 的 Stata 命令为,

```
xtlogit y x1 x2 x3, fe
```

## 17.4 面板二值选择模型的 Stata 实例

Chen(2014) 使用公元 25—1911 年期间的王朝面板 (dynamic panel) 来考察中国北方农民起义的决定因素, 横截面单位为王朝 (dyn), 而时间维度为年 (year)。数据集“uprising.dta”包括被解释变量 uprising (= 1, 发生农民起义; = 0, 其他), 以及以下解释变量: age (王朝年龄), pop (人口), temp (气温), sfamine (= 1, 发生严重饥荒; = 0, 其他), sfamine1 (sfamine 的一阶滞后), sfamine2 (sfamine 的二阶滞后), sfamine3 (sfamine 的三阶滞后), relief (= 1, 如果有政府救济; = 0, 其他), relief1 (relief 的一阶滞后), relief2 (relief 的二阶滞后), relief3 (relief 的三阶滞后)。由于面板 probit 无固定效应模型, 故使用面板 logit 模型。

```
. use uprising.dta, clear
```

首先, 进行固定效应面板 logit 估计。

```
. xtlogit uprising age pop temp sfamine sfamine1 sfamine2 sfamine3  
relief relief1 relief2 relief3, fe nolog  
. estimates store FE
```

note: multiple positive outcomes within groups encountered.						
note: 1 group (43 obs) dropped because of all positive or all negative outcomes.						
Conditional fixed-effects logistic regression						
Group variable: dyn			Number of obs	=	1561	
			Number of groups	=	10	
			Obs per group: min =		35	
			avg =		156.1	
			max =		287	
Log likelihood	= -416.20732		LR chi2(11)	=	191.55	
			Prob > chi2	=	0.0000	
uprising	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0119585	.0019157	6.24	0.000	.0082038	.0157131
pop	.0311968	.0145438	2.15	0.032	.0026914	.0597021
temp	.5265406	.1989067	2.65	0.008	.1366906	.9163906
sfamine	1.130333	.3002773	3.76	0.000	.5418006	1.718866
sfaminel	.5676738	.3271661	1.74	0.083	-.0735599	1.208907
sfamine2	.5911076	.3238025	1.83	0.068	-.0435335	1.225749
sfamine3	.577348	.3177628	1.82	0.069	-.0454556	1.200152
relief	-.5210128	.2390826	-2.18	0.029	-.9896061	-.0524195
relief1	-.6768912	.2469321	-2.74	0.006	-1.160869	-.1929132
relief2	-.8935897	.2521104	-3.54	0.000	-1.387717	-.3994624
relief3	-.7126752	.2436877	-2.92	0.003	-1.190294	-.235056

从上表第二行可知,有一个朝代(共43年的观测值)从未发生农民起义,故在估计固定效应模型时被去掉。检查数据可知,该朝代为第二个朝代(朝代变量dyn=2),即三国时期的曹魏。变量age, pop, temp的系数均至少在5%水平上显著为正,这意味着王朝后期更容易发生农民起义;人口压力越大,则农民起义的概率越高;而气温越高,也越易发生农民起义<sup>①</sup>。变量sfamine(严重饥荒)的当期系数在1%水平上显著为正,其前三阶滞后的系数也在10%水平上显著为正,这表明严重饥荒显著地增加农民起义的发生概率。另一方面,变量relief(政府救灾)的各期系数均至少在5%水平上显著为负,意味着政府救灾有助于缓解农民起义。

其次,估计混合回归logit模型。为保持与固定效应的样本一致,在混合回归中也去掉了“dyn=2”时的朝代。

```
. logit uprising age pop temp sfamine sfaminel sfamine2 sfamine3
relief relief1 relief2 relief3 if dyn ~ = 2, vce(cluster dyn) nolog
```

Logistic regression						
Number of obs						
Wald chi2(8)						
Prob > chi2						
Pseudo R2						
Log pseudolikelihood = -505.18692						
(Std. Err. adjusted for 10 clusters in dyn)						
uprising	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	.004675	.0021493	2.18	0.030	.0004623	.0088876
pop	.0099646	.0074699	1.33	0.182	-.0046761	.0246053
temp	.4547908	.3623444	1.26	0.209	-.2553911	1.164973
sfamine	1.234068	.4340109	2.84	0.004	.3834224	2.084714
sfaminel	.7412263	.363804	2.04	0.042	.0281836	1.454269
sfamine2	.787409	.2799314	2.81	0.005	.2387534	1.336065
sfamine3	.8675515	.2147783	4.04	0.000	.4465938	1.288509
relief	-.2743782	.1292077	-2.12	0.034	-.5276207	-.0211357
relief1	-.4558084	.2266392	-2.01	0.044	-.900013	-.0116037
relief2	-.5703093	.2113629	-2.70	0.007	-.984573	-.1560457
relief3	-.372896	.221107	-1.69	0.092	-.8062577	.0604657
_cons	-2.466138	.5024458	-4.91	0.000	-3.450914	-1.481362

<sup>①</sup> 进一步的研究发现,人口与气温的作用并不稳健。

为了在固定效应模型与混合回归之间进行选择,下面进行豪斯曼检验。但豪斯曼检验不允许在混合回归使用选择项“vce(cluster dyn)”,故重新进行混合回归。

```
. quietly logit uprising age pop temp sfamine sfaminel sfamine2
sfamine3 relief relief1 relief2 relief3 if dyn ~ = 2, nolog
.estimates store POOLED
.hausman FE POOLED
```

	<b>Coefficients</b>		<b>(b-B)</b> Difference	<b>sqrt(diag(V_b-V_B))</b> S.E.
	<b>(b)</b> FE	<b>(B)</b> POOLED		
age	.0119585	.004675	.0072835	.0015887
pop	.0311968	.0099646	.0212322	.0140593
temp	.5265406	.4547908	.0717497	.1283301
sfamine	1.130333	1.234068	-.1037347	.1321584
sfaminel	.5676738	.7412263	-.1735526	.149367
sfamine2	.5911076	.787409	-.1963013	.1502582
sfamine3	.577348	.8675515	-.2902035	.1540328
relief	-.5210128	-.2743782	-.2466346	.1026869
relief1	-.6768912	-.4558084	-.2210828	.1042203
relief2	-.8935897	-.5703093	-.3232804	.1141729
relief3	-.7126752	-.372896	-.3397792	.1134979

b = consistent under  $H_0$  and  $H_a$ ; obtained from xtlogit  
 B = inconsistent under  $H_a$ , efficient under  $H_0$ ; obtained from logit

Test:  $H_0$ : difference in coefficients not systematic

$$\begin{aligned} \text{chi2}(11) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 71.76 \\ \text{Prob}>\text{chi2} &= 0.0000 \end{aligned}$$

上表结果显示,强烈拒绝混合回归的原假设,认为应使用固定效应模型。下面进行随机效应面板 logit 估计。

```
. xtlogit uprising age pop temp sfamine sfaminel sfamine2 sfamine3
relief relief1 relief2 relief3 if dyn ~ = 2, re nolog
.estimates store RE
```

Random-effects logistic regression	Number of obs	=	1561		
Group variable: dyn	Number of groups	=	10		
Random effects u_i ~ Gaussian	Obs per group:	min =	35		
		avg =	156.1		
		max =	287		
Log likelihood = -458.18161	Wald chi2(11)	=	127.59		
	Prob > chi2	=	0.0000		
uprising	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	.0112875	.0018141	6.22	0.000	.007732 .014843
pop	.0242926	.0107208	2.27	0.023	.0032802 .0453051
temp	.5079557	.1956755	2.60	0.009	.1244387 .8914726
sfamine	1.147349	.2985703	3.84	0.000	.5621616 1.732536
sfamine1	.5874378	.3250784	1.81	0.071	-.0497041 1.22458
sfamine2	.6138714	.3219423	1.91	0.057	-.0171238 1.244867
sfamine3	.598085	.3153273	1.90	0.058	-.0199452 1.216115
relief	-.5251701	.2378131	-2.21	0.027	-.9912753 -.059065
relief1	-.684916	.2455628	-2.79	0.005	-1.16621 -.2036217
relief2	-.8914661	.2502494	-3.56	0.000	-1.381946 -.4009862
relief3	-.7093608	.242091	-2.93	0.003	-1.18385 -.2348712
_cons	-2.872401	.5282318	-5.44	0.000	-3.907716 -1.837085
/lnsig2u	.5281101	.5038209		-.4593608	1.515581
sigma_u	1.3022	.3280378		.7947876	2.133557
rho	.340125	.1130776		.1610808	.5804774
Likelihood-ratio test of rho=0: chibar2(01) = 94.01 Prob >= chibar2 = 0.000					

从上表最后一行可知,LR 检验的强烈拒绝原假设“ $H_0: \rho = 0$ ”,认为应使用面板随机效应模型,不宜进行混合回归。由于随机效应面板 Logit 的计算使用了数值积分,下面使用命令 quadchk 检验此数值积分的稳健性。

```
.quadchk,noout
```

其中,选择项“noout”表示不显示重估模型的结果。命令“xtlogit,re”默认在 12 个点上进行数值积分,而命令“quadchk”则在 8 个点与 16 个点上进行数值积分,重新估计此模型(refitting model),然后与默认的 12 点上数值积分结果相比较。一般来说,只要这几种数值积分的系数估计值的“相对差距”(relative difference)不超过  $10^{-4}$ ,则认为结果较为稳健。否则,应使用选项“intpoints(#)”来增加数值积分的计算点数,比如“xtlogit,re intpoints(24)”表示在 24 个点上进行数值积分。

Refitting model intpoints() = 8  
Refitting model intpoints() = 16

Quadrature check				
	Fitted quadrature 12 points	Comparison quadrature 8 points	Comparison quadrature 16 points	
Log likelihood	-458.18161	-458.18163 -.00001689 3.687e-08	-458.18161 9.789e-08 -2.136e-10	Difference Relative difference
uprising: age	.01128749	.01128749 -4.194e-12 -3.716e-10	.01128749 -1.647e-11 -1.459e-09	Difference Relative difference
uprising: pop	.02429264	.02429264 2.767e-11 1.139e-09	.02429264 4.978e-10 2.049e-08	Difference Relative difference
uprising: temp	.50795566	.50795566 1.938e-11 3.816e-11	.50795565 -2.265e-09 -4.458e-09	Difference Relative difference
uprising: sfamine	1.1473487	1.1473487 1.586e-10 1.382e-10	1.1473487 6.578e-09 5.733e-09	Difference Relative difference
uprising: sfaminel	.58743785	.58743785 1.730e-10 2.944e-10	.58743785 1.531e-09 2.606e-09	Difference Relative difference
uprising: sfamine2	.61387142	.61387142 2.203e-10 3.588e-10	.61387143 1.580e-09 2.574e-09	Difference Relative difference
uprising: sfamine3	.598085	.598085 3.440e-10 5.752e-10	.598085 6.962e-10 1.164e-09	Difference Relative difference
uprising: relief	-.52517014	-.52517014 3.558e-10 -6.775e-10	-.52517015 -5.434e-09 1.035e-08	Difference Relative difference
uprising: relief1	-.68491604	-.68491604 3.761e-10 -5.492e-10	-.68491604 -4.941e-09 7.214e-09	Difference Relative difference
uprising: relief2	-.89146607	-.89146607 3.875e-10 -4.347e-10	-.89146607 -5.759e-09 6.461e-09	Difference Relative difference
uprising: relief3	-.70936077	-.70936077 4.232e-10 -5.966e-10	-.70936078 -5.424e-09 7.646e-09	Difference Relative difference
uprising: _cons	-2.8724008	-2.8724008 -2.114e-08 7.361e-09	-2.8724008 -5.972e-08 2.079e-08	Difference Relative difference
lnsig2u: _cons	.5281101	.5281101 -1.799e-09 -3.407e-09	.52811025 1.540e-07 2.917e-07	Difference Relative difference

从上表可知,所有系数的相对差距均小于 $10^{-4}$ ,故数值积分的结果值得信赖。

为了在固定效应与随机效应模型之间进行选择,下面再次进行豪斯曼检验。

. hausman FE RE

	<u>Coefficients</u>			
	(b) FE	(B) RE	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
age	.0119585	.0112875	.000671	.0006156
pop	.0311968	.0242926	.0069041	.0098278
temp	.5265406	.5079557	.0185849	.0357068
sfamine	1.130333	1.147349	-.0170153	.0319724
sfamine1	.5676738	.5874378	-.0197641	.0369005
sfamine2	.5911076	.6138714	-.0227638	.0346586
sfamine3	.577348	.598085	-.020737	.0392664
relief	-.5210128	-.5251701	.0041573	.0246053
relief1	-.6768912	-.684916	.0080249	.0259682
relief2	-.8935897	-.8914661	-.0021236	.030576
relief3	-.7126752	-.7093608	-.0033144	.0278508

b = consistent under  $H_0$  and  $H_a$ ; obtained from xtlogit  
 B = inconsistent under  $H_a$ , efficient under  $H_0$ ; obtained from xtlogit

Test:  $H_0$ : difference in coefficients not systematic

chi2(11) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)  
 = 21.57  
 Prob>chi2 = 0.0280  
 (V\_b-V\_B is not positive definite)

上表结果表明,可以在5%水平上拒绝随机效应的原假设,认为应使用固定效应模型。

最后,为了演示的目的,估计面板probit的随机效应模型。

. xtprobit uprising age pop temp sfamine sfamine1 sfamine2 sfamine3  
 relief relief1 relief2 relief3 if dyn ~ =2, re nolog

Log likelihood	= -460.72517	avg = 156.1			
		max = 287			
		Wald chi2(11) = 135.18			
uprising	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	.0056746	.0009103	6.23	0.000	.0038904 .0074587
pop	.0134784	.0055652	2.42	0.015	.0025708 .024386
temp	.2441783	.1007821	2.42	0.015	.0466491 .4417075
sfamine	.6406771	.1649126	3.88	0.000	.3174543 .9638998
sfamine1	.3719557	.1765811	2.11	0.035	.0258631 .7180483
sfamine2	.349083	.1759537	1.98	0.047	.0042202 .6939458
sfamine3	.3321031	.1728483	1.92	0.055	-.0066732 .6708795
relief	-.2639172	.1241089	-2.13	0.033	-.5071662 -.0206682
relief1	-.3895312	.1271142	-3.06	0.002	-.6386704 -.140392
relief2	-.456395	.1278903	-3.57	0.000	-.7070553 -.2057347
relief3	-.3525801	.1251987	-2.82	0.005	-.5979651 -.1071952
_cons	-1.609203	.2701962	-5.96	0.000	-2.138778 -1.079628
/lnsig2u	-.8042879	.5053684		-1.794792	.1862159
sigma_u	.6688844	.1690165		.4076298	1.09758
rho	.309109	.1079268		.1424862	.5464199
Likelihood-ratio test of rho=0: chibar2(01) = 89.11 Prob >= chibar2 = 0.000					

上表最后一行的 LR 检验结果强烈拒绝混合 probit 回归,认为应使用随机效应模型。随机效应面板 probit 的回归结果在性质上与随机效应面板 logit 相似。

## 17.5 面板泊松回归

考虑被解释变量为计数变量的面板数据。比如,若干企业在一段时间内每年获得专利的个数,全国各省在几年内每年发生矿难的数目,世界各国近几十年每年发生示威游行的次数,一些病人在一段时间内的每期发病次数,等等。

对于个体  $i$ ,时期  $t$ ,记被解释变量为  $Y_{it}$ ,假设  $Y_{it} = y_{it}$  的概率由参数为  $\lambda_{it}$  的泊松分布所决定:

$$P(Y_{it} = y_{it} | \mathbf{x}_{it}) = \frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!} \quad (y_{it} = 0, 1, 2, \dots) \quad (17.19)$$

其中,  $\lambda_{it} > 0$  为泊松到达率。为了保证  $\lambda_{it}$  为非负,假设

$$\lambda_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + u_i) = \exp(\mathbf{x}'_{it}\boldsymbol{\beta}) \cdot \exp(u_i) \equiv v_i \exp(\mathbf{x}'_{it}\boldsymbol{\beta}) \quad (17.20)$$

其中,  $\mathbf{x}_{it}$  不含常数项,而  $v_i \equiv \exp(u_i)$  为乘积形式的个体效应(multiplicative individual effects)。如果  $v_1 = v_2 = \dots = v_n$ , 则不存在个体效应,为混合泊松回归(pooled Poisson),可将此面板数据作为截面数据处理。此时,可直接用截面数据的泊松回归 Stata 命令进行混合回归,但需使用聚类稳健标准误。

混合泊松回归的 Stata 命令为,

```
poisson y x1 x2 x3, vce(cluster id) irr
```

其中,  $id$  是用来确定面板单位的变量。选择项“`irr`”表示汇报发生率比(incidence rate ratio)。

更一般地,我们允许个体效应的存在,即不同个体拥有不同的  $v_i$ 。如果  $v_i$  与所有解释变量  $\mathbf{x}_{it}$  均不相关,则为“随机效应模型”。如果  $v_i$  与某个解释变量相关,则为“固定效应模型”。

首先考虑随机效应模型。由于这是非线性模型,故可使用最大似然估计(MLE)。记  $v_i$  的密度函数为  $g(v_i)$ 。假设样本为 iid,则在给定  $v_i$  的情况下,个体  $i$  的条件分布为,

$$f(y_{i1}, y_{i2}, \dots, y_{iT} | v_i) = \prod_{t=1}^T \left( \frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!} \right) = \prod_{t=1}^T \left( \frac{e^{-v_i \exp(\mathbf{x}'_{it}\boldsymbol{\beta})} [v_i \exp(\mathbf{x}'_{it}\boldsymbol{\beta})]^{y_{it}}}{y_{it}!} \right) \quad (17.21)$$

但上式中的  $v_i$  不可观测。为此,记  $(y_{i1}, y_{i2}, \dots, y_{iT}, v_i)$  的联合密度为  $f(y_{i1}, y_{i2}, \dots, y_{iT}, v_i)$ , 并进行如下分解,

$$f(y_{i1}, y_{i2}, \dots, y_{iT}, v_i) = f(y_{i1}, y_{i2}, \dots, y_{iT} | v_i) \cdot g(v_i) \quad (17.22)$$

将上式中的  $v_i$  积分积掉,即可得到  $(y_{i1}, y_{i2}, \dots, y_{iT})$  的边缘密度,

$$\begin{aligned} f(y_{i1}, y_{i2}, \dots, y_{iT}) &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT}, v_i) dv_i \\ &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT} | v_i) \cdot g(v_i) dv_i \end{aligned} \quad (17.23)$$

由于  $v_i = \exp(u_i) > 0$ , 故通常选择  $v_i$  服从 Gamma 分布(指数分布与卡方分布都是 Gamma 分布的特例)。特别地,假设  $v_i \sim \text{Gamma}(1/\alpha, \alpha)$ , 其中  $\alpha > 0$ , 则  $E(v_i) = 1$ ,  $\text{Var}(v_i) = \alpha$ 。将 Gamma  $(1/\alpha, \alpha)$  的概率密度代入方程(17.23),可以得到负二项分布的概率密度,然后进行 MLE 估计。由于参数  $\alpha$  为个体异质性  $v_i$  的方差,故可对“ $H_0: \alpha = 0$ ”(Stata 称为“`alpha = 0`”)进行 LR 检验来判

断是否存在个体异质性,即究竟应使用随机效应泊松回归还是混合泊松回归。

另一种选择是,假设  $u_i \sim N(0, \sigma_u^2)$ ,但得不到解析解,一般用“Gauss-Hermite quadrature”(高斯-汉密尔特正交法)方法对方程(17.23)进行数值积分;也需要使用 Stata 命令 quadchk 来检验此数值积分的精确度。类似地,可通过检验“ $H_0: \sigma_u = 0$ ”(Stata 称为“sigma\_u = 0”)来判断是否存在个体异质性。

下面考虑固定效应的面板泊松模型。类似于固定效应的面板二值选择模型(参见本章第 3 节),可以使用  $n_i \equiv \sum_{t=1}^T y_{it}$  作为  $v_i$  的充分统计量,并计算在给定  $n_i$  情况下的条件似然函数,然后进行条件最大似然估计(conditional MLE)。显然,如果  $n_i = 0$ ,则意味着  $y_{i1} = \dots = y_{iT} = 0$ ,故个体  $i$  的观测值对于条件似然函数的贡献为 0。因此,所有  $y_{it}$  观测值均为 0 的个体,在估计固定效应泊松回归时将被去掉,故可能损失样本容量。固定效应模型的另一缺点是,无法估计不随时间而变的变量系数。固定效应泊松回归的最大优点在于,它允许个体异质性  $v_i$  与解释变量  $x_{it}$  相关。在固定效应与随机效应泊松回归之间进行选择时,可借助于豪斯曼检验。

面板泊松回归的 Stata 命令为

```
xtpoisson y x1 x2 x3, fe normal irr
```

其中,选择项“fe”表示固定效应,默认为“re”(随机效应)。选择项“normal”表示在估计随机效应模型时,指定个体异质性服从正态分布,默认为服从 Gamma 分布。选择项“irr”表示汇报发生率比(incidence rate ratio)。

## 17.6 面板负二项回归

泊松回归的缺陷是,它假设均等分散,即方差等于期望。如果存在过度分散,即方差大于期望,则可考虑负二项回归。如果不存在个体异质性,则可直接进行混合负二项回归,其 Stata 命令为

```
nbreg y x1 x2 x3, vce(cluster id)
```

其中,id 为用来确定面板单位的变量。Stata 将输出过度分散参数  $\alpha$  的 95% 置信区间。根据这个置信区间,可以检验原假设“ $H_0: \alpha = 0$ ”(对应于泊松回归)。

对于随机效应的面板负二项模型,假设个体异质性服从 Beta( $r, s$ ) 分布,则可将个体异质性积分掉,然后进行 MLE 估计(此处理方法类似于随机效应的面板泊松模型)。对于固定效应的面板负二项模型,则可考虑在给定  $n_i \equiv \sum_{t=1}^T y_{it}$ (个体异质性的充分统计量)情况下的条件似然函数,然后进行条件最大似然估计(conditional MLE)。固定效应负二项回归的奇特之处在于,它可以估计不随时间而变的变量系数。在固定效应与随机效应负二项回归之间进行选择时,可使用豪斯曼检验。如何在泊松回归与负二项回归之间进行选择,涉及稳健性与有效性之间的权衡(前者稳健,而后者更有效),有关讨论参见第 13 章第 3 节。

面板负二项回归的 Stata 命令为

```
xtnbreg y x1 x2 x3, fe irr
```

其中,选择项“fe”表示固定效应,默认为“re”(随机效应)。选择项“irr”表示汇报发生率比(incidence rate ratio)。

## 17.7 面板计数模型的 Stata 实例

下面以数据集 mus18data.dta 为例<sup>①</sup>,介绍面板计数模型的 Stata 操作。此数据集来自“Rand Health Insurance Experiment”。该实验随机地给家庭分配不同的健康保险单(health insurance policy),然后跟踪调查几年,故为面板数据。该实验的主要目的是研究不同的共保率(coinsurance rate)如何影响对医疗服务的使用。如果共保率为 25%,则意味着投保人支付 25% 的医疗费用,而保险公司支付 75%。

在本例中,被解释变量 mdu 表示个体看医生的次数,为计数变量。解释变量包括 leoins(共保率加 1 再取对数,即  $\log(\text{coinsurance} + 1)$ ),ndisease(所患慢性病的数目),female(是否女性),age(年龄),lfam(家庭人数的对数),child(是否有孩子)。

### (1) 数据集的结构与统计特征。

```
. use mus18data.dta,clear  
. xtset id year
```

panel variable:	id (unbalanced)
time variable:	year, 1 to 5, but with gaps
delta:	1 unit

由上表可知,面板变量为 id(对应于每位个体),时间跨度为 5 年,为非平衡面板。

```
. xtdescribe
```

id:	125024, 125025, ..., 632167	n =	5908				
year:	1, 2, ..., 5	T =	5				
Delta(year)	= 1 unit						
Span(year)	= 5 periods						
	(id*year uniquely identifies each observation)						
Distribution of T_i:	min	5%	25%	50%	75%	95%	max
	1	2	3	3	5	5	5
Freq.	Percent	Cum.	Pattern				
3710	62.80	62.80	111..				
1584	26.81	89.61	11111				
156	2.64	92.25	1....				
147	2.49	94.74	11...				
79	1.34	96.07	..1..				
66	1.12	97.19	.11..				
33	0.56	97.75	..111				
33	0.56	98.31	.1111				
29	0.49	98.80	...11				
71	1.20	100.00	(other patterns)				
5908	100.00		XXXXX				

<sup>①</sup> 此数据集取自 Cameron and Trivedi (2010)。

由上表下部可知,89.61% 的个体在前三年或全部五年都在样本中。

```
. sum mdu lcoins ndisease female age lfam child
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mdu	20186	2.860696	4.504765	0	77
lcoins	20186	2.383588	2.041713	0	4.564348
ndisease	20186	11.2445	6.741647	0	58.6
female	20186	.5169424	.4997252	0	1
age	20186	25.71844	16.76759	0	64.27515
lfam	20186	1.248404	.5390681	0	2.639057
child	20186	.4014168	.4901972	0	1

也可使用命令 `xtsum` 得到更详细的面板数据统计特征。

(2) 进行混合泊松回归,并使用聚类稳健标准误。

```
. poisson mdu lcoins ndisease female age lfam child, vce (cluster id) nolog
```

Poisson regression		Number of obs = 20186			
		Wald chi2(6) = 476.93			
		Prob > chi2 = 0.0000			
Log pseudolikelihood = -62579.401		Pseudo R2 = 0.0609			
(Std. Err. adjusted for 5908 clusters in id)					
		Robust			
mdu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lcoins	-.0808023	.0080013	-10.10	0.000	-.0964846 -.0651199
ndisease	.0339334	.0026024	13.04	0.000	.0288328 .039034
female	.1717862	.0342551	5.01	0.000	.1046473 .2389251
age	.0040585	.0016891	2.40	0.016	.000748 .0073691
lfam	-.1481981	.0323434	-4.58	0.000	-.21159 -.0848062
child	.1030453	.0506901	2.03	0.042	.0036944 .2023961
_cons	.748789	.0785738	9.53	0.000	.5947872 .9027907

由上表可知,所有解释变量均至少在 5% 水平上显著。其中,lcoins 与 lfam 的系数显著为负,表明共保率越高、家庭成员越多,则看医生次数越少。ndisease,female,age 与 child 的系数显著为正,表明所患慢性病越多、女性、年龄越大、有子女,则看医生次数越多。

(3) 进行随机效应的面板泊松回归。由于命令“`xtpoisson,re`”不提供聚类稳健标准误,故使用选择项“`vce(boot)`”计算聚类自助标准误(有关自助法,参见第 19 章);运行以下命令将花费较长时间。

```
. xtpoisson mdu lcoins ndisease female age lfam child, re vce(boot,reps(50) seed(10101) nodots)
```

其中,“`reps(50)`”表示重复抽样 50 次,“`seed(10101)`”表示以 10101 作为随机数的种子(参见第 19 章附录),而“`nodots`”表示不显示重复抽样过程的点(dots)。

Random-effects Poisson regression Group variable: id	Number of obs = 20186 Number of groups = 5908				
Random effects u_i ~ Gamma	Obs per group: min = 1 avg = 3.4 max = 5				
Log likelihood = -43240.556	Wald chi2(6) = 476.31 Prob > chi2 = 0.0000				
(Replications based on 5908 clusters in id)					
mdu	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]
lcoins	-.0878258	.0080482	-10.91	0.000	-.1036 -.0720515
ndisease	.0387629	.0027883	13.90	0.000	.0332979 .044228
female	.1667192	.0383524	4.35	0.000	.0915499 .2418885
age	.0019159	.0017831	1.07	0.283	-.0015789 .0054107
lfam	-.1351786	.034328	-3.94	0.000	-.2024602 -.067897
child	.1082678	.0548715	1.97	0.048	.0007218 .2158139
_cons	.7574177	.0867116	8.73	0.000	.587466 .9273694
/lnalpha	.0251256	.0278546			-.0294685 .0797197
alpha	1.025444	.0285634			.9709615 1.082983
Likelihood-ratio test of alpha=0: chibar2(01) = 3.9e+04 Prob>chibar2 = 0.000					

上表最后一行的 LR 检验结果强烈拒绝“ $\alpha = 0$ ”的原假设(对应于混合泊松回归,其中  $\alpha$  为个体异质性  $v_i$  的方差),即拒绝混合泊松模型,认为应使用随机效应的面板泊松模型。命令“`xtpoisson,re`”默认假设个体异质性  $v_i$  服从 Gamma 分布。也可使用选择项“normal”指定个体异质性服从正态分布,为了提高速度,下面仅使用普通标准误(不使用自助标准误)。

```
. xtpoisson mdu lcoins ndisease female age lfam child, re normal nolog
```

Random-effects Poisson regression Group variable: id	Number of obs = 20186 Number of groups = 5908				
Random effects u_i ~ Gaussian	Obs per group: min = 1 avg = 3.4 max = 5				
Log likelihood = -43226.889	Wald chi2(6) = 798.96 Prob > chi2 = 0.0000				
(Replications based on 5908 clusters in id)					
mdu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lcoins	-.1145337	.0072788	-15.74	0.000	-.1287998 -.1002676
ndisease	.0408695	.0022941	17.81	0.000	.0363731 .045366
female	.208436	.0304848	6.84	0.000	.148687 .268185
age	.002674	.0011961	2.24	0.025	.0003297 .0050182
lfam	-.1443327	.0265365	-5.44	0.000	-.1963432 -.0923222
child	.0737146	.0344697	2.14	0.032	.0061553 .141274
_cons	.2872796	.0641625	4.48	0.000	.1615234 .4130359
/lnsig2u	.0549982	.0254991	2.16	0.031	.005021 .1049755
sigma_u	1.027881	.013105			1.002514 1.05389
Likelihood-ratio test of sigma_u=0: chibar2(01) = 3.9e+04 Pr>chibar2 = 0.000					

从上表可知,假设个体异质性服从正态分布的估计系数与默认 Gamma 分布的估计系数十分接近。而且,前者的对数似然函数略高于后者,说明前者拟合得更好。由于使用了数值积分,故需要检验数值积分的准确度。

. quadchk,nooutput

Quadrature check				
	Fitted quadrature 12 points	Comparison quadrature 8 points	Comparison quadrature 16 points	
Log likelihood	-43226.889	-43226.885 .00461531 -1.068e-07	-43226.885 .00462363 -1.070e-07	Difference Relative difference
mdu: lcoins	-.11453372	-.11458364 -.00004992 .00043585	-.11453628 -2.567e-06 .00002241	Difference Relative difference
mdu: ndisease	.04086953	.04088382 .00001429 .00034964	.04087027 7.382e-07 .00001806	Difference Relative difference
mdu: female	.20843598	.20851939 .0000834 .00040013	.20844042 4.440e-06 .0000213	Difference Relative difference
mdu: age	.00267396	.00267552 1.561e-06 .00058392	.00267402 6.599e-08 .00002468	Difference Relative difference
mdu: lfam	-.14433271	-.1443633 -.00003059 .00021195	-.14433429 -1.580e-06 .00001095	Difference Relative difference
mdu: child	.07371462	.07371141 -3.210e-06 -.00004355	.07371365 -9.678e-07 -.00001313	Difference Relative difference
mdu: _cons	.28727964	.28660796 -.00067167 -.00233805	.28726141 -.00001823 -.00006347	Difference Relative difference
lnsig2u: _cons	.05499824	.05529067 .00029244 .00531723	.05499982 1.580e-06 .00002873	Difference Relative difference

上表显示,对于某些系数,其相对差距(relative difference)超过了 $10^{-4}$ 。故下面使用选择项“intpoints(#)”将数值积分的计算点数增加到16(默认为12)。

```
. xtpoisson mdu lcoins ndisease female age lfam child,re normal nolog
intpoints(16)
```

Random-effects Poisson regression		Number of obs = 20186			
Group variable: id		Number of groups = 5908			
Random effects u_i ~ Gaussian		Obs per group: min = 1			
		avg = 3.4			
		max = 5			
		Wald chi2(6) = 799.05			
Log likelihood = -43226.885		Prob > chi2 = 0.0000			
mdu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lcoins	-.1145363	.0072785	-15.74	0.000	-.1288019 -.1002706
ndisease	.0408703	.0022941	17.82	0.000	.0363739 .0453666
female	.2084404	.0304838	6.84	0.000	.1486933 .2681875
age	.002674	.001196	2.24	0.025	.0003298 .0050182
lfam	-.1443343	.0265357	-5.44	0.000	-.1963433 -.0923253
child	.0737136	.0344692	2.14	0.032	.0061552 .1412721
_cons	.2872615	.0641609	4.48	0.000	.1615084 .4130146
/lnsig2u	.0549996	.0254986	2.16	0.031	.0050232 .1049759
sigma_u	1.027881	.0131048			1.002515 1.05389
Likelihood-ratio test of sigma_u=0: chibar2(01) = 3.9e+04 Pr>chibar2 = 0.000					

然后再检验数值积分的精确度。

```
. quadchk,nooutput
```

```
Refitting model intpoints() = 11
Refitting model intpoints() = 21
```

Quadrature check

	Fitted quadrature 16 points	Comparison quadrature 11 points	Comparison quadrature 21 points	
Log likelihood	-43226.885	-43226.894 -.00925502 2.141e-07	-43226.885 -.00029544 6.835e-09	Difference Relative difference
mdu: lcoins	-.11453627	-.11453627 -4.418e-11 3.858e-10	-.11453611 1.668e-07 -1.456e-06	Difference Relative difference
mdu: ndisease	.04087026	.04087026 1.112e-11 2.722e-10	.04087022 -4.246e-08 -1.039e-06	Difference Relative difference
mdu: female	.20844041	.20844041 6.809e-11 3.266e-10	.20844014 -2.617e-07 -1.256e-06	Difference Relative difference
mdu: age	.00267402	.00267402 1.133e-12 4.237e-10	.00267402 -4.253e-09 -1.591e-06	Difference Relative difference
mdu: lfam	-.14433431	-.14433431 -2.772e-11 1.921e-10	-.14433418 1.299e-07 -8.998e-07	Difference Relative difference
mdu: child	.07371365	.07371365 -3.389e-13 -4.598e-12	.07371368 3.289e-08 4.462e-07	Difference Relative difference
mdu: _cons	.28726149	.28726149 -3.947e-10 -1.374e-09	.28726306 1.564e-06 5.444e-06	Difference Relative difference
lnsig2u: _cons	.05499955	.05499955 2.615e-10 4.755e-09	.054999 -5.539e-07 -0.00001007	Difference Relative difference

上表显示,所有系数的相对差距均小于 $10^{-4}$ ,故可以认为结果是稳健的。

(4) 进行固定效应的面板泊松回归。但命令“`xtpoisson, fe`”并不提供聚类稳健标准误。解决方法之一是使用自助标准误的选择项,比如“`vce (boot, reps (50) seed (10101) nodots)`”。解决方法之二为下载非官方命令“`xtpqml`”(其中,`pqml` 表示 Poisson Quasi Maximum Likelihood),该命令的估计结果与“`xtpoisson, fe`”相同,但提供聚类稳健标准误。

```
.ssc install xtpqml (下载安装命令 xtpqml)
```

```
.xtpqml mdu lcoins ndisease female age lfam child,i(id)
```

其中,必选项“`i(id)`”表示以变量 `id` 来确定面板单位。

note: 265 groups (265 obs) dropped because of only one obs per group  
 note: 666 groups (2130 obs) dropped because of all zero outcomes  
 note: lcoins dropped because it is constant within group  
 note: ndisease dropped because it is constant within group  
 note: female dropped because it is constant within group

Iteration 0: log likelihood = -24182.852  
 Iteration 1: log likelihood = -24173.211  
 Iteration 2: log likelihood = -24173.211

Conditional fixed-effects Poisson regression  
 Number of obs = 17791  
 Group variable: id Number of groups = 4977  
 Obs per group: min = 2  
 avg = 3.6  
 max = 5  
 Wald chi2(3) = 19.24  
 Log likelihood = -24173.211 Prob > chi2 = 0.0002

mdu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	-.0112009	.0039024	-2.87	0.004	-.0188494 -.0035523
lfam	.0877134	.0554606	1.58	0.114	-.0209874 .1964141
child	.1059867	.0437744	2.42	0.015	.0201905 .1917829

Calculating Robust Standard Errors...

mdu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
mdu					
age	-.0112009	.0091493	-1.22	0.221	-.0291331 .0067314
lfam	.0877134	.1160837	0.76	0.450	-.1398064 .3152332
child	.1059867	.0786326	1.35	0.178	-.0481304 .2601037

Wald chi2(3) = 4.58 Prob > chi2 = 0.2051

上表前五行显示,有 265 组数据由于仅有一期观测值而被去掉,另有 666 组数据由于取值全为 0 而被去掉;而且,固定效应模型无法识别不随时间而变的变量系数(lcoins, ndisease, female),故这些变量也被去掉。回归结果的上半部提供普通标准误(事实上,复制了命令“xtpoisson, fe”的估计结果);而下半部的标准误则为聚类稳健标准误。可以看出,聚类稳健标准误大约是普通标准误的两倍,导致所有变量系数均不显著。

(5) 初步考察是否存在过度分散。

. sum mdu,detail

number face-to-face md visits					
Percentiles		Smallest			
1%	0	0			
5%	0	0			
10%	0	0	Obs	20186	
25%	0	0	Sum of Wgt.	20186	
50%	1		Mean	2.860696	
		Largest	Std. Dev.	4.504765	
75%	4	72			
90%	7	74	Variance	20.2929	
95%	10	76	Skewness	4.823598	
99%	21	77	Kurtosis	44.75799	

```
. dis r(Var)/r(mean)
```

7.093696

由此可知,被解释变量 mdu 的方差是平均值的 7 倍多,故可能存在过度分散的情况下,尽管泊松回归依然是一致的,但负二项回归可能更有效率。

#### (6) 进行混合负二项回归,并使用聚类稳健标准误。

```
. nbreg mdu lcoins ndisease female age lfam child, vce(cluster id) nolog
```

Negative binomial regression		Number of obs = 20186			
Dispersion = mean		Wald chi2(6) = 531.13			
Log pseudolikelihood = -43403.178		Prob > chi2 = 0.0000			
(Std. Err. adjusted for 5908 clusters in id)					
mdu	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
lcoins	-.0855356	.0079578	-10.75	0.000	-.1011325 -.0699387
ndisease	.036683	.0025083	14.62	0.000	.0317668 .0415993
female	.1746153	.0344054	5.08	0.000	.1071819 .2420486
age	.0035134	.0015948	2.20	0.028	.0003876 .0066392
lfam	-.1588963	.032843	-4.84	0.000	-.2232673 -.0945253
child	.1067837	.0523271	2.04	0.041	.0042244 .2093429
_cons	.7500703	.0805942	9.31	0.000	.5921086 .9080321
/lnalpha	.2621342	.0248113		.2135049	.3107634
alpha	1.299701	.0322473		1.23801	1.364466

上表最后一行显示,过度分散参数  $\alpha$  的 95% 置信区间为 [1.24, 1.36], 故拒绝 " $H_0: \alpha = 0$ ", 认为存在过度分散, 使用负二项回归可以提高效率。

(7) 进行随机效应的面板负二项回归, 并使用自助标准误; 运行此命令需较长时间。

```
. xtnbreg mdu lcoins ndisease female age lfam child, nolog vce(boot, reps(50) seed(10101) nodots)
```

Random-effects negative binomial regression		Number of obs = 20186	
Group variable: id		Number of groups = 5908	
Random effects u_i ~ Beta		Obs per group: min = 1 avg = 3.4 max = 5	
		Wald chi2(6) = 617.55	
Log likelihood = -40661.277		Prob > chi2 = 0.0000	
(Replications based on 5908 clusters in id)			
mdu	Observed Coef.	Bootstrap Std. Err.	z P> z  Normal-based [95% Conf. Interval]
lcoins	-.1072858	.0071451	-15.02 0.000 -.1212899 -.0932817
ndisease	.0333777	.0021562	15.48 0.000 .0291517 .0376038
female	.2039408	.0305068	6.69 0.000 .1441485 .2637331
age	.0023233	.0014846	1.56 0.118 -.0005865 .005233
lfam	-.1434145	.0281385	-5.10 0.000 -.1985651 -.088264
child	.1144578	.0410581	2.79 0.005 .0339853 .1949302
_cons	.8821153	.0791412	11.15 0.000 .7270014 1.037229
/ln_r	1.127977	.0287538	
/ln_s	.7258626	.0355438	1.071621 1.184334 .6561979 .7955272
r	3.089401	.0888319	2.920109 3.268508
s	2.066513	.0734518	1.92745 2.215609
Likelihood-ratio test vs. pooled: chibar2(01) = 5305.26 Prob>=chibar2 = 0.000			

上表最后一行的 LR 检验结果强烈拒绝了混合负二项回归的原假设, 认为应使用随机效应的面板负二项回归。

(8) 进行固定效应的面板负二项回归, 并使用自助标准误; 运行此命令需较长时间。

```
. xtnbreg mdu lcoins ndisease female age lfam child, fe nolog vce(boot, reps(50) seed(10101) nodots)
```

Conditional FE negative binomial regression	Number of obs	=	17791		
Group variable: id	Number of groups	=	4977		
	Obs per group: min =	2			
	avg =	3.6			
	max =	5			
	Wald chi2(6)	=	67.42		
Log likelihood = -21626.982	Prob > chi2	=	0.0000		
(Replications based on 4977 clusters in id)					
mdu	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]
lcoins	-.088532	.0199968	-4.43	0.000	-.127725    -.0493391
ndisease	.0153807	.0061465	2.50	0.012	.0033338    .0274276
female	.2459735	.0636349	3.87	0.000	.1212514    .3706956
age	-.0020748	.0026177	-0.79	0.428	-.0072054    .0030558
lfam	-.0007717	.0570776	-0.01	0.989	-.1126417    .1110983
child	.2032379	.0532196	3.82	0.000	.0989293    .3075464
_cons	.9243416	.1166414	7.92	0.000	.6957287    1.152955

由上表可知,固定效应的面板负二项回归依然可以估计不随时间而变的变量系数,这是固定效应负二项回归的优点。

(9) 最后,使用豪斯曼检验在固定效应与随机效应的负二项回归之间进行选择。

```
. quietly xtnbreg mdu lcoins ndisease female age lfam child
. est sto re
. quietly xtnbreg mdu lcoins ndisease female age lfam child,fe
. est sto fe
. hausman fe re
```

	Coefficients			
	(b) fe	(B) re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
lcoins	-.088532	-.1072858	.0187537	.0124662
ndisease	.0153807	.0333777	-.017997	.00347
female	.2459735	.2039408	.0420327	.0523672
age	-.0020748	.0023233	-.004398	.0016488
lfam	-.0007717	-.1434145	.1426429	.0406003
child	.2032379	.1144578	.0887801	.0382981

b = consistent under  $H_0$  and  $H_a$ ; obtained from xtnbreg  
B = inconsistent under  $H_a$ , efficient under  $H_0$ ; obtained from xtnbreg

Test:  $H_0$ : difference in coefficients not systematic

$$\begin{aligned} \text{chi2}(6) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 101.59 \\ \text{Prob}>\text{chi2} &= 0.0000 \end{aligned}$$

上表结果显示,豪斯曼检验强烈拒绝随机效应负二项回归,支持固定效应负二项回归。

## 17.8 面板 Tobit

考虑归并数据(censored data)的面板模型。假设

$$y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + u_i + \varepsilon_i \quad (17.24)$$

其中,  $y_i^*$  不可观测, 扰动项  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ , 而  $u_i$  为个体效应。如果  $u_1 = u_2 = \dots = u_n$ , 则可直接进行混合 Tobit 回归(参见第 14 章), 但须使用聚类稳健标准误。混合 Tobit 回归的 Stata 命令为

```
tobit y x1 x2 x3, ll(#) ul(#) vce(cluster id)
```

其中, 选择项“ll (#)”表示下限(lower limit), 选择项“ul (#)”表示上限(upper limit), 而 id 为指示面板单位的变量。

更一般地, 我们允许个体效应的存在。不失一般性, 为简便起见, 假定在 0 处存在左归并(left-censored)。假设可以观测到  $y_{it} = \begin{cases} y_{it}^*, & \text{若 } y_{it}^* > 0 \\ 0, & \text{若 } y_{it}^* \leq 0 \end{cases}$ 。如果  $u_i$  与解释变量  $\mathbf{x}_{it}$  不相关, 则为随机效应模型(RE); 反之, 则为固定效应模型(FE)。对于固定效应的 Tobit 模型, 由于找不到个体异质性  $u_i$  的充分统计量, 故无法像固定效应的 Logit 或计数模型那样进行条件最大似然估计。如果直接在混合 Tobit 回归中加入面板单位的虚拟变量(类似于 LSDV 法), 所得的固定效应估计量也是不一致的。因此, 下面仅考虑随机效应的 Tobit 模型。

首先, 在给定个体效应  $u_i$  的情况下, 个体  $i$  的条件分布为

$$\begin{aligned} & f(y_{i1}, y_{i2}, \dots, y_{iT} | u_i) \\ &= \prod_{t=1}^T [1 - \Phi((\mathbf{x}'_i \boldsymbol{\beta} + u_i) / \sigma_\varepsilon)]^{1(y_{it}=0)} \left[ \frac{1}{\sigma_\varepsilon} \phi((y_{it} - \mathbf{x}'_i \boldsymbol{\beta} - u_i) / \sigma_\varepsilon) \right]^{1(y_{it}>0)} \end{aligned} \quad (17.25)$$

上式的推导方法与第 14 章第 5 节类似。但上式中的个体异质性  $u_i$  不可观测。假设  $u_i \sim N(0, \sigma_u^2)$ , 记其概率密度函数为  $g(u_i)$ 。记  $(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$  的联合密度为  $f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$ , 并进行如下分解,

$$f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) = f(y_{i1}, y_{i2}, \dots, y_{iT} | u_i) \cdot g(u_i) \quad (17.26)$$

将上式中的  $u_i$  积分积掉, 即可得到  $(y_{i1}, y_{i2}, \dots, y_{iT})$  的边缘密度(或无条件分布),

$$\begin{aligned} f(y_{i1}, y_{i2}, \dots, y_{iT}) &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) du_i \\ &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT} | u_i) \cdot g(u_i) du_i \end{aligned} \quad (17.27)$$

上式的积分无解析解, 一般使用“Gauss-Hermite quadrature”方法进行数值积分; 也需要使用 Stata 命令 quadchk 来检验此数值积分的精确度。类似地, 可通过检验 “ $H_0: \sigma_u = 0$ ”(Stata 称为 “sigma\_u = 0”)来判断是否存在个体异质性。另外, 可以定义同一个体不同期扰动项的自相关系数为(Stata 记为 rho),

$$\rho \equiv \text{Corr}(u_i + \varepsilon_{it}, u_s + \varepsilon_{ts}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} \quad (t \neq s) \quad (17.28)$$

如果  $\rho$  越大, 则复合扰动项( $u_i + \varepsilon_{it}$ )中个体效应的部分( $u_i$ )越重要。特别地, 如果  $\rho = 0$ , 则说明  $\sigma_u^2 = 0$ , 即不存在个体随机效应, 而应选择混合回归。

随机效应的面板 Tobit 回归的 Stata 命令为

```
xttobit y x1 x2 x3, ll(#) ul(#) tobit
```

其中,选择项“ll(#)”表示下限(lower limit),选择项“ul(#)”表示上限(upper limit),而选择项“tobit”表示对原假设“ $H_0: \sigma_u = 0$ ”进行LR检验。

下面继续以数据集 mus18data.dta 为例<sup>①</sup>,介绍面板 Tobit 模型的 Stata 操作。考虑被解释变量 med(医疗支出),而解释变量与前面的面板计数模型实例相同。显然,医疗开支为非负,故在 0 处存在左归并。

首先,进行混合 Tobit 回归,并使用聚类稳健标准误。

```
. tobit med lcoins ndisease female age lfam child, ll(0) nolog vce  
(cluster id)
```

Tobit regression		Number of obs = 20186 F( 6, 20180) = 44.99 Prob > F = 0.0000 Pseudo R2 = 0.0033				
		(Std. Err. adjusted for 5908 clusters in id)				
med	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lcoins	-29.59847	3.603346	-8.21	0.000	-36.66132	-22.53562
ndisease	13.1485	1.509925	8.71	0.000	10.18892	16.10807
female	64.7522	14.45226	4.48	0.000	36.4246	93.0798
age	4.05935	.9118431	4.45	0.000	2.272063	5.846637
lfam	-52.63194	19.81777	-2.66	0.008	-91.47639	-13.78749
child	-54.00112	27.0619	-2.00	0.046	-107.0446	-.9575866
_cons	-99.85116	39.26419	-2.54	0.011	-176.8122	-22.89014
/sigma	795.977	77.26739			644.5266	947.4274
Obs. summary:		4453 left-censored observations at med<=0 15733 uncensored observations 0 right-censored observations				

其次,进行随机效应的面板 Tobit 回归。

```
. xttobit med lcoins ndisease female age lfam child, ll(0) nolog tobit
```

<sup>①</sup> 此例来自 Cameron and Trivedi (2010)。

Random-effects tobit regression	Number of obs	=	20186		
Group variable: id	Number of groups	=	5908		
Random effects u_i ~ Gaussian	Obs per group:	min =	1		
		avg =	3.4		
		max =	5		
Log likelihood = -130030.45	Wald chi2(6)	=	573.45		
	Prob > chi2	=	0.0000		
med	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lcoins	-31.10247	3.578498	-8.69	0.000	-38.1162 -24.08875
ndisease	13.49452	1.139156	11.85	0.000	11.26182 15.72722
female	60.10112	14.95966	4.02	0.000	30.78072 89.42152
age	4.075582	.7238253	5.63	0.000	2.656911 5.494254
lfam	-57.75023	14.68422	-3.93	0.000	-86.53077 -28.96968
child	-52.02314	24.21619	-2.15	0.032	-99.48599 -4.560284
_cons	-98.27203	36.05977	-2.73	0.006	-168.9479 -27.59618
/sigma_u	371.3134	8.64634	42.94	0.000	354.3668 388.2599
/sigma_e	715.1779	4.704581	152.02	0.000	705.9571 724.3987
rho	.2123246	.0086583			.1957541 .2296872
Likelihood-ratio test of sigma_u=0: chibar2(01)= 778.18 Prob>chibar2 = 0.000					
Observation summary: 4453 left-censored observations 15733 uncensored observations 0 right-censored observations					

上表底部的 LR 检验结果强烈拒绝“ $H_0: \sigma_u = 0$ ”，故认为存在个体效应，应使用随机效应的面板 Tobit 回归。

## 17.9 面板随机前沿模型

第 14 章介绍了横截面数据的随机前沿模型。将该模型用于估计生产函数可得：

$$\ln y_i = \beta_0 + \sum_{k=1}^K \beta_k \ln x_{ki} + \underbrace{v_i - u_i}_{\epsilon_i}, \quad u_i \geq 0 \quad (17.29)$$

其中， $u_i \geq 0$  为“无效率项”(inefficiency term)，而  $v_i$  为“idiosyncratic error”。为了得到一致估计，通常假设  $u_i$  与  $x_i$  不相关。但此假设未必成立，因为企业很可能知道自己的无效率项  $u_i$ ，并以此调整其最优投入水平  $x_i$ 。此难题在横截面数据的情况下很难解决；但如果用面板数据，则可以允许  $u_i$  与  $x_i$  相关。事实上，即使不考虑  $u_i$  与  $x_i$  相关性，使用横截面数据也得不到对无效率项  $u_i$  的一致估计，因为随着横截面单位  $n$  增大，待估计的无效率项也同步增多，故存在“伴生参数问题”(incidental parameters problem)。而如果用面板数据，则可在  $T \rightarrow \infty$  的条件下得到一致估计(这要求面板数据的时间维度较长，即为长面板)。

取决于无效率项  $u_i$  是否随时间而变，可将面板随机前沿模型分为以下两类。

## 1. 技术效率不随时间而变

考虑如下模型：

$$\ln y_{it} = \beta_0 + \sum_{k=1}^K \beta_k \ln x_{kit} + v_{it} - u_i, \quad u_i \geq 0 \quad (17.30)$$

其中，无效率项  $u_i$  不随时间而变，称为“time-invariant technical efficiency”。此模型与标准的线性面板模型十分类似（参见第15章），只是个体效应（ $-u_i$ ）一定是非正的（nonpositive）。如果  $u_i$  与  $x_i$  相关，则为固定效应模型；反之，则为随机效应模型。

对于固定效应模型，可将方程(17.30)写为

$$\ln y_{it} = \beta_{0i} + \sum_{k=1}^K \beta_k \ln x_{kit} + v_{it} \quad (17.31)$$

其中， $\beta_{0i} \equiv \beta_0 - u_i$  为企业  $i$  的特有截距项。显然，方程(17.31)可使用 LSDV 法进行估计，即引入  $(n-1)$  个企业虚拟变量进行 OLS 估计（参见第15章）。记截距项的估计结果为  $\hat{\beta}_{0i}$ ，并计算其最大值  $\hat{\beta}_0 \equiv \max_i \{\hat{\beta}_{0i}\}$ ，则可通过下式估计无效率项：

$$\hat{u}_i = \hat{\beta}_0 - \hat{\beta}_{0i} \quad (17.32)$$

显然， $\hat{u}_i \geq 0$ ，且至少有一家企业达到了效率边界；其他企业的效率水平则为离开此效率边界的相对距离。

对于随机效应模型，可对  $u_i$  与  $v_{it}$  的分布作出具体假定，然后进行 MLE 估计。这些假定与横截面数据的设定相似，即  $u_i, v_{it}$  均为 iid 且相互独立，而且独立于解释变量  $x_i$ ； $v_{it} \sim N(0, \sigma_v^2)$ ；而  $u_i$  服从断尾正态分布，即  $u_i \sim N^+(\mu, \sigma_u^2)$ ，且在原点左边断尾。在这些假定之下，可得出对数似然函数（推导过程类似于横截面数据的情形，参见 Kumbhakar and Lovell, 2000, p. 102–105）。类似地，可使用条件期望  $E(u_i | \varepsilon_i)$  作为无效率项  $u_i$  的估计，其中  $\varepsilon_{it} \equiv v_{it} - u_i$ 。

## 2. 技术效率随时间而变

如果面板数据的时间维度较大，则“技术效率不随时间而变”的假定可能不现实。在上述随机效应模型的框架下，Battese and Coelli (1992) 假设无效率项  $u_{it}$  同时随企业与时间而变：

$$u_{it} = e^{-\eta(t-T_i)} u_i \quad (17.33)$$

其中， $T_i$  为企业  $i$  的时间维度（允许非平衡面板）， $\eta$  为待估参数，而  $u_i \sim N^+(\mu, \sigma_u^2)$ 。上式表明， $u_{it}$  随时间而递减，直至到最后一期  $T_i$  时， $u_{iT_i} = u_i$ ；故称为“时变衰减模型”（time-varying decay model）。显然，如果  $\eta = 0$ ，则退化为技术效率不随时间而变的模型；对此，只要检验  $H_0: \eta = 0$ ，即可判断。估计方法仍为 MLE。当无效率项  $u_{it}$ （表示离生产前沿的距离）变化时，生产前沿本身也可能变化；这种技术进步（technical change）的效应可通过加入时间虚拟变量来捕捉。

对于技术效率不随时间而变的固定效应模型，可直接用 Stata 命令 reg 进行 LSDV 估计。对于随机效应模型，则有以下专门的 Stata 命令：

```
xtfrontier lny lnx1 lnx2 lnx3, cost ti tvd
```

其中，“lny”为对数形式的被解释变量，“lnx1 lnx2 lnx3”为对数形式的解释变量。选择项“cost”表示估计成本函数，默认估计生产函数。选择项“ti”表示估计技术效率不随时间而变的随机效应模型，而选择项“tvd”表示估计随机效应的时变衰减模型。

下面以 Stata 提供的数据集“xtfrontier1.dta”为例进行演示。该数据集的主要变量包括：lnwidgets（产量对数）、lnmachines（资本对数）、lnworkers（劳动力对数）、id（个体变量）以及 t（时间变量）。

首先，使用固定效应的 LSDV 法。

```
. use xtfrontier1.dta
```

```
. reg lnwidgets lnmachines lnworkers i.id, vce(cluster id)
```

Linear regression						Number of obs = 948
	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	F( 1, 90) = .
lnmachines	.2918565	.0152287	19.16	0.000	.2616021 .3221109	Prob > F = .
lnworkers	.2985121	.0174847	17.07	0.000	.2637756 .3332486	R-squared = 0.7243
						Root MSE = 1.006
(Std. Err. adjusted for 91 clusters in id)						
lnwidgets						
i.d						
2	-1.695635	.041544	-40.82	0.000	-1.778169 -1.613101	
3	1.425356	.0319736	44.58	0.000	1.361835 1.488877	
4	-.5109693	.0171278	-29.83	0.000	-.5449967 -.4769419	
5	-.0050908	.0215518	0.24	0.814	-.0377256 -.0479072	
6	-.6426216	.0216263	-29.71	0.000	-.6855859 -.5996572	
7	-.1.622634	.0164675	-98.54	0.000	-.1.65535 -.589918	
8	-.0976995	.0265487	-3.68	0.000	-.1504431 -.0440556	
9	-.9882372	.036626	-26.98	0.000	-.1.061001 -.9154732	
10	1.236901	.0299493	41.30	0.000	1.177401 1.2964	
11	-.1.732293	.0110949	-156.13	0.000	-.1.754335 -.1.710251	
12	-.2.131603	.017535	-121.56	0.000	-.2.166439 -.2.096767	
13	.9925154	.0365132	27.18	0.000	.9199754 1.065055	
14	.5199811	.0242669	21.43	0.000	.4717706 .5681916	
15	-.1.690762	.0318765	-53.04	0.000	-.1.75409 -.1.627434	
16	-.2.98755	.0237505	-124.16	0.000	-.2.99594 -.2.901571	
17	-.1.933026	.0235441	-82.10	0.000	-.1.9798 -.1.886251	
18	-.3.325573	.039951	-83.24	0.000	-.3.404943 -.3.246203	
19	-.581822	.0434022	-13.41	0.000	-.668048 -.495596	
20	-.1.204938	.0323211	-37.28	0.000	-.1.26915 -.1.140727	
21	-.1.363309	.0422181	-32.29	0.000	-.1.447182 -.1.279435	
22	.0796255	.0159948	4.98	0.000	.0478489 .1.11402	
23	1.216471	.0300302	40.51	0.000	1.156811 1.276131	
24	-.2.744233	.015575	-176.20	0.000	-.2.775176 -.2.713291	
25	1.040995	.0175595	59.28	0.000	1.00611 1.07588	
26	1.131819	.0422106	26.81	0.000	1.04796 1.215678	
27	-.1.467872	.0298633	-49.15	0.000	-.1.527201 -.1.408544	
28	.8157485	.0242732	33.61	0.000	.7675255 .8639714	
29	.7387448	.0197171	37.47	0.000	.6995733 .7779163	
30	-.1.487877	.0266645	-55.80	0.000	-.1.540851 -.1.434904	
31	1.009767	.0080488	125.46	0.000	.9937771 1.025758	
32	-.1.808957	.0210879	-85.78	0.000	-.1.850852 -.1.767062	
33	-.2.144338	.0310849	-68.98	0.000	-.2.206074 -.2.082563	
34	-.2.50431	.0074786	-334.86	0.000	-.2.519168 -.2.489453	
35	-.2.921074	.0276569	-105.64	0.000	-.2.976008 -.2.866141	
36	.4856123	.0300698	16.15	0.000	.4258731 .5435314	
37	-.2.318	.0296372	-78.21	0.000	-.2.376879 -.2.25912	
38	.9231861	.0114178	80.85	0.000	.9005026 .9458697	
39	-.2.132264	.0287247	-74.23	0.000	-.2.18933 -.2.075197	
40	1.46387	.0142943	102.41	0.000	1.435472 1.492268	
41	.2768429	.0148452	18.65	0.000	.2473565 .3063419	
42	1.049103	.020893	50.21	0.000	1.007595 1.090611	
43	-.7912524	.0346812	-22.82	0.000	-.8601526 -.7223521	
44	.9562655	.028856	33.14	0.000	.8989379 1.013593	
45	-.2.608002	.0232772	-112.04	0.000	-.2.654247 -.2.561758	
46	1.187144	.0343153	-34.60	0.000	-.1.255317 -.1.11897	
47	-.6.694994	.0386311	-43.88	0.000	-.1.771742 -.1.618247	
48	.4930057	.0279105	17.66	0.000	.4375566 .5484548	
49	-.8026196	.0343494	-23.37	0.000	-.8708608 -.734375	
50	.3637465	.028856	12.74	0.000	.3070071 .4204858	
51	-.3.566843	.0314315	-113.48	0.000	-.3.629287 -.3.504399	
52	.6163288	.0374223	16.47	0.000	.5419828 .6906748	
53	-.2.777265	.0280012	-99.18	0.000	-.2.832895 -.2.721636	
54	.6421726	.0356541	18.01	0.000	.5713394 .7130057	
55	1.260894	.0430793	29.27	0.000	1.17531 1.346479	
56	-.2.00337	.0597843	-33.51	0.000	-.2.122142 -.1.884598	
57	-.1.839978	.0266548	-6.90	0.000	-.2.369522 -.1.310434	
58	-.1.836363	.0325928	-56.34	0.000	-.1.901114 -.1.771611	
59	-.7331885	.0290314	25.26	0.000	.6755125 .7908644	
60	1.270419	.0194528	65.31	0.000	.2.31772 1.309065	
61	-.9638028	.0188156	-51.22	0.000	-.1.001183 -.9264224	
62	.0719316	.0253706	2.84	0.006	.0215285 .1223347	
63	-.1.591004	.0331893	-47.94	0.000	-.1.65694 -.1.525067	
64	.9536491	.0341483	28.54	0.000	.8872577 1.02004	
65	-.2.563908	.0253687	-101.07	0.000	-.2.614308 -.2.513509	
66	-.2.3831077	.0209696	-11.36	0.000	-.2.797663 -.1.964491	
67	-.1.145653	.0138836	-82.52	0.000	-.1.173235 -.1.11807	
68	1.218136	.0268255	45.41	0.000	.1.164843 1.27143	
69	.5956427	.0244782	24.33	0.000	.5470125 .644273	
70	.102732	.0190324	5.40	0.000	.0649208 .1405432	
71	-.2.271472	.0175619	-129.34	0.000	-.2.306362 -.2.236582	
72	.9855415	.0211339	46.63	0.000	.9435552 1.027528	
73	-.5846073	.0160393	-36.45	0.000	-.6164721 -.5527425	
74	-.1.471411	.0128814	-11.42	0.000	-.1.727322 -.1.12155	
75	-.2.15375	.0180408	-119.38	0.000	-.2.189591 -.2.117909	
76	.7950362	.0132214	60.18	0.000	.7687896 .8212829	
77	-.2138768	.0178099	-12.01	0.000	-.2.429594 -.1.784943	
78	-.1.7695	.0281078	-6.30	0.000	-.2.327911 -.1.211088	
79	.7954589	.0421389	18.88	0.000	.7117427 .8791751	
80	.500001	.0251092	19.91	0.000	.4501172 .5498848	
81	-.1.953061	.0306898	-63.64	0.000	-.2.014031 -.1.89209	
82	-.1.321939	.0342241	-3.86	0.000	-.2.00186 -.0642018	
83	-.0354772	.0249803	-1.42	0.159	-.0851048 .0141504	
84	-.1.027946	.0329124	-31.23	0.000	-.1.093332 -.9625596	
85	-.68322718	.0342629	-19.94	0.000	-.7513411 -.6152025	
86	-.4829249	.0199879	-24.16	0.000	-.5226343 -.4432155	
87	.307109	.0385769	7.96	0.000	.2304693 .3837487	
88	-.1.503912	.0278106	-54.08	0.000	-.1.559163 -.1.448661	
89	.0719344	.0375732	1.91	0.059	-.0027113 .1.46558	
90	-.1.147258	.0352705	-32.53	0.000	-.1.217329 -.1.077187	
91	.0047903	.0600026	0.08	0.937	-.1.144152 -.1239959	
_cons	1.692352	.0444898	38.04	0.000	1.603965 1.780738	

上表显示,绝大多数的个体虚拟变量均很显著,证明存在个体效应。

其次,估计技术效率不随时间而变的随机效应模型。

```
. xtfrontier lnwidgets lnmachines lnworkers, ti nolog
```

Time-invariant inefficiency model		Number of obs = 948			
Group variable: id		Number of groups = 91			
		Obs per group: min = 6			
		avg = 10.4			
		max = 14			
		Wald chi2(2) = 661.76			
Log likelihood = -1472.6069		Prob > chi2 = 0.0000			
lnwidgets	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnmachines	.2904551	.0164219	17.69	0.000	.2582688 .3226415
lnworkers	.2943333	.0154352	19.07	0.000	.2640808 .3245858
_cons	3.030983	.1441022	21.03	0.000	2.748548 3.313418
/mu	1.125667	.6479217	1.74	0.082	-.144236 2.39557
/lnsigma2	1.421979	.2672745	5.32	0.000	.898131 1.945828
/ilgtgamma	1.138685	.3562642	3.20	0.001	.4404204 1.83695
sigma2	4.145318	1.107938			2.455011 6.999424
gamma	.7574382	.0654548			.6083592 .8625876
sigma_u2	3.139822	1.107235			.9696821 5.309962
sigma_v2	1.005496	.0484143			.9106055 1.100386

下面,预测无效率项  $u_{it}$ ,将其记为 uhat,并罗列前 15 个观测值。

```
. predict uhat,u
.list id uhat in 1/15
```

	id	uhat
1.	1	1.340585
2.	1	1.340585
3.	1	1.340585
4.	1	1.340585
5.	1	1.340585
6.	1	1.340585
7.	1	1.340585
8.	1	1.340585
9.	1	1.340585
10.	1	1.340585
11.	1	1.340585
12.	2	2.977796
13.	2	2.977796
14.	2	2.977796
15.	2	2.977796

再次,估计随机效应的时变衰减模型。

```
. xtfrontier lnwidgets lnmachines lnworkers, tvd nolog
```

Time-varying decay inefficiency model		Number of obs	=	948
Group variable: id		Number of groups	=	91
Time variable: t		Obs per group: min =		6
		avg =		10.4
		max =		14
		Wald chi2(2) =		661.93
Log likelihood = -1472.5289		Prob > chi2 =		0.0000
lnwidgets	Coef.	Std. Err.	z	P> z  [95% Conf. Interval]
lnmachines	.2907555	.0164376	17.69	0.000 .2585384 .3229725
lnworkers	.2942412	.0154373	19.06	0.000 .2639846 .3244978
_cons	3.028939	.1436046	21.09	0.000 2.74748 3.310399
/mu	1.110831	.6452809	1.72	0.085 -.1538967 2.375558
/eta	.0016764	.00425	0.39	0.693 -.0066535 .0100064
/lnsigma2	1.410723	.2679485	5.26	0.000 .885554 1.935893
/ilgtgamma	1.123982	.3584243	3.14	0.002 .4214828 1.82648
sigma2	4.098919	1.098299		2.424327 6.930228
gamma	.7547265	.0663495		.603838 .8613419
sigma_u2	3.093563	1.097606		.9422943 5.244832
sigma_v2	1.005356	.0484079		.9104785 1.100234

上表的 eta 表示  $\eta$ ,估计值仅为 0.0017(十分接近 0),且 p 值高达 0.693;故可认为技术效率不随时间而变。由于该样本的平均时间跨度为 10.4 年,可能发生技术变迁,下面加入时间虚拟变量重新估计。

```
. xtfrontier lnwidgets lnmachines lnworkers i.t, tvd nolog
```

Time-varying decay inefficiency model			Number of obs	=	948
Group variable: id			Number of groups	=	91
Time variable: t			Obs per group:	min =	6
				avg =	10.4
				max =	14
				Wald chi2(15)	= 685.79
Log likelihood = -1465.4647				Prob > chi2	= 0.0000
lnwidgets	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnmachines	.2932471	.0163939	17.89	0.000	.2611156 .3253786
lnworkers	.2937614	.0153939	19.08	0.000	.2635899 .3239329
t					
2	-.2275212	.1483311	-1.53	0.125	-.5182449 .0632025
3	-.1717603	.1507034	-1.14	0.254	-.4671335 .1236129
4	-.2104515	.1546102	-1.36	0.173	-.5134819 .0925789
5	-.1665324	.1594565	-1.04	0.296	-.4790614 .1459966
6	-.3096068	.165391	-1.87	0.061	-.6337672 .0145536
7	-.2099839	.1758252	-1.19	0.232	-.554595 .1346272
8	-.1819103	.1875663	-0.97	0.332	-.5495335 .185713
9	-.1222577	.2016833	-0.61	0.544	-.5175498 .2730343
10	-.4271937	.2134572	-2.00	0.045	-.8455622 -.0088253
11	-.5210739	.2265704	-2.30	0.021	-.9651437 -.0770041
12	-.4829743	.2504117	-1.93	0.054	-.9737722 .0078236
13	-.2995339	.284717	-1.05	0.293	-.857569 .2585013
14	-.913919	.3452203	-2.65	0.008	-1.590538 -.2372996
_cons	3.264047	.1949113	16.75	0.000	2.882028 3.646066
/mu	1.079811	.5944632	1.82	0.069	-.0853158 2.244937
/eta	.0128377	.0073383	1.75	0.080	-.0015452 .0272206
/lnsigma2	1.314226	.2642142	4.97	0.000	.796376 1.832077
/ilgtgamma	1.015936	.3631487	2.80	0.005	.3041774 1.727694
sigma2	3.72187	.9833712			2.21749 6.246846
gamma	.7341802	.070872			.5754634 .8491172
sigma_u2	2.732523	.9825715			.8067188 4.658328
sigma_v2	.989347	.0476378			.8959787 1.082715

从上表可知,有三个时间虚拟变量在 5% 水平上显著。而且,时变系数  $\eta$ (eta) 也变得在 10% 水平上显著( $p$  值为 0.08)。

## 习题

**17.1** 使用 Stata 提供的数据集 union.dta 估计面板二值选择模型。被解释变量为 union(是否工会成员),解释变量包括 age(年龄)、grade(教育年限)、south(是否南方)、black(是否黑人)以及 not\_smsa(不在大城市)。

- (1) 进行混合 Logit 回归;
- (2) 估计随机效应的面板 Logit 模型;
- (3) 估计固定效应的面板 Logit 模型;

- (4) 通过豪斯曼检验,确定应使用以上哪种模型;  
(5) 估计随机效应的面板 Probit 模型。

17.2 使用数据集 patr7079.dta 估计面板计数模型。此数据集来自 Cameron and Trivedi (2005), 最初为 Hall et al (1986) 所使用。该数据集包括 346 家企业在 1970—1979 年所获专利数据。被解释变量为 PAT(企业于该年申请并最终获批的专利数), 主要解释变量为 LOGR(企业该年的 R&D 支出), 其他解释变量包括 SCISECT(企业是否属于科学部门, the scientific sector), 以及 LOGK(1972 年的企业账面资本额)。

- (1) 进行混合泊松回归;  
(2) 进行随机效应的面板泊松回归;  
(3) 进行固定效应的面板泊松回归;  
(4) 进行随机效应的面板负二项回归;  
(5) 进行固定效应的面板负二项回归;  
(6) 通过豪斯曼检验, 确定应使用以上哪个回归。

# 第 18 章 随机实验与自然实验

大多数实证分析的目的是为了揭示变量之间的因果关系。然而,一般的经济数据并不足以有力地证明因果关系,除非是实验数据。为此,经济学家越来越热衷于使用“随机实验”或“自然实验”数据。本章内容主要取自 Stock and Watson (2011)。

## 18.1 实验数据

实验方法是研究因果关系的有力工具。假设我们要研究变量  $x_1$  是否导致  $y$ ,以及此因果关系的强弱。假定  $\{x_1, x_2, \dots, x_k\}$  包含了所有可能影响  $y$  的因素。不同学科可能依条件的不同而采用不同的实验方法,大致可以分为以下几类。

(1) 控制实验 (controlled experiment)。比如,在理想的物理实验中,对除  $x_1$  以外的因素  $\{x_2, \dots, x_k\}$  全部控制不变,单独让  $x_1$  变化,然后观察  $y$  变化的情况。

(2) 随机(控制)实验 (randomized controlled experiment)。比如,医学上对新药  $x_1$  疗效的实验。由于参加实验者的体质与生活方式不同,不可能完全控制所有其他的因素  $\{x_2, \dots, x_k\}$  (即使用老鼠做实验,老鼠之间仍然有差异),故无法进行严格的控制实验。为此,当代统计学之父费舍尔 (Ronald Fischer) 提出了随机实验的概念。通常将实验人群(或个体)随机地分为两组,其中“实验组”或“处理组”(treatment group)服用真药,而“控制组”(control group,也称“对照组”)服用“安慰药”(placebo)<sup>①</sup>。所有被试者均不知道自己被分在哪一组,以避免心理作用的干扰。有时候,科研人员也不知道被试者究竟在哪一组。这被称为“双盲法”(double blind)。又比如,在农学中将地块随机地分成三组(因为很难找到土壤条件完全一样的地块),分别给予不同的施肥量,然后考察施肥的效果。在经济学中,“实验经济学”(experimental economics)所做的实验基本上属于随机实验。

(3) 自然实验或准实验 (natural experiment or quasi experiment)。由于某些并非为了实验目的而发生的外部突发事件,使得当事人仿佛被随机地分在了实验组或控制组。比如,一个州通过了一个法律,但相邻的另一个州没有通过这一法律。两个州的民众事先并不知道哪个州会通过这个法律,因此无法进行自我选择住哪个州。这样,从考察这个法律的效果而言,可以近似地认为民众是随机地选择住哪个州的,或者随机地被分在了哪个组,即实验组(通过法律)与控制组(没通过法律)。这种情形被称为“自然实验”或“准实验”。

(4) 思想实验 (thought experiment)。比如,弗里德曼 (Milton Friedman) 曾设想在一个与世隔绝的小岛上通过直升机投放货币,然后考察该岛的宏观经济会发生怎样的变化。所谓“思想实验”,其实并无实验数据,只是在理想情形下的逻辑推理。

<sup>①</sup> 比如,没有实际作用的糖浆。

## 18.2 理想的随机实验

在理想的随机实验(ideal randomized experiment)中,实验组与控制组成员的决定完全是随机的,比如,通过抛硬币或计算机随机数来决定。因此,个体究竟被分在哪一组或得到多大的实验“处理水平”(treatment level),与个体的特征或其他可能影响实验结果的因素是完全独立的。因此,解释变量“处理水平”与被遗漏的扰动项是不相关的。这样,就避免了遗漏变量偏差(omitted variable bias)或内生变量偏差(endogeneity bias)。这是随机实验的最大优点。

考虑以下回归模型:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (18.1)$$

其中, $x_i$ 是完全随机地决定的。由于 $x_i$ 与 $\varepsilon_i$ 相互独立,故 $\text{Cov}(x_i, \varepsilon_i) = 0$ ,因此无论遗漏了多少解释变量,OLS都是一致的。更进一步,由于 $x_i$ 与 $\varepsilon_i$ 相互独立,故 $E(\varepsilon_i | x_1, \dots, x_n) = 0$ ,也满足小样本理论中的严格外生性假定。因此,OLS也是无偏的。

在随机实验中,因果效应(causal effect)有着清晰的定义。在理想的随机实验中, $X$ 对 $y$ 的因果效应表现在条件期望的差别,即 $E(y | X=x) - E(y | X=0)$ ,也称为“处理效应”(treatment effect)。具体来看两个例子。

例  $x_i$ 为服用药量,比如, $x_i = \{0, 1, 2\}$ ,而 $y_i$ 为病情康复情况。

例  $x_i = \{0, 1\}$ 表示是否参加过某一就业培训项目(job training program), $y_i$ 为未来的就业状态,而扰动项 $\varepsilon_i$ 包含遗漏变量,比如个人能力、工作经验等。由于 $x_i$ 与 $\varepsilon_i$ 相互独立,故参加过培训的实验组成员的能力及工作经验等的概率分布与没有参加过的控制组成员是一样的。

如果 $x_i = \{0, 1\}$ 为虚拟变量,则方程(18.1)的OLS估计量为

$$\hat{\beta}_{\text{OLS}} = \bar{y}_{\text{treat}} - \bar{y}_{\text{control}} \quad (18.2)$$

其中, $\bar{y}_{\text{treat}}$ 为实验组的样本均值,而 $\bar{y}_{\text{control}}$ 为控制组的样本均值。这是因为,在回归方程中加入虚拟变量的效果就相当于给予实验组与控制组不同的截距项。另一方面,当 $\{y_i\}$ 对常数项进行回归时,其系数估计值便是 $\bar{y}$ 。因此, $\hat{\alpha}_{\text{OLS}} = \bar{y}_{\text{control}}$ ,而 $\hat{\alpha}_{\text{OLS}} + \hat{\beta}_{\text{OLS}} = \bar{y}_{\text{treat}}$ 。由于 $\hat{\beta}_{\text{OLS}}$ 等于实验组均值与控制组均值之差,故也被称为“差分估计量”(differences estimator)。

例 班级规模是否影响学习成绩?班级规模过大(师生比过低)可能影响任课老师给每位学生的注意力,从而影响学习效果。由于遗漏变量的存在,使用观测数据很难回答此问题。比如,规模较小的班级可能位于好学区,师资好,家庭也富有。为此,美国田纳西州进行了为期四年的随机实验(称为Project STAR,即Student-Teacher Achievement Ratio),将幼儿园至小学三年级的学生随机分为三组。第一组为普通班,每班22~25名学生;第二组为小班,每班13~17名学生;第三组也为小班,但配备一名教学助理(teacher's aide)。教师也随机分到这三类班级。实验结果发现,尽管班级规模对学习成绩的影响在统计上显著,但在经济上并不显著(即此效应本身比较小,普通班与小班的成绩差距类似于男生与女生的成绩差距)。有关此随机实验的更多信息,参见Stock and Watson(2011)第13章。

## 18.3 引入更多的解释变量

虽然在理想的随机实验条件下,OLS估计量(即差分估计量)是一致且无偏的,但由于遗漏

了较多变量,故  $\varepsilon_i$  的方差可能较大,OLS 估计可能效率不高。如果引入某些遗漏变量,则可以改善这个问题。另外,引入其他遗漏变量也提供了一个检验  $x_i$  是否完全随机的机会。假设引入的其他解释变量为  $\{z_{i1}, \dots, z_{iK}\}$ ,

$$y_i = \alpha + \beta x_i + \delta_1 z_{i1} + \dots + \delta_K z_{iK} + \varepsilon_i \quad (18.3)$$

如果  $x_i$  完全随机,则  $\{z_{i1}, \dots, z_{iK}\}$  应该对  $x_i$  没有解释力。因此,我们可以把  $x_i$  对  $\{z_{i1}, \dots, z_{iK}\}$  进行回归,并检验该回归方程的整体显著性。如果该方程显著,则认为  $x_i$  不完全随机。

**例** 家中有计算机能否改善教育效果(educational outcome)? 如果使用观测数据,会遇到遗漏变量偏差问题,比如在教育方面有追求的学生与家庭通常也积极地在家中添置计算机。为此,Fairlie and London (2012)在美国北加州的一个社区学院(community college)<sup>①</sup>的 286 名新生中进行了随机实验。这些学生被随机分成大致相等的两组,其中一组(实验组)得到免费计算机,而另一组为控制组;并进行问卷跟踪调查两年,同时从学院行政部门获取相关信息。该研究的被解释变量包括对教育效果的四种度量(课程通过率、课程中学分课的比例<sup>②</sup>、课程中可转学分课的比例、获得学位的比例)。其他解释变量(即上文的  $\{z_{i1}, \dots, z_{iK}\}$ )包括性别、种族、父母教育水平、高中成绩、是否自己有子女、是否与父母同住、家庭收入、是否工作、英语是否为母语等。Fairlie and London (2012)同时汇报了不包含其他解释变量的差分估计量(在文中称为 treatment-control difference),以及包括其他解释变量的差分估计量(在文中称为 regression-adjusted treatment-control difference)。结果发现,获赠计算机实验组的两种教育效果变量显著高于控制组,而其他两种教育效果变量的效应也为正(但不显著)。

## 18.4 随机实验执行过程中可能出现的问题

在随机实验的执行过程中,可能并不完全符合“理想的随机实验”的要求。可能出现的问题包括,“内部有效性问题”(internal validity)与“外部有效性问题”(external validity)。

### 1. 内部有效性问题

(1) 未能完全随机分组(failure to randomize)。比如,以姓氏字母在字母表的前半部或后半部来决定谁进入就业培训项目。然而,在美国,姓氏与种族有关,而种族又与就业机会有关。因此,这种分组其实不完全随机。此时,可把  $x_i$  对可观测的个体特征  $\{z_{i1}, \dots, z_{iK}\}$  进行回归,并检验该回归方程的整体显著性。如果该方程显著,则认为  $x_i$  不完全随机。

(2) 未能完全遵从实验设计(partial compliance)。比如,被指定参加就业培训者却因故没来,而未被指定参加者却自行来参加。即使研究者知道实际上谁参加了,但由于是否参加的决定部分地由个人做出,也违反了  $x_i$  的随机性假定( $x_i$  可能与不可观测的个体特性  $\varepsilon_i$  相关),导致不一致的估计。此时可使用工具变量法来解决  $x_i$  的内生性问题。具体来说,以“设计处理水平”(assigned treatment level,记为  $w_i$ )作为“实际处理水平”(actual treatment level,记为  $x_i$ )的工具变量。显然,只要有部分人遵循实验设计,则  $w_i$  就与  $x_i$  相关,满足相关性;另一方面,  $w_i$  完全由随机数确定,肯定与扰动项  $\varepsilon_i$  不相关,满足外生性;故  $w_i$  为有效工具变量。

使用工具变量法估计处理效应(instrumental variables estimation of the treatment effect)的前提

<sup>①</sup> 美国的社区学院通常为两年制,相当于中国的中专,其学生多来自低收入家庭。

<sup>②</sup> 能获得学分的课程称为“学分课”。有些课程不能获得学分。

是,研究者既知道设计处理水平,也知道实际处理水平。但研究者有时可能不知道实际处理水平,比如参加实验组者是否真吃了药。此时,对实际处理水平  $x_i$  的测量误差仍会导致不一致估计。

(3) 中途退出实验(attrition)。比如,参加就业培训的优秀者可能在项目进行过程中就找到工作而退出项目。那么,该项目只剩下较不优秀者,造成选择性偏差。但如果中途退出的原因与实验无关(比如,必须退出项目去照顾生病的家人),则不会造成选择性偏差。

(4) 实验效应或霍桑效应(experimental effect 或 Hawthorne effect)。参加实验本身可能改变个体的心理或行为,从而影响实验结果。对于药物疗效实验,可以通过使用双盲法来避免这种效应。但对于就业培训,双盲法并不可行,因为个体一定知道他是否参加了就业培训。这种效应也称为“霍桑效应”。20世纪二三十年代,美国通用汽车公司(General Motors)在霍桑(地名)的工厂做了一些实验,研究灯光、休息时间、车间布置、工时等因素对劳动生产率的影响。结果发现,无论这些条件如何改变,工人的劳动生产率始终在上升!其原因可能是,工人在实验过程中感到被关注而更加努力地工作。

(5) 样本过小。由于实验成本高,实验的样本容量可能较小。由于个体间差异可能较大,导致估计量的精确度不足。

## 2. 外部有效性问题

(1) 样本的代表性不足(non-representative sample)。比如,让即将获释的囚犯参与就业培训,其结论可能难以推广到普通人群,因为雇主可能歧视有犯罪记录的人。又比如,项目参加者为自愿报名的义工,而义工的素质通常高于普通人。

(2) 小型实验的条件与大规模推广时的现实条件不同。小型实验的条件与大规模推广时的现实条件可能很不同,而这种差异可能导致项目的结果截然不同。比如,英国《经济学家》杂志(Economist)在2008年6月12日报导了经济学家在非洲肯尼亚进行的实验,旨在推广蚊帐以防止疟疾。一种观点认为,免费发放蚊帐能够最快地推广蚊帐。另一种观点则认为,由于不知道蚊帐的价值,人们不会珍惜免费发放的蚊帐,而且免费发放使得蚊帐的长期供给变得困难。经济学家于是随机地对一些村庄提供免费蚊帐,而对另外一些村庄则有偿提供。实验结果证明,免费提供蚊帐更为有效。然而,这个实验结果如果大规模推广仍然有效吗?反驳者指出,在肯尼亚的那个实验地区,蚊帐的价值已广为人知;在实验过程中,蚊帐的供给也是有保证的;而这两点在大规模推广时不见得成立。

(3) 一般均衡效应(general equilibrium effect)。一个小型的就业培训项目可能不会改变社会上雇主的行为。但大面积推广后,雇主可能减少企业自行提供的员工培训,使得该项目给整个社会带来的净福利减少。

(4) 自我选择效应。在随机实验中,可以随机地安排实验组与控制组的成员。但在现实社会中,人们常常可以自由选择是否参与。这样,就会有自我选择问题。比如,在随机实验中,数据可能显示该项目并不有效。但在现实社会中,预期收益最大的人将最有积极性参加此项目。因此,这个项目可能在实际上很有效。

# 18.5 自然实验

随机实验虽然说服力强,但通常成本较高。另一方面,自然实验为自然发生(非为实验目的

而发生),几乎没有成本,而说服力类似于随机实验,故为实证研究者所青睐;但发现自然实验却需要一双慧眼。自然实验可分为两类。在第一类自然实验中,个体的分组或处理水平完全由自然实验所决定,可直接用 OLS 估计因果效应。在第二类自然实验中,个体的分组或处理水平只是部分地由自然实验所决定,此时应以自然实验带来的随机变动作为工具变量进行估计。不完美的自然实验也可能存在与不完美的随机实验类似的问题(参见上节),只是前者不可能存在实验效应,因为自然实验中的个体不会视自己为实验对象。

**例 最低工资对就业的影响。**提高法定最低工资(minimum wage)在多大程度上会影响对低技能工人的需求?由于工资与雇佣人数由劳动力市场的供求均衡所决定,故如果使用 OLS 估计劳动力需求函数,将导致联立方程偏差或内生变量偏差。设想这样一个随机实验:将参加实验的雇主分为两组,要求实验组的雇主提高最低工资,而控制组的雇主保持最低工资不变,然后比较这两组的劳动力需求变化。然而,这种实验很难执行,因为经济学家并没有立法的权力或实验成本太高。为此,Card and Krueger (1994)考虑了一个自然实验。在 1992 年,美国新泽西州通过法律将最低工资从每小时 \$4.25 提高到 \$5.05,但在相邻的宾夕法尼亚州最低工资却保持不变。在这种情况下,这两个州的雇主仿佛被随机地分配到实验组(新泽西州)与控制组(宾夕法尼亚州)。Card and Krueger (1994)收集了两个州的快餐店在实施新法前后雇佣人数的数据,并使用双重差分法(参见下节)进行估计。结果发现,提高最低工资对低技能工人的就业几乎没有影响。也就是说,对低技能工人的需求是没有弹性的,并不会因工资的上升而显著地减少雇佣人数。本例属于第一类自然实验,个体分组完全由自然实验所决定。

**例 服兵役是否影响退役后的长期收入?**如果使用观测数据比较服过兵役的人群与未服兵役人群的收入,则会产生选择偏差,因为能否参军取决于身体条件,以及个人选择(或许就业前景不佳者才选择参军)。为此,Angrist (1990)考察了越战期间的参军者,当时美国对全国年轻男子以生日抽签的方式进行征兵(draft)。尽管抽签结果完全随机,但是否参军还取决于体检结果,而且有些人得到豁免,另一些人虽未抽中却自愿参军。因此,本例属于第二类自然实验,即个体分组仅部分地由自然实验决定,故应使用抽签结果作为参军与否的工具变量进行估计。估计结果表明,服兵役会减少白人的长期收入,但不影响非白人的长期收入。

**例 经济增长对内战的影响。**显然,内战多发生于穷国。是否经济增长慢导致内战概率上升?另一方面,内战也会拖累经济增长,故存在内生性。Miguel et al (2004) 使用外生的降雨量变化(rainfall variation)作为自然实验,考察 41 个非洲国家在 1981—1999 年间,经济增长对内战概率的因果作用。由于灌溉系统与工业的缺乏,这些非洲国家的经济比较依赖于自然降雨的农业(rain-fed agriculture),故经济增长与降雨量有关,但并不完全取决于降雨量;故属于第二类自然实验。因此,Miguel et al (2004) 使用降雨量变化作为经济增长的工具变量,发现经济低增长显著地增加内战概率。这种利用大自然的外生变化而进行的自然实验被称为“大自然的自然实验”(natural natural experiments)<sup>①</sup>,在外生变化的随机性方面一般优于由于社会因素所引起的自然实验。大自然的自然实验的随机变化主要来自天气、自然灾害以及生物机制(比如,双胞胎、出生日期、性别)等。需要注意的是,尽管天气变化完全由大自然随机决定,也不一定在计量经济学意义上为外生,即与模型的扰动项不相关。这是因为,虽然经济系统内的变量一般不会影响天气,但天气却会影响经济系统内的变量(人们会根据天气而调整其行为)。因此,仍需要探讨工具变量是否满足排他性约束。为此,Miguel et al (2004)试图排除降雨量影响内战的其他渠道

<sup>①</sup> 参见 Rosenzweig and Wolpin (2000)。

(除了人均 GDP 这一渠道), 比如降雨变化可能影响收入分配、农村贫困率、政府税收、道路情况(水灾可破坏道路)、民众情绪(过分干燥使人易怒好斗), 而这些因素在理论上都有可能直接影响内战的概率(不通过人均 GDP 这一渠道)。

例(一个失败的例子) 京杭大运河流经省份的人均 GDP 平均而言高于其他省份。这是否可以归功于京杭大运河对区域经济增长的促进作用? 问题在于, 当隋炀帝确定京杭大运河的位置时, 他是在地图上随机画了一条线吗?

## 18.6 双重差分法

在做随机实验或自然实验时, 实验的效果常常需要一段时间才能显现出来, 而我们关心的恰恰是被解释变量实验前后的变化。为此, 考虑以下两期面板数据:

$$y_{it} = \alpha + \gamma D_t + \beta x_{it} + u_i + \varepsilon_{it} \quad (i=1, \dots, n; t=1, 2) \quad (18.4)$$

其中,  $D_t$  为实验期虚拟变量 ( $D_t = 1$ , 如果  $t=2$ , 实验后;  $D_t = 0$ , 如果  $t=1$ , 实验前),  $u_i$  为不可观测的个体特征, 而政策虚拟变量 (policy dummy)

$$x_{it} = \begin{cases} 1, & \text{若 } i \in \text{实验组, 且 } t=2 \\ 0, & \text{其他} \end{cases} \quad (18.5)$$

因此, 当  $t=1$  时(第一期), 实验组与控制组并没有受到任何不同对待,  $x_{it}$  都等于 0。当  $t=2$  时(第二期), 实验组  $x_{it}=1$ , 而控制组  $x_{it}$  依然等于 0。如果该实验未能完全地随机化(比如, 观测数据), 则  $x_{it}$  可能与被遗漏的个体特征  $u_i$  相关, 从而导致 OLS 估计不一致。由于是面板数据, 可以对方程(18.4)进行一阶差分(即第二期减去第一期), 以消掉  $u_i$ ,

$$\Delta y_{it} = \gamma + \beta x_{it} + \Delta \varepsilon_{it} \quad (18.6)$$

用 OLS 估计上式, 即可得到一致估计。根据与差分估计量(differences estimator)同样的推理可知

$$\hat{\beta}_{OLS} = \Delta \bar{y}_{treat} - \Delta \bar{y}_{control} = (\bar{y}_{treat,2} - \bar{y}_{treat,1}) - (\bar{y}_{control,2} - \bar{y}_{control,1}) \quad (18.7)$$

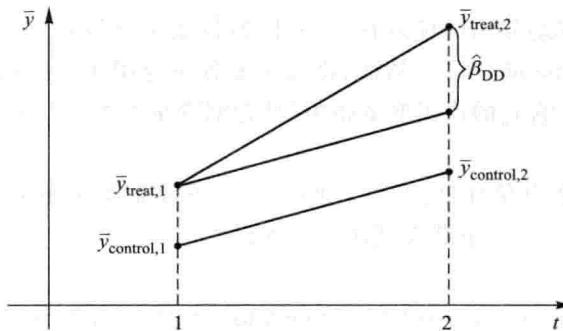


图 18.1 双重差分估计量示意图

因此, 这个估计法称为“双重差分估计量”(Difference-in-Differences estimator, 简记 DD), 记为  $\hat{\beta}_{DD}$ , 即实验组的平均变化与控制组的平均变化之差, 参见图 18.1。从图中可以直观地看出, 双重差分估计量已经剔除了实验组与控制组“实验前差异”(pretreatment differences)的影响。

对于双重差分估计量, 也可以引入其他解释变量  $\{z_{it}, \dots, z_{ik}\}$ :

$$\Delta y_i = \gamma + \beta x_{i2} + \delta_1 z_{i1} + \cdots + \delta_K z_{iK} + \Delta \varepsilon_i \quad (18.8)$$

然后用 OLS 估计上式。显然,以  $\Delta y_i$  为被解释变量的双重差分法(参见方程(18.6)与(18.8))不适用于多期的数据。此时,需要回到以  $y_{it}$  为被解释变量的面板模型。为了简洁,暂时忽略其他解释变量  $\{z_{i1}, \dots, z_{iK}\}$ ,并仍假设为两期数据。可以证明,方程(18.6)与以下两期面板模型等价:

$$y_{it} = \beta_0 + \beta_1 G_i \cdot D_t + \beta_2 G_i + \gamma D_t + \varepsilon_{it} \quad (i=1, \dots, n; t=1, 2) \quad (18.9)$$

其中,  $G_i$  为实验组虚拟变量 ( $G_i = 1$ , 如果个体  $i$  属于实验组;  $G_i = 0$ , 如果个体  $i$  属于控制组);  $D_t$  为实验期虚拟变量 ( $D_t = 1$ , 如果  $t = 2$ ;  $D_t = 0$ , 如果  $t = 1$ ), 而互动项  $G_i \cdot D_t = x_{it}$  (取值为 1, 若  $i \in$  实验组, 且  $t = 2$ ; 反之, 取值为 0)。在方程(18.9)中, 分组虚拟变量  $G_i$  刻画的是实验组与控制组本身的差异(即使不进行实验,也存在此差异),时间虚拟变量  $D_t$  刻画的是实验前后两期本身的差异(即使不进行实验,也存在此时间趋势),而互动项  $G_i \cdot D_t$  才能真正度量实验组的政策效应。如果有其他解释变量  $\{z_{i1}, \dots, z_{iK}\}$ , 可直接放入方程(18.9)。

当  $t = 1$  时, 方程(18.9)可以写为

$$y_{i1} = \beta_0 + \beta_2 G_i + \varepsilon_{i1} \quad (18.10)$$

当  $t = 2$  时, 方程(18.9)可以写为

$$y_{i2} = \beta_0 + \beta_1 G_i \cdot D_2 + \beta_2 G_i + \gamma + \varepsilon_{i2} \quad (18.11)$$

将方程(18.11)减去(18.10)可得

$$\Delta y_i = \gamma + \beta_1 G_i \cdot D_2 + (\varepsilon_{i2} - \varepsilon_{i1}) = \gamma + \beta_1 x_{i2} + \Delta \varepsilon_i \quad (18.12)$$

方程(18.12)与方程(18.6)完全相同。因此,对方程(18.9)进行 OLS 估计,得到的  $\hat{\beta}_1$ (即互动项  $G_i \times D_t$  的系数)就是双重差分估计量。使用方程(18.9)面板形式的双重差分估计量的好处在于,它可以很容易推广到多期数据的情形。比如,共有 4 期数据,则可估计如下方程:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 G_i + \gamma_1 D_{2t} + \gamma_2 D_{3t} + \gamma_3 D_{4t} + \varepsilon_{it} \quad (i=1, \dots, n; t=1, \dots, 4) \quad (18.13)$$

其中,  $D_{2t}, \dots, D_{4t}$  分别为对应于第 2~4 期的时间虚拟变量(以捕捉即使没有政策变化也可能存在的时间效应),而主要感兴趣的政策虚拟变量  $x_{it}$  定义为

$$x_{it} = \begin{cases} 1, & \text{若 } i \in \text{实验组, 且 } t \in \text{实验期} \\ 0, & \text{其他} \end{cases} \quad (18.14)$$

在方程(18.13)中,  $x_{it}$  的系数 OLS 估计值  $\hat{\beta}_1$  就是双重差分估计量。显然,在两期数据中,  $x_{it}$  就是交叉项  $G_i \times D_t$ 。双重差分法的优点在于,它同时控制了分组效应  $G_i$  (group-specific effects) 与时间效应  $D_t$  (time-specific effects)。然而,将双重差分法应用于观测数据时,如果政策虚拟变量  $x_{it}$  为内生,比如,  $x_{it}$  与影响  $y_{it}$  的遗漏变量相关(此遗漏变量为扰动项  $\varepsilon_{it}$  的一部分),则依然得不到一致估计。

在 Stata 中估计双重差分估计量,可手工进行,也可下载非官方命令 diff。

ssc install diff (下载安装命令 diff)

该命令的基本格式为

diff y, treat(varname) period(varname) cov(z1 z2) robust report test

其中,“y”为结果变量(outcome variable, 即上文的  $y_{it}$ ),必选项“treat(varname)”用来指定处理变量(即上文的  $G_i$ ),必选项“period(varname)”用来指定实验期虚拟变量(即上文的  $D_t$ )。选择项“robust”表示汇报稳健标准误。选择项“cov(z1 z2)”用来指定其他解释变量(即上文的  $z_{i1}, \dots, z_{iK}$ ),其中“cov”表示协变量(covariate);选择项“report”表示汇报对协变量系数的估计结果。选择项“test”表示检验在基期时,各变量在实验组与控制组的均值是否相等。

下面以数据集 cardkrueger1994.dta 为例。该数据集来自 Card and Kruger(1994) 的自然实验

(参见上节)。被解释变量为 fte (full time employment 人数), 实验组虚拟变量为 treated (treated = 1, 如果快餐店在新泽西州; treated = 0, 如果快餐店在宾夕法尼亚州), 实验期虚拟变量为  $t$  ( $t = 1$ , 如果时间为 1992 年 11 月;  $t = 0$ , 如果时间为 1992 年 2 月)。其他解释变量包括以下快餐品牌虚拟变量: bk (Burger King = 1), kfc (Kentucky Fried Chicken = 1), roys (Roy Rogers = 1), wendys (Wendy's = 1)。首先, 手工计算双重差分法。

```
. use cardkrueger1994,clear  
. gen gd=t*treated    (定义交叉项 gd)  
. reg fte gd treated t,r
```

Linear regression		Number of obs =	801
		F( 3, 797) =	1.43
		Prob > F =	0.2330
		R-squared =	0.0080
		Root MSE =	9.003
fte	Coef.	Robust Std. Err.	t P> t  [95% Conf. Interval]
gd	2.913982	1.736818	1.68 0.094 -.4952963 6.323261
treated	-2.883534	1.403338	-2.05 0.040 -5.638209 -.1288592
t	-2.40651	1.594091	-1.51 0.132 -5.535623 .7226031
_cons	19.94872	1.317281	15.14 0.000 17.36297 22.53447

从上表可知,处理效应(即交互项  $gd$ )的系数为 2.914,但仅在 10% 水平上显著( $p$  值为 0.094)。其次,使用命令 `diff` 来估计。

. diff fte,t(treated) p(t) robust

Number of observations in the DIFF-IN-DIFF: 801

	Baseline	Follow-up
Control:	78	155
Treated:	326	646
	404	397

R-square: 0.00805

DIFFERENCE IN DIFFERENCES ESTIMATION

	BASE LINE			FOLLOW UP		
Outcome Variable	Control	Treated	Diff(BL)	Control	Treated	Diff(FU)
DIFF-IN-DIFF						
fte	19.949	17.065	-2.884	17.542	17.573	0.030
Std. Error	2.914	1.317	0.484	1.403	0.898	1.023
t	1.737	1.68	15.14	13.99	-2.05	17.27
P> t	0.168	0.094*	0.000	0.000	0.040**	0.000
						0.976

\* Means and Standard Errors are estimated by linear regression  
\*\*Robust Std. Errors  
\*\*\*Inference: \*\*\* p<0.01; \*\* p<0.05; \* p<0.1

上表第一列显示,处理效应的系数估计值为 2.914,而  $p$  值为 0.094,这与手工回归的结果完全相同。上表将“ $t = 0$ ”的时期称为“Baseline”(基期),而将“ $t = 1$ ”的时期称为“Follow-up”(跟踪期)。再次,引入快餐品牌虚拟变量,再进行手工回归(以 wendys 作为参照虚拟变量)。

. reg fte gd treated t bk kfc roys, r

Linear regression					Number of obs = 801	
					F( 6, 794) = 57.30	Prob > F = 0.0000
					R-squared = 0.1878	Root MSE = 8.1617
<hr/>						
fte	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
gd	2.93502	1.543422	1.90	0.058	-.0946504	5.96469
treated	-2.323906	1.253701	-1.85	0.064	-4.784867	.1370549
t	-2.402678	1.410265	-1.70	0.089	-5.170966	.3656108
bk	.9168795	.9382545	0.98	0.329	-.9248729	2.758632
kfc	-9.204856	.8991089	-10.24	0.000	-10.96977	-7.439945
roys	-.8970458	1.041071	-0.86	0.389	-2.940623	1.146532
cons	21.16069	1.307146	16.19	0.000	18.59482	23.72656

从上表可知,引入其他解释变量(即快餐品牌虚拟变量)后,处理效应的系数估计值增大为2.935,并接近5%的显著性水平( $p$ 值为0.058)。下面再次使用命令 `diff`。

```
. diff fte,t(treated) p(t) cov(bk kfc roys) robust
```

DIFFERENCE-IN-DIFFERENCES WITH COVARIATES

Number of observations in the DIFF-IN-DIFF: 801

	Baseline	Follow-up
Control:	78	155
Treated:	326	646
	404	397

R-square: 0.18784

DIFFERENCE IN DIFFERENCES ESTIMATION

	BASE LINE		FOLLOW UP	
Outcome Variable	Control	Treated	Diff(BL)	Control
DIFF-IN-DIFF				Treated
fte	21.161	18.837	-2.324	18.758
Std. Error	2.935	1.307	0.886	0.908
t	1.543	1.619	1.254	0.900
P> t	1.90	18.54	-1.85	18.85
	0.058*	0.000	0.064*	0.000
				0.497

\* Means and Standard Errors are estimated by linear regression  
\*\*Robust Std. Errors  
\*\*\*Inference: \*\*\* p<0.01; \*\* p<0.05; \* p<0.1

上表第一列显示的处理效应估计结果与手工回归完全一样。最后,使用命令 `diff` 检验各变量在实验组与控制组的均值是否相等。

```
. diff fte,t(treated) p(t) cov(bk kfc roys wendys) test
```

TWO-SAMPLE T TEST					
Number of observations (baseline): 404					
	Baseline	Follow-up			
Control:	78	-	78		
Treated:	326	-	326		
	404	-			
t-test at period = 0:					
> _____					
Variable(s)	Mean Control	Mean Treated	Diff.	t	Pr( t )
> T> t )					
> _____					
fte	19.949	17.065	-2.884	2.44	0.015
> 0**					
bk	0.443	0.411	-0.032	0.52	0.603
> 5					
kfc	0.152	0.205	0.054	1.08	0.281
> 8					
roys	0.215	0.248	0.033	0.61	0.544
> 8					
wendys	0.190	0.136	-0.054	1.22	0.224
> 1					
> _____					
*** p<0.01; ** p<0.05; * p<0.1					

从上表可知,在基期,结果变量 `fte` 的均值在两组之间有显著差异,而各品牌虚拟变量的均值则在两组之间无显著差异。有关命令 `diff` 的更多功能(比如双重差分倾向得分匹配估计量),将在第 28 章介绍。

## 18.7 三重差分法

双重差分法的隐含假设是,即使没有政策变化,控制组与实验组的时间趋势也一样;在方程(18.9)中表现为共同的时间趋势项  $\gamma D_t$ 。然而,如果控制组与实验组的时间趋势不同,比如前者系数为  $\gamma_0$ ,而后者系数为  $\gamma_0 + \gamma_1$ ,则方程(18.9)应写为:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 G_i \cdot D_t + \beta_2 G_i + \gamma_0 D_t + \gamma_1 G_i \cdot D_t + \varepsilon_{it} \\ &= \beta_0 + (\beta_1 + \gamma_1) G_i \cdot D_t + \beta_2 G_i + \gamma_0 D_t + \varepsilon_{it} \end{aligned} \quad (18.15)$$

因此,用 OLS 估计上式只能得到对  $(\beta_1 + \gamma_1)$  的一致估计;无法得到对实验效应  $\beta_1$  的一致估计。此时,须进一步改进双重差分估计量。

来看一个具体的例子。假设美国 B 州针对 65 岁或以上老人(记为 E, 表示 `elderly`)引入了一项新的医疗保健政策,而该政策不适用于 65 岁以下人群(记为 N)。欲考察此政策对健康状况  $y$  的影响。一种做法是以 B 州 65 岁或以上人群作为实验组,而以该州 65 岁以下人群作为控制

组,然后进行双重差分估计。此种 DD 法的潜在缺陷是,年轻人相对于老年人的相对健康状况本身可能随时间发生变化(即使未推出此项新政策),比如由于联邦政府同时期的其他政策变化所引起。因此,另一种做法是以相邻 A 州的 65 岁或以上老年人作为控制组,然后进行双重差分估计。然而,第二种 DD 法同样有缺陷,因为它忽略了 A 州与 B 州老年人的健康状况,即使在没有新政策的情况下也可能有不同的时间趋势,比如两个州的州情不同。更稳健的做法则将以上两种 DD 法结合起来,估计如下方程(为了简洁,省略个体下标  $i$  与时间下标  $t$ ):

$$y = \beta_0 + \beta_1 B + \beta_2 E + \beta_3 B \cdot E + \gamma_0 D + \gamma_1 D \cdot B + \gamma_2 D \cdot E + \delta D \cdot B \cdot E + \varepsilon_{it} \quad (18.16)$$

其中, $B$  为 B 州虚拟变量( $B$  州 = 1,  $A$  州 = 0), $E$  为老年虚拟变量(老年 = 1, 其他年龄 = 0), $D$  为时间虚拟变量(第 2 期 = 1, 第 1 期 = 0)。互动项  $D \cdot B \cdot E$ (表示第 2 期 B 州老年人)的系数  $\delta$  即为政策效应。OLS 估计量  $\hat{\delta}$  可以表达为

$$\hat{\delta} = [(\bar{y}_{BE2} - \bar{y}_{BE1}) - (\bar{y}_{BN2} - \bar{y}_{BN1})] - [(\bar{y}_{AE2} - \bar{y}_{AE1}) - (\bar{y}_{AN2} - \bar{y}_{AN1})] \quad (18.17)$$

其中,  $\bar{y}_{BE2}$  表示第 2 期 B 州老年人的样本均值,  $\bar{y}_{BE1}$  表示第 1 期 B 州老年人的样本均值,以此类推。此估计量称为“三重差分估计量”(Difference-in-Difference-in-Differences estimator, 简记 DDD), 记为  $\hat{\delta}_{DDD}$ 。上式右边的第一个方括弧就是以 B 州年轻人为控制组的 DD 估计量, 但此 DD 估计量没有控制实验组与控制组(在未引入新政策情况下)的时间趋势差异(trend differential); 上式右边第二个方括弧通过 A 州来捕捉时间趋势差异; 如果 A 州与 B 州的时间趋势差异相似, 则  $\hat{\delta}_{DDD}$  就是对政策效应的可信估计。

下面对表达式(18.17)进行说明。根据方程(18.16), 对应于  $\bar{y}_{BE2}$  的总体期望为

$$E(y_{BE2}) = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \gamma_0 + \gamma_1 + \gamma_2 + \delta \quad (18.18)$$

类似地, 对应于  $\bar{y}_{BE1}$  的总体期望为

$$E(y_{BE1}) = \beta_0 + \beta_1 + \beta_2 + \beta_3 \quad (18.19)$$

将方程(18.19)减去(18.18)可得对应于  $(\bar{y}_{BE2} - \bar{y}_{BE1})$  的总体期望

$$E(y_{BE2}) - E(y_{BE1}) = \gamma_0 + \gamma_1 + \gamma_2 + \delta \quad (18.20)$$

根据同样推导可得,

$$E(y_{BN2}) - E(y_{BN1}) = \gamma_0 + \gamma_1 \quad (18.21)$$

$$E(y_{AE2}) - E(y_{AE1}) = \gamma_0 + \gamma_2 \quad (18.22)$$

$$E(y_{AN2}) - E(y_{AN1}) = \gamma_0 \quad (18.23)$$

综合方程(18.20)、(18.21)、(18.22)与(18.23)可知, 对应于方程(18.17)的总体期望为

$$\begin{aligned} & \{[E(y_{BE2}) - E(y_{BE1})] - [E(y_{BN2}) - E(y_{BN1})]\} - \\ & \{[E(y_{AE2}) - E(y_{AE1})] - [E(y_{AN2}) - E(y_{AN1})]\} \end{aligned} \quad (18.24)$$

$$= [(\gamma_0 + \gamma_1 + \gamma_2 + \delta) - (\gamma_0 + \gamma_1)] - [(\gamma_0 + \gamma_2) - (\gamma_0)] = [\gamma_2 + \delta] - [\gamma_2] = \delta$$

当然, 在方程(18.16)中, 也可加入其他解释变量  $\{z_1, \dots, z_{ik}\}$ , 或推广到多期数据。

## 18.8 观测数据的处理效应

在许多情况下, 并没有随机实验或自然实验的数据, 而只有观测数据(observational data)。比如, 政府提供的就业培训项目, 完全由个人决定是否参与。此时, 由于实际处理水平  $x_i$  存在自我选择(self selection), 并非随机分组, 故可能导致不一致的估计。因此, 对于观测数据, 需使用特别的计量方法来估计处理效应, 详见第 28 章。

## 习 题

**18.1** (选自 Stock and Watson, 2011) 考虑一项旨在评估在宿舍里安装互联网对大学生成绩的影响的研究。在一个大宿舍里, 随机地选择半个宿舍安装宽带互联网(实验组), 并收集所有宿舍成员的期末成绩。以下的哪种情况会对该随机实验的“内部有效性”产生威胁? (回答“会”或“不会”)

- (1) 在学年中间, 所有男性运动员都搬出了大学宿舍, 住进了兄弟会(fraternity), 因此退出了此项研究(观测不到他们的期末成绩);
- (2) 被分到控制组的工程专业学生自行安装了一个局部网, 分享一个私人无线互联网, 并共同付费;
- (3) 实验组中的艺术专业学生一直没有学会如何使用互联网账户;
- (4) 实验组中的经济专业学生为控制组成员提供互联网服务, 并收费。

# 第19章 蒙特卡罗法与自助法

## 19.1 蒙特卡罗法的思想与用途

通过计算机模拟从总体抽取大量随机样本的计算方法统称为“蒙特卡罗法”(Monte Carlo Methods,简记 MC)<sup>①</sup>。先来看一个例子。

为了计算圆周率  $\pi$ ,考虑下面的随机实验。在一个边长为 1 的正方形中内接一个  $1/4$  单位圆,参见图 19.1。显然,正方形的面积为 1,而  $1/4$  圆的面积为  $\pi/4$ 。因此,如果知道  $1/4$  单位圆面积占正方形面积的比例,就可以计算出  $\pi$ 。

向这个正方形随机地射箭,并假设其落点在该正方形上服从二维均匀分布(通过计算机模拟来实现)。重复实验  $n$  次( $n$  很大),其中有  $m$  次落在  $1/4$  圆内。根据大数定律, $m/n \xrightarrow{P} \pi/4$ ,故  $\pi \approx 4m/n$ 。

这个例子揭示了蒙特卡罗法的精神,但在蒙特卡罗法的具体应用上,由于目的不同可以有很多变化。在计量经济学中,常使用蒙特卡罗法来确定统计量的小样本性质。我们知道,许多统计量的精确分布没有解析解。解决方法之一是,使用大样本理论,用渐近分布来近似真实分布。然而,使用大样本理论的代价是要求样本容量较大<sup>②</sup>,而现实中的样本容量可能不够<sup>③</sup>。另外,有时统计量的分布即使在大样本的假设下,也推导不出来。此时,蒙特卡罗法的价值就更加凸显,比如第 21 章中的单位根检验。

例如,对于线性回归模型, $y_i = x'_i \beta + \varepsilon_i$  ( $i = 1, \dots, n$ ),希望对线性假设“ $H_0: R\beta = r$ ”进行显著性水平为 5% 的检验。如果不假定扰动项服从正态分布,则只能用  $\chi^2$  分布对此假设进行大样本检验,即检验统计量  $W \equiv n (\hat{R}\hat{\beta} - r)' [\widehat{R' Avar(\hat{\beta}) R}]^{-1} (\hat{R}\hat{\beta} - r) \xrightarrow{d} \chi^2(m)$ ,其中  $\hat{\beta}$  为  $\beta$  的 OLS 估计量,  $m$  为线性约束的个数。由于是大样本检验,故所使用的  $\chi^2$  分布其实只是统计量真实分布的近似(approximation)。因此,“5%”可能只是“名义显著性水平”(nominal size),而不是“真实显著性水平”(true or actual size),二者之差被称为“显著性水平扭曲”(size distortion)<sup>④</sup>。

此时,可以使用蒙特卡罗法来确定“真实显著性水平”。基本步骤如下。

第一步,给定参数  $\beta$  的具体取值,以及解释变量  $x$  与扰动项  $\varepsilon$  的概率分布。

第二步,从  $x$  与  $\varepsilon$  的分布中随机抽样,得到  $\{x_1, x_2, \dots, x_n\}$  与  $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ 。

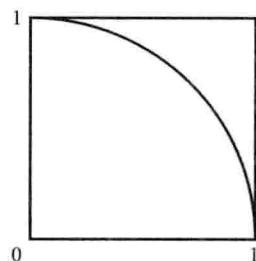


图 19.1 计算圆周率  
 $\pi$  的随机实验

<sup>①</sup> 这个名字来源于摩纳哥(在法国附近)的蒙特卡罗赌场(Monte Carlo Casino),这是最早使用这个方法的一位美国物理学家的叔叔常去的赌场。

<sup>②</sup> 理论上而言,只有当样本容量无穷大时,渐近分布才绝对地等于真实分布。

<sup>③</sup> 比如,中国的宏观经济年度数据,常常只有 1978 年改革开放以来才有,故只有三十几个数据。

<sup>④</sup> 也就是说,对第 I 类错误(Type I error)发生概率的衡量是不准确的。

第三步，根据方程“ $y_i = \mathbf{x}'\boldsymbol{\beta} + \varepsilon_i$ ”计算 $\{y_1, y_2, \dots, y_n\}$ 。

第四步，对这个样本进行 OLS 估计，计算检验统计量  $W$ ，并与 $\chi^2(m)$ 5% 的临界值比较，确定是否拒绝原假设“ $H_0: R\boldsymbol{\beta} = \mathbf{r}$ ”。

第五步，大量重复第二至第四步，得到  $M$  个随机样本（比如， $M = 1000$ ），进行  $M$  次检验，其中拒绝原假设的比例就是真实的显著性水平。

实现蒙特卡罗法的一个前提是，必须能够通过计算机产生随机数，并从任意概率分布进行随机抽样，参见本章附录。本章内容主要来自 Cameron and Trivedi (2005, 2010)。

## 19.2 蒙特卡罗法实例：模拟中心极限定理

根据中心极限定理(CLT)，当样本容量  $n \rightarrow \infty$  时，样本均值的分布收敛于正态分布。假设  $x$  服从在  $(0, 1)$  上的均匀分布，样本容量为 30，我们希望用蒙特卡罗法直观地“看到”样本均值  $\bar{x}_{30}$  的分布，并与正态分布相比较。为此，从  $(0, 1)$  上的均匀分布抽取 10 000 个样本容量为 30 的随机样本，得到 10 000 个样本均值  $\bar{x}_{30}$  的观测值，然后画其直方图。

相应的 Stata 程序如下<sup>①</sup>。基本步骤为，用命令 `program` 定义一个叫“`onesample`”的程序，从均匀分布抽取一个随机样本，并计算  $\bar{x}_{30}$ ；然后用命令 `simulate` 重复此程序 10 000 次，得到 10 000 个  $\bar{x}_{30}$  的观测值；最后用命令 `histogram` 画  $\bar{x}_{30}$  的直方图。

```
program onesample, rclass
    (定义程序 onesample，并以 r() 形式储存结果)
    drop _all
    (删去内存中已有数据)
    set obs 30
    (确定随机抽样的样本容量为 30)
    gen x = runiform()
    (得到在(0,1)上均匀分布的一个随机样本)
    sum x
    (使用命令“sum”得到样本均值)
    return scalar mean_sample = r(mean)
    (将样本均值记为 r(mean_sample))
end
(程序“onesample”结束)

simulate xbar = r(mean_sample), seed(101) reps(10000): onesample
```

其中，选择项“`reps(10000)`”表示，命令 `simulate` 将运行“`onesample`”程序 10 000 遍，并生成变量 `xbar` 来记录这 10 000 个样本均值。其中，选择项“`seed(101)`”用来确定随机数的初始值，以便再次模拟或别人运行此程序时，也能得到完全一样的结果，参见本章附录。

```
hist xbar, normal xtitle("xbar from many sample")
```

其中，选择项“`normal`”表示画出相应的正态分布，选择项“`xtitle`”表示横轴标题（如图 19.2）。

从上图的结果来看， $\bar{x}_{30}$  (`xbar`) 的直方图与相应的正态分布十分接近，从而验证了中心极限定理适用于均匀分布的随机变量。从这个例子也可以看出蒙特卡罗法的一个局限，即仅仅这个蒙特卡罗模拟本身并不能说明中心极限定理也适用于具有其他分布的随机变量。

<sup>①</sup> 由于 Stata 版本不同，如果发现本章少数命令无法运行，可输入命令“`update all`”，更新 Stata 命令库。

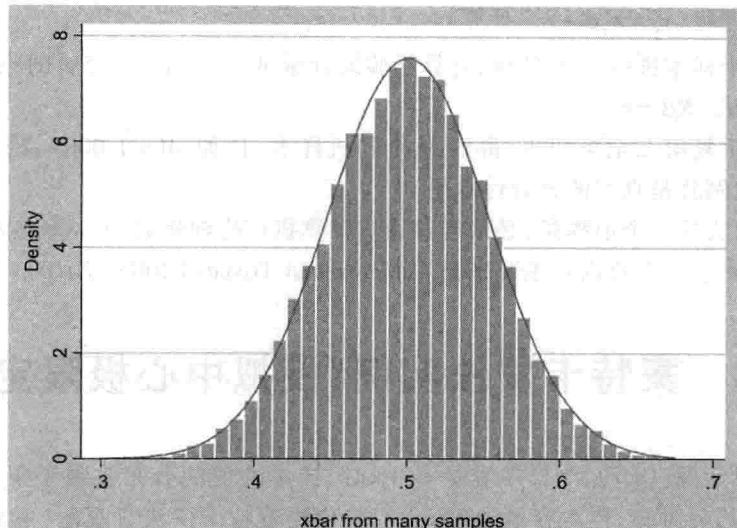


图 19.2 模拟中心极限定理

### 19.3 蒙特卡罗法实例：服从卡方分布的扰动项

考虑下面的数据生成过程(DGP)：

$$y = \alpha + \beta x + \varepsilon, \quad \varepsilon \sim \chi^2(1) - 1, \quad x \sim \chi^2(1) \quad (19.1)$$

其中,  $\alpha = 1, \beta = 2$ , 样本容量为  $n = 30$ 。解释变量  $x$  服从  $\chi^2(1)$  分布, 而扰动项  $\varepsilon$  为经过位移后的  $\chi^2(1)$  分布(以保证其期望为零), 而且  $x$  与  $\varepsilon$  相互独立。由于小样本理论要求扰动项服从正态分布, 这个模型显然不满足小样本理论的假定, 但符合大样本理论的要求。

我们将抽取 1 000 个样本容量为 30 的随机样本, 计算 1 000 个  $\hat{\beta}_{OLS}$ , 然后以显著性水平 5% 对原假设 “ $H_0: \beta = 2$ ” 进行 1 000 次  $t$  检验。由于  $t$  检验是大样本检验, 故 5% 只是“名义显著性水平”, 而“真实显著性水平”等于在 1 000 次蒙特卡罗模拟中, 拒绝 “ $H_0: \beta = 2$ ” 的比例。

为此, 先用命令 `program` 定义一个叫“`chi2data`”程序进行一次抽样与检验; 然后, 用命令 `simulate` 来重复此程序 1 000 次:

```

program chi2data,rclass          (定义程序 chi2data, 并以 r() 形式储存结果)
drop _all                         (删去内存中已有数据)
set obs 30                          (确定随机抽样的样本容量为 30)
gen double x = rchi2(1)            (生成服从  $\chi^2(1)$  分布的解释变量)
gen y = 1 + 2 * x + rchi2(1) - 1   (生成被解释变量)
reg y x                            (线性回归)
return scalar t2 = (_b[x] - 2) / _se[x] (计算 t 统计量)
return scalar r2 = abs(return(t2)) > invttail(28, 0.025)
                                         (是否拒绝原假设)
end                                (程序“chi2data”结束)
simulate reject = r(r2), reps(1000) seed(101):chi2data (模拟 1 000 次)

```

mean reject (计算拒绝原假设的比例)

在上述程序中,“invttail(28,0.025)”给出了  $t(28)$  分布所对应的显著性水平为 5% 的双边检验的临界值。该模拟结果显示(参见下表),拒绝原假设的比例为 4.6%, 比较接近于 5% 的名义显著性水平。

Mean estimation		Number of obs = 1000		
		Mean	Std. Err.	[95% Conf. Interval]
reject		.046	.0066278	.032994 .059006

## 19.4 蒙特卡罗积分

蒙特卡罗法的另一常见用途是用来计算复杂或高维的积分, 称为“蒙特卡罗积分”(Monte Carlo integration)。考虑计算定积分  $\int_a^b f(x) dx$ , 其中  $a, b$  为有限值。显然, 可以通过变量替换, 将积分的下限与上限分别变为 0 与 1。因此, 不失一般性, 仅考虑  $I = \int_0^1 f(x) dx$ 。假设随机变量  $x$  服从在  $[0, 1]$  上的均匀分布, 则随机变量函数  $f(x)$  的期望值为

$$E[f(x)] = \int_0^1 f(x) \cdot 1 dx = I \quad (19.2)$$

如果抽取随机变量  $x$  的样本容量为  $S$  ( $S$  代表 Simulation) 的随机样本, 记为  $\{x_1, \dots, x_s, \dots, x_s\}$ , 则蒙特卡罗积分估计值即为  $f(x)$  的样本均值:

$$\hat{I}_{MC} = \frac{1}{S} \sum_{s=1}^S f(x_s) \quad (19.3)$$

根据大数定律, 当  $S \rightarrow \infty$  时, 样本均值  $\hat{I}_{MC}$  依概率收敛于总体均值  $E[f(x)] = I$ 。上述方法的缺点是, 如果积分上限  $a$  或下限  $b$  为无穷, 则不适用, 因为均匀分布的定义域一定是有限的。为此, 考虑从某个适当的概率密度  $g(x)$  中抽取随机样本  $\{x_1, \dots, x_s, \dots, x_s\}$ 。原积分总可以写为

$$\int_a^b f(x) dx = \int_a^b \left[ \frac{f(x)}{g(x)} \right] g(x) dx \equiv \int_a^b w(x) g(x) dx = E[w(x)] \quad (19.4)$$

其中,  $w(x) = \frac{f(x)}{g(x)}$ 。此时, 蒙特卡罗积分估计值为

$$\hat{I}_{MC} = \frac{1}{S} \sum_{s=1}^S w(x_s) \quad (19.5)$$

从密度函数  $g(x)$  中抽样的方法被称为“重要性抽样”(importance sampling), 因为函数  $w(x)$  决定了每个样本点的权重或重要性。

作为实例, 考虑计算  $E\{\exp[-\exp(y)]\}$ , 其中  $y \sim N(0, 1)$ 。该期望的表达式为

$$E\{\exp[-\exp(y)]\} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp[-\exp(y)] \exp(-y^2/2) dy \quad (19.6)$$

此积分虽然存在, 但没有解析解。为此, 从标准正态分布中抽取  $S$  个随机观测值, 记其第  $s$  个抽样值为  $y^s$ , 则可使用以下估计值:

$$\hat{E}\{\exp[-\exp(y)]\} = \frac{1}{S} \sum_{s=1}^S \exp[-\exp(y^s)] \quad (19.7)$$

相应的 Stata 命令如下(令  $S = 100$ )：

```
. clear all
. set obs 100
. set seed 10101
. gen double y = rnormal(0,1)
. gen double gy = exp(-exp(y))
. sum gy
```

Variable	Obs	Mean	Std. Dev.	Min	Max
gy	100	.3679463	.2408939	4.50e-11	.8852481

上表显示,当  $S = 100$  时,表达式(19.6)的蒙特卡罗估计值为 0.367 946 3。但我们并不知道  $S = 100$  是否能达到足够的精确度。为此,令  $S = 1 000$ ,重复上述命令。

```
. clear all
. set obs 1000
. set seed 10101
. gen double y = rnormal(0,1)
. gen double gy = exp(-exp(y))
. sum gy
```

Variable	Obs	Mean	Std. Dev.	Min	Max
gy	1000	.3821273	.2577481	2.05e-36	.956477

上表显示,当  $S = 1 000$  时,表达式(19.6)的蒙特卡罗估计值为 0.382 127 3,与  $S = 100$  时结果有较大差异;故应考虑将  $S$  增大为 10 000,以此类推。

## 19.5 最大模拟似然法与模拟矩估计

蒙特卡罗法也可用于最大似然法或矩估计。使用 MLE 的前提是,必须能写出似然函数  $f(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ 。但在有些情况下,该似然函数的表达式可能包含无法求解的积分。比如,在随机效应的非线性面板模型中(参见第 17 章),一般需要将个体效应  $u_i$  积分掉(因为  $u_i$  不可观测),才能写出其似然函数。记  $u_i$  的密度函数为  $g(u_i)$ ,并假设第  $i$  个观测值的似然函数为

$$f(y_i | \mathbf{x}_i, \boldsymbol{\theta}) = \int h(y_i | \mathbf{x}_i, \boldsymbol{\theta}, u_i) g(u_i) du_i \quad (19.8)$$

如果上式中的积分无解析解,则可使用蒙特卡罗积分进行估计。从分布  $g(u_i)$  中随机抽取  $S$  个观测值,记为  $\{u_i^1, \dots, u_i^S\}$ ,则上式的估计值可写为

$$\hat{f}(y_i | \mathbf{x}_i, \boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^S h(y_i | \mathbf{x}_i, \boldsymbol{\theta}, u_i^s) \quad (19.9)$$

假设样本为 iid,则整个样本的对数似然函数估计值为

$$\ln \hat{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \ln \hat{f}(y_i | \mathbf{x}_i, \boldsymbol{\theta}) \quad (19.10)$$

其中,  $n$  为样本容量。最大化上式所得到的估计量  $\hat{\boldsymbol{\theta}}_{\text{MSL}}$  称为“最大模拟似然估计量”(Maximum Simulated Likelihood Estimator, 简记 MSL)。在一定的正则条件下, 当模拟抽样的次数  $S \rightarrow \infty$  时,  $\hat{f}$  对  $f$  的近似程度越来越好, 即  $(\hat{f} - f) \xrightarrow{P} 0$ , 则 MSL 为一致估计量。进一步, 如果  $\sqrt{n}/S \rightarrow 0$  (即  $S$  的增长速度快于  $\sqrt{n}$ ), 则 MSL 为渐近有效估计量(渐近等价于 MLE), 且服从渐近正态分布。

类似地, 在进行矩估计时, 如果矩条件中包含无解析解的积分, 也可使用蒙特卡罗积分来估计此矩条件, 然后再进行矩估计。此方法称为“模拟矩估计”(Method of Simulated Moments), 简记 MSM。在一定的正则条件下, MSM 为一致估计量且渐近正态。

## 19.6 自助法的思想与用途

蒙特卡罗法虽然威力很大, 但缺点是必须对总体模型(即数据生成过程)做很具体的假定, 比如确定所有参数的取值以及所有解释变量与扰动项的概率分布; 而且严格来说, 所得到的结论仅适用于那个特定的数据生成过程, 并不清楚它在多大意义上能够推广。为此, Efron (1979) 提出了一种对原始样本进行“再抽样”(resampling)的方法, 即“自助法”或“自举法”(bootstrap)<sup>①</sup>。

假设从总体抽得样本容量为  $n$  的随机样本。显然, 来自总体的这个样本带有总体的信息。在一定程度上, 可以将此样本看作是一个总体, 再进行“有放回”(with replacement)地抽样, 样本容量仍然为  $n$ 。这种样本被称为“自助样本”(bootstrap sample)。由于是有放回地抽样, 原来的某些观测值可能不出现, 而有些观测值则可能多次出现。自助法的优点是, 可以通过计算机模拟毫不费力地获得许多自助样本, 然后利用这些自助样本对总体进行统计推断。

假设  $\{x_1, x_2, \dots, x_n\}$  是来自总体  $F$  的一个随机样本, 可以定义总体  $F$  的经验分布函数(empirical distribution function)  $F_n$  如下

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i \leq x), \quad -\infty < x < \infty \quad (19.11)$$

其中,  $\mathbf{1}(\cdot)$  为示性函数, 而  $\sum_{i=1}^n \mathbf{1}(x_i \leq x)$  表示样本中小于或等于  $x$  的个数。经验分布函数的形状为阶梯函数(step function), 在每个  $x_i$  处向上跳一个台阶。可以证明(参见本章附录), 对任意  $x$ ,  $F_n(x) \xrightarrow{P} F(x)$ , 即当样本容量  $n$  越来越大时, 经验分布函数  $F_n$  与真实分布函数  $F$  越来越像, 乃至最终收敛。这正是自助法能够成立的前提, 因为自助法可以看成是从经验分布函数中不断地抽样。

自助法在计量经济学中的用途主要有两个方面。首先, 对于某些统计量(比如, 样本中位数), 用常规方法很难得到其标准误的估计量。此时可以使用自助法, 计算每个自助样本的样本中位数, 从而得到样本中位数的分布, 并计算其标准误。其次, 对研究者而言, 还可以使用自助法得到更加渐近有效的估计量(asymptotic refinement)<sup>②</sup>。

<sup>①</sup> “Bootstrap”一词的英文原意为“鞋带”。在英国有句谚语, 人应该自己绑鞋带, 意即人应该“自助”。由于“bootstrap”法利用现有样本再抽样进行推断(而不依赖于另外的样本或数据), 故也是一种“自助”的表现。

<sup>②</sup> 即在大样本下, 估计量能以更快的速度收敛到真实值。

## 19.7 自助法的分类

根据自助法的操作程序,可以将自助法分为以下几类。

(1) 非参数自助法 (nonparametric bootstrap), 也称为“经验分布自助法”(empirical distribution function bootstrap)。这种方法就是前面所介绍的将原始样本进行有放回地随机抽样。在回归模型中,这意味着将被解释变量与所有解释变量,即 $(y_i, \mathbf{x}_i)$ , 成对地抽样,故也称为“成对自助法”(paired bootstrap)。这是最简单,也是最常见的自助法。

(2) 参数自助法 (parametric bootstrap)。假设总体分布函数的形式已知,为 $F(x, \theta)$ ,其中 $\theta$ 为未知参数。则可以先得到 $\theta$ 的估计量 $\hat{\theta}$ (比如,使用最大似然估计法),然后从总体 $F(x, \hat{\theta})$ 中重复抽样。这个方法的前提是必须对总体分布函数的形式比较确信。在此前提下,参数自助法通常比非参数自助法更有效率。在回归模型中,则需要先确定条件分布的具体形式,即 $y|x \sim F(x, \theta)$ 。具体来说,一种方法是,在得到估计量 $\hat{\theta}$ 后,给定原先的 $\mathbf{x}_i$ ,从 $F(\mathbf{x}_i, \hat{\theta})$ 中随机抽样得到对应的 $y_i$ 。这相当于“固定解释变量”(fixed regressors)的情形<sup>①</sup>。另一种方法是,先从 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 中进行再抽样(resample),得到 $\mathbf{x}_i^*$ ,然后再从 $F(\mathbf{x}_i^*, \hat{\theta})$ 中随机抽样得到对应的 $y_i^*$ 。这相当于“随机解释变量”(stochastic regressors)的情形。

(3) 残差自助法(residual bootstrap)。对于回归模型 $y_i = g(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i$ ,首先通过估计得到残差 $\hat{\varepsilon}_i = y_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ ,然后对残差 $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n\}$ 使用自助法,得到残差的自助样本 $\{\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \dots, \hat{\varepsilon}_n^*\}$ ,然后计算对应的 $y_i^* = g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) + \hat{\varepsilon}_i^*$ ,进而得到自助样本 $\{(y_i^*, \mathbf{x}_i), \dots, (y_n^*, \mathbf{x}_n)\}$ 。

## 19.8 使用自助法估计标准误

假设原始样本为 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 。对于未知参数 $\theta$ 的估计量 $\hat{\theta} = \hat{\theta}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ ,需要计算其标准差 $\sigma_{\hat{\theta}} \equiv \sqrt{\text{Var}(\hat{\theta})}$ ,也称为估计量 $\hat{\theta}$ 的“标准误”(standard error)。但有时 $\sqrt{\text{Var}(\hat{\theta})}$ 没有解析表达式,故希望用计算机模拟的方法来估计 $\sqrt{\text{Var}(\hat{\theta})}$ 。

如果可以从真实总体 $F$ 产生样本容量为 $n$ 的随机样本(比如, $B$ 个样本),对每个样本都可以计算 $\hat{\theta}$ ,从而得到 $B$ 个估计值 $\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_B\}$ ,则可以使用

$$s_{\hat{\theta}} \equiv \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}_i - \bar{\theta})^2} \quad (19.12)$$

来估计 $\sigma_{\hat{\theta}}$ ,其中 $\bar{\theta} = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i$ 。但真实总体 $F$ 的分布常常未知,无法从中抽取随机样本,而从实际总体中进行多次实地抽样的成本可能很高。为此,考虑以经验分布函数 $F_n$ 来近似真实分布函数 $F$ ,并从 $F_n$ 中大量抽取随机样本,即在原始样本 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 中每次有放回地随机抽样,得

<sup>①</sup> 在计量经济学的本科课程中,有时这样假设。

到样本容量为  $n$  的自助样本  $\{x_1^*, x_2^*, \dots, x_n^*\}$ , 并以此计算  $\hat{\theta}^* = \hat{\theta}(x_1^*, x_2^*, \dots, x_n^*)$ 。如此重复, 共抽取  $B$  个自助样本, 则得到  $\theta$  的  $B$  个自助估计值  $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*\}$ 。可以定义标准误的自助估计为

$$s_{\hat{\theta}}^* = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}_i^* - \bar{\theta}^*)^2} \quad (19.13)$$

其中,  $\bar{\theta}^* = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^*$ 。

## 19.9 使用自助法进行区间估计

考虑用自助法对  $\theta$  进行置信度为  $(1 - \alpha)$  的区间估计, 有以下三种方法。

(1) 百分位法 (percentile method)。根据上一节, 我们已经得到自助估计量  $\hat{\theta}^*$  的经验分布  $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*\}$ 。将  $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*\}$  按从小到大的顺序排列, 并记其  $\alpha/2$  与  $(1 - \alpha/2)$  上分位数 (upper quantile)<sup>①</sup> 分别为  $\hat{\theta}_{\alpha/2}^*$  与  $\hat{\theta}_{1-\alpha/2}^*$ , 则  $\theta$  的置信区间为

$$[\hat{\theta}_{1-\alpha/2}^*, \hat{\theta}_{\alpha/2}^*] \quad (19.14)$$

(2) 基于正态的置信区间 (normal-based confidence interval)。也可以使用标准正态分布来估计置信区间, 即

$$[\hat{\theta} - 1.96 \times s_{\hat{\theta}}^*, \hat{\theta} + 1.96 \times s_{\hat{\theta}}^*] \quad (19.15)$$

其中,  $s_{\hat{\theta}}^*$  是用自助法估计的标准误, 并假定置信度为 95%。

(3) 百分位  $t$  法 (percentile- $t$ method)。根据每个自助样本计算对应的自助  $t$  统计量

$$t_i^* = \frac{\hat{\theta}_i^* - \hat{\theta}}{s_{\hat{\theta}}^*}, \quad i = 1, \dots, B \quad (19.16)$$

其中,  $\hat{\theta}$  为根据原始样本计算的  $\theta$  估计量, 而  $s_{\hat{\theta}}^*$  是根据  $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*\}$  计算的标准误。如此, 即得到自助  $t$  统计量的经验分布  $\{t_1^*, t_2^*, \dots, t_B^*\}$ , 并记其  $\alpha/2$  与  $(1 - \alpha/2)$  上分位数 (upper quantile) 分别为  $t_{\alpha/2}^* > 0$  与  $t_{1-\alpha/2}^* < 0$ , 则  $\theta$  的置信区间为  $[\hat{\theta} + t_{1-\alpha/2}^* \times s_{\hat{\theta}}^*, \hat{\theta} + t_{\alpha/2}^* \times s_{\hat{\theta}}^*]$ , 其中  $s_{\hat{\theta}}^*$  是根据原始样本计算的标准误。百分位  $t$  法比百分位法更渐近有效, 但在小样本中不一定有优势。

## 19.10 使用自助法进行假设检验

考虑用自助法进行如下双边检验,  $H_0: \theta = \theta_0$  vs  $H_1: \theta \neq \theta_0$ 。一种方法是, 如果  $\theta_0 \in [\hat{\theta}_{1-\alpha/2}^*, \hat{\theta}_{\alpha/2}^*]$ , 则接受原假设  $H_0$ ; 反之, 则拒绝。这就是“百分位法”(percentile method)。

另一方法是, 在假设  $H_0$  成立的情况下, 计算原始样本的  $t$  统计量

<sup>①</sup> “上分位数”指的是从分布的上限往下计算的分位数, 而不是通常的从下限往上计算的(下)“分位数”。

$$t \equiv \frac{\hat{\theta} - \theta_0}{\hat{s}_{\hat{\theta}}} \quad (19.17)$$

如果  $t \in [\hat{\theta} + t_{1-\alpha/2}^* \times \hat{s}_{\hat{\theta}}, \hat{\theta} + t_{\alpha/2}^* \times \hat{s}_{\hat{\theta}}]$ , 则接受原假设  $H_0$ ; 反之, 则拒绝。其中,  $t_{\alpha/2}^*$  与  $t_{1-\alpha/2}^*$  的定义如上。这就是百分位  $t$  法, 它比百分位法更渐近有效。

## 19.11 自助法的一致性(选读)

自助法是一致估计量吗? 考虑以下一般的自助法。随机样本  $\{x_1, x_2, \dots, x_n\}$  来自总体  $F(x, \theta)$ 。我们希望知道统计量  $T_n = T_n(x_1, x_2, \dots, x_n)$  的概率分布  $G_n(t, F) \equiv P(T_n \leq t)$ , 其中  $G_n(t, F)$  的下标  $n$  表示该分布依赖于样本容量  $n$ 。另外,  $G_n(t, F)$  也依赖于真实的总体分布  $F$ 。

大样本理论(渐近理论)通过让  $n \rightarrow \infty$ , 即用  $G_\infty(t, F)$  来近似  $G_n(t, F)$ ; 对于  $G_n(t, F)$  中未知的  $F$ , 则使用其一致估计量  $\hat{F} = F(x, \hat{\theta})$  来替代, 其中  $\hat{\theta}$  是  $\theta$  的一致估计量。自助法则不需要让  $n \rightarrow \infty$ , 而是使用经验分布函数  $F_n$  来替代  $F$ , 即用  $G_n(t, F_n)$  来近似  $G_n(t, F)$ 。尽管  $G_n(t, F_n)$  一般没有解析解, 但可以把样本  $F_n$  近似地看作总体  $F$ , 并从  $F_n$  中进行再抽样, 获得自助样本。从一个自助样本  $\{x_1^*, x_2^*, \dots, x_n^*\}$ , 就可以得到一个统计量  $T_n^* = T_n(x_1^*, x_2^*, \dots, x_n^*)$ 。独立地重复抽样  $B$  次, 得到  $B$  个统计量  $\{T_{n,1}^*, T_{n,2}^*, \dots, T_{n,B}^*\}$ , 从中可以得到  $T_n$  分布函数的自助估计(bootstrap estimate)  $\hat{G}_n(t, F_n)$ , 即  $\{T_{n,1}^*, T_{n,2}^*, \dots, T_{n,B}^*\}$  的经验分布函数:

$$\hat{G}_n(t, F_n) \equiv \frac{1}{B} \sum_{i=1}^B \mathbf{1}(T_{n,i}^* \leq t) \quad (19.18)$$

其中,  $\mathbf{1}(\cdot)$  为示性函数。显然, 当  $B \rightarrow \infty$  时,  $\hat{G}_n(t, F_n) \xrightarrow{P} G_n(t, F_n)$ , 因为前者是后者的经验分布函数(参见本章附录)。然而, 当  $n \rightarrow \infty$  时, 是否  $G_n(t, F_n) \xrightarrow{P} G_n(t, F)$ ? 显然, 这要求  $F_n \xrightarrow{P} F$  (即  $F_n$  是  $F$  的一致估计)。另外, 这还要求总体分布函数  $F(x, \theta)$  与统计量  $T_n$  的分布函数  $G_n(t, F)$  都是光滑的。这些条件在实践中一般都成立, 故自助估计量是一致的。

## 19.12 异方差情况下的自助法

如果回归模型的扰动项存在条件异方差, 通常的解决方法是使用稳健标准误, 参见第 5 章。尽管稳健标准误在大样本下是一致的, 但在小样本中可能不够精确。使用自助法可能得到更精确的估计。

由于条件异方差的存在并不影响观测数据  $(y_i, x_i)$  仍然为独立同分布<sup>①</sup>, 故仍可以使用成对自助法(paired bootstrap)。然而, 残差自助法(residual bootstrap)却不成立, 因为在条件异方差的情况下, 扰动项不是独立同分布的, 故经验分布函数  $F_n$  不是总体分布函数  $F$  的一致估计, 自助估计量也就不是一致估计。

然而, 如果对残差  $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n\}$  先进行某种线性变换后再进行抽样, 则可能满足从异方差

<sup>①</sup> 独立同分布的假设表明, 扰动项的条件方差函数  $\text{Var}(\varepsilon_i | x_i)$  的形式相同, 但其具体取值可以依  $x_i$  的不同而变化。

的扰动项进行抽样的要求。为此, Wu (1986) 与 Liu (1988) 提出了如下“野自助法”(wild bootstrap)。定义具有两点分布的新残差为

$$\hat{\varepsilon}_i^* = \begin{cases} \frac{1-\sqrt{5}}{2}\hat{\varepsilon}_i, & \text{以 } \frac{\sqrt{5}+1}{2\sqrt{5}} \text{ 的概率} \\ \frac{1+\sqrt{5}}{2}\hat{\varepsilon}_i, & \text{以 } \frac{\sqrt{5}-1}{2\sqrt{5}} \text{ 的概率} \end{cases} \quad (19.19)$$

可以证明(参见习题),  $E(\hat{\varepsilon}_i^*) = 0$ ,  $\text{Var}(\hat{\varepsilon}_i^*) = \hat{\varepsilon}_i^2$ 。因此, 对  $\{\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \dots, \hat{\varepsilon}_n^*\}$  进行再抽样, 就是从异方差的扰动项进行抽样。Horowitz (1997, 2001) 的模拟显示, 在存在异方差的情况下, 野自助法比成对自助法更有效率; 即使在同方差的情况下, 也不逊色。

## 19.13 面板数据与时间序列的自助法

对于  $n$  大  $T$  小的短面板(short panel), 考虑其离差形式

$$\bar{y}_{it} = \bar{x}'_{it} \beta + \bar{\varepsilon}_{it} \quad (19.20)$$

其中, 固定效应已被去掉(参见第 15 章)。假设对于不同的个体  $i$ , 残差  $\bar{\varepsilon}_{it}$  是相互独立的; 但给定个体  $i$ , 则  $\bar{\varepsilon}_{it}$  可能存在异方差或自相关。此时, 可以使用成对自助法对个体  $i$  进行再抽样, 而不对时间  $t$  进行再抽样, 即如果抽中个体  $i$ , 则个体  $i$  在所有时间的观测值都同时被抽中。这种方法称为“分块自助法”(block bootstrap), 对于非线性面板数据或聚类数据(clustered data)也适用。

由于自助法假定样本为独立同分布的, 而时间序列数据通常存在自相关(故不是独立同分布的), 因此针对时间序列的自助法更为复杂。

## 19.14 自助法的 Stata 命令

使用自助法的一个具体问题是, 如何确定自助样本的个数  $B$ 。为了节省计算时间, Stata 的默认值为  $B = 50$ 。然而, 对于正式的研究而言, 这一般不够, 故必须使用选择项来指定自助样本个数  $B$ 。Cameron and Trivedi (2010) 建议, 针对 5% 的显著性水平, 如果使用自助法计算标准误差, 则取  $B = 400$ ; 如果进行区间估计或假设检验, 则取  $B = 999$ 。

在 Stata 中有三种方法可以实现自助法。

### 1. 使用选择项“vce(bootstrap)”

如果只是用自助法估计标准误, 对于横截面数据而言, 大多数 Stata 命令支持选择项“vce(bootstrap)”进行成对自助法, 比如,

```
reg y x1 x2 x3, vce(boot, reps(400) seed(10101) nodots)
```

其中, “boot”为“bootstrap”的简写, 选择项“reps(400)”表示  $B = 400$ , “seed(10101)”表示抽样“种子”为“10101”以保证可重复性<sup>①</sup>, “nodots”表示在执行命令过程中不显示代表抽样进程的“点”(可节省时间)。Stata 的输出结果包括了基于正态的置信区间(normal-based

<sup>①</sup> 只要每次均使用相同的抽样“种子”(seed), 即可保证自助抽样结果的可重复性。

confidence interval)。

对于面板数据而言,某些 Stata 命令也支持选择项“vce(bootstrap)”,但进行的是针对个体  $i$  的分块自助法,而非针对观测值  $(i, t)$  的成对自助法。

在执行了含选择项“vce(bootstrap)”的 Stata 命令之后,还可以使用以下命令获得自助法的更多信息:

```
estat boot, all
```

其中,选择项“all”表示显示根据不同方法计算的置信区间。

### 2. 使用命令“bootstrap”

对于一般的统计量使用自助法,可以使用命令 bootstrap, 其格式为

```
bootstrap explist[, options]: command
```

其中,“expList”提供了需要进行自助法估计的统计量。比如,使用如下命令可以得到与上面使用选择项“vce(bootstrap)”相同的结果。

```
bootstrap, reps(400) seed(10101) nodots: reg y x1 x2 x3
```

这个命令只对回归系数用了自助法,如果对标准误也使用自助法,则可以用如下命令。

```
bootstrap _b _se, reps(400) seed(10101) nodots: reg y x1 x2 x3
```

其中,“\_b”表示回归系数(得到回归系数估计量的自助分布),“\_se”表示标准误(得到标准误估计量的自助分布)。

在使用命令 bootstrap 之后,也可以用“estat boot”得到进一步的自助信息。

### 3. 使用命令“bsample”

使用命令 bootstrap 的局限是,自助法的整个计算过程必须能够用一个命令(比如, reg y x1 x2 x3)来表达。对于更复杂的情形(比如,参数自助法、残差自助法、野自助法),则可以用命令 bsample 进行一次有放回地再抽样,计算相应的统计量,然后用命令 simulate 进行重复抽样与计算。比如,使用命令“bsample 50”将得到样本容量为 50 的有放回再抽样样本(原始样本的样本容量必须大于或等于 50),这个再抽样的样本数据将替代内存中已有的数据。如果不指定再抽样的样本容量,则默认值为原始样本的样本容量  $n$ 。

## 19.15 使用自助法进行稳健的豪斯曼检验

考虑对解释变量的内生性进行豪斯曼检验(参见第 10 章)<sup>①</sup>。对于模型  $y_i = \mathbf{x}'\boldsymbol{\beta} + \varepsilon_i$ , 记  $\hat{\boldsymbol{\beta}}_{OLS}$  与  $\hat{\boldsymbol{\beta}}_{IV}$  分别为 OLS 与 IV 估计量。如果原假设 “ $H_0$ :所有解释变量均为外生变量” 成立, 则  $\hat{\boldsymbol{\beta}}_{OLS}$  与  $\hat{\boldsymbol{\beta}}_{IV}$  都是一致的, 但  $\hat{\boldsymbol{\beta}}_{OLS}$  更有效率。反之, 如果  $H_0$  不成立, 则  $\hat{\boldsymbol{\beta}}_{OLS}$  不一致, 而  $\hat{\boldsymbol{\beta}}_{IV}$  一致。因此, 在  $H_0$  成立的情况下,  $(\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{OLS}) \xrightarrow{P} \mathbf{0}$ , 而统计量

$$(\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{OLS})' [\widehat{\text{Var}(\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{OLS})}]^{-1} (\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{OLS}) \xrightarrow{d} \chi^2(r) \quad (19.21)$$

其中,  $r$  为内生解释变量的个数。为了计算  $\text{Var}(\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{OLS})$ , 需要知道  $\text{Cov}(\hat{\boldsymbol{\beta}}_{IV}, \hat{\boldsymbol{\beta}}_{OLS})$ 。为此, Hausman(1978) 假设, 在  $H_0$  成立的情况下, OLS 是最有效率的(fully efficient)。在此条件下, 可

<sup>①</sup> 对于使用豪斯曼检验来决定应该在面板模型中使用固定效应还是随机效应估计量, 以下讨论也大致适用。

以证明  $\text{Var}(\hat{\beta}_{IV} - \hat{\beta}_{OLS}) = \text{Var}(\hat{\beta}_{IV}) - \text{Var}(\hat{\beta}_{OLS})$ 。但如果存在异方差，则 OLS 并不是最有效率的（最有效率的估计量为 GLS）。因此，该检验在异方差情况下不适用。

为了得到异方差稳健的豪斯曼检验，可以使用自助法来直接计算  $\text{Var}(\hat{\beta}_{IV} - \hat{\beta}_{OLS}^*)$ ，而不需要假设 OLS 在  $H_0$  成立的情况下是最有效率的。自助法的操作步骤为，对样本进行  $B$  次自助抽样，得到  $B$  个自助估计量  $\{\hat{\beta}_{IV,b}^* - \hat{\beta}_{OLS,b}^*, b=1, \dots, B\}$ ，进而得到对  $\text{Var}(\hat{\beta}_{IV} - \hat{\beta}_{OLS}^*)$  的自助估计量：

$$\widehat{\text{Var}(\hat{\beta}_{IV} - \hat{\beta}_{OLS}^*)} = \frac{1}{B-1} \sum_{b=1}^B (\hat{\beta}_{IV,b}^* - \hat{\beta}_{OLS,b}^* - \bar{\beta}_{Diff}^*) (\hat{\beta}_{IV,b}^* - \hat{\beta}_{OLS,b}^* - \bar{\beta}_{Diff}^*)' \quad (19.22)$$

其中， $\bar{\beta}_{Diff}^* \equiv \frac{1}{B} \sum_{b=1}^B (\hat{\beta}_{IV,b}^* - \hat{\beta}_{OLS,b}^*)$ 。

以数据集 grilic.dta 为例。第 10 章曾使用该数据集估计教育投资的回报率，并进行了传统（非稳健）的豪斯曼检验。为了进行稳健的豪斯曼检验，先定义一个叫“hausmantest”的程序，来计算  $(\hat{\beta}_{OLS} - \hat{\beta}_{IV})$ ：

```
. use grilic.dta, clear
. program hausmantest, eclass
  1. tempname b bols biv
  2. reg lw iq s expr tenure rns smsa, r
  3. matrix `bols'=e(b)
  4. ivregress 2sls lw s expr tenure rns smsa (iq = med kww), r
  5. matrix `biv'=e(b)
  6. matrix `b'=`bols'-`biv'
  7. ereturn post `b'
  8. end
```

然后，用命令“bootstrap”进行 400 次自助抽样，计算  $(\hat{\beta}_{OLS}^* - \hat{\beta}_{IV}^*)$  400 次，并得到其自助协方差矩阵：

```
. bootstrap _b, reps(400) seed(10101) nodots nowarn: hausmantest
```

运行结果为

Bootstrap results		Number of obs = 758			
		Replications = 400			
	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]
iq	-.0106493	.0060772	-1.75	0.080	-.0225603 .0012618
s	.032007	.0182011	1.76	0.079	-.0036664 .0676804
expr	-.0039794	.0036892	-1.08	0.281	-.0112101 .0032513
tenure	.0045648	.003909	1.17	0.243	-.0030967 .0122262
rns	-.0310054	.0199684	-1.55	0.120	-.0701428 .008132
smsa	.0095145	.0128142	0.74	0.458	-.0156008 .0346298
cons	.6771285	.3887073	1.74	0.082	-.0847239 1.438981

然后检验所有解释变量的显著性：

```
. test s iq expr tenure rns smsa _cons
```

```
(1) s = 0
(2) iq = 0
(3) expr = 0
(4) tenure = 0
(5) rns = 0
(6) smsa = 0
(7) _cons = 0

chi2( 7) =      3.19
Prob > chi2 =    0.8673
```

需要注意的是,上述检验的统计量(3.19)是正确的,但 $\chi^2$ 分布的自由度应该为1(内生解释变量的个数)而非7(所有解释变量的个数)。为此,手工计算 $\chi^2(1)$ 的p值,

```
. dis chi2tail(1,3.19)
. 07409001
```

结果显示,p值为0.07,故可在10%的显著性水平上拒绝“所有解释变量均为外生变量”的原假设。

有关蒙特卡罗模拟与自助法的进一步知识及Stata操作,可参考Cameron and Trivedi (2005, 2010)的相应章节。

## 习 题

**19.1** 对于野自助法,证明(1) $E(\hat{\varepsilon}_i^*) = 0$ ; (2) $\text{Var}(\hat{\varepsilon}_i^*) = \hat{\varepsilon}_i^2$ 。

**19.2** 针对数据集nerlove.dta(参见第4章),使用自助标准误估计成本函数。

**19.3** (选做)针对面板数据集traffic.dta(参见第15章),使用自助法进行稳健的豪斯曼检验,以确定应该使用固定效应模型,还是随机效应模型。

## 附 录

### A19.1 随机数的产生

服从在 $(0,1)$ 区间上均匀分布的随机变量的样本值,称为“均匀分布随机数”,简称“随机数”(random number)。有了均匀分布的随机数后,就可以产生几乎任何分布的随机数了。然而,由电脑产生的所谓“随机数”,其实只是“伪随机数”(pseudo random number),因为它仍然来自确定性的序列(由递推公式产生)。计算机其实并不知道如何掷骰子!尽管如此,这些“伪随机数”都能通过独立性与均匀分布的统计检验,故可以作为随机数来使用。

比如,一个简单的“随机数产生器”(random number generator)可使用以下递推公式:

$$x_j = (kx_{j-1} + c) \bmod m \quad (19.23)$$

其中, $k, c$ 为常数,算子“ $a \bmod b$ ”表示 $a$ 除以 $b$ 后的余数,而 $m$ 是计算机所能表示的最大数<sup>①</sup>。这个序列就是介于0与 $m$ 之间的整数,而 $r_j \equiv x_j/m$ 就是介于0与1之间的随机数。初始值 $x_0$ ,称为种子(seed),决定了该随机序列的初始值 $r_0 \equiv x_0/m$ 。为了使随机样本具有可重复性,保证以后抽样或别人抽样也能得到完全一样的样本,需要设定种子。如果不给定种子,则按照“计算机内置钟表”(computer clock)来自动选择种子,这样每次抽样的样本就会不同,模拟的结果也就不能完全复制。

在Stata中,使用命令“set seed #”来确定“种子”,例如:

① 比如,如果计算机使用32位数进行计算,则 $m = 2^{32} - 1$ 。

```
set seed 10101      (确定“种子”为 10101)
set obs 30          (确定样本容量为 30)
```

```
gen x = runiform()  (产生均匀分布的随机变量 x)
```

### A19.2 离散型随机变量的抽样

假设离散型随机变量的分布律为

$$P(X = x_i) = p_i, \quad i = 1, 2, \dots, \quad \sum_{i=1}^{\infty} p_i = 1 \quad (19.24)$$

其中,  $x_i$  按下标  $i$  从小到大的顺序排列。显然, 随机变量  $X$  的累积分布函数是一个值域在  $[0, 1]$  之间的阶梯函数, 参见图 19.3。

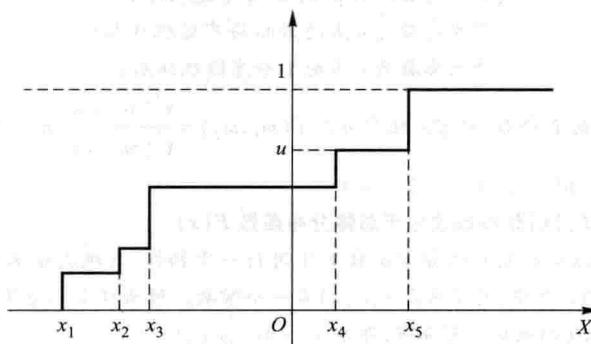


图 19.3 离散型随机变量的抽样

首先产生一个均匀分布的随机数  $u \in (0, 1)$ , 然后在纵轴上沿着累积分布函数找到  $u$  在横轴上对应的  $x_i$  取值(在图 19.3 中为  $x_4$ )。在代数上, 可以定义  $X$  为

$$X = \begin{cases} x_1, & u < p_1 \\ x_2, & p_1 \leq u < p_1 + p_2 \\ \dots \\ x_i, & \sum_{j=1}^{i-1} p_j \leq u < \sum_{j=1}^i p_j \\ \dots \end{cases} \quad (19.25)$$

可以证明, 这个随机变量  $X$  的分布律就是  $P(X = x_i) = p_i$ , 因为

$$P(X = x_i) = P\left\{\sum_{j=1}^{i-1} p_j \leq u < \sum_{j=1}^i p_j\right\} = \sum_{j=1}^i p_j - \sum_{j=1}^{i-1} p_j = p_i \quad (19.26)$$

其中, 第二个等号用到了均匀分布 “ $u \sim U(0, 1)$ ” 的性质。

在 Stata 中, 对离散型随机变量进行随机抽样的命令有

```
gen x = rpoisson(m)      (产生参数为 m 的泊松分布随机样本)
```

```
gen x = rbinomial(n,p)    (产生二项分布 b(n,p) 的随机样本)
```

### A19.3 连续型随机变量的抽样

上述对离散型随机变量抽样的思想, 对于连续型随机变量也是适用的。假设连续型随机变量  $X$  的累积分布函数为  $F(x)$ , 在产生一个均匀分布的随机数  $u \in (0, 1)$  后, 仍然可以在纵轴上沿着累积分布函数找到  $u$  在横轴上对应的  $x$  取值。其理论基础是下面的这个命题。

**命题** 设随机变量  $U \sim U(0, 1)$ ,  $F(x)$  是随机变量  $X$  的累积分布函数, 且  $F(x)$  是一个严格单调递增的连续函数, 则随机变量  $F^{-1}(U)$  具有分布函数  $F(x)$ , 其中  $F^{-1}(\cdot)$  是  $F(\cdot)$  的反函数。

**证明:** 由于  $F(x)$  是一个严格单调递增的连续函数, 故其反函数  $F^{-1}(x)$  存在, 而且  $F^{-1}(x)$  也是一个严格单调递增的连续函数。因此, 随机变量  $F^{-1}(U)$  的分布函数为

$$\begin{aligned} P\{F^{-1}(U) \leq x\} &= P\{F[F^{-1}(U)] \leq F(x)\} && (\text{因为 } F(x) \text{ 严格单调递增}) \\ &= P\{U \leq F(x)\} && (\text{反函数的性质}) \end{aligned}$$

$$= F(x) \quad (\text{因为 } U \sim U(0,1))$$

因此,  $F^{-1}(U)$  的分布函数为  $F(x)$ 。

这个命题表明,要产生以  $F(x)$  为分布函数的随机变量  $X$ ,只要先产生随机数  $U \sim U(0,1)$ ,然后再令  $X = F^{-1}(U)$  即可。这种方法称为“逆概率变换”(inverse-probability transformation)。

在 Stata 中,对连续型随机变量进行随机抽样的命令有

gen x = rnormal()	(产生标准正态分布 $N(0,1)$ 的随机样本)
gen x = rnormal(m,s)	(产生正态分布 $N(m,s^2)$ 的随机样本)
gen x = rt(m)	(产生自由度为 $m$ 的 $t$ 分布随机样本)
gen x = rchi2(m)	(产生自由度为 $m$ 的 $\chi^2$ 分布随机样本)
gen x = rbeta(a,b)	(产生参数为 $a,b$ 的 Beta 分布随机样本)
gen x = rgamma(a,b)	(产生参数为 $a,b$ 的 $\Gamma$ 分布随机样本)

对于自由度为  $(m_1, m_2)$  的  $F$  分布,则可以使用公式  $F(m_1, m_2) = \frac{\chi^2(m_1)/m_1}{\chi^2(m_2)/m_2}$  来产生其随机样本,即

$$\text{gen x} = (\text{rchi2(m1)/m1}) / (\text{rchi2(m2)/m2})$$

#### A19.4 经验分布函数 $F_n(x)$ 依概率收敛于总体分布函数 $F(x)$

证明:任意给定  $x$ ,  $-\infty < x < \infty$ ,从总体分布函数  $F$  中进行一次抽样,则观测值落入区间  $(-\infty, x]$  的概率为  $F(x)$ (根据分布函数的定义)。其中,对于给定  $x$ ,  $F(x)$  是一个常数。因为样本容量为  $n$ ,故在  $n$  次抽样中,观测值落在区间  $(-\infty, x]$  的次数  $S(x)$  服从二项分布,即  $S(x) \sim b(n, F(x))$ 。

根据二项分布的性质,  $E[S(x)] = nF(x)$ ,而  $\text{Var}[S(x)] = nF(x)[1 - F(x)]$ 。

根据定义,经验分布函数  $F_n(x) = \frac{1}{n}S(x)$ ,故  $E[F_n(x)] = \frac{1}{n}E[S(x)] = F(x)$ ,  $\text{Var}[F_n(x)] = \frac{1}{n^2}\text{Var}[S(x)] = \frac{1}{n}F(x)[1 - F(x)] \xrightarrow{P} 0$ 。

因此,  $F_n(x)$  依均方收敛于  $F(x)$ 。由于依均方收敛是依概率收敛的充分条件(参见第 5 章),故随机变量  $F_n(x)$  依概率收敛于常数  $F(x)$ 。

# 第 20 章 平稳时间序列

第 20~22 章介绍“时间序列数据”(time series data)的计量模型。根据时间序列的随机过程特性,可以分为“平稳序列”(stationary)与“非平稳序列”(non-stationary)两大类,需要使用不同的计量方法。本章介绍平稳序列<sup>①</sup>,下一章介绍非平稳序列。

## 20.1 时间序列的数字特征

时间序列指的是同一个人在不同时点上的数据。对于离散时间  $\{1, 2, \dots, T\}$ , 记随机变量  $y$  的相应观测值为  $\{y_1, y_2, \dots, y_T\}$ , 并假设其为(严格)平稳过程。因此,该时间序列的期望、方差、自协方差、自相关系数等数字特征均不随时间推移而改变。考察时间序列的数字特征,有助于辨别应该使用怎样的时间序列模型进行估计。

期望  $\mu \equiv E(y)$  反映的是该序列的平均水平,可以用样本均值  $\bar{y} \equiv \frac{1}{T} \sum_{t=1}^T y_t$  来估计。

方差  $Var(y)$  反映的是该序列的波动幅度,可以用  $\frac{1}{T-1} \sum_{t=1}^T (y_t - \bar{y})^2$  来估计。

通常,不同期的观测值之间存在相关性<sup>②</sup>。为此,引入以下两个概念。

定义 时间序列  $\{y_t\}$  的“ $k$  阶自协方差”(autocovariance of order  $k$ ) 为

$$\gamma_k \equiv Cov(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)] \quad (20.1)$$

其中,  $\mu \equiv E(y)$ 。它反映的是同一变量( $y$ )相隔  $k$  期之间的自相关程度。显然,当  $k=0$  时,  $\gamma_0 = Var(y)$ 。对  $\gamma_k$  的估计值为“样本自协方差”:

$$\hat{\gamma}_k \equiv \frac{1}{T-k} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) \quad (20.2)$$

定义 时间序列  $\{y_t\}$  的“ $k$  阶自相关系数”(autocorrelation of order  $k$ ) 为

$$\rho_k \equiv \text{Corr}(y_t, y_{t+k}) \equiv \frac{Cov(y_t, y_{t+k})}{\sqrt{Var(y_t)Var(y_{t+k})}} \quad (20.3)$$

它将自协方差标准化为介于  $[-1, 1]$  之间的量。显然,对于严格平稳过程,  $\rho_k$  不依赖于时间,而仅仅是滞后阶数  $k$  的函数,故被称为“自相关函数”(Autocorrelation Function,简记 ACF)。将  $(k, \rho_k)$  画成图,即为“自相关图”(correlogram)。由于自相关函数关于原点对称 ( $\rho_k = \rho_{-k}$ ),故一般只画自相关图的正的半边。对  $\rho_k$  的估计值为“样本自相关系数”:

$$\hat{\rho}_k \equiv \hat{\gamma}_k / \hat{\gamma}_0 \quad (20.4)$$

然而,  $y_t$  与  $y_{t+k}$  之间的相关性可能由二者之间的变量  $\{y_{t+1}, \dots, y_{t+k-1}\}$  所引起,而  $\rho_k \equiv \text{Corr}(y_t,$

① 如无特别说明,本章的时间序列均为“严格平稳”(strictly stationary) 过程。

② 如果将每个观测值比作一颗“珍珠”,则可以将时间序列想象为将“珍珠”串在一起的项链,即相邻的“珍珠”之间由线连在一起而“相关”。如果时间序列比较短,则每个观测值对于研究者的价值,大概也如“珍珠”般珍贵。

$y_{t+k}$ )并未对 $\{y_{t+1}, \dots, y_{t+k-1}\}$ 的作用进行控制。为此,引入以下概念。

定义 时间序列 $\{y_t\}$ 的“ $k$ 阶偏自相关系数”(partial autocorrelation of order  $k$ )为

$$\rho_k^* \equiv \text{Corr}(y_t, y_{t+k} | y_{t+1}, \dots, y_{t+k-1}) \quad (20.5)$$

即给定 $\{y_{t+1}, \dots, y_{t+k-1}\}$ 条件下, $y_t$ 与 $y_{t+k}$ 的条件相关系数。显然, $\rho_k^*$ 也只是 $k$ 的函数,故称为“偏自相关函数”(Partial Autocorrelation Function,简记 PACF)。为了得到对 $\rho_k^*$ 的样本估计值 $\hat{\rho}_k^*$ ,对以下 $k$ 阶自回归方程进行 OLS 估计:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_k y_{t-k} + \varepsilon_t \quad (20.6)$$

则 $\hat{\beta}_k$ 就是“ $k$ 阶样本偏自相关系数” $\hat{\rho}_k^*$ 。以上这些时间序列的数字特征,都是时间序列固有的特征,不依赖于模型的设定。另一方面,我们在设定模型时,希望尽可能地与这些数字特征相一致。

## 20.2 自回归模型

到目前为止,我们强调的是以回归模型来推断因果关系。如果从一个客户的角度仅关心某变量(比如股价)的未来值,则可以用该变量的过去值来预测其未来值(因为时间序列一般存在自相关)。这种模型被称为“单变量时间序列”(univariate time series)。此时,可以不必理会因果关系,只考虑相关关系即可。比如,看到街上有人带伞,可以预测今天要下雨,但行人带伞并不导致下雨。对于样本数据 $\{y_1, y_2, \dots, y_T\}$ ,最简单的预测方法为一阶自回归(AR(1)):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \quad (t=2, \dots, T) \quad (20.7)$$

其中,扰动项 $\varepsilon_t$ 为白噪声,满足零期望 $E(\varepsilon_t) = 0$ ,同方差 $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ ,且无自相关 $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0, t \neq s$ 。假设 $|\beta_1| < 1$ ,则 $\{y_t\}$ 为渐近独立的平稳过程(ergodic stationary)。由于 $y_{t-1}$ 依赖于 $\{\varepsilon_{t-1}, \dots, \varepsilon_1\}$ ,而 $\varepsilon_t$ 与 $\{\varepsilon_{t-1}, \dots, \varepsilon_1\}$ 不相关,故 $y_{t-1}$ 与 $\varepsilon_t$ 不相关,因此用 OLS 估计方程(20.7)是一致的。然而,使用 OLS 只能用观测值 $t=2, \dots, T$ 进行回归,将损失一个样本容量。

为了提高估计效率,考虑最大似然估计。为此,进一步假设 $\{\varepsilon_t\}$ 为独立同分布,且服从正态分布 $N(0, \sigma_\varepsilon^2)$ 。由于 $\{y_t\}$ 为平稳过程,其期望与方差均不随时间而变。对方程(20.7)两边同时取期望可得

$$E(y) = \beta_0 + \beta_1 E(y) \quad (20.8)$$

经移项整理可知, $\{y_t\}$ 的无条件期望为 $\frac{\beta_0}{1 - \beta_1}$ 。对方程(20.7)两边同时取方差可得

$$\text{Var}(y) = \beta_1^2 \text{Var}(y) + \sigma_\varepsilon^2 \quad (20.9)$$

故 $\{y_t\}$ 的无条件方差为 $\frac{\sigma_\varepsilon^2}{1 - \beta_1^2}$ 。因此, $y_1$ 服从正态分布 $N\left(\frac{\beta_0}{1 - \beta_1}, \frac{\sigma_\varepsilon^2}{1 - \beta_1^2}\right)$ ,其(无条件)密度函数为

$$f_{y_1}(y_1) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2/(1-\beta_1^2)}} \exp\left\{-\frac{[y_1 - (\beta_0/(1-\beta_1))]^2}{2\sigma_\varepsilon^2/(1-\beta_1^2)}\right\} \quad (20.10)$$

在给定 $y_1$ 的条件下, $y_2$ 的条件分布为, $y_2 | y_1 \sim N(\beta_0 + \beta_1 y_1, \sigma_\varepsilon^2)$ ,其条件密度为

$$f_{y_2|y_1}(y_2 | y_1) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{(y_2 - \beta_0 - \beta_1 y_1)^2}{2\sigma_\varepsilon^2}\right\} \quad (20.11)$$

则 $y_1$ 与 $y_2$ 的联合分布密度为

$$f_{y_1, y_2}(y_1, y_2) = f_{y_1}(y_1) f_{y_2|y_1}(y_2 | y_1) \quad (20.12)$$

依此类推得,  $y_t | y_{t-1} \sim N(\beta_0 + \beta_1 y_{t-1}, \sigma_e^2)$ ,  $t = 2, \dots, T$ 。写出整个样本数据  $\{y_1, y_2, \dots, y_T\}$  的联合概率密度(即似然函数)为

$$f_{y_1, \dots, y_T}(y_1, \dots, y_T) = f_{y_1}(y_1) \prod_{t=2}^T f_{y_t|y_{t-1}}(y_t | y_{t-1}) \quad (20.13)$$

取对数可得对数似然函数:

$$\ln L(\beta_0, \beta_1, \sigma_e^2; y_1, \dots, y_T) = \ln f_{y_1}(y_1) + \sum_{t=2}^T \ln f_{y_t|y_{t-1}}(y_t | y_{t-1}) \quad (20.14)$$

代入  $f_{y_1}(y_1)$  与  $f_{y_t|y_{t-1}}(y_t | y_{t-1})$  的表达式可得

$$\begin{aligned} \ln L = & -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln [\sigma_e^2 / (1 - \beta_1^2)] - \frac{[y_1 - (\beta_0 / (1 - \beta_1))]^2}{2\sigma_e^2 / (1 - \beta_1^2)} \\ & - \frac{T-1}{2} \ln 2\pi - \frac{T-1}{2} \ln \sigma_e^2 - \sum_{t=2}^T \frac{(y_t - \beta_0 - \beta_1 y_{t-1})^2}{2\sigma_e^2} \end{aligned} \quad (20.15)$$

寻找最优参数  $(\beta_0, \beta_1, \sigma_e^2)$ , 使得  $\ln L$  最大化, 这个估计量被称为“精确最大似然估计量”(Exact MLE)。由于  $\ln L$  为  $(\beta_0, \beta_1, \sigma_e^2)$  的非线性函数, 故需要使用迭代法进行数值计算。

如果样本容量  $T$  较大, 则第一个观测值对似然函数的贡献较小, 可以忽略。此时, 相当于考虑在第一个观测值  $y_1$  给定的情况下,  $\{y_2, \dots, y_T\}$  的条件分布, 则对数似然函数简化为

$$\ln L = -\frac{T-1}{2} \ln 2\pi - \frac{T-1}{2} \ln \sigma_e^2 - \sum_{t=2}^T \frac{(y_t - \beta_0 - \beta_1 y_{t-1})^2}{2\sigma_e^2} \quad (20.16)$$

最大化上式所得到的估计量称为“条件最大似然估计量”(conditional MLE)。显然, (20.16)式右边第一项不含任何参数(可视为常数), 第二项仅与  $\sigma_e^2$  有关, 而在第三项中  $\sigma_e^2$  仅出现于分母。因此, 为了使上式最大化, 必须让上式右边第三项的分子最小化(参见第6章), 即

$$\min_{(\beta_0, \beta_1)} \sum_{t=2}^T (y_t - \beta_0 - \beta_1 y_{t-1})^2 \quad (20.17)$$

其结果与对方程(20.7)进行 OLS 估计是一样的。因此, 与“精确 MLE”相比, “条件 MLE”的计算更为简便。更一般地, 可以考虑  $p$  阶自回归模型, 记为 AR( $p$ ):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t \quad (20.18)$$

对 AR( $p$ ) 的估计与 AR(1) 类似。在使用“条件 MLE”时, 考虑在给定  $\{y_1, \dots, y_p\}$  的情况下,  $\{y_{p+1}, \dots, y_T\}$  的条件分布。由于条件 MLE 等价于 OLS, 而后者并不依赖于正态性假定, 故条件 MLE 的一致性也不依赖于正态性假定。

然而, 通常我们并不知道滞后期  $p$ 。因此, 如何估计  $\hat{p}$  在实践中有着重要的意义。

方法之一: 设一个最大滞后期  $p_{\max}$ , 然后令  $\hat{p} = p_{\max}$  进行估计, 并对最后一个滞后期系数的显著性进行  $t$  检验。如果接受该系数为 0, 则令  $\hat{p} = p_{\max} - 1$ , 重新进行估计, 再对(新的)最后一个滞后期的系数进行  $t$  检验, 如果显著, 则停止; 否则, 令  $\hat{p} = p_{\max} - 2$ ; 依此类推。这个准则被称为“由大到小的序贯  $t$  规则”(general-to-specific sequential  $t$  rule)。

方法之二: 使用信息准则, 选择  $\hat{p}$  使得 AIC, BIC 或 HQIC 最小化, 分别记为  $\hat{p}_{\text{AIC}}$ ,  $\hat{p}_{\text{BIC}}$  与  $\hat{p}_{\text{HQIC}}$ 。比如,

$$\min_p \text{AIC} \equiv \ln(\mathbf{e}'\mathbf{e}/T) + \frac{2}{T}(p+1) \quad (20.19)$$

其中,  $\mathbf{e}'\mathbf{e}$  为残差平方和。可以证明,  $\hat{p}_{\text{BIC}}$  与  $\hat{p}_{\text{HQIC}}$  是真实滞后阶数  $p$  的一致估计, 但  $\hat{p}_{\text{AIC}}$  在大样本中可能高估  $p$ 。然而, 这几种信息准则在小样本中难分优劣。

在使用信息准则时, 通常也需要设定最大滞后期  $p_{\max}$ , 否则可能得到很大(不合理)的  $\hat{p}$ 。事

实际上,如果令  $\hat{p} = T - 1$ , 则该模型只剩下一个方程:

$$y_T = \beta_0 + \beta_1 y_{T-1} + \cdots + \beta_{T-1} y_1 + \varepsilon_T \quad (20.20)$$

显然,对这个方程用 OLS, 将得到唯一的残差  $e_T = 0$ 。因此, 残差平方和  $e'e = 0$ ,  $\ln(e'e/T) = -\infty$ 。也就是说,  $\hat{p} = T - 1$  使得 AIC 最小化, 但这显然不合理。

在实践中, 可以结合以上两种方法来确定  $\hat{p}$ 。如果二者不一致, 为了保守起见(即尽量避免遗漏变量偏差), 可取二者滞后阶数的大者。

## 20.3 移动平均模型

另一类时间序列模型为“移动平均过程”(Moving Average Process, 简记 MA)。记一阶移动平均过程为 MA(1):

$$y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1} \quad (20.21)$$

其中,  $\{\varepsilon_t\}$  为白噪声, 而  $\varepsilon_t$  的系数被标准化为 1<sup>①</sup>。由于  $y_t$  可以被看成是白噪声的移动平均<sup>②</sup>, 故得名。考虑使用条件 MLE 估计。为此, 假设  $\{\varepsilon_t\}$  为独立同分布且服从正态分布  $N(0, \sigma_\varepsilon^2)$ 。如果已经知道  $\varepsilon_{t-1}$ , 则

$$y_t | \varepsilon_{t-1} \sim N(\mu + \theta \varepsilon_{t-1}, \sigma_\varepsilon^2) \quad (20.22)$$

假设  $\varepsilon_0 = 0$ , 则可以知道  $\varepsilon_1 = y_1 - \mu$ 。给定  $\varepsilon_1$ , 则可以知道  $\varepsilon_2 = y_2 - \mu - \theta \varepsilon_1$ 。以此类推, 则可以使用递推公式 “ $\varepsilon_t = y_t - \mu - \theta \varepsilon_{t-1}$ ” 计算出全部  $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T\}$ 。这样, 在给定  $\varepsilon_0 = 0$  的条件下, 就可以写下样本数据  $\{y_1, y_2, \dots, y_T\}$  的似然函数, 然后使用数值方法求解其最大化问题。

更一般地, 对于  $q$  阶移动平均过程, 记为 MA( $q$ ):

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (20.23)$$

也可以进行条件 MLE 估计, 即在给定 “ $\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{-q+1} = 0$ ” 的条件下, 最大化样本数据的似然函数。

## 20.4 ARMA

为了使模型更好地拟合数据, 可以将 AR( $p$ ) 与 MA( $q$ ) 结合起来, 得到 ARMA( $p, q$ ):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (20.24)$$

其中,  $\{\varepsilon_t\}$  为白噪声。在给定  $\{y_1, \dots, y_p\}$  与 “ $\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{-q+1} = 0$ ” 的条件下, 可以使用条件 MLE 来估计 ARMA( $p, q$ )。为了估计 ARMA( $p, q$ ), 首先必须确定  $(p, q)$ , 而经济理论通常不能提供这些信息, 故只能根据数据来估计  $(\hat{p}, \hat{q})$ 。

在实践中, 常常先考察数据的自相关函数(ACF)与偏自相关函数(PACF), 以判断是否存在  $p = 0$  或  $q = 0$  的情形。如果  $p = 0$ , 则为 MA( $q$ ):

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (20.25)$$

显然, 如果  $j > q$ , 则  $\text{Cov}(y_t, y_{t-j}) = 0$ , 因为产生  $y_t$  的扰动项  $\{\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}\}$  与产生  $y_{t-j}$  的扰动

<sup>①</sup> 如果  $\varepsilon_t$  的系数不为 1, 可以在方程两边同时除以该系数。

<sup>②</sup> 其实是一种广义的“移动平均”, 因为并不要求权重之和为 1, 即  $(1 + \theta)$  不一定等于 1。

项 $\{\varepsilon_{t-j}, \varepsilon_{t-j-1}, \dots, \varepsilon_{t-j-q}\}$ 没有重叠的部分(无交集)。这意味着,对于 $MA(q)$ 模型,ACF函数在 $j > q$ 时都等于零,即出现“截尾”。另一方面,对于 $AR(1)$ ,我们知道其ACF函数呈指数衰减,被称为“拖尾”(tails off to zero),但不存在截尾。

如果 $q=0$ ,则 $ARMA(p,q)$ 简化为 $AR(p)$ 模型:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t \quad (20.26)$$

假设真实模型为 $AR(p)$ ,却用OLS来估计 $AR(p+1)$ ,即 $y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \beta_{p+1} y_{t-p-1} + \varepsilon_t$ ,则 $\lim_{T \rightarrow \infty} \hat{\beta}_{p+1} = 0$ ,因为 $\beta_{p+1} = 0$ 。而 $\hat{\beta}_{p+1}$ 正是对 $(p+1)$ 阶偏自相关函数的估计。由此可知,对于 $AR(p)$ 模型,PACF函数在 $j > p$ 时都等于零,即出现截尾。另一方面,对于 $MA(q)$ 模型,PACF函数逐渐衰减,即拖尾,但不存在截尾。

综上所述,对于 $AR(p)$ 模型,其ACF函数拖尾,而PACF函数截尾。如果出现这种情形,则可判断其为 $AR(p)$ ,不包含移动平均的部分。

另一方面,对于 $MA(q)$ 模型,其ACF函数截尾,而PACF函数拖尾。如果出现这种情形,则可判断其为 $MA(q)$ ,不包含自回归的部分。

如果以上两种情形均不符合,即ACF函数与PACF函数都拖尾<sup>①</sup>,则要考虑一般的 $ARMA(p,q)$ 模型,其中 $p,q$ 均不为零。时间序列分析的鼻祖,Box, Jenkins and Reinsel(1994)认为,对大多数情况, $p \leq 2$ 与 $q \leq 2$ 就足够了。当然,为了保险起见,可以让 $p_{\max}$ 与 $q_{\max}$ 更大些。具体如何确定 $p,q$ ,可以依据信息准则或由大到小的序贯 $t$ 规则。

在估计完模型之后,仍然需要进行一些诊断性分析(diagnostic checking),以确定 $ARMA(p,q)$ 模型的假定是否成立。其中,最重要的假定是,扰动项 $\{\varepsilon_t\}$ 为白噪声。如果模型过小(inadequate),即 $\hat{p} < p$ 或 $\hat{q} < q$ ,则相当于遗漏解释变量。这些被遗漏的解释变量被纳入扰动项中,导致扰动项出现自相关,不再是白噪声。为此,可以使用Q检验来检验模型的残差是否存在自相关。如果残差存在自相关,则应考虑使模型更大些(增加自回归或移动平均的阶数),重新对模型进行估计,然后再检验新模型的残差是否为白噪声,如此反复,直至确认残差为白噪声。

## 20.5 自回归分布滞后模型

在自回归模型中,也可以引入其他变量来构成“自回归分布滞后模型”(Autoregressive Distributed Lag Model,简记ADL( $p,q$ )):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \gamma_1 x_{t-1} + \dots + \gamma_q x_{t-q} + \varepsilon_t \quad (20.27)$$

比如,记通货膨胀率为 $\pi_t$ ,失业率为 $un_t$ ,则一个可能的预测模型为

$$\Delta \pi_t = \beta_0 + \beta_1 \Delta \pi_{t-1} + \beta_2 \Delta \pi_{t-2} + \beta_3 \Delta \pi_{t-3} + \beta_4 \Delta \pi_{t-4} + \gamma_1 un_{t-1} + \varepsilon_t \quad (20.28)$$

其中, $\Delta \pi_t$ 相当于“ $y_t$ ”。这个ADL(4,1)反映的是宏观经济学中的经验菲利普斯曲线(empirical Phillip Curve)。如果 $\gamma_1 < 0$ ,则失业率越低,物价越有上涨的压力。另一方面,通胀的调整也受到过去通胀变化的滞后作用。

也可以引入更多的解释变量。比如,共有 $K$ 个解释变量 $\{x_{1t}, \dots, x_{Kt}\}$ ,其中第 $j$ 个解释变量 $x_j$ 共有 $q_j$ 个滞后值被包括在模型中, $j = 1, \dots, K$ 。

<sup>①</sup> 不太可能二者均截尾。

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_{11} x_{1,t-1} + \cdots + \gamma_{1q_1} x_{1,t-q_1} + \cdots + \gamma_{K1} x_{K,t-1} + \cdots + \gamma_{Kq_K} x_{K,t-q_K} + \varepsilon_t \quad (20.29)$$

如果自回归分布滞后模型满足以下假定，则万事大吉，可以用 OLS 来估计它。

(i)  $E(\varepsilon_t | y_{t-1}, y_{t-2}, \dots, x_{1,t-1}, x_{1,t-2}, \dots, x_{K,t-1}, x_{K,t-2}, \dots) = 0$ 。这个假定类似于第 3 章的严格外生性假设，它意味着扰动项  $\varepsilon_t$  与所有解释变量的整个历史全部无关。这保证了对滞后期数  $(p, q_1, \dots, q_K)$  的设定是正确的。如果滞后期数的设定不正确，比如，真实模型还应该包括  $y_{t-(p+1)}$ ，但该项  $\beta_{p+1} y_{t-(p+1)}$  却被纳入扰动项  $\varepsilon_t$  中，则扰动项  $\varepsilon_t$  便与解释变量相关，导致 OLS 不一致。

(ii)  $\{y_t, x_{1t}, \dots, x_{Kt}\}$  为渐近独立的平稳序列。

(iii)  $\{y_t, x_{1t}, \dots, x_{Kt}\}$  有非零的有限四阶矩。

(iv) 解释变量无完全多重共线性。

对滞后期数的选择可以使用信息准则（最小化 AIC 或 BIC），或使用  $t, F$  检验来检验最后一期系数的显著性。更一般地，可以在 ARMA 模型中引入其他变量，称为“ARMAX”模型。

## 20.6 ARMA 模型的 Stata 命令及实例

### 1. 自相关与偏自相关

`corrgram y, lags(#)` （计算第 1 至第#阶 ACF 与 PACF，以及检验自相关的 Q 统计量）

`ac y, lags(#)` （将 ACF 画成自相关图，并给出置信区间）

`pac y, lags(#)` （将 PACF 画成偏自相关图，并给出置信区间）

### 2. ARMA

`arima y, ar(1/#) ma(1/#)`

其中，选择项“`ar(1/#)`”表示第 1 至第#阶自回归，而“`ma(#)`”表示第 1 至第#阶移动平均。

ARMA 的另一等价命令格式为

`arima y, arima(#p,#d,#q)`

其中，“`#p`”表示自回归的阶数，“`#q`”表示移动平均的阶数，而“`#d`”表示原序列  $\{y_t\}$  需要经过几次差分才是平稳过程（参见第 21 章）。下面检验其残差项是否存在自相关：

`predict e1,res` （计算残差，并将其命名为 `e1`）

`corrgram e1, lags(#)` （检验残差是否存在第 1 至第#阶自相关的 Q 检验）

### 3. ADL 与 ARMAX

ARMAX 的 Stata 命令为

`arima y x1 x2 x3, ar(#) ma(#)`

下面以数据集 `pe.dta` 为例<sup>①</sup>。该数据集的主要变量 `logpe` 为 1871—2002 年美国标准普尔股指（S&P）的市盈率（price earning ratio）的对数。由于 `logpe` 为非平稳序列（参见第 21 章），故对其差分建立 ARMA 模型。首先，定义其一阶差分为“`d_logpe`”：

`. use pe.dta, clear`

`. tsset year`

<sup>①</sup> 此例来自 Verbeek (2004)。

```
time variable: year, 1871 to 2002
delta: 1 unit
```

```
. g d_logpe = d.logpe
```

计算其前 10 阶自相关与偏自相关系数:

```
. corrgram d_logpe, lags(10)
```

LAG	AC	PAC	Q	Prob>Q	[Autocorrelation]	[Partial Autocor]
1	0.0651	0.0652	.56741	0.4513		
2	-0.1661	-0.1723	4.2951	0.1168		
3	-0.0611	-0.0393	4.8031	0.1868		
4	-0.1796	-0.2117	9.2305	0.0556		
5	-0.0530	-0.0525	9.6196	0.0868		
6	0.0897	0.0232	10.741	0.0967		
7	0.1288	0.0927	13.07	0.0704		
8	-0.0050	-0.0425	13.074	0.1093		
9	-0.0163	0.0132	13.112	0.1576		
10	0.0583	0.0917	13.601	0.1920		

上表显示,直至第 4 阶的 Q 统计量较为显著( $p$  值为 0.055 6)。更直观地,可以考察自相关图(结果如图 20.1)与偏自相关图(结果如图 20.2)。

```
. ac d_logpe, lags(10)
```

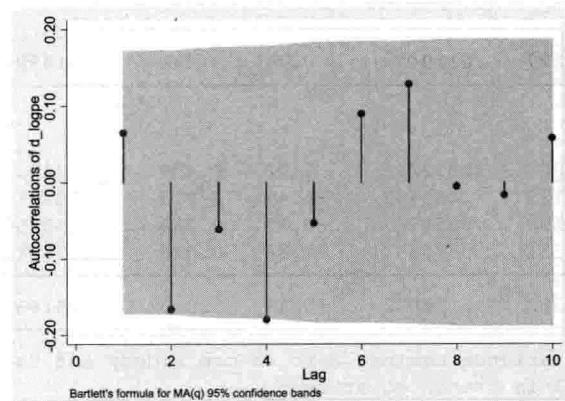


图 20.1 自相关图

```
. pac d_logpe, lags(10)
```

以上两图显示,第 4 阶自相关与偏自相关系数均在 5% 水平显著地不为 0(落在 95% 的置信区间之外),而可以认为 4 阶以上的自相关与偏自相关系数为 0。由于自相关系数与偏自相关系数均存在断尾,故分别考虑 AR(4) 与 MA(4) 模型。首先,估计 AR(4) 模型:

```
. arima d_logpe, ar(1/4) nolog
```

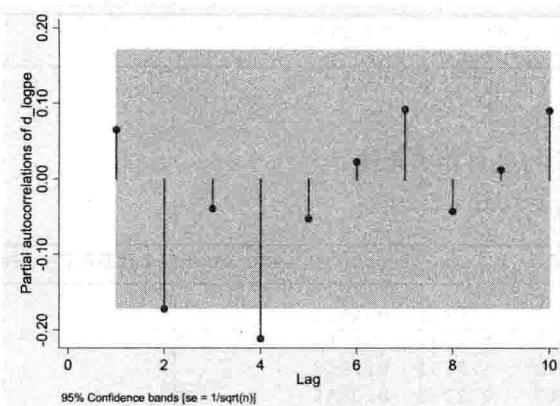


图 20.2 偏自相关图

ARIMA regression						
Sample:	1872 - 2002			Number of obs	=	131
				Wald chi2(4)	=	10.14
Log likelihood =	43.35448			Prob > chi2	=	0.0382
<hr/>						
d_logpe	OPG	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
d_logpe						
_cons	.0084227	.0119439	0.71	0.481	-.0149868	.0318322
<hr/>						
ARMA						
ar						
L1.	.0601002	.0832216	0.72	0.470	-.1030112	.2232117
L2.	-.2029129	.081432	-2.49	0.013	-.3625167	-.0433091
L3.	-.0228198	.0854787	-0.27	0.789	-.1903549	.1447153
L4.	-.206651	.0875637	-2.36	0.018	-.3782728	-.0350293
/sigma	.1736352	.0112072	15.49	0.000	.1516695	.195601
<hr/>						
Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.						

计算信息准则：

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	131	.	43.35448	6	-74.70897	-57.45779
Note: N=Obs used in calculating BIC; see [R] BIC note						

检验其残差项是否存在自相关：

```
. predict e1,res
(1 missing value generated)
. corrgram e1, lags(10)
```

LAG	AC	PAC	Q	Prob>Q	[Autocorrelation]	[Partial Autocor]
1	-0.0100	-0.0100	.01327	0.9083		
2	-0.0003	-0.0004	.01329	0.9934		
3	0.0103	0.0106	.0276	0.9988		
4	-0.0033	-0.0031	.0291	0.9999		
5	-0.0328	-0.0341	.17765	0.9993		
6	0.0256	0.0260	.26879	0.9996		
7	0.0912	0.0965	1.4376	0.9844		
8	-0.0211	-0.0190	1.5005	0.9927		
9	-0.0429	-0.0481	1.7638	0.9947		
10	0.0783	0.0801	2.6469	0.9886		

结果表明,可以接受残差项“无自相关”的原假设(检验至第 10 阶自相关)。

其次,估计 MA(4) 模型:

. arima d\_logpe,ma(1/4) nolog

ARIMA regression						
Sample: 1872 - 2002				Number of obs	=	131
				Wald chi2(4)	=	8.78
Log likelihood = 42.69439				Prob > chi2	=	0.0669
d_logpe	OPG	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
d_logpe						
_cons	.0080427	.0111621		0.72	0.471	-.0138347 .02992
ARMA						
ma						
L1.	.0478866	.0841641		0.57	0.569	-.1170719 .2128451
L2.	-.1875636	.0868965		-2.16	0.031	-.3578777 -.0172496
L3.	-.0400797	.0922776		-0.43	0.664	-.2209405 .1407812
L4.	-.1462202	.0904237		-1.62	0.106	-.3234474 .0310071
/sigma	.1745448	.0108632		16.07	0.000	.1532533 .1958363
Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.						

下面计算信息准则:

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	131		42.69439	6	-73.38878	-56.1376
Note: N=Obs used in calculating BIC; see [R] BIC note						

检验其残差项是否存在自相关:

. predict e2,res  
(1 missing value generated)  
. corrgram e2, lags(10)

LAG	AC	PAC	Q	Prob>Q	[Autocorrelation]	[Partial Autocor]
1	0.0082	0.0082	.00898	0.9245		
2	-0.0137	-0.0139	.03448	0.9829		
3	-0.0033	-0.0031	.03595	0.9982		
4	-0.0276	-0.0282	.14015	0.9977		
5	-0.0502	-0.0522	.48898	0.9925		
6	0.0873	0.0908	1.5509	0.9560		
7	0.0956	0.0986	2.8338	0.8999		
8	0.0024	0.0025	2.8346	0.9443		
9	-0.0390	-0.0426	3.0521	0.9622		
10	0.0687	0.0774	3.7308	0.9587		

上表显示,MA(4)的残差项也不存在自相关。根据信息准则,AR(4)模型似乎略优于MA(4)模型。然而,两个模型中的第1阶与第3阶系数均不显著。为此,考虑更为简洁的模型,将AR(4)与MA(4)模型中的第1阶与第3阶变量略去:

```
. arima d_logpe,ar(2 4) nolog
```

ARIMA regression						
Sample: 1872 - 2002					Number of obs	= 131
Log likelihood = 43.02852					Wald chi2(2)	= 11.61
					Prob > chi2	= 0.0030
d_logpe	OPG					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
d_logpe						
_cons	.0083801	.0114639	0.73	0.465	-.0140888	.0308489
ARMA						
ar						
L2.	-.201582	.0788054	-2.56	0.011	-.3560378	-.0471263
L4.	-.2119179	.0841825	-2.52	0.012	-.3769126	-.0469231
/sigma	.1740682	.0110923	15.69	0.000	.1523277	.1958086
Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.						

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	131	.	43.02852	4	-78.05704	-66.55625

Note: N=Obs used in calculating BIC; see [R] BIC note

```
. arima d_logpe,ma(2 4) nolog
```

ARIMA regression						
Sample: 1872 - 2002			Number of obs = 131			
			Wald chi2(2) = 9.66			
Log likelihood = 42.4702			Prob > chi2 = 0.0080			
<hr/>						
d_logpe	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
d_logpe		OPG				
_cons	.0080273	.0108818	0.74	0.461	-.0133006	.0293552
ARMA						
ma						
L2.	-.1809963	.0866252	-2.09	0.037	-.3507785	-.011214
L4.	-.1563501	.0865606	-1.81	0.071	-.3260057	.0133055
/sigma	.1748415	.0107209	16.31	0.000	.1538288	.1958541

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	131	.	42.4702	4	-76.9404	-65.43961

Note: N=Obs used in calculating BIC; see [R] BIC note

上表显示,略去第1阶与第3阶变量的简洁模型,其信息准则比完整模型更优。根据这个标准,似乎略去第1阶与第3阶变量的AR(4)模型最为合适。

## 20.7 误差修正模型

ADL是一种动态模型。从经济理论而言,相关的变量之间可能存在长期的均衡关系,而变量的短期变动则是向着这个长期均衡关系的部分调整。“误差修正模型”(Error Correction Model,简记 ECM)正是这一思想在计量经济学的体现。

考虑最简单的AR(1)模型:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \quad (20.30)$$

其中,  $|\beta_1| < 1$ , 故  $\{y_t\}$  为平稳过程。对方程(20.30)两边求期望, 并令长期均衡值  $y^* = E(y_t) = E(y_{t-1})$ , 则可得  $y^* = \frac{\beta_0}{1 - \beta_1}$ 。将  $\beta_0 = (1 - \beta_1)y^*$  代入方程(20.30), 并在方程两边同时减去  $y_{t-1}$  可得

$$\Delta y_t = (1 - \beta_1)y^* - (1 - \beta_1)y_{t-1} + \varepsilon_t \quad (20.31)$$

$$\Delta y_t = \underbrace{(1 - \beta_1)(y^* - y_{t-1})}_{\text{error correction}} + \varepsilon_t \quad (20.32)$$

这就是 AR(1) 的误差修正模型, 它将  $\Delta y_t$  表达为对长期均衡的偏离( $y^* - y_{t-1}$ )的部分调整(即误

差修正)加上扰动项。

考虑略微复杂的 ADL 模型:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + \varepsilon_t \quad (20.33)$$

其中,  $|\beta_1| < 1$ 。假设经济理论认为  $(y, x)$  之间存在长期均衡关系  $y = \phi + \theta x$ 。对方程(20.33)两边求期望, 并令  $y^* = E(y_t) = E(y_{t-1})$ ,  $x^* = E(x_t) = E(x_{t-1})$ , 可得

$$y^* = \beta_0 + \beta_1 y^* + \gamma_0 x^* + \gamma_1 x^* \quad (20.34)$$

$$(1 - \beta_1) y^* = \beta_0 + (\gamma_0 + \gamma_1) x^* \quad (20.35)$$

$$y^* = \frac{\beta_0}{(1 - \beta_1)} + \frac{(\gamma_0 + \gamma_1)}{(1 - \beta_1)} x^* \quad (20.36)$$

由此可得,  $\phi = \frac{\beta_0}{1 - \beta_1}$ ,  $\theta = \frac{\gamma_0 + \gamma_1}{1 - \beta_1}$ 。其中,  $\theta = \frac{\gamma_0 + \gamma_1}{1 - \beta_1}$  称为“长期乘数”(long-run multiplier), 它衡量的是当  $x$  永久性地变化一单位时, 将会导致  $y$  的永久性变化幅度。显然,  $\beta_0 = (1 - \beta_1)\phi$ ,

$\gamma_0 + \gamma_1 = (1 - \beta_1)\theta$ 。对方程(20.33)作如下变形, 即可得到误差修正模型。在方程(20.33)两边同时减去  $y_{t-1}$ , 并在方程右边加上、减去  $\gamma_0 x_{t-1}$ , 可得

$$\Delta y_t = \beta_0 - (1 - \beta_1) y_{t-1} + \gamma_0 (x_t - x_{t-1}) + (\gamma_0 + \gamma_1) x_{t-1} + \varepsilon_t \quad (20.37)$$

代入  $\beta_0 = (1 - \beta_1)\phi$ , 可得

$$\Delta y_t = (1 - \beta_1)\phi - (1 - \beta_1) y_{t-1} + \gamma_0 \Delta x_t + (\gamma_0 + \gamma_1) x_{t-1} + \varepsilon_t \quad (20.38)$$

代入  $\gamma_0 + \gamma_1 = (1 - \beta_1)\theta$ , 则有

$$\Delta y_t = \gamma_0 \Delta x_t + \underbrace{(\beta_1 - 1)(y_{t-1} - \phi - \theta x_{t-1})}_{\text{error correction}} + \varepsilon_t \quad (20.39)$$

这就是误差修正的形式。其中,  $(\beta_1 - 1)(y_{t-1} - \phi - \theta x_{t-1})$  称为“误差修正项”(error correction term)。参数  $\{\phi, \theta\}$  称为“长期参数”, 而参数  $\{\gamma_0, \beta_1 - 1\}$  称为“短期参数”。误差修正模型的形式看似非线性, 但仍是线性回归, 即可以把它还原为 ADL 模型, 然后用 OLS 来估计。一般来说, ADL 模型都可以转换成 ECM 模型(参见习题)。误差修正模型的优点是, 经济含义十分明确, 而且可以分别考察长期效应与短期效应。

## 20.8 MA( $\infty$ )与滞后算子

我们知道, 对于  $MA(q)$ , 其自相关函数(ACF)存在截尾。如果希望其自相关系数永远不为 0, 可以考虑  $MA(\infty)$ :

$$y_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \quad (20.40)$$

其中,  $\theta_0 = 1$ 。问题是, 这个无穷多个随机变量之和, 能收敛到某个随机变量吗? 一个常用的充分条件是, 序列  $\{\theta_j\}_{j=0}^{\infty}$  为“绝对值可加总”(Absolutely Summable, 简记 AS), 即  $\sum_{j=0}^{\infty} |\theta_j| < \infty$ (有限)。在 AS 的条件下, 不仅  $MA(\infty)$  有定义, 而且还是一个弱平稳过程(可通过计算二阶矩来验证); 更进一步, 如果  $\{\varepsilon_t\}$  为独立同分布的, 则  $MA(\infty)$  是一个严格平稳过程。虽然样本容量  $T$  通常有限, 而我们也不可能追溯到无穷远的过去, 但  $MA(\infty)$  在理论上仍然有着重要意义。

为了计算方便, 在时间序列分析中常引入“滞后算子”(lag operator 或 backshift operator)。

定义  $L y_t = y_{t-1}$ ,  $L^2 y_t = L(L y_t) = y_{t-2}, \dots, L^p y_t = y_{t-p}$ 。特别地,  $L^0 y_t = 1 \cdot y_t = y_t$ 。

显然, 滞后算子的运算相当于幂函数。比如, 两个滞后算子相乘满足交换律, 其乘积等于指

数相加,即  $L^p \cdot L^q = L^{p+q}$ 。由于  $\Delta y_t = y_t - y_{t-1} = (1-L)y_t$ ,故差分算子  $\Delta \equiv 1 - L$ 。

对于 AR( $p$ ),  $y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$ , 可以用滞后算子简洁地表示:

$$y_t = \beta_0 + \beta_1 L y_t + \cdots + \beta_p L^p y_t + \varepsilon_t \quad (20.41)$$

移项后可得

$$y_t - \beta_1 L y_t - \cdots - \beta_p L^p y_t = \beta_0 + \varepsilon_t \quad (20.42)$$

向右提取公因子  $y_t$

$$(1 - \beta_1 L - \cdots - \beta_p L^p) y_t = \beta_0 + \varepsilon_t \quad (20.43)$$

定义“滞后多项式”(lag polynomial)  $\beta(L) \equiv 1 - \beta_1 L - \cdots - \beta_p L^p$ , 则

$$\beta(L) y_t = \beta_0 + \varepsilon_t \quad (20.44)$$

上式容易让人产生一种愿望,即如果存在  $\beta(L)^{-1}$ , 则可以在两边同时左乘  $\beta(L)^{-1}$ , 从而得到 AR( $p$ ) 的一个“解”。

**定义** 对于任意一个实数序列  $(\alpha_0, \alpha_1, \alpha_2, \dots)$ , 定义其对应的“滤波”(filter)为  $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$ 。因此,滤波的实质就是一个(无穷多项)的滞后多项式。

**命题** 假设  $\{x_t\}$  为弱平稳过程, 序列  $\{\theta_j\}_{j=0}^{\infty}$  为绝对值可加总(AS), 则  $y_t = \alpha + \sum_{j=0}^{\infty} \theta_j x_{t-j} = \alpha + \sum_{j=0}^{\infty} \theta_j L^j x_t$  有定义,且为弱平稳。进一步,如果  $\{x_t\}$  为独立同分布,则  $\{y_t\}$  为严格平稳过程。

这个命题告诉我们,弱平稳过程经过 AS 滤波的作用后,仍为弱平稳过程。

**定义** 对于两个滤波 “ $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$ ”与“ $\beta(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots$ ”, 定义其乘积为

$$\begin{aligned} \delta(L) &\equiv \alpha(L)\beta(L) \equiv (\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots)(\beta_0 + \beta_1 L + \beta_2 L^2 + \cdots) \\ &= \alpha_0\beta_0 + (\alpha_0\beta_1 + \alpha_1\beta_0)L + (\alpha_2\beta_0 + \alpha_1\beta_1 + \alpha_0\beta_2)L^2 + \cdots \end{aligned} \quad (20.45)$$

显然,滤波的乘积满足交换律。我们最感兴趣的情形是  $\delta(L) = 1$ , 即令上式第二行中的常数项  $\alpha_0\beta_0$  为 1, 而其余各项的系数均为 0。此时,我们称  $\beta(L)$  为  $\alpha(L)$  的“逆”(inverse), 并记为  $\alpha(L)^{-1}$ 。只要  $\alpha_0 \neq 0$ , 则  $\alpha(L)^{-1}$  都有定义且唯一, 因为可以得到满足  $\alpha(L)\beta(L) = 1$  的唯一解  $(\beta_0, \beta_1, \beta_2, \dots)$ , 即  $\beta_0 = 1/\alpha_0$ ,  $\beta_1 = -\alpha_1\beta_0/\alpha_0$ , 等等。

**例**  $(1 - L)^{-1} = 1 + L + L^2 + L^3 + \cdots$

**证明:** 显然,  $(1 - L)(1 + L + L^2 + L^3 + \cdots) = 1$ 。需要注意的是,  $(1 - L)^{-1}$  不再是 AS。

**例**  $(1 - \beta L)^{-1} = 1 + \beta L + \beta^2 L^2 + \beta^3 L^3 + \cdots$

**证明:** 只要将上例中的“ $L$ ”换成“ $\beta L$ ”即可。如果  $|\beta| < 1$ , 则  $(1 - \beta L)^{-1}$  为 AS。

**命题** 对于  $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$ , 假设  $|\beta_1| < 1$ , 则此 AR(1) 是 MA( $\infty$ )。

**证明:**

方法一(迭代法)

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \\ &= \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= (\beta_0 + \beta_0\beta_1) + \beta_1^2 y_{t-2} + \beta_1 \varepsilon_{t-1} + \varepsilon_t \\ &= (\beta_0 + \beta_0\beta_1) + \beta_1^2(\beta_0 + \beta_1 y_{t-3} + \varepsilon_{t-2}) + \beta_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \beta_0(1 + \beta_1 + \beta_1^2) + \beta_1^3 y_{t-3} + \beta_1^2 \varepsilon_{t-2} + \beta_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \cdots \\ &= \beta_0(1 + \beta_1 + \beta_1^2 + \cdots) + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_1^2 \varepsilon_{t-2} + \beta_1^3 \varepsilon_{t-3} + \cdots \end{aligned}$$

$$= \frac{\beta_0}{1 - \beta_1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_1^2 \varepsilon_{t-2} + \beta_1^3 \varepsilon_{t-3} + \dots$$

方法二(滤波求逆法)

由于  $(1 - \beta_1 L) y_t = \beta_0 + \varepsilon_t$ , 故

$$\begin{aligned} y_t &= (1 - \beta_1 L)^{-1} (\beta_0 + \varepsilon_t) \\ &= (1 + \beta_1 L + \beta_1^2 L^2 + \dots) \beta_0 + (1 + \beta_1 L + \beta_1^2 L^2 + \dots) \varepsilon_t \\ &= \beta_0 (1 + \beta_1 + \beta_1^2 + \dots) + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_1^2 \varepsilon_{t-2} + \beta_1^3 \varepsilon_{t-3} + \dots \\ &= \frac{\beta_0}{1 - \beta_1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_1^2 \varepsilon_{t-2} + \beta_1^3 \varepsilon_{t-3} + \dots \end{aligned}$$

因此,可以将平稳的 AR(1)看成是过去所有扰动项的总效应之和,而且离现在越远的扰动项其影响力呈几何级数递减。从证明方法二,可以看出滤波求逆方法的威力。从 AR(1) 的 MA( $\infty$ ) 表达式可以看出

$$IRF(j) \equiv \frac{\partial y_{t+j}}{\partial \varepsilon_t} = \beta_1^j \quad (20.46)$$

$(\partial y_{t+j}/\partial \varepsilon_t)$  表示的是,当第  $t$  期的扰动项  $\varepsilon_t$  变化 1 单位时(而其他期的扰动项均不变),对相隔  $j$  期的  $y_{t+j}$  的影响,被称为“动态乘子”(dynamic multiplier)。显然,动态乘子与绝对时间  $t$  无关,而是时间间隔  $j$  的函数。将  $(\partial y_{t+j}/\partial \varepsilon_t)$  视为  $j$  的函数,则称为“脉冲响应函数”(Impulse Response Function,简记 IRF),它刻画的是  $y_{t+j}$  对  $\varepsilon_t$  的 1 单位脉冲(impulse)的响应(response)。将  $(j, \partial y_{t+j}/\partial \varepsilon_t)$  画图,就可以得到对 IRF 的直观认识。有时也考察  $\varepsilon_t$  从第 0 期(即当前期)到第  $k$  期对  $y_{t+j}$  的累积效应:

$$CIRF(k) \equiv \sum_{j=0}^k \frac{\partial y_{t+j}}{\partial \varepsilon_t} \quad (20.47)$$

上式称为“累积脉冲响应函数”(Cumulative Impulse Response Function,简记 CIRF)。类似地,AR( $p$ )也是 MA( $\infty$ )。将 AR( $p$ )“ $y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t$ ”写为“ $\beta(L)y_t = \beta_0 + \varepsilon_t$ ”,如果  $\beta(L)^{-1}$  为 AS,则

$$y_t = \beta(L)^{-1} \beta_0 + \beta(L)^{-1} \varepsilon_t \quad (20.48)$$

显然,这是 MA( $\infty$ )的形式。对于 AR( $p$ )的脉冲响应函数,可以通过计算  $\beta(L)^{-1}$  得到。另一方法是,直接对 AR( $p$ )进行如下模拟(simulation):令  $y_{-1} = y_{-2} = \dots = y_{-p} = 0, \varepsilon_0 = 1$ , 而  $\varepsilon_t = 0 (\forall t \neq 0)$ , 使用 AR( $p$ )方程“ $y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t$ ”(令  $t = 0$ )可得  $y_0$ 。然后,将  $\{y_0, y_{-1}, \dots, y_{-p+1}\}$  代入 AR( $p$ )方程(令  $t = 1$ )可得  $y_1$ 。依此类推,可以计算出  $y_j$ ,就是相隔  $j$  期的动态乘子  $(\partial y_{t+j}/\partial \varepsilon_t)$ 。

同样道理,更一般的 ARMA( $p, q$ )也是 MA( $\infty$ ):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (20.49)$$

$$y_t - \beta_1 L y_{t-1} - \dots - \beta_p L^p y_{t-p} = \beta_0 + \varepsilon_t + \theta_1 L \varepsilon_{t-1} + \dots + \theta_q L^q \varepsilon_t \quad (20.50)$$

$$\beta(L)y_t = \beta_0 + \theta(L)\varepsilon_t \quad (20.51)$$

其中,  $\theta(L) \equiv 1 + \theta_1 L + \dots + \theta_q L^q$ 。假设  $\beta(L)^{-1}$  为 AS, 则

$$y_t = \beta(L)^{-1} \beta_0 + \beta(L)^{-1} \theta(L) \varepsilon_t \quad (20.52)$$

这还是 MA( $\infty$ )的形式。总之,ARMA( $p, q$ )在本质上都可以归结为 MA( $\infty$ )。

## 20.9 向量自回归过程

我们常常同时关心几个经济变量的预测,比如GDP增长率与失业率。一种方法是用单变量时间序列的方法对每个变量分别作预测。另一种方法则是将这些变量放在一起,作为一个系统来预测,以使得预测相互自洽(mutually consistent),这称为“多变量时间序列”(multivariate time series)。由Sims(1980)所提倡的“向量自回归”(Vector Autoregression,简记VAR)正是这样一种方法。

假设有两个时间序列变量 $\{y_{1t}, y_{2t}\}$ ,分别作为两个回归方程的被解释变量;而解释变量为这两个变量的 $p$ 阶滞后值,构成一个二元(bivariate)的 $VAR(p)$ 系统:

$$\begin{cases} y_{1t} = \beta_{10} + \beta_{11}y_{1,t-1} + \cdots + \beta_{1p}y_{1,t-p} + \gamma_{11}y_{2,t-1} + \cdots + \gamma_{1p}y_{2,t-p} + \varepsilon_{1t} \\ y_{2t} = \beta_{20} + \beta_{21}y_{1,t-1} + \cdots + \beta_{2p}y_{1,t-p} + \gamma_{21}y_{2,t-1} + \cdots + \gamma_{2p}y_{2,t-p} + \varepsilon_{2t} \end{cases} \quad (20.53)$$

其中, $\{\varepsilon_{1t}\}$ 与 $\{\varepsilon_{2t}\}$ 均为白噪声过程(故不存在自相关),但允许两个方程的扰动项之间存在“同期相关性”(contemporaneous correlation):

$$\text{Cov}(\varepsilon_{1t}, \varepsilon_{2s}) = \begin{cases} \sigma_{12}, & \text{若 } t=s \\ 0, & \text{其他} \end{cases} \quad (20.54)$$

注意到上面两个方程的解释变量完全一样。将两个方程写在一起:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \gamma_{11} \\ \beta_{21} & \gamma_{21} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \beta_{1p} & \gamma_{1p} \\ \beta_{2p} & \gamma_{2p} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (20.55)$$

将同期变量写成列向量,并把相应的系数合并为矩阵,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \gamma_{11} \\ \beta_{21} & \gamma_{21} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \beta_{1p} & \gamma_{1p} \\ \beta_{2p} & \gamma_{2p} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (20.56)$$

记 $\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$ , $\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$ ,则有

$$\mathbf{y}_t = \underbrace{\begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix}}_{\Gamma_0} + \underbrace{\begin{pmatrix} \beta_{11} & \gamma_{11} \\ \beta_{21} & \gamma_{21} \end{pmatrix}}_{\Gamma_1} \mathbf{y}_{t-1} + \cdots + \underbrace{\begin{pmatrix} \beta_{1p} & \gamma_{1p} \\ \beta_{2p} & \gamma_{2p} \end{pmatrix}}_{\Gamma_p} \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (20.57)$$

定义相应的系数矩阵为 $\boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_p$ ,可得

$$\mathbf{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (20.58)$$

这个形式与 $AR(p)$ 很相似,故名“ $VAR(p)$ ”。其中, $\{\boldsymbol{\varepsilon}_t\}$ 是一维白噪声过程的推广,称为“向量白噪声过程”(vector white noise process)。

**定义** 向量白噪声过程 $\{\boldsymbol{\varepsilon}_t\}$ 是一个弱平稳的随机过程,满足以下条件:

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}, \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) = \boldsymbol{\Sigma}, \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_s) = \mathbf{0} \quad (t \neq s) \quad (20.59)$$

其中, $\boldsymbol{\Sigma}$ 为正定矩阵(positive definite)。由于不要求 $\boldsymbol{\Sigma}$ 为对角矩阵,故向量白噪声过程的各个分量之间可以存在同期相关。有时也把白噪声过程称为“新息过程”(innovation process)。

由于 $VAR(p)$ 系统中的解释变量 $\{y_{t-1}, \dots, y_{t-p}\}$ 依赖于 $\{\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots\}$ ,而 $\boldsymbol{\varepsilon}_t$ 与 $\{\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots\}$ 不相关,故可视所有解释变量为前定变量(predetermined),与当期扰动项 $\boldsymbol{\varepsilon}_t$ 不相关,故

可以用 OLS 对每个方程分别进行估计。由此可见，“扰动项  $\{\epsilon_t\}$  不存在自相关”是保证 OLS 一致性的主要条件。如果假设扰动项服从正态分布，则可以推导出条件 MLE 估计量。条件 MLE 等价于对每个方程分别用 OLS 进行估计。由于方程 (20.58) 右边的解释变量均为滞后变量  $\{y_{t-1}, \dots, y_{t-p}\}$ ，不包含同期变量  $y_t$ ，故也称为“简化式 VAR”(reduced-form VAR)。如果解释变量包括同期变量  $y_t$ ，则称为“结构 VAR”(structural VAR)，需要附加额外的约束条件以解决内生性问题，参见第 24 章。

在进行 VAR 建模时，需要确定变量的滞后阶数，以及 VAR 系统中包含几个变量。

### 1. 滞后阶数的选择

方法之一是使用信息准则。根据残差  $\hat{\epsilon}_t$  可以估计协方差矩阵  $\Sigma$ ，记为  $\hat{\Sigma}$ 。矩阵  $\hat{\Sigma}$  的  $(i, j)$  元素为  $\hat{\Sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_{it} \hat{\epsilon}_{jt}$ ，其中  $T$  为样本容量。则 VAR 模型的 AIC 与 BIC 分别为

$$AIC(p) = \ln |\hat{\Sigma}| + n(np+1) \frac{2}{T} \quad (20.60)$$

$$BIC(p) = \ln |\hat{\Sigma}| + n(np+1) \frac{\ln T}{T} \quad (20.61)$$

其中， $n$  为 VAR 系统中变量的个数， $p$  为滞后阶数， $|\hat{\Sigma}|$  为  $\hat{\Sigma}$  的行列式，而  $n(np+1)$  为 VAR 模型中待估系数之总数（每个方程共有  $(np+1)$  个系数，共有  $n$  个方程）。将  $|\hat{\Sigma}|$  视为多维情形下的残差平方和，则以上表达式为单一方程的信息准则向多方程情形的推广。比如，当  $n=1$  时， $\hat{\Sigma}$  为 1 阶矩阵，故  $|\hat{\Sigma}| = \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2 = SSR/T$ ，故  $AIC(p) = \ln(SSR/T) + (p+1)\frac{2}{T}$ ，这正是单一方程的 AIC 表达式（其中， $p+1$  为解释变量个数，含常数项）。

方法之二是检验最后一阶系数的显著性（类似于由大到小的序贯  $t$  规则）。在上面的例子中，假设要确定使用  $VAR(p)$  还是  $VAR(p-1)$ ，则可以检验原假设“ $H_0: \beta_{1p} = \beta_{2p} = \gamma_{1p} = \gamma_{2p} = 0$ ”。

方法之三是检验 VAR 模型的残差是否为白噪声，即是否存在自相关。如果真实模型为  $VAR(p)$ ，但被错误地设置为  $Var(p-1)$ ，则解释变量的最后一阶滞后  $y_{t-p}$  被纳入扰动项  $\epsilon_t$ ，导致扰动项出现自相关。更糟糕的是，由于  $\{y_t\}$  的相关性，包含  $y_{t-p}$  的扰动项  $\epsilon_t$  将与解释变量  $\{y_{t-1}, \dots, y_{t-(p-1)}\}$  相关，导致 OLS 估计不一致。为此，需要检验 VAR 模型的残差是否存在自相关。如果存在自相关，则可能意味着应该在解释变量中加入更高阶的滞后变量。

### 2. VAR 变量个数的选择

VAR 系统中包含的变量个数越多，则需要估计的系数越多。假设有 5 个变量，滞后 4 期，则每个方程中共有 21 个待估系数（含截距项），而整个 VAR 系统共有 105 个待估系数！待估系数过多将使得样本容量过小，增大估计误差，降低预测精度。因此，VAR 模型通常仅包含为数不多的几个变量。

另一方面，如果 VAR 模型太小，则可能存在遗漏变量偏差。比如，早期的 VAR 模型常发现一个奇怪结果，即紧缩货币政策之后通货膨胀反而上升，称为“物价之谜”(price puzzle)。Sims (1992) 认为，这是由于美联储在调整基准利率时，会考虑许多因素，而有些因素未包括在 VAR 模型中。如果某遗漏变量的变化预示着通货膨胀加剧，而美联储据此提高利率，则所谓的货币冲击 (monetary shocks) 不过是预期通胀的反映。此后，在 VAR 模型常使用“大宗商品价格”(commodity price) 来控制未来的预期通胀。

在设定 VAR 模型时，主要应根据经济理论来确定哪些变量应在 VAR 模型中。比如，经济理论

告诉我们,通货膨胀率、失业率、短期利息率互相关联,可以构成一个三变量的 VAR 模型。如果 VAR 模型包含不相关的变量,则会增大估计量方差,降低预测能力。另外,也可以在 VAR 系统中引入其他外生变量,比如  $\{z_{1t}, z_{2t}, \dots, z_{kt}\}$ ,与扰动项不相关。

## 20.10 VAR 的脉冲响应函数

由于 VAR 模型包含许多参数,而这些参数的经济意义很难解释,故将注意力集中于脉冲响应函数。为了研究 VAR 的脉冲响应函数,首先定义向量移动平均过程。将 MA( $\infty$ ) 向多维推广,可以定义  $n$  维“无穷阶向量移动平均过程”(Vector Moving Average, 简记 VMA( $\infty$ )):

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\psi}_0 \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2 \boldsymbol{\varepsilon}_{t-2} + \dots = \boldsymbol{\alpha} + \sum_{j=0}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\varepsilon}_{t-j} \quad (20.62)$$

其中  $\boldsymbol{\psi}_0 = \mathbf{I}_n$ ,  $\boldsymbol{\psi}_j$  皆为  $n$  维方阵。定义“多维滤波”(multivariate filter)为(无穷项)“滞后矩阵多项式”(lag matrix polynomial):

$$\boldsymbol{\psi}(L) \equiv \boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 L + \boldsymbol{\psi}_2 L^2 + \dots \quad (20.63)$$

因此,可以将 VMA( $\infty$ ) 简洁地写为  $\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\psi}(L) \boldsymbol{\varepsilon}_t$ 。对于两个多维滤波,可以类似于一维滤波那样地定义其乘积<sup>①</sup>,以及多维滤波  $\boldsymbol{\psi}(L)$  的“逆”(inverse)  $\boldsymbol{\psi}(L)^{-1}$ 。

使用滞后算子,可以把上述 VAR( $p$ ) 系统 “ $\mathbf{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \dots + \boldsymbol{\Gamma}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$ ” 写成 VMA( $\infty$ ) 的形式:

$$(\mathbf{I} - \boldsymbol{\Gamma}_1 L - \dots - \boldsymbol{\Gamma}_p L^p) \mathbf{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\varepsilon}_t \quad (20.64)$$

$$\boldsymbol{\Gamma}(L) \mathbf{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\varepsilon}_t \quad (20.65)$$

其中,  $\boldsymbol{\Gamma}(L) \equiv \mathbf{I} - \boldsymbol{\Gamma}_1 L - \dots - \boldsymbol{\Gamma}_p L^p$ 。在方程(20.65)两边同时左乘  $\boldsymbol{\Gamma}(L)^{-1}$  可得

$$\mathbf{y}_t = \boldsymbol{\Gamma}(L)^{-1} \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}(L)^{-1} \boldsymbol{\varepsilon}_t \quad (20.66)$$

记  $\boldsymbol{\Gamma}(L)^{-1} \equiv \boldsymbol{\psi}(L) = \boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 L + \boldsymbol{\psi}_2 L^2 + \dots$ ,  $\boldsymbol{\Gamma}(L)^{-1} \boldsymbol{\Gamma}_0 \equiv \boldsymbol{\alpha}$ , 则得到 VAR 模型的 VMA 表示法 (Vector Moving Average Representation):

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\psi}_0 \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2 \boldsymbol{\varepsilon}_{t-2} + \dots = \boldsymbol{\alpha} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t-i} \quad (20.67)$$

其中,  $\boldsymbol{\psi}_0 \equiv \mathbf{I}_n$ , 而其余的  $\boldsymbol{\psi}_i$  可通过递推公式  $\boldsymbol{\psi}_i = \sum_{j=1}^i \boldsymbol{\psi}_{i-j} \boldsymbol{\Gamma}_j$  来确定(参见习题)。根据向量微分法则可知

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}'_t} = \boldsymbol{\psi}_s \quad (20.68)$$

其中,  $(\partial \mathbf{y}_{t+s} / \partial \boldsymbol{\varepsilon}'_t)$  为  $n$  维列向量  $\mathbf{y}_{t+s}$  对  $n$  维行向量  $\boldsymbol{\varepsilon}'_t$  求偏导数, 故得到  $n \times n$  矩阵  $\boldsymbol{\psi}_s$ 。矩阵  $\boldsymbol{\psi}_s$  是一维情形下相隔  $s$  期的动态乘子向多维的推广, 其第  $i$  行、第  $j$  列元素等于  $(\partial y_{i,t+s} / \partial \varepsilon'_{jt})$ 。它表示的是, 当第  $j$  个变量在第  $t$  期的扰动项  $\varepsilon_{jt}$  增加 1 单位时(而其他变量与其他期的扰动项均不变), 对第  $i$  个变量在第  $(t+s)$  期的取值  $y_{i,t+s}$  的影响。将  $(\partial y_{i,t+s} / \partial \varepsilon'_{jt})$  视为时间间隔  $s$  的函数, 就是“脉冲响应函数”(IRF)。

有时也考察累积脉冲响应函数(CIRF), 即经过  $k$  期后的累积效应:

$$\sum_{s=0}^k \frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}'_t} = \sum_{s=0}^k \boldsymbol{\psi}_s \quad (20.69)$$

<sup>①</sup> 需要注意的是, 多维滤波的乘积不满足交换律, 正如矩阵乘法不满足交换律一样。

在方程(20.69)中,让 $k\rightarrow\infty$ ,则得到长期效应(long-run effect):

$$\sum_{s=0}^{\infty} \frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}'_t} = \sum_{s=0}^{\infty} \boldsymbol{\psi}_s = \boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 + \boldsymbol{\psi}_2 + \cdots = \boldsymbol{\psi}(1) \quad (20.70)$$

根据定义, $\boldsymbol{\Gamma}(L)\boldsymbol{\psi}(L)=\mathbf{I}$ 。令 $L=1$ 可得, $\boldsymbol{\Gamma}(1)\boldsymbol{\psi}(1)=\mathbf{I}$ 。因此,长期效应可以写为:

$$\boldsymbol{\psi}(1) = \boldsymbol{\Gamma}(1)^{-1} = (\mathbf{I} - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_p)^{-1} \quad (20.71)$$

脉冲响应函数的缺点是,它假定在计算( $\partial y_{t+s}/\partial \varepsilon_{jt}$ )时,只让 $\varepsilon_{jt}$ 变动,而所有其他同期扰动项均不变。此假定只有当扰动项的协方差矩阵 $\boldsymbol{\Sigma} \equiv E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t)$ 为对角矩阵时才成立(即同期扰动项之间正交);否则,当 $\varepsilon_{jt}$ 变动时,必然伴随着其他方程的同期扰动项发生相应的变动。为此,需要考虑“正交化的脉冲响应函数”(Orthogonalized Impulse Response Function,简记OIRF),即从扰动项 $\boldsymbol{\varepsilon}_t$ 中分离出相互正交的部分,记为 $\mathbf{u}_t$ ,然后计算当 $\mathbf{u}_t$ 中的某个分量变动时,对各变量在不同时期的影响。可以证明<sup>①</sup>,对任何对称正定矩阵 $\boldsymbol{\Sigma}_{n \times n}$ 均可进行以下“三角分解”(triangular factorization):

$$\boldsymbol{\Sigma} = \mathbf{A} \mathbf{D} \mathbf{A}' \quad (20.72)$$

其中, $\mathbf{D}$ 为 $n$ 级对角矩阵,其所有元素均为正数; $\mathbf{A}$ 为 $n$ 阶下三角矩阵(lower triangular matrix),其主对角线元素均为1;而且,矩阵 $\mathbf{A}$ 与 $\mathbf{D}$ 唯一确定。对扰动项的协方差矩阵 $\boldsymbol{\Sigma}$ 进行三角分解,并定义“正交化的扰动项”为

$$\mathbf{u}_t \equiv \mathbf{A}^{-1} \boldsymbol{\varepsilon}_t \quad (20.73)$$

可以证明,新扰动项 $\mathbf{u}_t$ 不再存在同期相关,因为其协方差矩阵为对角矩阵:

$$\begin{aligned} E(\mathbf{u}_t \mathbf{u}'_t) &= E(\mathbf{A}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{A}^{-1}) = \mathbf{A}^{-1} E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) \mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}'^{-1} = \underbrace{\mathbf{A}^{-1} \mathbf{A} D}_{=I} \underbrace{\mathbf{A}'^{-1}}_{=I} = \mathbf{D} \end{aligned} \quad (20.74)$$

其中, $\mathbf{D}$ 主对角线上的每一元素即为新扰动项 $\mathbf{u}_t$ 相应分量的方差。在实践中,常使用以下与三角分解相关联的“乔利斯基分解”(Cholesky decomposition)。由于 $\boldsymbol{\Sigma} = \mathbf{A} \mathbf{D} \mathbf{A}'$ ,而 $\mathbf{D}$ 为对角矩阵且所有元素均为正数,故可进一步分解为

$$\boldsymbol{\Sigma} = \mathbf{A} \mathbf{D}^{1/2} \mathbf{D}^{1/2} \mathbf{A}' \equiv \mathbf{P} \mathbf{P}' \quad (20.75)$$

其中, $\mathbf{P} \equiv \mathbf{A} \mathbf{D}^{1/2}$ 。由于 $\mathbf{A}$ 为下三角矩阵, $\mathbf{D}^{1/2}$ 为对角矩阵,故 $\mathbf{P}$ 依然为下三角矩阵;而 $\mathbf{P}'$ 依然为上三角矩阵。定义经乔利斯基分解所得的正交化扰动项为

$$\mathbf{v}_t \equiv \mathbf{P}^{-1} \boldsymbol{\varepsilon}_t \quad (20.76)$$

容易证明,新扰动项 $\mathbf{v}_t$ 的协方差矩阵为单位矩阵:

$$\begin{aligned} E(\mathbf{v}_t \mathbf{v}'_t) &= E(\mathbf{P}^{-1} \boldsymbol{\varepsilon}_t \mathbf{P}'^{-1}) = \mathbf{P}^{-1} E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) \mathbf{P}'^{-1} \\ &= \mathbf{P}^{-1} \boldsymbol{\Sigma} \mathbf{P}'^{-1} = \mathbf{P}^{-1} \mathbf{P} \mathbf{P}' \mathbf{P}'^{-1} = \mathbf{I} \end{aligned} \quad (20.77)$$

显然,新扰动项 $\mathbf{v}_t$ 各分量正交,且方差均被标准化为1;因此,新扰动项 $\mathbf{v}_t$ 的一个分量变化一单位,也就是变化一个标准差。利用(20.75)的乔利斯基分解,可将VAR模型的VMA表示法(20.67)写为:

$$\mathbf{y}_t = \boldsymbol{\alpha} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t-i} = \boldsymbol{\alpha} + \sum_{i=0}^{\infty} \underbrace{\boldsymbol{\psi}_i \mathbf{P} \mathbf{P}'^{-1}}_{\boldsymbol{\Phi}_i} \underbrace{\boldsymbol{\varepsilon}_{t-i}}_{\mathbf{v}_t} = \boldsymbol{\alpha} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \mathbf{v}_{t-i} \quad (20.78)$$

其中, $\boldsymbol{\Phi}_i \equiv \boldsymbol{\psi}_i \mathbf{P}$ 。由向量微分法则可知:

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{v}'_t} = \boldsymbol{\Phi}_s \quad (20.79)$$

<sup>①</sup> 参见 Hamilton(1994), p. 87。

其中,矩阵 $\Phi_i$ 的第*i*行、第*j*列元素等于( $\partial y_{i,t+s}/\partial v_{jt}$ ),即正交化的脉冲响应函数(OIRF)。在实际估计过程中,可根据残差 $\hat{\epsilon}_t$ ,估计 $\hat{\Sigma}$ ,对 $\hat{\Sigma}$ 进行乔利斯基分解得到 $\hat{P}$ 以及 $\hat{v}_t = \hat{P}^{-1}\hat{\epsilon}_t$ 。

然而,正交化的脉冲响应函数依然有缺点。首先,正交化冲击(orthogonalized shocks) $v_t$ 的经济含义不易解释。其次,由于乔利斯基分解所用矩阵 $P$ 为下三角矩阵,故正交化的脉冲响应函数依赖于变量的次序(order of variables);如果改变变量次序,则可能得到很不相同的结果。将方程(20.76)两边同时左乘 $P$ 可得 $Pv_t = \epsilon_t$ ,即

$$\begin{pmatrix} p_{11} & 0 & \cdots & 0 \\ p_{21} & p_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{nt} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{nt} \end{pmatrix} \quad (20.80)$$

作为一个思想实验,假设 $y_{2t}$ 受到一个单位的正交化冲击,即 $v_{2t}=1$ ,而向量 $v_t$ 的其他分量均为0。由于矩阵 $P$ 的第一行只有一个非零元素 $p_{11}$ ,故对 $v_{2t}$ 的冲击对于 $\epsilon_{1t}$ 没有影响,也不影响 $y_{1t}$ 。表达式(20.80)的第二个方程可写为

$$\epsilon_{2t} = p_{21}v_{1t} + p_{22}v_{2t} = p_{21} \times 0 + p_{22} \times 1 = p_{22} \quad (20.81)$$

上式表明, $v_{2t}$ 对 $y_{2t}$ 的冲击为 $p_{22}$ 。类似地,可以证明 $\epsilon_{3t} = p_{32}, \epsilon_{4t} = p_{42}$ ,等等。总之,对 $v_{2t}$ 的单位冲击会影响 $\epsilon_{2t} = \epsilon_{3t} = \cdots$ ,进而影响 $y_{2t} = y_{3t} = \cdots$ ,但不会影响 $\epsilon_{1t}$ 与 $y_{1t}$ 。更一般地,对 $v_{jt}$ 的一单位冲击,只对第*t*期的第*j*至第*n*个变量有影响,而对第*t*期的第1至第(*j*-1)个变量无影响。因此,按照一定变量顺序进行乔利斯基分解,意味着在VAR模型之上附加了一种“递归结构”(recursive structure),故也称为“递归VAR”(recursive VAR)。

在递归VAR中,对第一个方程的正交化冲击会在同期影响所有方程;对第二个方程的正交化冲击会在同期影响除第一个方程之外的所有方程;依此类推,对最后一个方程的正交化冲击则只会在同期影响最后一个方程本身。这个结果完全是由于在对协方差矩阵 $\hat{\Sigma}$ 进行乔利斯基分解时选择了某种变量次序;如果选择另一次序,就会得到不同的结果。因此,OIRF虽然使得因果关系更清楚,但代价是需要对变量起作用的次序作较强的先验假设,而经济理论通常无法对变量次序给出明确的指南。

在实践中,可借助以下三种方法来确定OIRF的变量次序。

(1) 通过考察两个变量之间的“交叉相关图”(cross correlogram),来确定两个变量的次序。对于变量 $y_t$ 与 $x_t$ ,定义其“交叉相关函数”(cross-correlation function)为

$$\rho_{yx}(k) \equiv \text{corr}(y_t, x_{t+k}) \quad (20.82)$$

即 $y_t$ 与 $x_{t+k}$ 的相关系数。将 $\rho_{yx}(k)$ 视为*k*的函数,并画图,即得到交叉相关图。需要注意的是, $\rho_{yx}(k)$ 并不关于原点对称。一般来说, $\rho_{yx}(k) \neq \rho_{yx}(-k)$ ;但 $\rho_{yx}(k) \equiv \text{corr}(y_t, x_{t+k}) = \text{corr}(x_t, y_{t-k}) \equiv \rho_{xy}(-k)$ 。通过交叉相关图,可看出使得 $|\rho_{yx}(k)|$ (绝对值)最大的*k*值,记为 $k^*$ 。如果 $k^* > 0$ ,则意味着 $y_t$ 与未来的 $x_{t+k^*}$ 最相关(变量排序应为 $y_t, x_t$ );反之,如果 $k^* < 0$ ,则意味着 $y_t$ 与过去的 $x_{t+k^*}$ 最相关(变量排序应为 $x_t, y_t$ )。

(2) 借助格兰杰因果检验来确定两个变量之间的排序,参见本章第12节。

(3) 在难以确定变量次序的情况下,为了稳健起见,对于不同的变量排序,分别画其正交化脉冲响应图,然后进行比较。如果不同变量排序的OIRF差别不大,则结果较可信;反之,则值得怀疑。

## 20.11 预测误差的方差分解

VAR 模型的用途之一是预测。对于  $\text{VAR}(p)$  模型  $\mathbf{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$ , 在得到参数估计后, 很容易进行向前一期的预测 (one-step-ahead forecast) :

$$\hat{\mathbf{y}}_{t+1} = \hat{\boldsymbol{\Gamma}}_0 + \hat{\boldsymbol{\Gamma}}_1 \mathbf{y}_t + \cdots + \hat{\boldsymbol{\Gamma}}_p \mathbf{y}_{t-p+1} \quad (20.83)$$

其中, 假设  $\mathbf{y}_t$  为已知, 即当前时间为第  $t$  期。为了预测  $\mathbf{y}_{t+2}$ , 则可将  $\hat{\mathbf{y}}_{t+1}$  视为  $\mathbf{y}_{t+1}$ , 即

$$\hat{\mathbf{y}}_{t+2} = \hat{\boldsymbol{\Gamma}}_0 + \hat{\boldsymbol{\Gamma}}_1 \hat{\mathbf{y}}_{t+1} + \hat{\boldsymbol{\Gamma}}_2 \mathbf{y}_t + \cdots + \hat{\boldsymbol{\Gamma}}_p \mathbf{y}_{t-p+2} \quad (20.84)$$

类似地, 可计算向前  $h$  期的预测  $\hat{\mathbf{y}}_{t+h}$ 。利用 VAR 模型的 VMA 表示法, 可将向前  $h$  期的预测误差 ( $h$ -step forecast error) 写为 (Lutkepohl, 2005, sec. 2.2.2) :

$$\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h} = \boldsymbol{\psi}_0 \boldsymbol{\varepsilon}_{t+h} + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t+h-1} + \cdots + \boldsymbol{\psi}_{h-1} \boldsymbol{\varepsilon}_{t+1} = \sum_{i=0}^{h-1} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t+h-i} \quad (20.85)$$

我们希望度量各方程的扰动项对预测误差的单独贡献。由于  $\boldsymbol{\varepsilon}_{t+h-i}$  的各分量存在同期相关 (contemporaneously correlated), 故利用乔利斯基分解将(20.85)式写为

$$\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h} = \sum_{i=0}^{h-1} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t+h-i} = \sum_{i=0}^{h-1} \underbrace{\boldsymbol{\psi}_i \mathbf{P} \mathbf{P}^{-1}}_{\boldsymbol{\Phi}_i} \boldsymbol{\varepsilon}_{t+h-i} = \sum_{i=0}^{h-1} \boldsymbol{\Phi}_i \boldsymbol{v}_{t+h-i} \quad (20.86)$$

其中,  $\boldsymbol{v}_{t+h-i}$  为正交化冲击。记矩阵  $\boldsymbol{\Phi}_i$  的  $(m, k)$  元素为  $\phi_{i,mk}$ , 向量  $\boldsymbol{v}_{t+h-i}$  的第  $k$  个元素为  $v_{k,t+h-i}$ , 则根据表达式(20.86), 可将  $\mathbf{y}_{t+h}$  的第  $j$  个分量  $y_{j,t+h}$  的预测误差写为:

$$\begin{aligned} y_{j,t+h} - \hat{y}_{j,t+h} &= \sum_{i=0}^{h-1} (\phi_{i,j1} v_{1,t+h-i} + \cdots + \phi_{i,jn} v_{n,t+h-i}) \\ &= \sum_{i=0}^{h-1} \sum_{k=1}^n \phi_{i,jk} v_{k,t+h-i} \\ &= \sum_{k=1}^n \sum_{i=0}^{h-1} \phi_{i,jk} v_{k,t+h-i} \\ &= \sum_{k=1}^n (\phi_{0,jk} v_{k,t+h} + \cdots + \phi_{h-1,jk} v_{k,t+1}) \end{aligned} \quad (20.87)$$

在上式的推导中交换了求和的次序, 即从“先对  $k$  求和, 再对  $i$  求和”变为“先对  $i$  求和, 再对  $k$  求和”。由于  $\boldsymbol{v}_{t+h-i}$  的每个正交化分量均不相关, 且方差为 1, 故很容易计算对  $y_{j,t+h}$  预测的均方误差 (MSE) :

$$\text{MSE}(\hat{y}_{j,t+h}) \equiv E[(y_{j,t+h} - \hat{y}_{j,t+h})^2] = \sum_{k=1}^n (\phi_{0,jk}^2 + \cdots + \phi_{h-1,jk}^2) \quad (20.88)$$

根据方程(20.88), 可以计算第  $l$  个变量的正交化冲击对  $y_{j,t+h}$  预测均方误差的贡献比例:

$$\frac{\phi_{0,jl}^2 + \cdots + \phi_{h-1,jl}^2}{\sum_{k=1}^n (\phi_{0,jk}^2 + \cdots + \phi_{h-1,jk}^2)} \quad (20.89)$$

这是度量第  $l$  个变量对第  $j$  个变量的影响的又一方法 (另一方法为脉冲响应函数)。显然, 由表达式(20.89)计算的贡献比例是预测时期  $h$  的函数, 故可将其列表或画图。而且, 所有变量对预测误差均方误差的贡献比例之和为 1。此性质称为“预测误差的方差分解” (Forecast-error Variance Decomposition, 简记 FEVD)。由于它将预测误差的来源分别归因于各变量的正交化新息 (orthogonalized innovation), 故也称为“新息核算” (innovation accounting)。

由于在推导过程中使用了乔利斯基分解, 故 FEVD 的结果也依赖于变量排序。这是 FEVD 与 IRF 的共同缺陷。为了稳健起见, 一般建议比较不同变量排序的 FEVD 结果; 如果结果对于变量排序不敏感, 则较为可信。

## 20.12 格兰杰因果检验

经济学中常常需要确定因果关系究竟是从  $x$  到  $y$ , 还是从  $y$  到  $x$ , 抑或双向因果关系。格兰杰 [Granger(1969)] 提出的检验方法基于以下思想。如果  $x$  是  $y$  的因, 但  $y$  不是  $x$  的因, 则  $x$  的过去值可以帮助预测  $y$  的未来值, 但  $y$  的过去值却不能帮助预测  $x$  的未来值。考虑以下时间序列模型<sup>①</sup>:

$$y_t = \gamma + \sum_{m=1}^p \alpha_m y_{t-m} + \sum_{m=1}^p \beta_m x_{t-m} + \varepsilon_t \quad (20.90)$$

其中, 滞后阶数  $p$  可根据“信息准则”或“由大到小的序贯  $t$  规则”来确定。检验原假设 “ $H_0: \beta_1 = \dots = \beta_p = 0$ ”, 即  $x$  的过去值对预测  $y$  的未来值没有帮助。如果拒绝  $H_0$ , 则称  $x$  是  $y$  的“格兰杰因”(Granger cause)。将以上回归模型中  $x$  与  $y$  的位置互换, 则可以检验  $y$  是否为  $x$  的格兰杰因。在实际操作中, 常将  $(x, y)$  构成一个二元 VAR 系统, 然后在 VAR 的框架使用 Stata 命令 vargranger 进行格兰杰因果检验。也可使用非官方命令 gcause, 下载方法为“ssc install gcause”。

需要指出的是, 格兰杰因果关系并非真正意义上的因果关系。它充其量只是一种动态相关关系, 表明的是一个变量是否对另一变量有“预测能力”(predictability)。从某种意义上来说, 它顶多是因果关系的必要条件(如果不考虑非线性的因果关系)。另外, 格兰杰因果关系也可能由第三个变量所引起。在经济学的实证研究中, 由于通常不可能进行“控制实验”, 能够最有说服力地说明因果关系的当属随机实验与自然实验(参见第 18 章)。

另外, 格兰杰因果检验仅适用于平稳序列, 或者有协整关系的单位根过程(参见第 21 章)。对于不存在协整关系的单位根变量, 则只能先差分, 得到平稳序列后再进行格兰杰因果检验。

## 20.13 面板格兰杰因果检验

对于面板数据, 可进行面板格兰杰因果检验。考虑以下面板模型:

$$y_{it} = \gamma + \sum_{m=1}^p \alpha_m y_{i,t-m} + \sum_{m=1}^p \beta_m x_{i,t-m} + u_i + \varepsilon_{it} \quad (i=1, \dots, n; t=1, \dots, T) \quad (20.91)$$

其中,  $u_i$  为个体  $i$  的异质性(个体效应), 可能与解释变量相关。显然, 这是一个动态面板模型。如果时间维度较短, 应使用差分 GMM 或系统 GMM 进行估计; 反之, 如果时间维度较长, 可考虑偏差校正 LSDV 法(参见第 16 章)。Hartwig(2009, 2010) 提供了进行面板格兰杰因果检验的两个实例。

## 20.14 VAR 的 Stata 命令及实例

与 VAR 相关的 Stata 命令包括(假设变量为  $x, y, z$ )

<sup>①</sup> 更一般地, 可以考虑  $y_t = \gamma + \sum_{l=1}^p \alpha_l y_{t-l} + \sum_{m=1}^q \beta_m x_{t-m} + \varepsilon_t$ , 其中  $p$  与  $q$  不一定相等。

```
varsoc x y z,maxlag (#)
```

此命令用来计算不同滞后期的信息准则,其中“`soc`”表示“Selection-Order Criteria”,选择项“`maxlag (#)`”表示最大滞后期,默认值为 4。在未使用选择项“`maxlag (#)`”的情况下,如果恰好选择最优滞后 4 期,则应通过选择项“`maxlag (#)`”增加最大滞后期数。

```
varbasic x y z,lags (numlist) irf fevd
```

这是估计 VAR 模型的便捷命令。选择项“`lags (numlist)`”表示滞后阶数,默认为“`lags (1 2)`”,即滞后二阶。选择项“`irf`”表示画(未正交化)脉冲响应图,选择项“`fevd`”表示画预测方差分解图,默认为“`oirf`”(画正交化脉冲响应图)。

更正式而深入的 VAR 命令为

```
var x y z,lags (numlist) dfk small exog (w1 w2)
```

其中,选择项“`lags (numlist)`”表示滞后阶数,默认为“`lags (1 2)`”,即滞后二阶。如果样本容量较小,可使用选择项“`dfk`”进行自由度调整。选择项“`small`”表示显示小样本的  $t$  或  $F$  统计量,而非大样本的标准正态或  $\chi^2$  统计量。选择项“`exog (w1 w2)`”表示在 VAR 模型中引入外生变量  $w_1, w_2$ 。

在使用命令 `var` 完成 VAR 模型的估计之后,Stata 提供了一系列“估计后命令”(postestimation commands)来检验 VAR 的性质与假设。

```
varlmar
```

估计 VAR 后,对残差是否存在自相关进行 LM 检验。

```
varnorm
```

估计 VAR 后,检验残差是否服从正态分布。

```
varstable,graph
```

估计 VAR 后,通过特征值检验该 VAR 系统是否为平稳过程。如果所有特征值都在单位圆内部,则为平稳过程,参见第 21 章。选择项“`graph`”表示画出特征值的几何分布图。

```
varwle
```

估计 VAR 后,对每个方程以及所有方程的各阶系数的联合显著性进行 Wald 检验,其中“`wle`”表示 Wald lag-exclusion statistics。

```
vargranger
```

估计 VAR 后,进行格兰杰因果检验。

```
irf create irfname, set (filename) step (#) replace
```

估计 VAR 后,将有关脉冲响应的结果存为“`irfname`”(可自行命名)。选择项“`set (filename)`”表示建立脉冲文件“`filename`”,使之成为当前的脉冲文件(`make filename active`),并将脉冲结果“`irfname`”存入此脉冲文件“`filename`”(若未使用选择项“`set (filename)`”指定脉冲文件,则将脉冲响应结果存入当前的脉冲文件);选择项“`step (#)`”表示考察#期的脉冲响应函数,默认值为“`step (8)`”;选择项“`replace`”表示替代已有的同名脉冲响应结果 `irfname`(如果有)。一个脉冲文件“`filename`”可存储多个脉冲响应结果“`irfname`”。

```
irf graph irf, impulse (varname) response (varname) noci
```

画脉冲响应图(未正交化)。其中,选择项“`impulse (varname)`”用于指定脉冲变量,而选择项“`response (varname)`”用来指定反应变量,默认画出所有变量的脉冲响应图。选择项“`noci`”表示不画置信区间,默认画置信区间。

```
irf graph cirf, impulse (varname) response (varname)
```

画累积脉冲响应图(未正交化)

```
irf graph oirf, impulse(varname) response(varname)
```

画正交化的脉冲响应图

```
irf graph coirf, impulse(varname) response(varname)
```

画正交化的累积脉冲响应图

```
irf graph fevd, impulse(varname) response(varname)
```

画预测方差分解图

如果将以上命令中的“`irf graph`”改为“`irf table`”, 则将相应信息列表而非画图。

```
fcast compute prefix, step(#)
```

估计 VAR 后, 计算被解释变量未来#期的预测值, 并把预测值赋予被解释变量加上前缀“prefix”(自行确定)的变量名。

```
fcast graph varlist, observed
```

运行命令“`fcast compute`”后, 将变量“`varlist`”的预测值画图, 其中选择项“`observed`”表示与实际观测值相比较。

下面以数据集 `varexample.dta` 为例演示<sup>①</sup>。该数据集包含自 1960 年第 1 季以来的美国宏观经济季度数据, 最初来自 Stock and Watson(2001), 后经 Becketti(2013)拓展至 2012 年第 1 季。主要变量包括:`date`(季度), `federal_funds`(联邦基金利率), `inflation`(通货膨胀率)<sup>②</sup>, 以及 `unrate`(失业率)<sup>③</sup>。我们希望估计一个关于利率、通胀率与失业率的三元 VAR 系统。

```
. use varexample.dta, clear
```

首先, 看一下这三个变量的时间趋势图, 结果参见图 20.3。

```
. tsline inflation unrate fedfunds, lpattern(" - " _) xline(169)
```

其中, 选择项“`xline(169)`”表示在 x 轴上  $t = 169$ (即 2002 年第 1 季)的位置画一条垂直线。

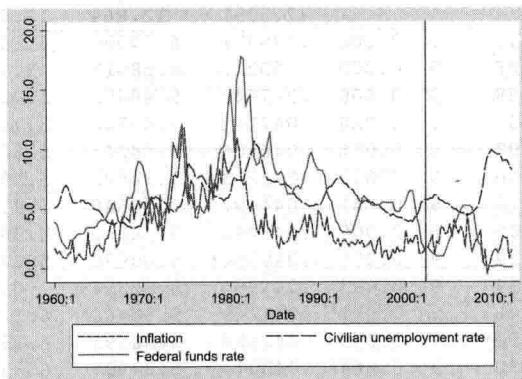


图 20.3 利率、通胀率与失业率的时间趋势图

在本例中, 预留最近十年的数据(即 2002 年第 2 季—2012 年第 1 季)作为“样本外预测”(out-of-sample forecasts)。图 20.3 中的垂直线为样本内(in-sample)与样本外(out-of-sample)的分

<sup>①</sup> 本例主要借鉴了 Becketti(2013)。

<sup>②</sup> 以年化的季度 GDP 平减指数的百分比变化(annualized quarterly percentage change in GDP price deflator)来衡量, 计算公式为  $400 \times \ln(P_t/P_{t-1})$ , 其中  $P_t$  为 GDP 平减指数。

<sup>③</sup> 其中, 联邦基金利率与失业率的季度值为其相应月度值的平均。