

第十四章 LISREL

§14.1 LISREL 操作使用

LISREL 是线性结构方程模型的专用软件包，这里的介绍基于LISREL VI，它对结构模型的表达形式与LISREL IV、V有所不同，它能自动计算初值，给出修正指数、增加模型的参数、结果绘图，可以自由格式读取数据。用五种方法(IV、2SLS、OLS、GLS和ML法) 进行模型估计。结构方程模型是典型的确证型分析。

§14.1.1 PC LISREL 运行环境

LISREL 运行不需数学协处理器，但可以自动使用它，应用协处理器可以大大加快软件的运行速度。在没有协处理器的AT 机上运行时建议设置环境变量NO87 =FALSE。建议在配置文件CONFIG.SYS 中使用DEVICE=[d:][PATH]ANSI.SYS 和BREAK=ON 设置。PC LISREL VI 系统包括以下文件：

README.DOC	有关LISREL的信息
CONFIG.SYS	LISREL 系统配置文件
LISREL.BAT	执行LISREL 的批文件
PCLIS1.EXE	第一阶段LISREL 执行文件
PCLIS2.EXE	第二阶段LISREL 执行文件
INPUT	检验程序
DATA	检验数据
OUTPUT1	第一阶段检测输出
OUTPUT2	第二阶段检测输出
STACKED.BAT	运行批处理的样本程序
KLEIN-1.INP	Klein 美国经济模型，ULS 估计
KLEIN-2.INP	Klein 美国经济模型，ML 估计
KLEIN.DAT	Klein 模型数据
MTMM-CSA.INP	MTMM (Multi-Trait, Multi-Method) 心理学分析示例

辅助程序DISP.EXE 用于在屏幕上逐屏显示运行结果，FPRINT.EXE 用于在打印机上输出运行结果。DISP.EXE 与DOS中MORE.COM执行不同之处在于它滤掉了回车控制符。

STACKED.BAT 内容为：

```
ECHO OFF
PCLIS1 KLEIN-1.INP KLEIN-1.UT1 KLEIN-1.UT2
PCLIS2
PCLIS1 KLEIN-2.INP KLEIN-2.UT1 KLEIN-2.UT2
PCLIS2
PCLIS1 MTMM-CSA.INP MTMM-CSA.UT1 MTMM-CSA.UT2
PCLIS2
ECHO ON
```

可见PCLIS1.EXE 可以带有三个命令行参数, 第一个为程序, 第二个和第三个是生成模型文件的结果文件。ECHO OFF/ON 是DOS内部命令, 用于批处理执行时屏幕显示的关闭与开启。

§14.1.2 进入系统

运行LISREL 批处理文件LISREL.BAT:

```
C:\LISREL>LISREL <Enter>
```

这时提示程序名、模型信息文件名和结果文件名。运行结束后, 用浏览或打印程序显示和打印结果文件。

也可以直接在命令行上同时给出三个文件名, 如:

```
C:\LISREL>LISREL PEER.INP FILE1 FILE2
```

表示利用程序PEER.INP 把运算结果放在FILE1 和FILE2 文件中。在软件提示运行程序名时直接打<Enter> 键则显示当前目录中的所有文件名。

使用命令C:\LISREL>LISREL INPUT OUT1 OUT2 <Enter> 产生结果文件OUT1 和OUT2 与OUTPUT1 和OUTPUT2 进行比较。

在IBM 机上的版本使用以下的运行格式:

```
//EXEC LISREL, SIZE=nnn 其中nnn 为占用空间的千字节数目。
```

§14.1.3 LISREL 建模过程

使用LISREL 一般有以下步骤:

- . 构思通相互关联的结构或通径图;
- . 根据结构, 写出结构方程组;
- . 收集资料;
- . 利用收集的数据进行实算;
- . 结果判断;
- . 模型的选择;
- . 实际意义的解释。

LISREL 的运行分为两个阶段, 第一阶段读入并检查程序和数据, 计算初值和待分析的矩阵。第一阶段的结果存入第一个结果文件, 默认为OUTPUT1, 利用它可以检查设定是否与期望相符。第二阶段接受第一阶段生成的二进制的中间文件并且进行其余的计算, 若有输出, 则隐含存入文件OUTPUT2中。微机系统LISREL大型计算机上的版本语句略有不同。

§14.1.4 LISREL 控制卡和用例

使用LISREL软件, 首先要熟悉由希腊字母表示的矩阵记号, 完整的程序应当指示分析的标题、数据、模型、输出, 每种指示用两个字母做为关键字, 称为控制卡, 一个卡一行写不下时, 可在参数的位置放以"C"做标记。对于每个分析, 可以使用一个标题, 使用的方法是在第80列上放非空的字符, 则标题出现在LISREL每页的输出上, 对应标题、数据、模型和输出。LISREL使用四种最基本的控制卡, 以下加以介绍:

1. TTitle 标题卡. 是LISREL 程序的第一卡。

2. DATA 数据卡. 指定LISREL 分析的数据。

DA 卡有参数NGroup=组数, 默认为1; NInputvar=输入变量数, 默认为0; NObs=样本大小, 默认为100。在N 未知时, 设定NO=0, 程序自动计算N值; MA=矩阵类型: MM 矩统计量(moment) 矩阵、CM 协方差阵、KM 相关阵、AM 增广矩统计量矩阵(即交叉乘积矩阵), 默认为CM。

在DA 卡的后面, 还可进一步使用四种信息, 即卡的标号、数据格式、矩阵的形式以及变量与数据的对应关系。用LAbels 指定标号。数据格式与Fortran数据格式说明类似, 如: (4F6.2), * 表示自由格式。矩阵形式有: RA 原始数据阵、MM 矩的矩阵、CM 协方差阵、KM 相关阵、FU 满元矩阵、SY 对称阵、ME 均值、SD 标准差。每个卡均可以带上UNIT=n,Fomat=ON REwind的三个参数。给隐式变量的标号是LK 与LE 卡, 相应于 ξ 与 η 量。数据顺序由SE 卡指示, 如: SE 1 3 4/ 表示数据顺序是1、3、4, 其中的反斜杠表示舍弃后面的量。

3. MOdel 模型卡. 指定LISREL 分析的模型, 内容包括: 模型中观察变量(X 与 Y) 的数目, 模型中的隐含变量数目(ξ 与 η), 每个待分析矩阵的形式(完整矩阵、对角阵、三角阵、单位阵、零矩阵)、每个矩阵估计的状态(如各元素取值, 是固定还是不固定)。仅不分析LISREL模型时使MO 卡缺失, 此时仅给出待分析的矩阵。

MO 卡可带有NY,NX,NE,NK量指示 y, x, ξ 及 η 的数目 p, q, m, n 。

4. OUput 输出卡. 指定LISREL 的输出。

OU 卡是针对五种方法的, 其组合为:

IV 仅计算辅助变量(IV) 估计;

TS 两步最小二乘(TSLS) 估计, 类似于IO(Initial only 仅仅初值);

UL 给出IV 及无约束最小二乘(ULS) 估计;

GL 给出TSLS 与广义最小二乘(GLS) 估计;

ML 给出TSLS 与ML 估计。

未加指示时用ML 方法, 指示方法很多时使用最末一个。与之有关的是NS 卡, 指示用给定的初值开始最速下降法的计算。其它的参数有:

PT 打印技术方面的输出(Technical Output);

SE 打印标准误;

TV 打印T-值;

PC 打印估计量的相关;

RS 打印与协方差阵有关的信息、正态化残差及 $Q-Q$ 图;

EF 打印总的效应;

VA 打印方差与协方差;

MR 作用等价于RS、EF 与VA;

MI 打印修正指数;

FS 打印因子分数的回归;

FD 打印一阶导数;

SS 打印标准解;

AL 给出所有输出;

NS 不用自动初值

TO 改动默认的132 字符行为80 字符;

ND 输出的小数位数(0-8), 默认为3。

TM 解此问题的最大CPU 秒数, 默认为60;

OU 卡应为求解问题时的最后一个卡, 其打印输出采用了Fortran 语言对打印机的控制方法。

软件运行时的错误信息可由多种原因引起[3]: 使用的关键字与LISREL 的习惯不符, 省略了必需的斜杠、等号、逗号或空格, 括号和引号不匹配, TI、DA、MO 和OU 卡的顺序写错, 输入相关阵漏掉了1, 使用了小写, 初值与实际值相差太远, 在使用自己的初值时在OU 中遗漏了NS。

LISREL 用几种手段进行适合度的评价, 注意的量是: 参数估计值、标准误(ML)、复相关的平方、决定系数、参数估计的相关(ML)。其中出现不合理的值则提示模型根本上是错误的, 如方差为负、相关大于1、方差或协方差阵非正定、复相关为负、估计量的标准误太大, 参数值高度相关等。当所有观察为多元正态分布、分析使用样本协方差阵和样本量足够大时, 可使用程序提供的 χ^2 检验, 它是一个似然比统计量, 此值对于样本含量的大小是敏感的, 模型假设(线性假设、共线性和可加性) 不合理也是常见的原因。另外可以使用拟合度指数(GFI)、调整拟合度指数(AGFI) 和剩余平方和(RMR)。GFI表示模型所解释的方差协方差的相对大小, AGFI 与GFI 的区别仅仅在于后者调整了模型的自由度。RMR 是样本协方差阵与假设协方差阵元素间平均的差异, 因此对于拟合较好的模型来说, 此值应该很小, 如小于0.05。利用这些指标可以比较针对于同一批资料两个模型的好坏。

LISREL 给出了一些指标来评价参数的取值。第一项是 t 值, 一般说来, $t > 2$ 则认为显著, 若 t 值不显著, 则参数可在以后的拟合中设为固定。第二项指标是标化残差, 它是样本协方差阵和假设协方差阵的差, 大于2 的值指示模型指示的问题。为了进一步研究模型的拟合, LISREL 提供了 $Q-Q$ 图, 它是标化残差的图, 在残差几乎重合于 45° 的线时表示拟合效果较好。最后, 针对每个固定参数, 软件给出了修正指数(MI), 它表示一个固定参数在松弛时期望使模型 χ^2 值下降多少, 是拟合函数关于固定和约束参数的导数。对于自由参数, 修正指数为零。修正指数可与自由度为1 的卡方比较, 若此值比较大, 则表明松弛此参数可以使拟合最大限度地得到改善, 修正的方法通常的逐个参数地进行而且参数的松弛应该有意义。使用修正指数来选择模型并不是一个值得推荐的方法。

现把模型指示综合列于下表:

模型参数可能的取值列于下表:

其中:

ZE= 0 是零矩阵

ID= I 是单位阵

IZ=[I 0] 或[I 0]' 为分块单位阵与零

ZI=[0 I] 或[0 I]' 为分块零与单位阵

DI=对角阵

SD=子对角阵, 即对角元为0 的下三角阵

SY=非对角的对称阵

ST=对称阵, 对角元为1的相关阵

FU=方阵或不对称方阵

表 14.1 LISREL 常见的几种模型及指示方法

模 型	指 示	默 认	参 数
$x = \Lambda_x \xi + \delta$	NX,NK	NY,NE	$\Lambda_x, \Phi, \Theta_\delta$
$y = By + \Gamma x + \zeta$	NY,NX	NE,NK	B, Γ, Ψ
$y = \Gamma x + \zeta$	NY,NX	NE,NK	$B = 0$
$y = By + \zeta$	NY	NX,NE,NK	B, Ψ
$y = \Lambda_y(1 - B)^{-1}(\Gamma\xi + \zeta) + \varepsilon$	NY,NE,NK	NX	$\Lambda_y, B, \Gamma, \Phi, \Psi, \Theta_\varepsilon$
$y = \Lambda_y(\Gamma\xi + \zeta) + \varepsilon$	NX,NE,NK	NX	$B = 0$
$y = \Lambda_y(1 - B)^{-1}\zeta + \varepsilon$	NY,NE	NX,NK	$\Lambda_y, B, \Psi, \Theta_\varepsilon$
$y = \Lambda_y\zeta + \varepsilon$	NY,NE	NX,NK	$B = 0$

表 14.2 LISREL 各种模型参数可能的取值

名称	LISREL 记号	LISREL 名称	阶数	可能的形式	默认值	默认为固 定或自由
Lambda-y	Λ_y	LY	NY,NE	ID,IZ,ZI,DI,FU	FU	FI
Lambda-x	Λ_x	LX	NX,NK	ID,IZ,ZI,DI,FU	FU	FI
BEta	B	BE	NE,NE	ZE,SD,FU	ZE	FI
GAmma	Γ	GA	NE,NK	ID,IZ,ZI,DI,FU	FU	FR
PHi	Φ	PH	NK,NK	ID,DI,SY,ST	SY	FR
PSi	Ψ	PS	NE,NE	ZE,DI,SY	SY	FR
Theta-epsilon	Θ_ε	TE	NY,NY	ZE,DI,SY	DI	FR
Theta-Delta	Θ_δ	TD	NX,NX	ZE,DI,SY	DI	FR

施加约束的方法：如：FR PS(1)、PS(2) PS(6-8) 可指示第一个、第二个及第六至第八个为自由参数。还可以结合相等条件的约束，见EQ 卡。

对于TE 与TD 指示，默认是对角和自由的。TD=SY 示对称阵、对角线自由，否则固定，TE=SY,FR,TD=SY,FR 示整个阵是自由的。

EQ 卡指示相等的约束

VA 卡与ST卡用于给出初值，如：ST 0.5 BE(1,1)-BE(3,3)、ST 0.5 ALL。

MA 卡即矩阵卡，用自由格式读入时可用斜杠中止数据读取。

【例14.1】Blau and Duncan(1967) 资料，有关的说明见第二节。设数据存放在文件BLAU.DAT 中，其排列为下三角阵：

```
1.000
0.516  1.000
0.453  0.438  1.000
0.332  0.417  0.538  1.000
0.322  0.405  0.596  0.541  1.000
```

现用以下程序调用这个数据进行分析：

```
Model for Blau and Duncan Stratification Data
DA NI=5 NO=20700
LA
*
'X1' 'X2' 'Y3' 'Y4' 'Y5'
KM SY FILE=Blau.DAT
SE
3 4 5 1 2
MO NX=2 NY=3 BE=SD PS=DI
FI GA 2 1 GA 3 1
OU TV RS MI FS EF
```

程序指定标题为Model for Blau and Duncan Stratification Data，数据卡指明共有五个输入变量，观察例数为20700。变量标号格式为自由格式，记为通常的 x_1, x_2, y_1, y_2 ，读取相关阵的数据，文件名为blau.dat。变量顺序是3, 4, 5, 1, 2，模型指定B矩阵是下三角阵 Ψ 阵为对角阵。 Γ 阵的两个元素应指明为固定的，分析输出内生变量及外变量影响的总效应等。

极大似然法计算结果表明，各量的符号正如期望， t -值均有显著性，从修正指数来看， Y_4 与 X_1 间的修正指数最大，但同时它的正态化残差也最大。

【例14.2】Duncan, Haller, and Portes(1968) 抱负的影响数据，相关阵为：

```
Respondent
x2 1.
x1 .1839 1.
x3 .2220 .0489 1.
y1 .4105 .2137 .3240 1.
```

```

y2 .4043 .2742 .4047 .6247 1.
Best friend
x5 .3355 .0782 .2302 .2995 .2863 1.
y6 .1021 .1147 .0931 .0760 .0702 .2087 1.
x4 .1861 .0186 .2707 .2930 .2407 .2950 -.0438 1.
y4 .2598 .0839 .2786 .4216 .3275 .5007 .1988 .3607 1.
y3 .2903 .1124 .3054 .3269 .3669 .5191 .2784 .4105 .6404 1.

```

设放于文件PEER.DAT 中, 分析程序为:

```

PEER INFLUENCES ON AMBITION(MODEL 2)
DA NI=10 NO=329
LA
*
'x2' 'x1' 'x3' 'y1' 'y2' 'x5' 'x6' 'x4' 'y4' 'y3'
KM SY FILE=PEER.DAT
SE
  4 5 10 9 2 1 3 8 6 7
MO NY=4 NE=2 NX=6 FIXEDX BE=FU
FR LY 2 1 BE 1 2
FI GA 1 5 GA 1 6 GA 2 1 GA 2 2
ST 1.0 LY 1 1 LY 4 2
EQ BE 1 2 BE 2 1
EQ LY 2 1 LY 3 2
EQ GA 1 1 GA 2 6
EQ GA 1 2 GA 2 5
EQ GA 1 3 GA 2 4
EQ GA 1 4 GA 2 3
EQ PS 1 1 PS 2 2
EQ TE 1 1 TE 4 4
EQ TE 2 2 TE 3 3
OU SE TV EF SS

```

输入变量数为10, 4 个为Y 变量, 6 个为X 变量, 观察数为329, 结果要求输出标准误、 T 值和总效应、标化解。

	Y1	Y2	Y3	Y4	X1	X2
	-----	-----	-----	-----	-----	-----
Y1	1.000					
Y2	.625	1.000				
Y3	.327	.367	1.000			
Y4	.422	.328	.640	1.000		
X1	.214	.274	.112	.084	1.000	

X2	.411	.404	.290	.260	.184	1.000
X3	.324	.405	.305	.279	.049	.222
X4	.293	.241	.411	.361	.019	.186
X5	.300	.286	.519	.501	.078	.336
X6	.076	.070	.278	.199	.115	.102

	X3	X4	X5	X6
	-----	-----	-----	-----
X3	1.000			
X4	.271	1.000		
X5	.230	.295	1.000	
X6	.093	-.044	.209	1.000
DETERMINANT = .700505D-01				

因为原始相关矩阵中变量顺序的需要变化, 使用卡片SE 把它们调整过来。软件给出相关阵的行列式值为 ‘700505D-01。在LISREL 的第一个输出文件中包括了待求解的参数, 在 $\Lambda_y, B, A, \Psi, \Theta$ 阵相应的位置指示出来, 本例 Φ 是固定的。

本例中使用两阶段最小二乘法(TSLS) 进行初值估计。

最后的极大似然估计下Y 的总决定系数为.937, 结构方程复相关系数的平方为: .564 和.571。结构方程总的决定系数为.729。模型 χ^2 值为30. 63, 自由度为24, $P=0.165$, 拟合度指数为0.982, 调整拟合度指数为0.900, 均方误差为0.023。在设定 Ψ 为对角矩阵、 $i_{21} = 0, \beta_1 = \beta_2$ 的条件下, $\chi^2 = 26.90$, 自由度为17, $p = 0.06$ 。

【例14.3】Klein 宏观经济学模型的ULS程序:

```
KLEIN'S MODEL I ESTIMATED BY IV AND ULS
DA NI=15 NO=21
LA file=klein.dat
RA file=klein.dat
SE
1 4 3 10 14 11 13 12 7 8 15 9 2 5 6
MO NY=8 NX=7 BE=FU GA=FI PS=FI
FR BE(1,4) BE(1,7) BE(2,4) BE(3,8)
FR GA(1,5) GA(2,5) GA(2,6) GA(3,4) GA(3,7)
FR PS(1,1)-PS(3,3)
VA 1 BE(4,5) BE(5,1) BE(5,2) BE(6,2) BE(7,3) BE(8,5) C
GA(5,3) GA(6,6) GA(7,1) GA(8,2)
VA -1 BE(4,7) GA(5,2) GA(8,1)
OU UL
```

Σ 矩阵在使用使用所有十五个参数时是奇异的, 因此ML法不可行。由于五个 y 变量是冗余的, 去掉后估计可行。

```
KLEIN'S MODEL I ESTIMATED BY ML
DA NI=15 NO=21
```


表 14.3 Klein 氏I 类模型估计结果

参数	TSLS	ULS	ML
a_1	0.02	0.04	-0.23
a_2	0.22	0.21	0.38
a_3	0.81	0.81	0.80
b_1	0.15	0.05	-0.80
b_2	0.62	0.68	1.05
b_3	-0.16	-0.17	-0.15
c_1	0.44	0.39	0.24
c_2	0.15	0.20	0.28
c_3	0.13	0.15	0.23

```

LA file=klein.dat
RA file=klein.dat
SE
1 4 3 7 8 15 9 2 5 6 /
MO NY=3 NE=8 NX=7 FI LY=IZ BE=FU GA=FI PS=FI TE=ZE
FR BE(1,4) BE(1,7) BE(2,4) BE(3,8)
FR GA(1,5) GA(2,5) GA(2,6) GA(3,4) GA(3,7)
FR PS(1,1)-PS(3,3)
VA 1 BE(4,5) BE(5,1) BE(5,2) BE(6,2) BE(7,3) BE(8,5) C
GA(5,3) GA(6,6) GA(7,1) GA(8,2)
VA -1 BE(4,7) GA(5,2) GA(8,1)
ST 5 PS(1,1) PS(2,2) PS(3,3)
OU NS

```

程序由文件KLEIN.DAT文件读取标号和数据(Fortran格式), 用到了上述各种卡控制。运行结果见下表:

§14.2 线性结构方程模型简介

单方程的线性模型和多元线性模型考虑了多个自变量间的相关, 但是它们将各个自变量平均地加以处理。线性结构方程模型(Linear Structural Relationship Model) 则是处理多个变量相互关联结构的一种方法, 它能对观察相关及就量间的相互影响给予解释, 模型已广泛用于计量经济学、心理学和社会学等学科。该方法常结合通径分析(path analysis) 和确证型因子分析(confirmatory factor analysis) 见于标准的统计专著中。

LISREL 模型由Joreskog 于1973年引入, 模型归结为以下三个方程: 结构子模型: $\eta =$

$$B\eta + \Gamma\xi + \zeta$$

$$y \text{ 测量子模型: } y = \Lambda_y \eta + \varepsilon$$

$$x \text{ 测量子模型: } x = \Lambda_x \xi + \delta$$

结构模型:

$B_{(n \times n)}$ 是关联内生变量的回归矩阵

$\Gamma_{(m \times n)}$ 是关联 m 个外生变量与 n 个内生变量的回归矩阵

$\Phi_{(m \times m)}$ 是 m 个外生变量 ξ 的对称方差协方差阵

$\Psi_{(n \times n)}$ 是 n 个内生变量剩余 ζ 的对称方差协方差阵

$\eta_{(n \times 1)}$ 是内生变量向量

$\xi_{(m \times 1)}$ 是外生变量向量, $\xi \sim N(0, \Phi)$

$\zeta_{(n \times 1)}$ 是剩余向量, $\zeta \sim N(0, \Psi)$

测量模型:

$\Lambda_{x(p \times m)}$ 是外生变量与 p 个测量变量间的回归矩阵

$\Lambda_{y(q \times n)}$ 是内生变量与 q 个测量变量间的回归矩阵

$\Theta_{\delta(p \times p)}$ 是外生变量测量误差间的对称方差协方差阵

$\Theta_{\varepsilon(q \times q)}$ 是内生变量测量误差间的对称方差协方差阵

假设是: ζ 与 ξ 不相关

ε 与 η 不相关

δ 与 ξ 不相关

ζ, ε 与 δ 互不相关

B 的对角元是零, 但 $I - B$ 非奇异

结构方程指示隐含变量间的关系, 测量模型把隐含变量同其示性量关联起来, 结构方程模型的假设通常有两大类: 一种是因果联系的假设, 一种是关于误差的分布。模型分析所用变量一般是标准化的, 但在说明时常常不专门区分。

模型有三类变量, 内生的(endogenous)、外生的(exogenous)量以及扰动量(disturbance)。上式中 y 为内生变量, 是由模型内确定。 x 为外生变量, 是事先给定的。内生变量可受其它内生变量、外生变量和扰动量的影响, 内生量在模型中可作为原因出现, 但不作为结果出现。除了这种区分外, 模型变量还可区分为可测的量与不可测的量。表示模型的结构关系, 还可以用通径分析那样的记号, 表达变量间的关系。如 \square 表示可测量、 \rightarrow 表示变量间的作用、 \circ 表示不可测量, 等等。不同的误差或扰动(disturbances)不相关即它们之间没有双箭头, 因果结构是单向的, 则模型是可逆归的(recursive)。用矩阵的记号, 没有双箭头意味着误差的协方差阵是对角的, 作用是单向的表明 B 矩阵是下三角阵。把这个条件稍微放松一些, 若模型不符合逆归模型的条件, 但是对于内生变量和误差的子集之间存在, 这时称做块逆归的(block-recursive)。

模型一旦被指定, 就涉及参数的估计问题, 也称做识别问题(identification problem)。在单一方程线性模型中, 若设计矩阵是满秩的, 模型的一般假设就能保证其参数被估计; 结构方程模型中一些分布的假设一般来说不能保证模型能够识别, 这时要施加一些额外的限制。这种限制采取的形式通常有两种: 一是指定一些结构上为零, 这与对数线性模型有些类似; 另外就是对误差协方差阵的限制。一个参数若能估计就称为识别(identified), 否则就是不能识别(underidentified)或过识别(overidentified), 过识别是指一个参数能有几个解, 只有当一个参数恰好有一个解时, 才是恰好识别(exactly identified)。模型识别的条件是待估计参数的数目不能超于已知量的数目。

有许多方法能用于结构方程模型的识别, LISREL 提供了工具变量法、两阶段最小二乘法等几种估计方法, 从模型正态误差出发进行模型参数估计在计量经济学中常称为充分信息极大似然估计(FIML), LISREL 模型指定时已限定某些参数为 0 或 1, 所以是一个约束最优化

方法, 往往要采用数值方法。

记 F 是目标函数, k 是观测 x, y 变量数, d 是模型自由度, t 是被估计的独立参数数目, 有关公式如下:

$$F(\text{ULS}) = 0.5 \text{tr}[(S - \Sigma)^2]$$

$$\text{GFI} = 1 - \text{tr}(S - \Sigma)^2 / \text{tr}(S^2)$$

$$F(\text{ML}) = \log |\Sigma| + \text{tr}(S\Sigma^{-1}) - \log |S| - (p + q)$$

$$\text{GFI} = 1 - \text{tr}(\hat{\Sigma}^{-1}S - I)^2 / (\hat{\Sigma}^{-1}S)^2$$

$$\text{AGFI} = 1 - [k(k+1)/2d](1 - \text{GFI})$$

$$\text{RMR} = \sqrt{2 \sum_{i=1}^k \sum_{j=1}^i (s_{ij} - \hat{\sigma}_{ij})^2 / k(k+1)}$$

$$\text{AIC} = \chi^2 + 2t$$

$$\text{CAIC} = \chi^2 + (1 + \log N)t$$

$$\text{ECVI} = \chi^2 / (N - 1) + 2t / (N - 1)$$

RMR 是均方误差。

模型 χ^2 值是最小目标函数的拟合值 $\times (N - 1)$, 自由度取作 $0.5(p + q)(p + q - 1) - t$ 或 $0.5k(k + 1) - t$ 。

据R.A. Johnson(1988), 模型模型的拟合战略有: 1. 如果可能, 用几种准则估计模型中的参数, 然后注意: 各量的符号与大小是一致的吗? 所有的方差估计为正吗? 剩余矩阵是否对称? 2. 利用协方差阵和相关阵分别进行计算, 标化观察变量后, 结果有何影响; 3. 把一个大的数据集分成两部分, 分别重复 1、2. 做法, 然后观察两部分结果之间以及与全部数据之间的稳定性。

LISREL 方法的优点在于它能给出变量作用的间接效应、把变量对间统计关系区分成因果和非因果的组分, 这一点与通径分析有相同之处, 除了与通径分析的关联外, 与因子分析也有密切的关系, LISREL 进行的因子分析模型通常称作确证型因子分析与探索性因子分析(Exploratory Factor Analysis) 相区分。因为后者并不是先验地假定变量间有给定的关系存在。

另外要注意的是, 结构方程模型虽用于描述多个变量关联及相互作用的大小, 但有时称作Causal Model 并不恰当, 因为许多研究旨在发现因果关联, 很少是把简单的相关或预测作为目的。同时, 进行因果推断时, 同样存在着固有的缺陷[2]。

Blau and Duncan stratification 模型[2,3] (观察例数为20,700)

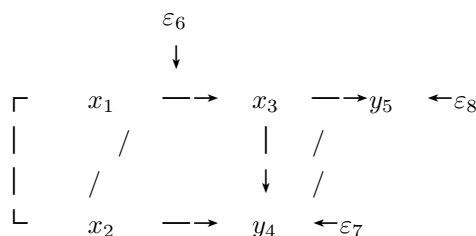


图12.1 递归结构方程模型

变量含义: x_1 -父亲的教育, x_2 -父亲的职业, y_3 -子女的教育, y_4 -子女的第一个工作, y_5 -子女在1962年的职业。

$$\begin{aligned}
 y_3 &= \gamma_{31}x_1 + \gamma_{32}x_2 + \varepsilon_6 \\
 y_4 &= \beta_{43}y_3 + \gamma_{42}x_2 + \varepsilon_7 \\
 y_5 &= \beta_{53}y_3 + \beta_{54}y_4 + \gamma_{52}x_2 + \varepsilon_8
 \end{aligned}$$

比较一般模型 $y = By + \Gamma x + E$, 有:

$$B = \begin{pmatrix} 0 & 0 & 0 \\ \beta_{43} & 0 & 0 \\ \beta_{53} & \beta_{54} & 0 \end{pmatrix} \quad \Gamma = \begin{pmatrix} \gamma_{31} & \gamma_{32} \\ 0 & \gamma_{42} \\ 0 & \gamma_{52} \end{pmatrix} \quad E = \text{Diag}(\varepsilon_6 \quad \varepsilon_7 \quad \varepsilon_8)$$

现在, 观察相关矩阵为:

$$\begin{array}{cccccc}
 1.000 & & & & & \\
 0.516 & 1.000 & & & & \\
 0.453 & 0.438 & 1.000 & & & \\
 0.332 & 0.417 & 0.538 & 1.000 & & \\
 0.322 & 0.405 & 0.596 & 0.541 & 1.000 &
 \end{array}$$

极大似然法结果:

$$\begin{aligned}
 y_3 &= 0.309x_1 + 0.278x_2 + 0.738\varepsilon_6 \\
 y_4 &= 0.224x_2 + 0.440y_3 + 0.670\varepsilon_7 \\
 y_5 &= 0.115x_2 + 0.395y_4 + 0.281y_4 + 0.566\varepsilon_8
 \end{aligned}$$

Duncan、Haller、Portes 的peer influence 模型[2,3]. 内生变量具有多个示性量并假设外生变量度量无误差, 相互作用关系图示如下:

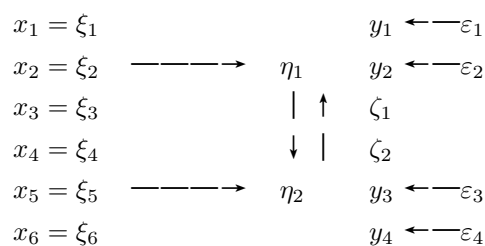


图12.2 含有隐含变量(或伪变量、哑变量)的模型

变量含义:

x_1 -应答者父母的抱负, x_2 -应答者的智力

x_3 -应答者的社会经济状况

x_4 -最好朋友的社会经济状况, x_5 -最好朋友的智力

x_6 -最好朋友父母的就业愿望

y_1 -应答者的就业愿望, y_2 -应答者的教育愿望

y_3 -最好朋友的智力教育愿望, y_4 -最好朋友的就业愿望

η_1 -应答者的雄心, η_2 -最好朋友的雄心

外生量的关系为: $x = \xi$ (即 $\Lambda_x = I_6, \delta = 0$), $\Phi = \Sigma_{xx}, \theta = 0$

y 测量子模型是:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ 0 & 1 \\ 0 & \lambda_{42} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

$\Theta_{\varepsilon 4 \times 4} = \text{diag}[\theta_{11} \quad \theta_{22} \quad \theta_{33} \quad \theta_{44}]$ 即内生变量的度量误差不相关。

结构子模型为:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & 0 & 0 \\ 0 & 0 & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

现在, $[yx]' \sim N(0, \Sigma)$, 模型求解的结果是:

$$\hat{\Lambda}_y^* = \begin{pmatrix} .7667 & 0 \\ .8148 & 0 \\ 0 & .8299 \\ 0 & .7716 \end{pmatrix} \quad \hat{\Theta}_{\varepsilon(4 \times 4)}^* = \text{diag} \begin{bmatrix} .4121 & .3361 & .3112 & .4046 \\ (.0512) & (.0521) & (.0459) & (.0462) \end{bmatrix}$$

$$\hat{B}_y^* = \begin{pmatrix} 0 & -0.1994 \\ & (.1027) \\ -0.2176 & 0 \\ (.1103) \end{pmatrix} \quad \hat{\Gamma}^* = \begin{pmatrix} .2103 & .3256 & .2848 & .0937 & 0 & 0 \\ (.0506) & (.0574) & (.0576) & (.0648) & & \\ 0 & 0 & .0746 & .2758 & .4205 & .1922 \\ & & (.0624) & (.0532) & (.0546) & (.0468) \end{pmatrix}$$

可见 η_1 与 η_2 的许多解是可比的。

Klein 关于美国经济的 I 类模型是一个经典计量经济学模型, LISREL 讨论了它的实现方法, 其它的如Greene, W.H.(1990) 也做了介绍, 它是根据两次世界大战间美国每年的经济数据而构造的八个方程的方程组。数据是一个时间序列, 所以模型是动态的。模型有三个行为方程:

$$\begin{array}{lll} C_t = & a_1 P_t + a_2 P_{t-1} + a_3 W_t & \text{集团消费} \\ I_t = & b_1 P_t + b_2 P_{t-1} + b_3 K_{t-1} & \text{净投资} \\ W_{t*} = & c_1 E_t + c_2 E_{t-1} + c_3 A_t & \text{个人工资} \\ \text{五个量} & P_t = Y_t - W_t & \text{总利润} \\ & Y_t = C_t + I_t + G_t - T_t & \text{总收入} \\ & K_t = K_{t-1} + I_t & \text{年末资本存量} \\ & W_t = W_{t*} + W_{t**} & \text{总工资} \\ & E_t = Y_t + T_t - W_{t**} & \text{私有工业总产值} \end{array}$$

左边是内生变量, 含义在式子右端给出。 $C_t, I_t, W_t, P_t, Y_t, K_t, W_t, E_t$ 分别记做 $y_1 - y_8$ 。预先定义的量是外生变量。 W_{t**} = 政府工资, T_t = 税收, G_t = 政府非工资支出, A_t = 1931年以来所过年数。衍生内生变量是 $P_{t-1}, K_{t-1}, E_{t-1}$, 分别记作 x_5, x_6, x_7 。Klein 使用了1921-1941年间的数据进行分析, 该模型常用于检验计量经济学估计量。

B, Γ 及 Ψ 矩阵分别是:

$$\begin{array}{c}
 C_t \quad I_t \quad W_{t*} \quad P_t \quad Y_t \quad K_t \quad W_t \quad E_t \\
 C_t \quad I_t \quad W_{t*} \quad P_{t*} \quad Y_t \quad K_t \quad W_t \quad E_t \\
 \left(\begin{array}{cccccccc}
 0 & 0 & 0 & a_1 & 0 & 0 & a_3 & 0 \\
 0 & 0 & 0 & b_1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 \\
 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0
 \end{array} \right)
 \end{array}$$

及

$$\left(\begin{array}{cccccccc}
 i_{11} & & & & & & & \\
 i_{21} & i_{22} & & & & & & \\
 i_{31} & i_{32} & i_{33} & & & & & \\
 0 & 0 & 0 & 0 & & & & \\
 0 & 0 & 0 & 0 & 0 & & & \\
 0 & 0 & 0 & 0 & 0 & 0 & & \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{array} \right)$$

SAS/STAT 的PROC CALIS; SPSS LISREL; BMDP EQS; 1989 年版的SYSTAT EzPATH 均可用于结构方程建模。SAS 技术报告[7]详细解释了LISREL、EQS 及COSAN 等的联系, 并配有实例。第4 章介绍了CALIS 用于计量经济学分析的例子。SAS/ETS PROC SYSLIN 等也常用于计量经济学模型分析。

附Klein.dat 文件内容:

```

(15A4)          INPUT LABELS FOR KLEIN'S MODEL
      C P-1  W*   I K-1 E-1 W**   T   A   P   K   E   W   Y   G
(15F5.1)        INPUT DATA FOR KLEIN'S MODEL
41.9 12.7 25.5 -0.2182.8 44.9  2.7  7.7-10.0 12.4182.6 45.6 28.2 40.6  6.6
45.0 12.4 29.3  1.9182.6 45.6  2.9  3.9 -9.0 16.9184.5 50.1 32.2 49.1  6.1
49.2 16.9 34.1  5.2184.5 50.1  2.9  4.7 -8.0 18.4189.7 57.2 37.0 55.4  5.7
50.6 18.4 33.9  3.0189.7 57.2  3.1  3.8 -7.0 19.4192.7 57.1 37.0 56.4  6.6
52.6 19.4 35.4  5.1192.7 57.1  3.2  5.5 -6.0 20.1197.8 61.0 38.6 58.7  6.5
55.1 20.1 37.4  5.6197.8 61.0  3.3  7.0 -5.0 19.6203.4 64.0 40.7 60.3  6.6
56.2 19.6 37.9  4.2203.4 64.0  3.6  6.7 -4.0 19.8207.6 64.4 41.5 61.3  7.6
57.3 19.8 39.2  3.0207.6 64.4  3.7  4.2 -3.0 21.1210.6 64.5 42.9 64.0  7.9
57.8 21.1 41.3  5.1210.6 64.5  4.0  4.0 -2.0 21.7215.7 67.0 45.3 67.0  8.1
55.0 21.7 37.9  1.0215.7 67.0  4.2  7.7 -1.0 15.6216.7 61.2 42.1 57.7  9.4
50.9 15.6 34.5 -3.4216.7 61.2  4.8  7.5  0.0 11.4213.3 53.4 39.3 50.7 10.7
45.6 11.4 29.0 -6.2213.3 53.4  5.3  8.3  1.0  7.0207.1 44.3 34.3 41.3 10.2
46.5  7.0 28.5 -5.1207.1 44.3  5.6  5.4  2.0 11.2202.0 45.1 34.1 45.3  9.3

```

48.7	11.2	30.6	-3.02	02.0	45.1	6.0	6.8	3.0	12.3	199.0	49.7	36.6	48.9	10.0
51.3	12.3	33.2	-1.3	199.0	49.7	6.1	7.2	4.0	14.0	197.7	54.4	39.3	53.3	10.5
57.7	14.0	36.8	2.1	197.7	54.4	7.4	8.3	5.0	17.6	199.8	62.7	44.2	61.8	10.3
58.7	17.6	41.0	2.0	199.8	62.7	6.7	6.7	6.0	17.3	201.8	65.0	47.7	65.0	11.0
57.5	17.3	38.2	-1.9	201.8	65.0	7.7	7.4	7.0	15.3	199.9	60.9	45.9	61.2	13.0
61.6	15.3	41.6	1.3	199.9	60.9	7.8	8.9	8.0	19.0	201.2	69.5	49.4	68.4	14.4
65.0	19.0	45.0	3.3	201.2	69.5	8.0	9.6	9.0	21.1	204.5	75.7	53.0	74.1	15.4
69.7	21.1	53.3	4.9	204.5	75.7	8.5	11.6	10.0	23.5	209.4	88.4	61.8	85.3	22.3

