第十章　Panel Data模型

经济分析时，经常会遇到时间序列和横截面两者相结合的数据。

例如，企业投资需求分析中，会遇到“多个企业的若干指标的不同月度或季度时间序列”；在城镇居民消费分析中，会遇到不同省市地区的反映居民消费和居民收入的年度时间序列。

本章将前述的企业或地区等统称为个体，这种具有“三维（指标 x、个体 i、时间 t）信息的数据结构”称为“**面板数据**”（panel data），也可称为平行数据。

本章将利用面板数据的计量模型，简称为Panel Data 模型。

经典线性计量经济学模型，在分析时只利用了面板数据中的某些二维数据信息；例如，使用若干经济指标的时间序列建模或利用横截面数据建模。

然而，实际经济分析中，仅利用二维信息的模型在很多时候往往不能满足分析问题的需要。

例如，在生产函数分析中：

仅利用横截面数据，只能对规模经济进行分析；

仅利用混有规模经济和技术革新信息的时间序列数据，只有在假设规模收益不变的条件下、才能实现技术革新的分析；

而利用面板数据，可同时分析企业的规模经济（选择同一时期的不同规模的企业数据作为样本观测值）和技术革新（选择同一企业不同时期的数据作为样本观测值），可实现规模经济和技术革新的综合分析。

面板数据含有“ **指标 *x*、横截面 *i*、时间 *t*”三维信息**，利用面板数据模型可以构造和检验比以往单独使用横截面数据或时间序列数据更为真实的行为方程，可进行更加深入的分析。

正是基于实际经济分析的需要，作为非经典计量经济学问题，同时利用横截面和时间序列数据的模型已经成为近年来计量经济学理论方法的重要发展之一。

10.1 Panel Data模型概述

EViews对Panel Data模型的估计，是通过含有Pool对象的工作文件和具有面板结构的工作文件来实现的。

处理面板数据的EViews对象，称为**Pool**。通过Pool对象可实现对各种变截距、变系数时间序列模型的估计，但Pool对象侧重分析“截面成员较少、而时期较长的数据”[即侧重时间序列分析的“窄而长”的数据，故仍在“[时序结构Dated]的workfile”中定义Pool对象（如中日韩，金砖五国）]，。

对截面成员较多、时期较少的[侧重截面分析的“宽而短”的数据]，一般通过 直接含有“面板结构的工作文件”（**Panel workfile**）进行分析。利用面板结构的工作文件，可实现变截距Panel Data模型以及动态Panel Data模型的估计。

10.1.1 含有Pool对象的工作文件

Pool对象在EViews中扮演着两种角色。

首先，Pool对象中包含了一系列的标识名。这些标识名描述了工作文件中的面板数据的数据结构。在这个角色中，Pool对象在管理和处理面板数据上的功能、与组对象有些相似。

其次，利用Pool对象中的过程，可实现对各种Panel Data模型的估计及对估计结果的检验和处理。在这个角色中，Pool对象、与方程对象有些相似 。

创建Pool对象的核心，是建立表示截面成员的名称表。为简便起见，名称要相对较短。例如，国家作为截面成员时，可使用USA代表美国，CAN代表加拿大，UK代表英国。

定义了Pool的截面成员名称，就等于告诉了EViews模型的数据结构。在上面的例子中，EViews会自动把这个Pool理解成对每个国家使用单独的时间序列。

必须注意，Pool对象本身不包含序列或数据。一个Pool对象只是对基本数据结构的一种描述。

因此，删除一个Pool、并不会同时删除它所使用的序列，但修改Pool使用的原序列、会同时改变Pool中的数据。

1. 创建Pool对象

本章中，使用的是一个研究投资需求的例子，包括了“三个变量、[五家企业、20个年度观测值]”的时间序列：

例10.4　研究汽车制造商投资需求模型

5家企业：

GM：通用汽车公司

CH：克莱斯勒公司

GE：通用电器公司

WE：西屋公司

US：美国钢铁公司

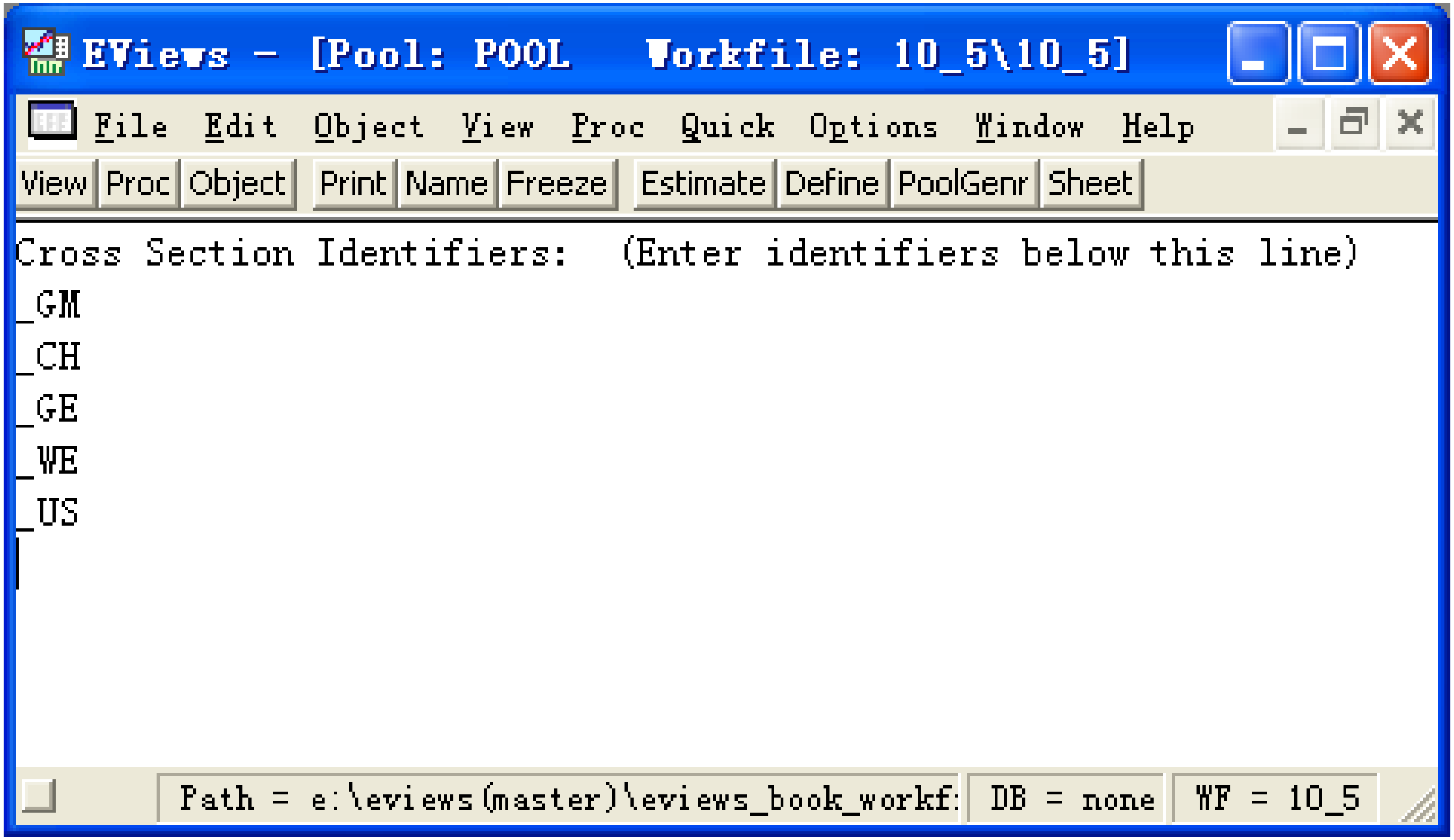
3个变量：

I：总投资

M：前一年企业的市场价值（反映企业的预期利润）

K：前一年末工厂存货和设备的价值（反映企业必要重置投资期望值）

创建Pool对象，选择Objects/New Object/Pool…并在编辑窗口中输入截面成员的识别名称：



对截面成员的识别名称没有特别要求，但必须能使用这些识别名称建立合法的EViews序列名称。

此处推荐*在每个识别名中使用“_”字符*，它不是必须的，但把它作为序列名的一部分，可很容易找到识别名称。

2. Pool序列命名

Pool中使用序列的关键，是序列组合命名：

"基本名（指标）**+ 截面识别名称**（个体）**"**

例如，现有一个Pool对象含有识别名 __JPN， __USA，__UK，想建立每个截面成员的GDP的时间序列，我们就使用"GDP"作为序列的基本名。

可把识别名称、放在基本名的后面，此时序列名为：GDP__JPN，GDP__USA，GDP__UK；

或者把识别名称、放在基本名的前面，此时序列名为：JPN__GDP，USA__GDP，UK__GDP。

把识别名称放在序列名的前面，中间或后面并没什么关系，只要易于识别就行了——但是必须注意要保持一致。

不能这样命名序列：JPNGDP，GDPUSA，UKGDP1，因为EViews无法在Pool对象中识别这些序列（变量名与个体名，混在一起了！）。

3. Pool序列概念

一个Pool序列，实际就是**“一组序列”**：其序列名，是由基本指标名、和所有截面识别名构成的。

Pool序列名，使用 [指标名+“?”占位符]，其中“?”代表截面识别名：如果序列名为GDPJPN，GDPUSA，GDPUK，相应的Pool序列为[GDP?]；

如果序列名为JPNGDP，USAGDP，UKGDP，相应的Pool序列为[?GDP]。

当使用一个Pool序列名时，EViews认为将使用Pool序列中的所有序列。EViews会自动循环查找所有截面识别名称、并用识别名称替代“?”。然后，会按指令使用这些替代后的名称了。

Pool序列、必须通过Pool对象来定义，因为如果没有截面识别名称，占位符“?”就没有意义。

4．观察或编辑Pool定义

要显示Pool中的截面成员识别名称，单击工具条的Define按钮，或选择View/Cross-Section Identifiers。如果需要，也可对识别名称列进行编辑。

5．Pool序列数据

Pool中使用的数据，都存在普通EViews序列中。

这些序列，可按通常方式使用：可列表显示，图形显示，产生新序列，或用于估计。也可使用Pool对象来处理各单独序列。

10.1.2 输入Pool数据

有很多种输入数据的方法，在介绍各种方法之前，首先要理解面板数据的结构，区别堆积数据和非堆积数据形式。

面板数据的数据信息用三维表示：时期，截面成员，变量。例如：1950年，通用汽车公司，投资数据。

使用三维数据比较困难，一般要转化成二维数据。有几种常用的方法。

1. 非堆积数据

存在工作文件的数据都是这种非堆积数据，在这种形式中，给定截面成员、给定变量的观测值放在一起，但和其他变量、其他截面成员的数据分开。例如，假定我们的数据文件为下面的形式：

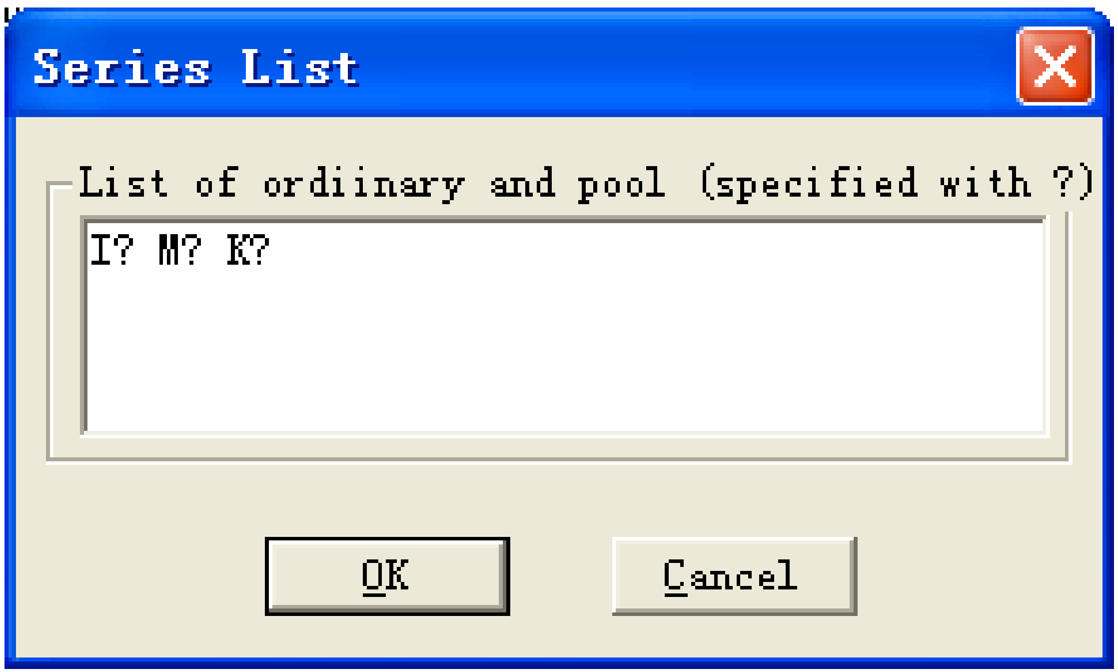
obs	I_GM	I_CH	I_GE	I_WE	I_US	M_GM	M_CH
1935	317.6000	40.29000	33.10000	12.93000	209.9000	3078.500	417.5000
1936	391.8000	72.76000	45.00000	25.90000	355.3000	4661.700	837.8000
1937	410.6000	66.26000	77.20000	35.05000	469.9000	5387.100	883.9000
1938	257.7000	51.60000	44.60000	22.89000	262.3000	2792.200	437.9000
1939	330.8000	52.41000	48.10000	18.84000	230.4000	4313.200	679.7000
1940	461.2000	69.41000	74.40000	28.57000	261.6000	4643.900	727.8000
1941	512.0000	68.35000	113.0000	48.51000	472.8000	4551.200	643.6000
1942	448.0000	46.80000	91.90000	43.34000	445.6000	3244.100	410.9000
1943	499.6000	47.40000	61.30000	37.02000	361.6000	4053.700	588.4000
1944	547.5000	59.57000	56.80000	37.81000	288.2000	4379.300	698.4000
1945	561.2000	88.78000	93.60000	39.27000	258.7000	4840.900	846.4000
1946	688.1000	74.12000	159.9000	53.46000	420.3000	4900.900	893.8000
1947	568.9000	62.68000	147.2000	55.56000	420.5000	3526.500	579.0000
1948	529.2000	89.36000	146.3000	49.56000	494.5000	3254.700	694.6000
1949	555.1000	78.98000	98.30000	32.04000	405.1000	3700.200	590.3000
1950	642.9000	100.6600	93.50000	32.24000	418.8000	3755.600	693.5000
1951	755.9000	160.6200	135.2000	54.38000	588.2000	4833.000	809.0000
1952	891.2000	145.0000	157.3000	71.78000	645.2000	4924.900	727.0000
1953	1304.400	174.9300	179.5000	90.08000	641.0000	6241.700	1001.500
1954	1486.700	172.4900	189.6000	68.60000	459.3000	5593.600	703.2000

其中基本名 I 代表企业总投资、M 代表前一年企业的市场价值、K 代表前一年末工厂存货和设备的价值。每个企业，都有“单独的 I、M、K 数据”。

EViews会自动按附录A中 介绍的标准输入程序，读取非堆积数据。并把每个截面变量看作一个单独序列。注意要按照上述的Pool命名规则命名。

2. 堆积数据**

选择View/Spreadsheet（stacked data），EViews会要求输入序列名列表



确认后EViews会打开新建序列的堆积式数据表。我们看到的是按截面成员堆积的序列，Pool序列名在每列表头，截面成员/年代识别符标识每行：

Pool数据排列成堆积形式，一个变量的所有数据放在一起、和其他变量的数据分开。

大多数情况下，不同截面成员的数据从上到下依次堆积，**每一列、代表一个变量**：

Pool: POOL Workfile: 10_5\10_5

View | Proc | Object | Print | Name | Freeze | Edit+/- | Order+/- | Smpl+/- | Fc

obs	I?	M?	K?
_GM-1935	317.6000	3078.500	2.800000
_GM-1936	391.8000	4661.700	52.60000
_GM-1937	410.6000	5387.100	156.9000
_GM-1938	257.7000	2792.200	209.2000
_GM-1939	330.8000	4313.200	203.4000
_GM-1940	461.2000	4643.900	207.2000
_GM-1941	512.0000	4551.200	255.2000
_GM-1942	448.0000	3244.100	303.7000
_GM-1943	499.6000	4053.700	264.1000
_GM-1944	547.5000	4379.300	201.6000
_GM-1945	561.2000	4840.900	265.0000
_GM-1946	688.1000	4900.900	402.2000
_GM-1947	568.9000	3526.500	761.5000
_GM-1948	529.2000	3254.700	922.4000
_GM-1949	555.1000	3700.200	1020.100
_GM-1950	642.9000	3755.600	1099.000
_GM-1951	755.9000	4833.000	1207.700
_GM-1952	891.2000	4924.900	1430.500
_GM-1953	1304.400	6241.700	1777.300
_GM-1954	1486.700	5593.600	2226.300
_CH-1935	40.29000	417.5000	10.50000
_CH-1936	72.76000	837.8000	10.20000
_CH-1937	66.26000	883.9000	34.70000
_CH-1938	51.60000	437.9000	51.80000
_CH-1939	52.41000	679.7000	64.30000

我们称上表数据是<u>以截面成员堆积的</u>，单击Order+/-实现堆积方式转换，也可<u>以按日期堆积数据</u>：

Pool: POOL Workfile: 10_5\10_5

View | Proc | Object | Print | Name | Freeze | Edit+/- | Order+/- | Smpl+/- | Form

obs	I?	M?	K?
_GM-1935	317.6000	3078.500	2.800000
_CH-1935	40.29000	417.5000	10.50000
_GE-1935	33.10000	1170.600	97.80000
_WE-1935	12.93000	191.5000	1.800000
_US-1935	209.9000	1362.400	53.80000
_GM-1936	391.8000	4661.700	52.60000
_CH-1936	72.76000	837.8000	10.20000
_GE-1936	45.00000	2015.800	104.4000
_WE-1936	25.90000	516.0000	0.800000
_US-1936	355.3000	1807.100	50.50000
_GM-1937	410.6000	5387.100	156.9000
_CH-1937	66.26000	883.9000	34.70000
_GE-1937	77.20000	2803.300	118.0000
_WE-1937	35.05000	729.0000	7.400000
_US-1937	469.9000	2676.300	118.1000
_GM-1938	257.7000	2792.200	209.2000
_CH-1938	51.60000	437.9000	51.80000

<u>每一列、代表一个变量</u>，每一列内数据都是按年排列的。

若数据按年排列，要确保各年内截面成员的排列顺序要一致。

3. 手工输入/剪切和粘贴

可通过手工输入数据，也可使用剪切和粘贴工具输入：

(1) 通过确定工作文件样本，来指定堆积数据表中要包含哪些时间序列观测值。

(2) 打开Pool，选择View/Spreadsheet（stacked data），EViews会要求输入序列名列表，可以输入普通序列名或Pool序列名。如果是已有序列，EViews会显示序列数据；如果这个序列不存在，EViews会使用已说明的Pool序列的截面成员识别名称建立新序列或序列组。

(3) 打开Pool序列的堆积式数据表。需要的话，还可单击Order+/-按钮、进行按截面成员堆积和按日期堆积之间的转换。

(4) 单击Edit+/-按钮，打开数据编辑模式输入数据。

如果有一个Pool包含识别名 _CM，_CH，_GE，_WE，_US，通过输入：I? M? K?，指示EViews来创建如下序列：I_CM，I_CH，I_GE，I_WE，I_US；M_CM，M_CH，M_GE，M_WE，M_US；K_CM，K_CH，K_GE，K_WE，K_US。

4. 文件输入（略讲）

可使用Pool对象从文件输入堆积数据到各单独序列。当文件数据按截面成员或时期堆积成时，EViews要求：

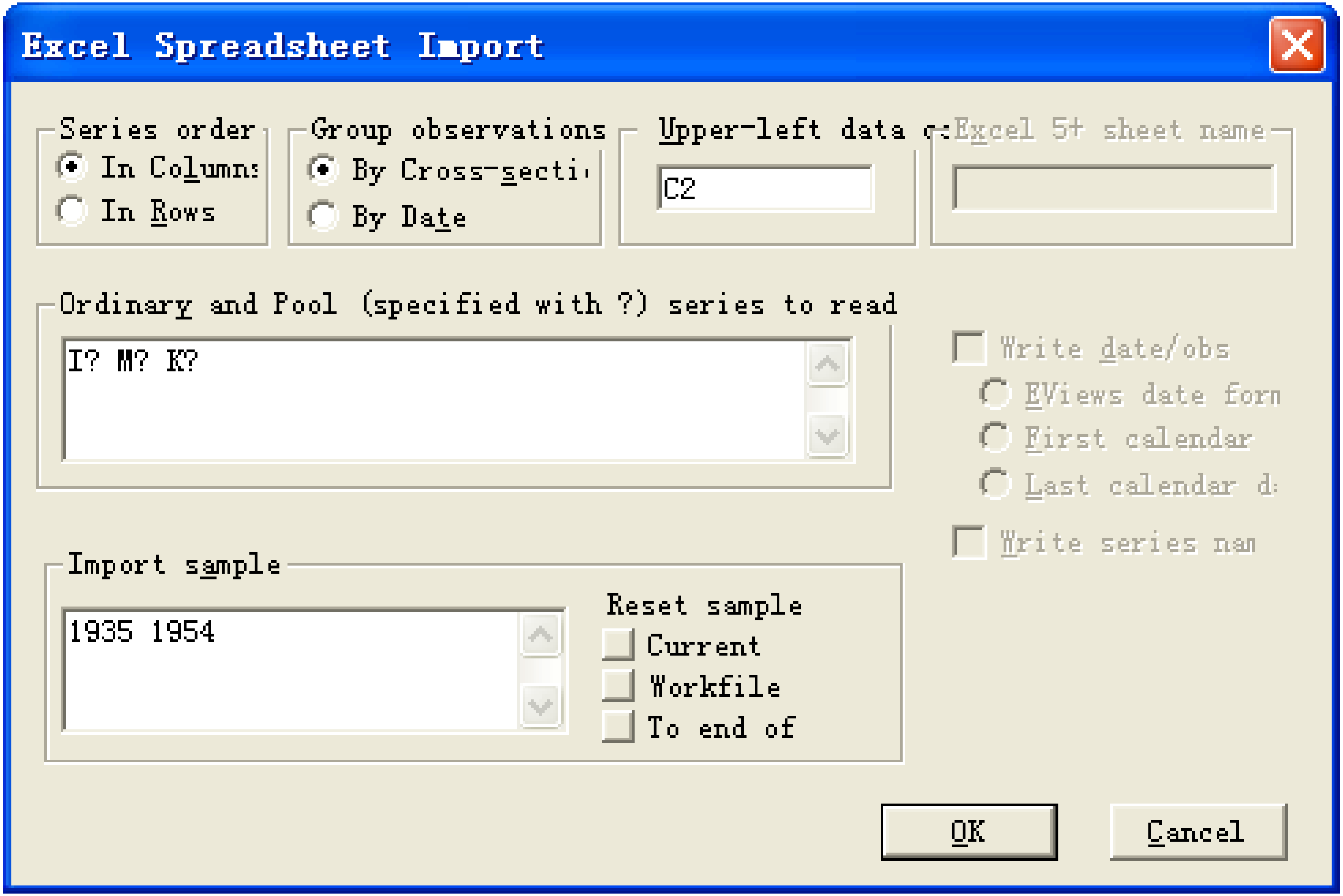
(1) 堆积数据是平衡的；

(2) 截面成员在文件中、和在Pool中的排列顺序相同。

平衡数据的意思是：若按截面成员堆积数据，每个截面成员、应包括正好相同的时期；若按日期堆积数据，每个日期、应包含相同数量的截面成员观测值，并按相同顺序排列。

特别要指出的是，基础数据并不一定是平衡的，只要在输入文件中有表示即可。如果观测值中有缺失数据，一定要保证文件中给这些缺失值留有位置。

要使用Pool对象从文件读取数据，先打开Pool，然后选择Procs/Import Pool Data(ASCII,.XLS,.WK?)…，要使用与Pool对象对应的输入程序。


Excel Spreadsheet Import
Series order
In Columns
In Rows
Group observations
By Cross-secti
By Date
Upper-left data c
C2
Excel 5+ sheet name
Ordinary and Pool (specified with ?) series to read
I? M? K?
Write date/obs
EViews date form
First calendar
Last calendar d
Write series nam
Import sample
1935 1954
Reset sample
Current
Workfile
To end of
OK
Cancel

通过附录A的学习，大家对这个对话框应该比较熟悉，填写说明如下：

注明Pool序列是按行还是按列排列，数据是按截面成员堆积还是按日期堆积。

在编辑框输入序列的名称。这些序列名应该是普通序列名或者是Pool名。

填入样本信息，起始格位置和表单名（可选项）。

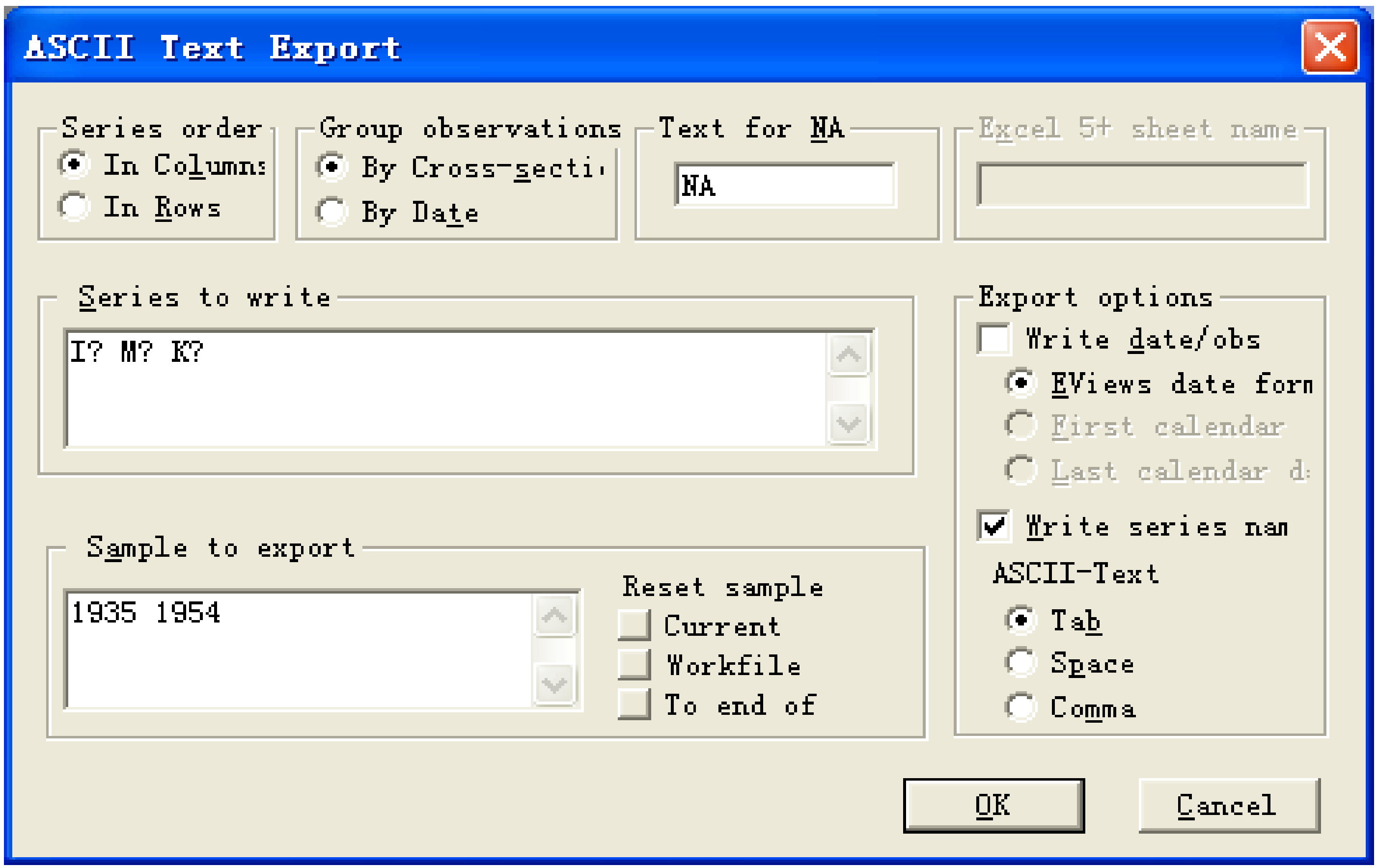
如果输入序列用Pool序列名，EViews会用截面成员识别名创建和命名序列。如果用普通序列名，EViews会创建单个序列。

EViews会使用样本信息读入文件到说明变量中。如果输入的是普通序列名，EViews会把多个数据值输入到序列中，直到从文件中读入的最后一组数据。

10.1.3 输出Pool数据

按照和上面数据输入相反的程序，可进行数据输出。

由于EViews可输入/输出“**非堆积数据、**按截面成员堆积数据、按日期堆积数据”，因此可利用EViews按照需要调整数据结构。



10.1.4　使用Pool数据

每个截面成员的基础序列都是普通序列，因此EViews中对各单个截面成员序列适用的工具都可使用。

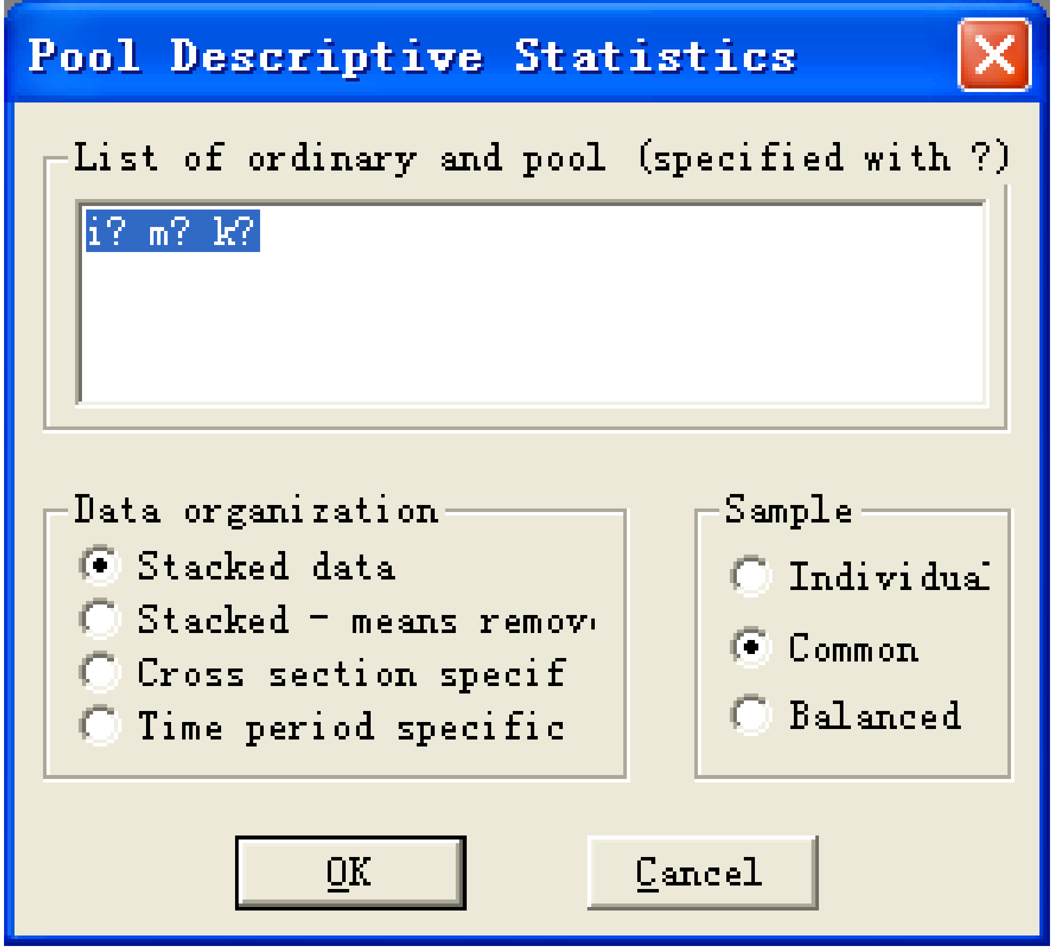
另外，EViews还有专门**适用于Pool数据的专用工具**。可以使用EViews对与一特定变量对应的所有序列进行类似操作。

1. 检查数据

用数据表形式查看堆积数据。选择View/Spreadsheet (stacked data)，列出要显示的序列。序列名包括普通序列名和Pool序列名。

2. 描述数据

可以使用Pool对象计算序列的描述统计量。在Pool工具栏选择View/Descriptive Statistics…，EViews会打开如下对话框：



在编辑框内，输入计算描述统计量的序列：EViews可计算序列的平均值，中位数，最小值，最大值，标准差，偏度，峰度，和Jarque-Bera统计量。

下一步选择样本选项：

（1）Individual（单独的）：利用所有的有效观测值，即使某一变量的观测值仅是针对某一截面成员的（非平衡数据），也计算在内。

（2）Common（**截面共同的**）：使用的有效观测值，必须是某一截面成员的数据、在同一期“对所有变量 j ”都有数值。而不管同期其他截面成员的变量，是否有值。

（3）Balanced（平衡的）：使用的有效观测值，必须是对所有截面成员 i 、所有变量 j 在同一期 t 都有数值。

最后，还必须选择与计算方法相对应的数据结构：

（1）堆积数据（Stacked data）：计算表中每一变量所有截面成员，所有时期的统计量。如果忽略数据的pool性质，得到的就是变量的描述统计量。

（2）去掉均值的堆积数据（Stacked-means removed）：计算除去截面平均值之后的描述统计量值。

（3）截面成员变量（Cross-section specific）：计算每个截面变量所有时期的描述统计量。是通过对各单独序列计算统计量而得到的。

（4）时期变量（Time period specific）: 计算时期特性描述统计量。对每一时期，使用pool中所有截面成员的变量数据计算的统计量。

注意，后两种方法可能产生很多输出结果。如截面成员描述计算，会对每一变量/截面成员组合产生一系列结果：如果有三个Pool变量，20个截面成员，EViews就会计算60个序列的描述统计量。

3. 生成数据***

(1) 使用PoolGenr (panelgenr) 程序、生成或者修改Pool序列。点击Pool工具栏的Poolgenr并输入要生成的方程式，使用正确的Pool名称。例如上面的例子，输入：

r? = I?/I_US

相当于输入下面五个命令：

r_CM = I_CM/I_US

r_CH = I_CH/I_US

r_GE = I_GE/I_US

r_WE = I_WE/I_US

r_US = I_US/I_US

PoolGenr按照输入的方程、在各截面成员间进行循环计算，生成新的序列或修改已有序列。

(2) 可联合使用PoolGenr和Genr，生成新的变量。

例如，要生成一个虚拟变量，在美国钢铁（US）时取1，其他企业时取0：

先选择PoolGenr，然后输入：dum?=0，从而初始化所有虚拟变量序列为0。

然后，把US值设置为1，在主菜单选择Genr，然后输入：dum_US=1。

(3) 还可利用数据的内在循环特性，进行给定时期的截面成员间的计算。

例如，建立一普通序列总投资IS，在主菜单选择Genr，然后输入：IS=0，即初始值设为0，然后选PoolGenr并输入：

IS=IS+ I?

相当于对普通序列从Genr输入下列计算：

IS= I_GM+I_CH+I_GE+I_WE+I_US

这个例子，用来说明内在循环这个概念。

4. 生成Pool组（略）

如果希望使用EViews的组对象工具处理一系列Pool序列，选择Procs/Make Group…输入普通序列和Pool序列名称，EViews就会生成一个包含这些序列的未命名组对象。

5. 删除和存取数据

Pool可用来删除和存取序列。

只需选择Procs/Delete pool series…，Procs/Store pool series(DB)…，Procs/Fetch pool series(DB)…，输入普通序列和Pool序列名称即可。

10.1.5　Pool对象估计的模型形式

设有单一因变量 y_{it} 与$k\times1$维解释变量向量$x_{it}=(x_{1,it}, x_{2,it}, \ldots, x_{k,it})'$，满足线性关系：

$$y_{it} = \alpha_{it} + \boldsymbol{x}'_{it}\boldsymbol{\beta}_{it} + u_{it}$$

$i=1, 2, \ldots, N$，　　$t=1, 2, \ldots, T$　　　(10.1.1)

式(10.1.1)是考虑 k 个经济指标在 N 个个体及 T 个时间点上的变动关系。其中 N 表示个体截面成员的个数，T 表示每个截面成员的观测时期总数，参数 α_{it} 表示模型的常数项，β_{it} 表示对应于解释变量向量 x_{it} 的 $k\times1$ 维系数向量，k 表示解释变量个数。

随机误差项 $\boldsymbol{u_{it}}$ 相互独立，且满足零均值、等方差为 σ_u^2 的假设。

在式(10.1.1)描述的模型中，自由度（NT）、远远小于参数个数，这使得模型无法估计。

为实现模型的估计，可分别建立以下两类模型：

1）从个体成员角度考虑，建立含有 N 个个体成员方程的Panel Data模型；

2）在时间点上截面，建立含有 T 个时间点截面方程的Panel Data模型。

1．含有“N个个体”成员方程的Panel Data模型

Panel Data模型简化为如下形式：

$$y_i = \alpha_i + x_i \beta_i + u_i \qquad i=1, 2, \ldots, N \qquad (10.1.2)$$

其中：$\boldsymbol{y}_i$是$T \times 1$维被解释变量向量，$\boldsymbol{x}_i$是$T \times k$维解释变量矩阵，y_i和x_i的各分量是个体成员的经济指标时间序列，例如若个体成员代表各不同地区，则$\boldsymbol{y}_i$和$\boldsymbol{x}_i$代表i地区的消费、收入/物价等指标的经济时间序列。

截距项α_i和$k \times 1$维系数向量$\boldsymbol{\beta}_i$，其取值受不同个体i的影响、但不随时期t的变化而变化。u_i是$T \times 1$维扰动项向量，满足均值为零、方差为σ_u^2的假设。

式（10.1.2）写成矩阵的回归形式为：

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{x}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{x}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_N \end{pmatrix} \quad (10.1.3)$$

式（10.1.3）含有N个截面方程。

2．含有“T 期时间”截面方程的Panel Data模型

Panel Data 模型简化为如下形式：

$$\mathbf{y}_t = \mu_t + \mathbf{x}_t \boldsymbol{\gamma}_t + \mathbf{v}_t$$

$$t = 1, 2, \ldots, T \qquad (10.1.4)$$

其中：y_t 是$N\times 1$维被解释变量向量，x_t 是$N\times k$维解释变量矩阵，y_t 和 x_t 的各分量、是对应于某个时间点 t 的各个体成员的经济指标序列，例如2003年各不同地区的消费、收入/物价等的经济指标序列。

截距项μ_t和$k\times 1$维系数向量 γ_t，其取值受不同时期 t 的影响、但不随个体观测对象 i 的变化而变化。v_t 是$N\times 1$维扰动项向量，满足均值为零、方差为σ_v^2的假设。

式（10.1.4）写成矩阵的回归形式为：

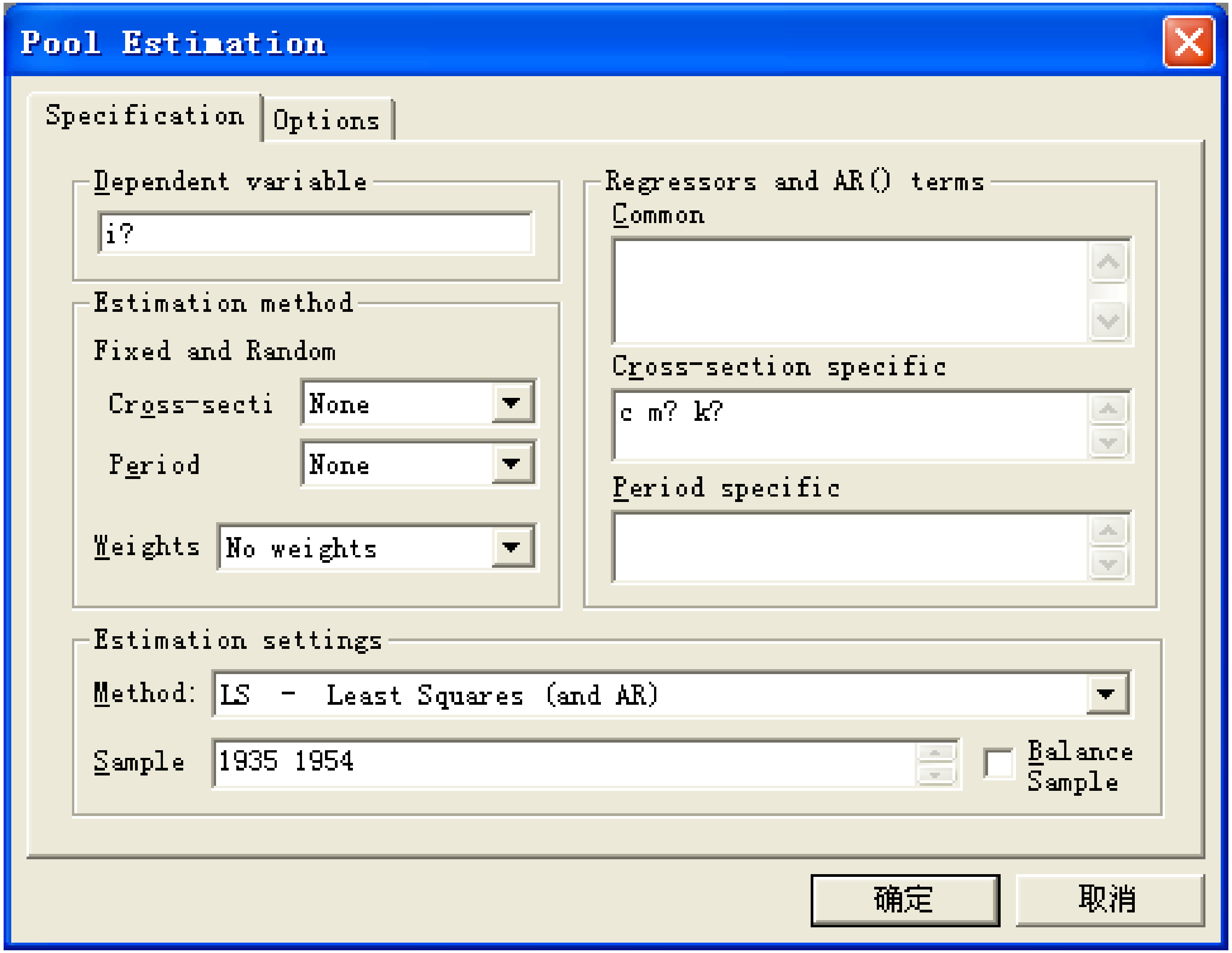
$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_T \end{pmatrix} + \begin{pmatrix} \boldsymbol{x}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{x}_T \end{pmatrix} \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \vdots \\ \boldsymbol{\gamma}_T \end{pmatrix} + \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_T \end{pmatrix} \quad (10.1.5)$$

式(10.1.5)含有T个时间截面方程。

由于含有 N 个个体成员方程的式（10.1.2）和含有 T个时间截面方程的式（10.1.4）两种形式的模型、在估计方法上类似，因此本章主要讨论含有 N 个个体成员方程的Panel Data模型的估计方法。

10.1.6 如何估计Pool方程

单击Pool工具栏的Estimate选项打开如下对话框：



1. 因变量

在因变量对话框中，输入Pool变量或Pool变量表达式。

2. 样本

在下面的编辑窗口中输入样本说明。样本的缺省值是各截面成员中的最大样本值。如果得不到某时期截面成员的解释变量或因变量的值，那么此观测值会被排除掉。

复选框**Balanced Sample**，可在各截面成员间进行数据排除。只要某一时期数据对任何一个截面成员无效，此时期就被排除。这种排除保证得到的样本区间，对所有截面成员都是有效的。

如果某截面成员的所有观测值都没有，那么Pool在进行估计时就排除这个截面成员。同时EViews会在输出中告诉漏掉的截面成员。

3. 解释变量

在三个编辑框中输入解释变量。

(1) Common :——此栏中输入的变量对所有截面成员有相同的系数，并用一般名称或Pool名称输出结果。

(2) Cross-section specific :——此栏中输入的变量对Pool中每个截面成员的系数不同。EViews会对每个截面成员估计不同的系数，并使用截面成员识别名后跟一般序列名，中间用“_”连接进行标签。

(3) Period specific :——此栏中输入的变量对Pool中每个时期的系数不同。EViews会对每个时期估计不同的系数，并使用变量名后跟时期，中间用“_”连接进行标签。

如果在截面系数编辑框中、输入pool变量M?和K?，会输出M? 和 K? 的估计系数：_GM--M_GM，_CH--M_CH，_GE--M_GE，_WE--M_WE，_US--M_US 和 _GM--K_GM，_CH--K_CH，_GE--K_GE，_WE--K_WE，_US--K_US，等等。

如果在**时期系数编辑框**中、输入pool变量M?和K?，会输出M?和K?的各时期的估计系数：C_1935, C_1936 , … , K?_1935, K?_1936, …, M?_1935，M?_1936 , … , 等等。

注意，使用截面成员特定系数法估计模型会生成很多系数，等于Pool中截面成员数和所列变量数的乘积。

10.2 模型形式设定检验**

模型(10.1.2)常用的有如下三种情形：

情形1： $\alpha_i = \alpha_j, \beta_i = \beta_j$ (不变系数模型)

情形2： $\alpha_i \neq \alpha_j, \beta_i = \beta_j$ (变截距模型)

情形3： $\alpha_i \neq \alpha_j, \beta_i \neq \beta_j$ (变参数/变系数模型)

对情形1，在横截面上无个体影响、无结构变化，则普通最小二乘法估计给出了α 和 β 的一致有效估计。相当于将多个时期的截面数据放在一起作为样本数据。

对情形2，称为变截距模型，在横截面上个体影响不同，个体影响表现为模型中“被忽略的反映个体差异的变量”的影响，又分为“固定影响和随机影响”两种情况。

对情形3，称为变系数模型，除了存在个体影响外，在横截面上还存在变化的经济结构，因而结构参数在不同横截面上是不同的。

在对Panel Data模型进行估计时，使用的样本数据包含了个体、指标、时间3个方向上的信息。如果模型形式设定不正确，估计结果将与所要模拟的经济现实偏离甚远。

因此，建立Panel Data模型的第一步——便是检验被解释变量 y_{it} 的参数 α_i 和 β_i，是否对所有个体样本点、或所有时期都是一样的，即检验“样本数据究竟符合上面哪种Panel Data模型”形式，从而避免模型设定的偏差，改进参数估计有效性。

经常使用的检验是协方差分析检验，主要检验如下两个假设：

$$H_1: \quad \beta_1 = \beta_2 = \cdots = \beta_N$$

$$H_2: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_N$$

$$\beta_1 = \beta_2 = \cdots = \beta_N$$

如果接受假设 H_2 则可认为样本数据符合情形1，即模型为不变参数模型，无需进行进一步的检验。

如果拒绝假设H_2，则需进一步检验假设H_1。

若接受H_1，则认为样本数据符合情形2，即模型为变截距模型，反之拒绝H_1，则认为样本数据符合情形3，即模型为变参数模型。

下面介绍假设检验的 F 统计量的计算方法。首先计算情形3(变参数模型)的残差平方和，记为 S_1；情形2(变截距模型)的残差平方和记为 S_2；情形1(不变参数模型)的残差平方和记为 S_3 。计算 F_2 统计量

$$F_2 = \frac{(S_3 - S_1)/[(N-1)(k+1)]}{S_1/(NT - N(k+1))} \sim F[(N-1)(k+1), N(T-k-1)] \tag{10.2.7}$$

首先，在假设 H_2 下、检验统计量 F_2 服从相应自由度下的F分布。若计算所得到的统计量 F_2 的值不小于给定置信度下的相应临界值，则拒绝假设 H_2，继续检验假设 H_1。反之，接受 H_2则认为样本数据符合模型情形1，即不变参数模型。

若H_2被拒绝，再在假设H_1下、检验统计量F_1也服从相应自由度下的F分布，即：

$$F_1 = \frac{(S_2 - S_1)/[(N-1)k]}{S_1/(NT - N(k+1))} \sim F[(N-1)k, N(T-k-1)] \tag{10.2.8}$$

若计算所得到的统计量F_1的值、不小于给定置信度下的相应临界值，则拒绝假设H_1。

如果接受H_1，则认为样本数据符合情形2，即模型为变截距模型，反之拒绝H_1 ，则认为样本数据符合情形3，即模型为变参数模型。

例10.5　研究企业投资需求的固定影响变系数模型

建立一个研究五家企业投资需求状况的Panel Data模型：

$$I_t = \alpha_i + M_t \beta_1 + K_t \beta_2 + u_t$$

$t = 1, 2, \ldots, 20$

其中：企业标识数字从1 ~ 5，分别对应通用汽车（GM）、克莱斯勒（CH）、通用电气（GE）、西屋（WE）和美国钢铁（US）。

被解释变量：$I_t = (I_{1,t}, I_{2,t}, I_{3,t}, I_{4,t}, I_{5,t})$ 分别是5家企业的总投资。

解释变量：$M_t = (M_{1,t}, M_{2,t}, M_{3,t}, M_{4,t}, M_{5,t})$ 分别是5家企业前一年企业市场价值（反映企业的预期利润）；

$K_t = (K_{1,t}, K_{2,t}, K_{3,t}, K_{4,t}, K_{5,t})$ 分别是5家企业前一年末工厂存货及设备价值（反映企业必要重置投资期望值）。

变参数模型：

Dependent Variable: I?
Method: Pooled Least Squares
Date: 11/21/07 Time: 10:09
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100
White cross-section standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
_GM--C	-149.7825	97.26706	-1.539909	0.1273
_CH--C	-6.189961	10.41058	-0.594583	0.5537
_GE--C	-9.956306	21.67945	-0.459251	0.6472
_WE--C	-0.509390	8.433365	-0.060402	0.9520
_US--C	-30.36853	114.6879	-0.264793	0.7918
_GM--M_GM	0.119281	0.024722	4.824800	0.0000
_CH--M_CH	0.077948	0.016840	4.628836	0.0000
_GE--M_GE	0.026551	0.011791	2.251880	0.0269
_WE--M_WE	0.052894	0.015826	3.342304	0.0012
_US--M_US	0.156571	0.054437	2.876178	0.0051
_GM--K_GM	0.371445	0.044306	8.383545	0.0000
_CH--K_CH	0.315718	0.021796	14.48498	0.0000
_GE--K_GE	0.151694	0.017913	8.468579	0.0000
_WE--K_WE	0.092406	0.053010	1.743205	0.0849
_US--K_US	0.423866	0.154779	2.738529	0.0075

R-squared	0.952260	Mean dependent var	248.9570
Adjusted R-squared	0.944396	S.D. dependent var	267.8654
S.E. of regression	63.16379	Akaike info criterion	11.26682
Sum squared resid	339121.5	Schwarz criterion	11.65760
Log likelihood	-548.3410	F-statistic	121.1043
Durbin-Watson stat	0.978693	Prob(F-statistic)	0.000000

变截距模型：

（平均水平 ＋

个体效应）

Dependent Variable: I?
Method: Pooled Least Squares
Date: 11/21/07 Time: 10:12
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100
White cross-section standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M?	0.105980	0.017070	6.208430	0.0000
K?	0.346660	0.032259	10.74604	0.0000
_GM--C	-76.06675	68.04195	-1.117939	0.2665
_CH--C	-29.37358	11.98102	-2.451676	0.0161
_GE--C	-242.1708	39.43717	-6.140674	0.0000
_WE--C	-57.89941	14.87369	-3.892741	0.0002
_US--C	92.53854	36.17211	2.558284	0.0121

R-squared	0.937454	Mean dependent var	248.9570
Adjusted R-squared	0.933419	S.D. dependent var	267.8654
S.E. of regression	69.11798	Akaike info criterion	11.37694
Sum squared resid	444288.4	Schwarz criterion	11.55930
Log likelihood	-561.8468	F-statistic	232.3194
Durbin-Watson stat	0.806789	Prob(F-statistic)	0.000000

不变参数模型：

Dependent Variable: I?
Method: Pooled Least Squares
Date: 11/21/07 Time: 10:13
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100
White cross-section standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-48.02974	11.67694	-4.113213	0.0001
M?	0.105085	0.008604	12.21286	0.0000
K?	0.305366	0.044863	6.806564	0.0000

R-squared	0.778856	Mean dependent var	248.9570
Adjusted R-squared	0.774296	S.D. dependent var	267.8654
S.E. of regression	127.2583	Akaike info criterion	12.55986
Sum squared resid	1570884.	Schwarz criterion	12.63801
Log likelihood	-624.9928	F-statistic	170.8140
Durbin-Watson stat	0.228055	Prob(F-statistic)	0.000000

例10.5中系数α 和β 取何种形式，可以利用模型形式设定检验方法来确定。

(1) 首先分别计算3种形式的模型：变参数模型、变截距模型和不变参数模型，在每个模型的回归统计量里可以得到相应的残差平方和$S1$=339121.5、$S2$ = 444288.4 和$S3$ = 1570884。

(2) 按(10.2.7)式和(10.2.8)式计算F统计量，其中N=5、k=2、T=20，得到的两个F统计量分别为：

$$F1=((S2-S1)/8)/(S1/85) = 3.29$$

$$F2=((S3-S1)/12)/(S1/85) = 25.73$$

利用函数 @qfdist(d,k_1,k_2) 得到F分布的临界值，其中d 是临界点，k_1和k_2是自由度。在给定5%的显著性水平下(d=0.95)，得到相应的临界值为：

$$F_{\alpha 2}(12, 85) = 1.87 \qquad F_{\alpha 1}(8, 85) =2.049$$

由于 F_2>1.87，所以拒绝H_2；又由于 F_1>2.049，所以也拒绝H_1。因此，例10.5的模型应采用变系数的形式。

Panel Data模型估计方法（略讲）

使用Panel Data模型数据结构信息，有很多种方法进行方程估计。可估计固定截距模型，随机截距模型，或者模型变量对各截面成员的系数不同，以及估计单独的AR项系数；也可以为各个截面成员分别估计一个方程。

EViews的Pool对象估计模型使用的方法有：最小二乘法，估计截面权重的加权最小二乘法或似不相关回归。这些方法的使用，都不改变原数据的排序。

下面将介绍怎样使用Pool和系统估计更一般和复杂的模型，包括二阶段最小二乘估计和非线性模型，以及有复杂截面系数限制的模型。

下面讨论Pool模型的计算方法。设有N个观测值互相堆积。为讨论方便，把堆积方程表示为：

$$\boldsymbol{y}_i = \alpha_i + \boldsymbol{x}_i \boldsymbol{\beta}_i + \boldsymbol{u}_i \quad , \quad i=1, 2, \ldots, N \qquad (10.3.1)$$

其中 y_i 是第 i 个截面成员的$T\times 1$维因变量向量，x_i 是第 i 个截面成员的$T\times k$ 维解释变量矩阵。β_i 是第 i 个截面成员的$k\times 1$维未知参数向量，u_i 是第 i 个截面成员的$T\times 1$维扰动项向量。

用分块矩阵形式表示如下：

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{x}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_2 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \boldsymbol{x}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_N \end{bmatrix}$$

并且方程的残差协方差矩阵为：

$$\boldsymbol{\Omega}=E(\boldsymbol{u}\boldsymbol{u}')=E\begin{bmatrix}\boldsymbol{u}_1\boldsymbol{u}_1' & \boldsymbol{u}_1\boldsymbol{u}_2' & \cdots & \boldsymbol{u}_1\boldsymbol{u}_N' \\ \boldsymbol{u}_2\boldsymbol{u}_1' & \boldsymbol{u}_2\boldsymbol{u}_2' & \cdots & \boldsymbol{u}_2\boldsymbol{u}_N' \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_N\boldsymbol{u}_1' & \boldsymbol{u}_N\boldsymbol{u}_2' & \cdots & \boldsymbol{u}_N\boldsymbol{u}_N'\end{bmatrix}$$

基本说明把Pool作为联立方程系统，并使用“系统最小二乘法”估计模型。

不变参数模型（所有截面截距相同、系数相同）

当残差同期不相关，并且时期和截面同方差时，

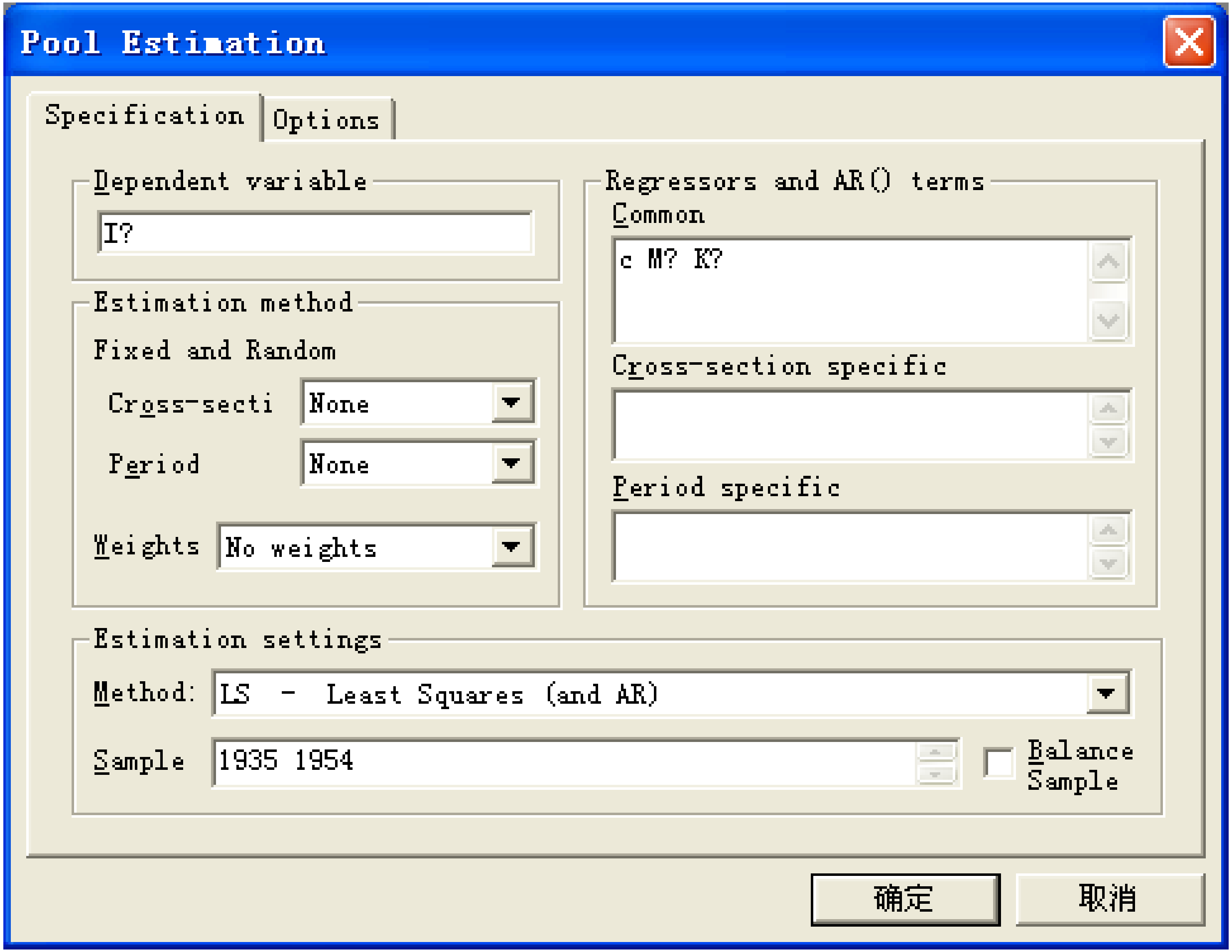
$$\boldsymbol{\Omega}=\sigma^2\boldsymbol{I}_N\otimes\boldsymbol{I}_T=\begin{pmatrix}\sigma^2\boldsymbol{I}_T & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \sigma^2\boldsymbol{I}_T & \cdots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \cdots & \sigma^2\boldsymbol{I}_T\end{pmatrix}$$

对堆积数据模型使用普通最小二乘法估计系数和协方差。相当于情形3：$\alpha_i=\alpha_j$，$\beta_i=\beta_j$，在横截面上无个体影响、无结构变化，则普通最小二乘法估计给出了 α 和 β 的一致有效估计。相当于将多个时期的截面数据放在一起、作为样本数据。

$$\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\\ \vdots\\ \boldsymbol{y}_T\end{pmatrix}=\alpha+\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \vdots\\ \boldsymbol{x}_T\end{pmatrix}\boldsymbol{\beta}+\boldsymbol{u}$$

其中 y_i 和 x_i 分别是 各时期的因变量向量和解释变量矩阵。

以例10.5为例:


Pool Estimation
Specification
Options
Dependent variable
I?
Estimation method
Fixed and Random
Cross-secti
None
Period
None
Weights
No weights
Regressors and AR() terms
Common
c M? K?
Cross-section specific
Period specific
Estimation settings
Method:
LS - Least Squares (and AR)
Sample
1935 1954
Balance Sample
确定
取消

Dependent Variable: I?
Method: Pooled Least Squares
Date: 11/22/06 Time: 09:33
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-48.02974	21.48017	-2.236004	0.0276
M?	0.105085	0.011378	9.235980	0.0000
K?	0.305366	0.043508	7.018637	0.0000

R-squared	0.778856	Mean dependent var	248.9570
Adjusted R-squared	0.774296	S.D. dependent var	267.8654
S.E. of regression	127.2583	Akaike info criterion	12.55986
Sum squared resid	1570884.	Schwarz criterion	12.63801
Log likelihood	-624.9928	F-statistic	170.8140
Durbin-Watson stat	0.228055	Prob(F-statistic)	0.000000

所有的截面的系数相等，和将5个公司的数据接到一起，用OLS的估计结果相同。

10.3 变截距模型

10.3.1 固定影响变截距模型

(1) 固定影响 (Fixed Effects) (情形2：$\alpha_i \neq \alpha_j$，$\beta_i = \beta_j$)

固定影响估计量，通过为每个截面成员估计不同常数项使α_i不同。模型对应的向量形式如下：

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{e} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \alpha_1 + \begin{pmatrix} 0 \\ \boldsymbol{e} \\ \vdots \\ 0 \end{pmatrix} \alpha_2 + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \boldsymbol{e} \end{pmatrix} \alpha_N + \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_N \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_N \end{pmatrix} \qquad (10.3.2)$$

其中：y_i, e, u_i是$T\times 1$维向量，x_i是$T\times k$维矩阵，即

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}_{T\times 1}, \quad \boldsymbol{e} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T\times 1}, \quad \boldsymbol{x}_i = \begin{pmatrix} x_{i,11} & x_{i,12} & \cdots & x_{i,1k} \\ x_{i,21} & x_{i,22} & \cdots & x_{i,2k} \\ \vdots & & & \\ x_{i,T1} & x_{i,T2} & \cdots & x_{i,Tk} \end{pmatrix}_{T\times k}, \quad \boldsymbol{u}_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}_{T\times 1}$$

EViews将每个变量减去平均值，并用转换后的数据，通过最小二乘估计来计算固定影响。

$$\hat{\beta}_{FE}=\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_{i})(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_{i})'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_{i})(y_{it}-\bar{y}_{i})\right] \quad (10.3.3)$$

其中 $\bar{y}_i=\sum_t y_{it}/T$， $\bar{x}_i=\sum_t x_{it}/T$ ， $x_{it}=(x_{1,it}, x_{2,it}, \ldots, x_{k,it})'$

<u>固定影响本身、不是直接估计的</u>，**计算公式为：**

$$\hat{\alpha}_i=\bar{y}_i-\bar{\boldsymbol{x}}_i'\hat{\beta}_{FE} \quad , \qquad i=1, 2, \ldots, N \qquad (10.3.4)$$

(2) 引进总体均值截距项的固定影响变截距模型

如果引进总体均值截距项（m），可以将模型（10.3.1）写成如下的等价形式：

$$y_i = m + x_i\beta + \alpha_i^* + u_i$$

$$i = 1, 2, \ldots, N \quad (10.3.10)$$

在该形式下，模型（10.3.1）中的反映个体影响的跨成员方程变化的截距项被分解成在各个体成员方程中都相等的总体均值截距项（m）和跨成员方程变化的表示个体对总体均值偏离的个体截距项（α_i^*）。个体截距项α_i^* 表示的是个体成员 i 对总体平均状态的偏离，所有偏离之和应该为零，即

$$\sum_{i=1}^{N} \alpha_i^* = 0 \quad (10.3.11)$$

在该约束下，可得到模型(10.3.10)中的各参数的最优线性无偏估计：

$$\hat{\boldsymbol{\beta}}_{FE}=\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i)\right] \tag{10.3.12}$$

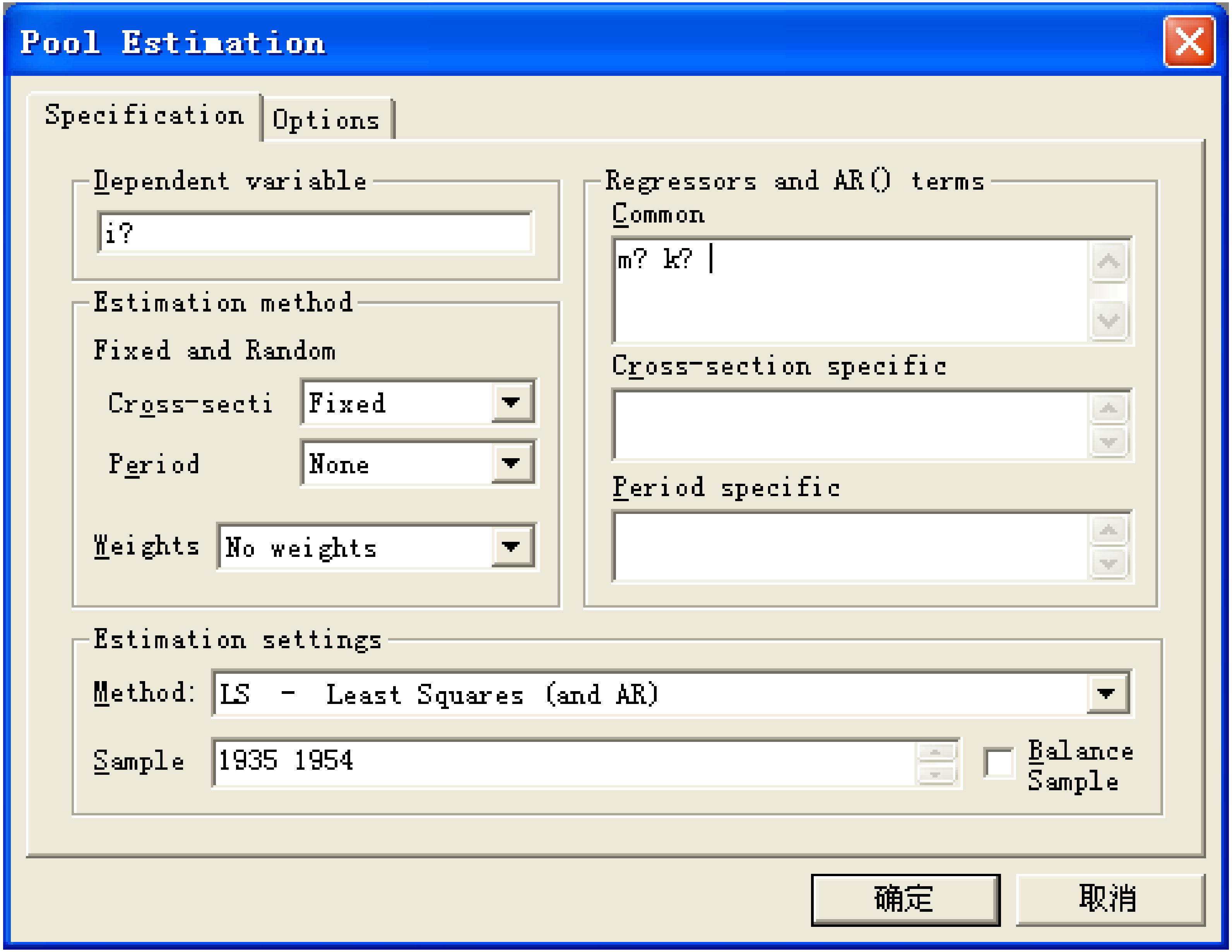
$$\hat{m}=\bar{y}-\bar{\boldsymbol{x}}'\hat{\boldsymbol{\beta}}_{FE} \tag{10.3.13}$$

$$\hat{\alpha}_i^{*}=\bar{y}_i-\hat{m}-\bar{\boldsymbol{x}}_i'\hat{\boldsymbol{\beta}}_{FE} \tag{10.3.14}$$

其中：$\bar{\boldsymbol{x}}=\dfrac{1}{NT}(\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it})$ ，$x_{it}=(x_{1,it},x_{2,it},\ldots,x_{k,it})'$，$\bar{y}=\dfrac{1}{NT}(\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it})$

$\bar{y}_i=\sum_t y_{it}/T$，$\bar{x}_i=\sum_t x_{it}/T$ 。

EViews计算固定影响，是包含总体均值截距项的变截距模型，以例10.5为例：


Pool Estimation
Specification
Options
Dependent variable
i?
Estimation method
Fixed and Random
Cross-secti
Fixed
Period
None
Weights
No weights
Regressors and AR() terms
Common
m? k?
Cross-section specific
Period specific
Estimation settings
Method:
LS - Least Squares (and AR)
Sample
1935 1954
Balance Sample
确定
取消

Dependent Variable: I?
Method: Pooled Least Squares
Date: 11/22/06 Time: 09:56
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-62.59439	29.44191	-2.126030	0.0361
M?	0.105980	0.015891	6.669182	0.0000
K?	0.346660	0.024161	14.34781	0.0000
Fixed Effects (Cross)				
_GM--C	-13.47235			
_CH--C	33.22081			
_GE--C	-179.5764			
_WE--C	4.694980			
_US--C	155.1329			

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.937454	Mean dependent var	248.9570
Adjusted R-squared	0.933419	S.D. dependent var	267.8654
S.E. of regression	69.11798	Akaike info criterion	11.37694
Sum squared resid	444288.4	Schwarz criterion	11.55930
Log likelihood	-561.8468	F-statistic	232.3194
Durbin-Watson stat	0.806789	Prob(F-statistic)	0.000000

(3) 包含时期个体恒量的固定影响变截距模型

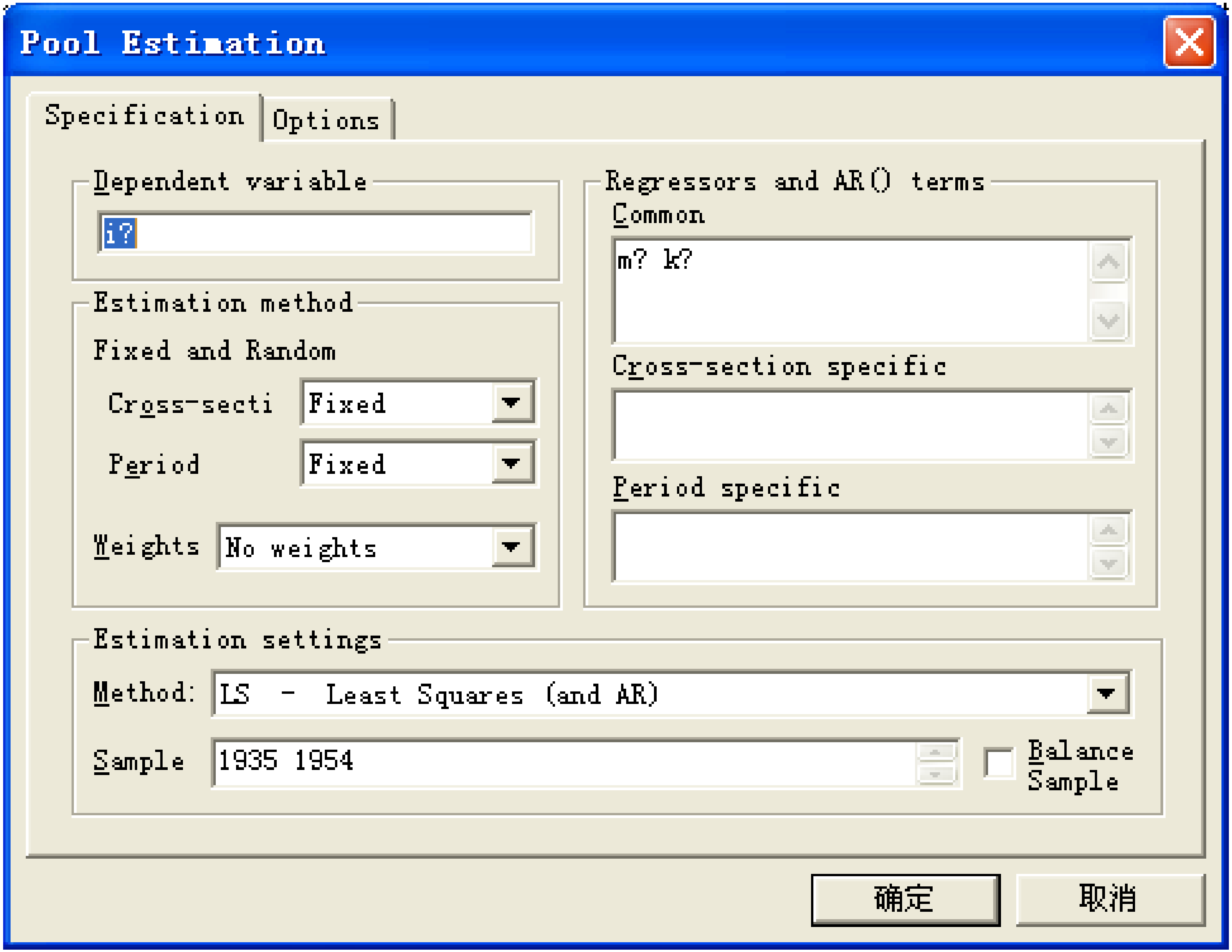
模型(10.3.1)还可以推广为包含时期个体恒量的形式，即模型形式为：

$$\mathbf{y}_{it} = m + \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i^* + \gamma_t + \mathbf{u}_{it}$$

$$i = 1, 2, \ldots, N\text{，}\quad t = 1, 2, \ldots, T \qquad (10.3.15)$$

其中：γ_t 为时期 t 的个体恒量，反映时期 t 的特有影响。

类似地，通过引进相应的个体成员和时期虚拟变量，利用普通最小二乘法可以得到各参数的OLS估计。


Pool Estimation
Specification
Options
Dependent variable
i?
Estimation method
Fixed and Random
Cross-secti
Fixed
Period
Fixed
Weights
No weights
Regressors and AR() terms
Common
m? k?
Cross-section specific
Period specific
Estimation settings
Method:
LS - Least Squares (and AR)
Sample
1935 1954
Balance Sample
确定
取消

Dependent Variable: I?
Method: Pooled Least Squares
Date: 11/22/06 Time: 16:44
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-105.8386	43.00599	-2.461021	0.0162
M?	0.126031	0.023174	5.438392	0.0000
K?	0.361776	0.035986	10.05336	0.0000
Fixed Effects (Cross)				
_GM--C	-66.92696			
_CH--C	60.73287			
_GE--C	-181.3062			
_WE--C	33.19241			
_US--C	154.3079			
Fixed Effects (Period)				
1935--C	59.74633			
1936--C	20.19311			
1937--C	-28.40345			
1938--C	-9.042101			
1939--C	-59.39754			
1940--C	-29.29686			
1941--C	36.63555			
1942--C	46.02723			
1943--C	5.469688			
1944--C	2.774454			
1945--C	-13.43547			
1946--C	28.67775			
1947--C	29.29740			
1948--C	24.27390			
1949--C	-26.10428			
1950--C	-21.52140			
1951--C	-4.145028			
1952--C	-2.895486			
1953--C	-9.129487			
1954--C	-49.72433			

Effects Specification

Cross-section fixed (dummy variables)
Period fixed (dummy variables)

R-squared	0.948772	Mean dependent var	248.9570
Adjusted R-squared	0.931466	S.D. dependent var	267.8654
S.E. of regression	70.12471	Akaike info criterion	11.55732
Sum squared resid	363893.1	Schwarz criterion	12.23467
Log likelihood	-551.8661	F-statistic	54.82119
Durbin-Watson stat	0.775172	Prob(F-statistic)	0.000000

3. 固定影响变截距模型的广义最小二乘估计

(1) 截面加权(个体成员截面异方差情形的GLS估计)

利用OLS参数估计，我们得到5个公司的方程残差的方差σ_i^2，具有截面异方差性。

残差的方差

通用汽车公司（**GM**）	**9410.9**
克莱斯勒公司（**CH**）	**755.85**
通用电器公司（**GE**）	**34288.49**
西屋公司（**WE**）	**633.42**
美国钢铁公司（**US**）	**33455.51**

当残差具有<u>截面异方差性和同期不相关</u>时，最好进行截面加权回归：

$$\boldsymbol{\Omega}=E(\boldsymbol{uu}')=\begin{bmatrix}\sigma_1^2\boldsymbol{I}_{T_1} & 0 & \cdots & 0\\ 0 & \sigma_2^2\boldsymbol{I}_{T_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \cdots & \sigma_N^2\boldsymbol{I}_{T_N}\end{bmatrix}$$

EViews进行可行广义最小二乘（FGLS）。

首先从一阶段Pool最小二乘回归，得到方差 σ_i^2 的估计值 s_i^2，计算公式为：

$$s_i^2=\sum_{t=1}^{T_i}(y_{it}-\hat{y}_{it})^2/T_i\ ,\quad i=1,2,\ldots,N \qquad (10.3.24)$$

其中 $\hat{y}_{it}$ 是OLS的拟合值。

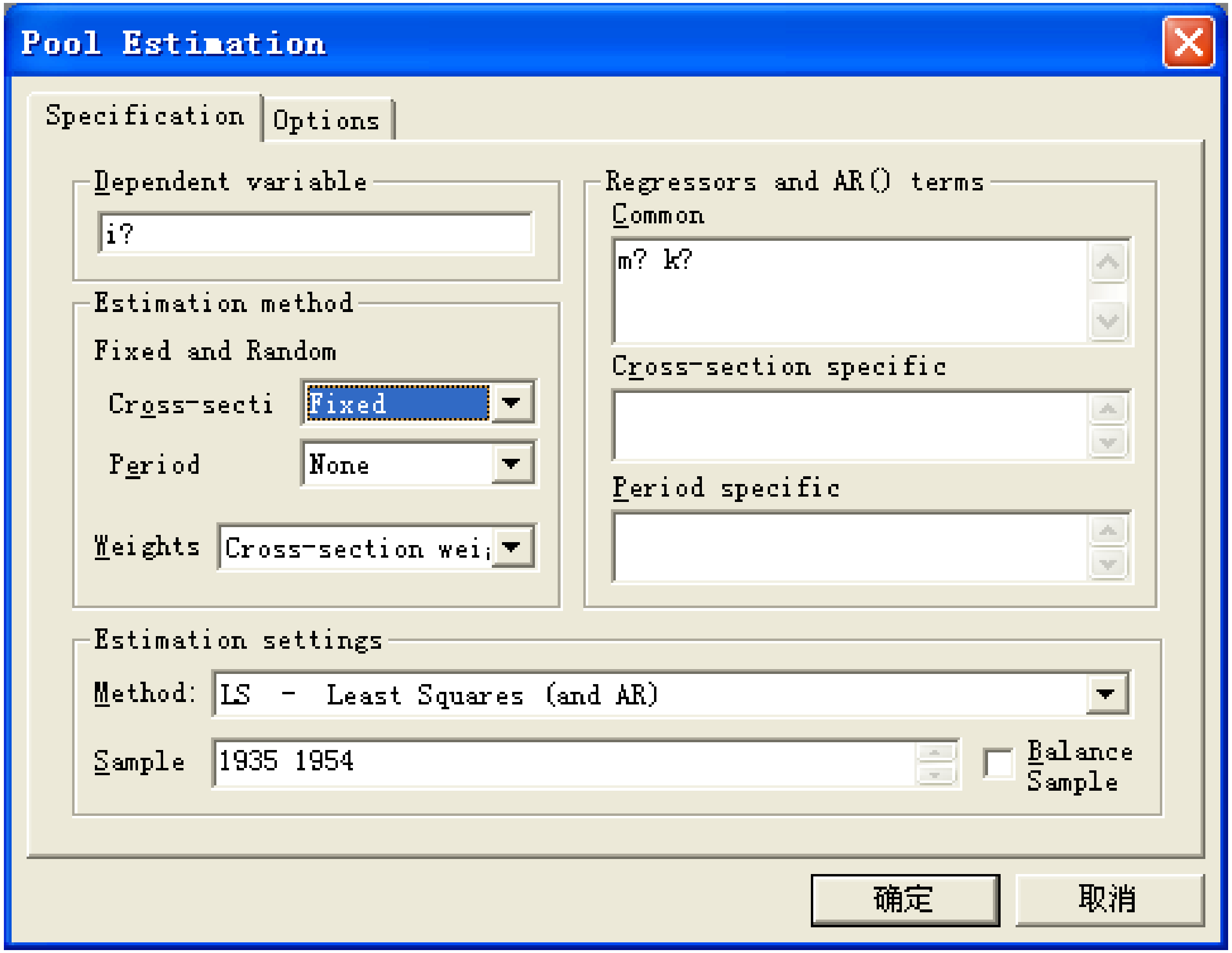
个体成员方程截面异方差的协方差矩阵的估计为

$$\boldsymbol{\Sigma}_N = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_N^2 \end{pmatrix}$$

然后，用得到的样本方差估计作为各个体成员的权重，即加权矩阵为$\Sigma_N \otimes I_T$，利用加权最小二乘方法得到相应的GLS估计。类似地，可以得到含有T个时间截面方程情形下的时期异方差的GLS估计。

其次系数值 β 由标准GLS估计量估计，是有效估计量。

$$\hat{\beta}_{FE}=\left[X'(\Sigma_N\otimes I_T)^{-1}X\right]^{-1}\left[X'(\Sigma_N\otimes I_T)^{-1}y\right]=\left[\sum_{i=1}^{N}\frac{1}{s_i^2}x_i'x_i\right]^{-1}\left[\sum_{i=1}^{N}\frac{1}{s_i^2}x_i'y_i\right]$$



Dependent Variable: I?
Method: Pooled EGLS (Cross-section weights)
Date: 11/25/06 Time: 16:39
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.310744	23.06175	0.143560	0.8862
M?	0.075996	0.012685	5.991264	0.0000
K?	0.320075	0.020388	15.69935	0.0000
Fixed Effects (Cross)				
_GM--C	67.80600			
_CH--C	-8.676003			
_GE--C	-176.6351			
_WE--C	-38.81703			
_US--C	156.3221			

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.908860	Mean dependent var	273.9123
Adjusted R-squared	0.902980	S.D. dependent var	210.0050
S.E. of regression	65.41248	Sum squared resid	397927.7
F-statistic	154.5677	Durbin-Watson stat	0.904337
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.932530	Mean dependent var	248.9570
Sum squared resid	479269.1	Durbin-Watson stat	0.741946

(2) 同期相关协方差情形的SUR估计

当残差具有**“截面异方差性和同步相关性”**时，SUR(似不相关回归)加权最小二乘是可行的GLS估计量：

$$\boldsymbol{\Omega} = E(\boldsymbol{u}\boldsymbol{u}') = \begin{bmatrix} \sigma_{11}\boldsymbol{I}_T & \sigma_{12}\boldsymbol{I}_T & \cdots & \sigma_{1N}\boldsymbol{I}_T \\ \sigma_{21}\boldsymbol{I}_T & \sigma_{22}\boldsymbol{I}_T & & \sigma_{2N}\boldsymbol{I}_T \\ \vdots & & \ddots & \vdots \\ \sigma_{N1}\boldsymbol{I}_T & \cdots & & \sigma_{NN}\boldsymbol{I}_T \end{bmatrix} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T$$

其中Σ 是同步相关的对称阵：

$$\boldsymbol{\Sigma}_N = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2N} \\ \vdots & & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix} \qquad (10.3.28)$$

一般项 $\sigma_{ij} = E(u_{jt}u_{it})$ ，在所有的 t 时为常数。

EViews估计SUR模型时使用的 σ_{ij}，是由一阶段Pool最小二乘回归得到：

$$s_{ij}=\frac{[(\boldsymbol{y}_i-\hat{\alpha}_i-\boldsymbol{x}_i\hat{\boldsymbol{\beta}}_{FE})'(\boldsymbol{y}_j-\hat{\alpha}_j-\boldsymbol{x}_j\hat{\boldsymbol{\beta}}_{FE})]}{T}$$

$$i, j=1, 2, \ldots, N \quad (10.3.30)$$

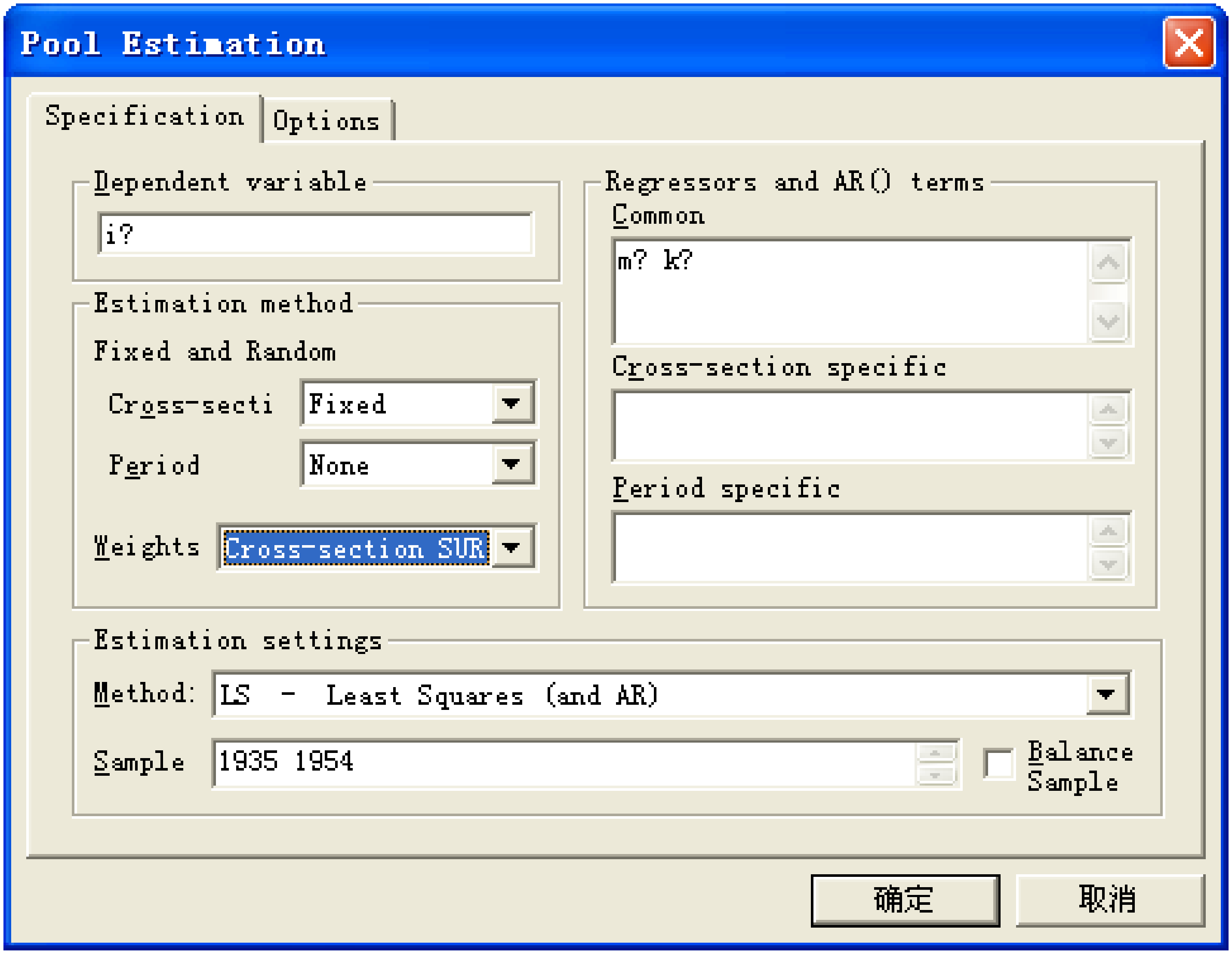
其中：$\hat{\boldsymbol{\beta}}_{FE}$ 和 $\hat{\alpha}_i$ 可由式（10.3.3）和式（10.3.4）得到。计算后，再进行广义最小二乘估计（GLS），此时 β 的SUR估计为：

$$\hat{\boldsymbol{\beta}}_{SUR}=\left[(\boldsymbol{X}-\bar{\boldsymbol{X}})'(\hat{\boldsymbol{\Sigma}}\otimes\boldsymbol{I}_T)^{-1}(\boldsymbol{X}-\bar{\boldsymbol{X}})\right]^{-1}\left[(\boldsymbol{X}-\bar{\boldsymbol{X}})'(\hat{\boldsymbol{\Sigma}}\otimes\boldsymbol{I}_T)^{-1}(\boldsymbol{Y}-\bar{\boldsymbol{Y}})\right] \quad (10.3.31)$$

其中:

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\\ \vdots\\ \boldsymbol{y}_N\end{pmatrix},\quad \boldsymbol{X}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \vdots\\ \boldsymbol{x}_N\end{pmatrix},\quad \bar{\boldsymbol{Y}}=\begin{pmatrix}\bar{y}_1\\ \bar{y}_2\\ \vdots\\ \bar{y}_N\end{pmatrix},\quad \bar{\boldsymbol{X}}=\begin{pmatrix}\bar{\boldsymbol{x}}_1\\ \bar{\boldsymbol{x}}_2\\ \vdots\\ \bar{\boldsymbol{x}}_N\end{pmatrix}$$

此时β的SUR估计为：



Dependent Variable: I?
Method: Pooled EGLS (Cross-section SUR)
Date: 11/27/06 Time: 09:41
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-34.88009	16.62984	-2.097439	0.0387
M?	0.091285	0.008868	10.29415	0.0000
K?	0.348374	0.017892	19.47071	0.0000
Fixed Effects (Cross)				
_GM--C	21.38870			
_CH--C	15.48556			
_GE--C	-179.4484			
_WE--C	-13.30695			
_US--C	155.8811			

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.949898	Mean dependent var	4.603362
Adjusted R-squared	0.946666	S.D. dependent var	4.425242
S.E. of regression	1.021971	Sum squared resid	97.13155
F-statistic	293.8718	Durbin-Watson stat	1.046036
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.936821	Mean dependent var	248.9570
Sum squared resid	448785.1	Durbin-Watson stat	0.782194

10.3.2 随机影响变截距模型 (Random Effects)

随机影响模型假设α_{it}项，是共同系数 α 和不随时间改变的截面说明“随机变量 v_i ”的和， v_i 和残差 u_i 是不相关的。

$$\mathbf{y}_i = \alpha + v_i + \mathbf{x}_i\boldsymbol{\beta} + \mathbf{u}_i \quad , \quad i=1, 2, \ldots, N \qquad (10.3.36)$$

为分析方便，可将模型(10.3.36)写成如下形式：

$$Y = \tilde{X}\delta + w \qquad (10.3.43)$$

其中： $\tilde{x}_i = (e, x_i)_{T\times(k+1)}$ ， $\delta = (\alpha, \beta')'$ ， $w_i = v_i + u_i$。

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad X = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_N \end{pmatrix}, \quad e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix}$$

EViews按下列步骤估计随机影响模型：

(1) 使用固定影响模型的残差，估计 u_i 的方差：

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - \bar{y}_i) - (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\hat{\boldsymbol{\beta}}_{FE}]^2}{NT - N - k} \tag{10.3.52}$$

使用包含总体均值截距项的变截距模型的残差，估计 v_i 的方差，

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{N}(\bar{y}_i - \hat{\alpha}_i - \bar{\boldsymbol{x}}_i'\hat{\boldsymbol{\beta}}_{FE})^2}{N - k - 1} - \frac{\hat{\sigma}_u^2}{T} \tag{10.3.52}$$

(2) 由于

$$\boldsymbol{\Omega}=\begin{pmatrix}\sigma_u^2+\sigma_v^2 & \sigma_v^2 & \cdots & \sigma_v^2\\ \sigma_v^2 & \sigma_u^2+\sigma_v^2 & \cdots & \sigma_v^2\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_v^2 & \sigma_v^2 & \cdots & \sigma_u^2+\sigma_v^2\end{pmatrix}$$

有了成分方差$\hat{\sigma}_u^2$和$\hat{\sigma}_v^2$的估计，可求出模型(10.3.42)中参数δ的GLS估计量：

$$\hat{\boldsymbol{\delta}}_{GLS}=\left[\sum_{i=1}^{N}\tilde{\boldsymbol{X}}_i'\hat{\boldsymbol{\Omega}}^{-1}\tilde{\boldsymbol{X}}_i\right]^{-1}\left[\sum_{i=1}^{N}\tilde{\boldsymbol{X}}_i'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}_i\right]\qquad(10.3.50)$$

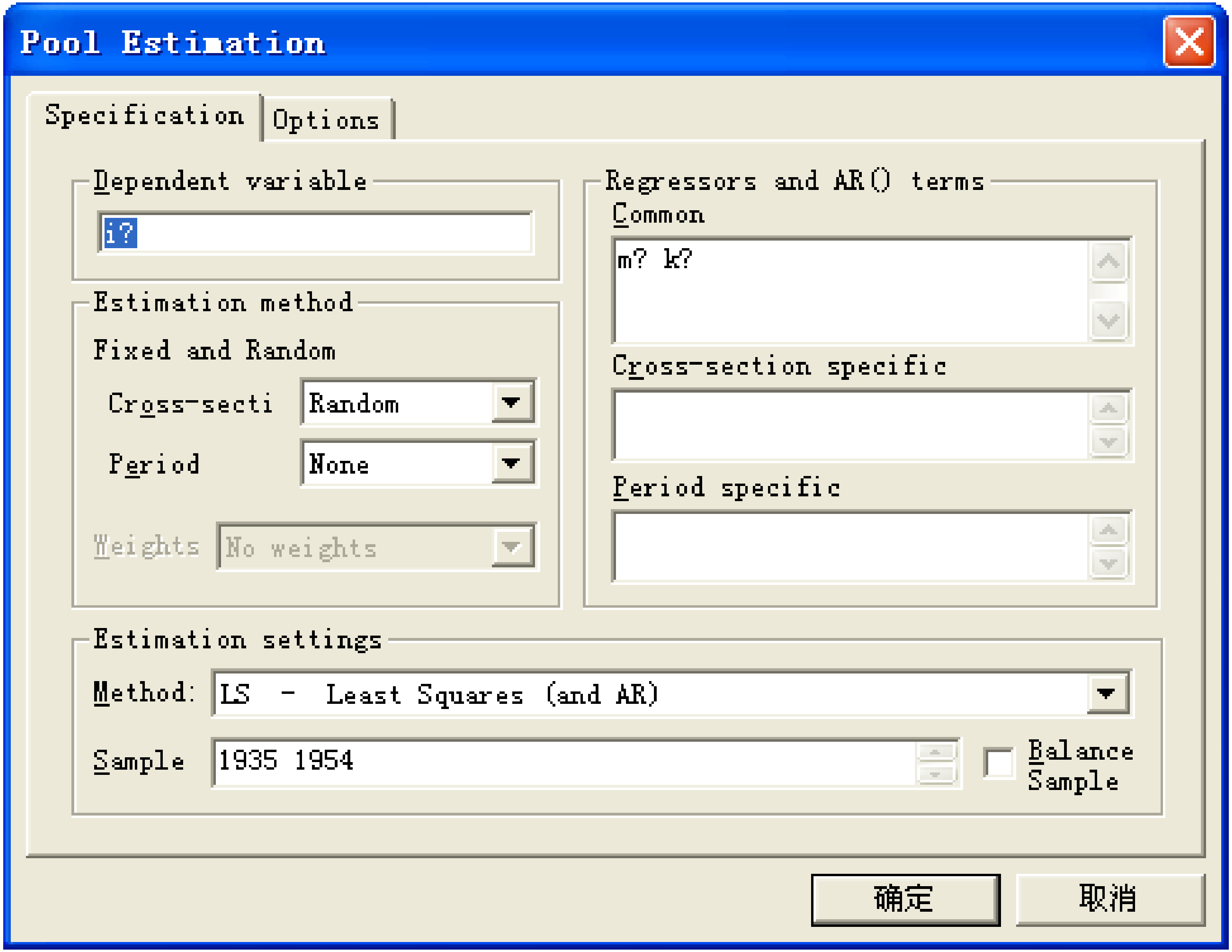
其中：$\tilde{\boldsymbol{X}}_i=(\tilde{\boldsymbol{x}}_i,\tilde{\boldsymbol{x}}_i,\ldots,\tilde{\boldsymbol{x}}_i)'$ 。

个体随机影响 v_i 相应的估计为

$$\hat{v}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_B^2} \sum_{t=1}^{T} (y_{it} - \tilde{\boldsymbol{x}}_{it}' \hat{\boldsymbol{\delta}}_{GLS}) \qquad i = 1, 2, \ldots, N \quad (10.3.56)$$

其中：

$$\hat{\sigma}_B^2 = \frac{\sum_{i=1}^{N} (\bar{y}_i - \hat{\alpha}_i - \bar{\boldsymbol{x}}_i' \hat{\boldsymbol{\beta}}_{FE})^2}{N - k - 1} \qquad (10.3.57)$$


Pool Estimation
Specification
Options
Dependent variable
i?
Estimation method
Fixed and Random
Cross-secti
Random
Period
None
Weights
No weights
Regressors and AR() terms
Common
m? k?
Cross-section specific
Period specific
Estimation settings
Method:
LS - Least Squares (and AR)
Sample
1935 1954
Balance Sample
确定
取消

Dependent Variable: I?
Method: Pooled EGLS (Cross-section random effects)
Date: 11/27/06 Time: 10:40
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-60.29050	54.16656	-1.113058	0.2684
M?	0.104886	0.014711	7.129710	0.0000
K?	0.346016	0.024112	14.35019	0.0000
Random Effects (Cross)				
_GM--C	-10.38936			
_CH--C	31.07585			
_GE--C	-175.6668			
_WE--C	3.112561			
_US--C	151.8678			

Effects Specification

Cross-section random S.D. / Rho	104.6527	0.6963
Idiosyncratic random S.D. / Rho	69.11798	0.3037

Weighted Statistics

R-squared	0.798727	Mean dependent var	36.37182
Adjusted R-squared	0.794577	S.D. dependent var	153.3923
S.E. of regression	69.52289	Sum squared resid	468842.9
F-statistic	192.4666	Durbin-Watson stat	0.762426
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.775749	Mean dependent var	248.9570
Sum squared resid	1592956.	Durbin-Watson stat	0.224399

10.3.3 Hausman检验

Hausman（1978）等学者认为应总是把个体影响处理为随机的，即随机影响模型优于固定影响模型，其主要原因为：固定影响模型将个体影响设定为跨截面变化的常数、使得分析过于简单，并且从实践的角度看，在估计固定影响模型时将损失较多的自由度，特别是对“宽而短”的面板数据。

但相对于固定影响模型，随机影响模型也存在明显的不足：在随机影响模型中是假设随机变化的个体影响与模型中的解释变量不相关，而在实际建模过程中这一假设很有可能由于模型中省略了一些变量而不满足，从而导致估计结果出现不一致性。

对于如何检验模型中个体影响与解释变量之间是否相关，Hausman（1978）提出了一种严格的统计检验方法——Hausman检验。该检验的原假设是：

随机影响模型中，个体影响、与解释变量不相关

检验过程中所构造的统计量（W）形式如下：

$$W = \left[b - \hat{\beta}\right]' \hat{\Sigma}^{-1}\left[b - \hat{\beta}\right] \tag{10.3.68}$$

其中b为固定影响模型中回归系数的估计结果，$\hat{\beta}$为随机影响模型中回归系数的估计结果。$\hat{\Sigma}$为两类模型中回归系数估计结果之差的方差，即

$$\hat{\Sigma} = \operatorname{var}\left[b - \hat{\beta}\right] \tag{10.3.69}$$

Hausman证明在原假设下，式（10.3.68）给出的统计量W服从自由度为 k 的χ^2分布，k为模型中解释变量的个数。

Hausman检验的操作

EViews中可实现检验模型中个体影响与解释变量之间是否相关的Hausman检验。

为了实Hausman检验，必须首先估计一个随机效应模型。然后，选择View/Fixed/Random Effects Testing/Correlated Random Effects - Hausman Test，EViews将自动估计相应的固定效应模型，计算检验统计量，显示检验结果和辅助回归结果。

例10.3 城镇居民消费行为的区域差异分析

本例按照国家有关部门的划分标准将29个省市自治区划分为三个区域——东部、中部和西部，其中：

东部区域包括：北京市、天津市、河北省、辽宁省、上海市、江苏省、浙江省、福建省、山东省、广东省和海南省；

中部区域包括：山西省、吉林省、黑龙江省、安徽省、江西省、河南省、湖北省和湖南省；

西部区域包括：内蒙古自治区、四川省、广西自治区、贵州省、云南省、陕西省、甘肃省、青海省、宁夏自治区和新疆自治区。

利用29个省市的居民收入、消费数据，分别建立东部、中部和西部的城镇居民的消费模型。各模型中的被解释变量为城镇居民人均全年消费*CS*，解释变量为城镇居民人均全年可支配收入*YD*，样本区间为1991～1994年。东部、中部和西部的城镇居民消费模型形式设定检验的结果由表10.4给出。

中部地区模型的Hausman Test结果：

Correlated Random Effects - Hausman Test
Pool: ZHONG_10_3
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	0.290584	1	0.5898

Cross-section random effects test comparisons:

Variable	Fixed	Random	Var(Diff.)	Prob.
?YD	0.778298	0.777074	0.000005	0.5898

由（10.3.68）式构造的中部地区模型的Hausman Test统计量（W） 是0.29，p值是0.59，接受原假设: 随机影响模型中、个体影响与解释变量不相关，可将模型设定为随机模型。

以中部地区为例，根据(10.2.7),(10.2.8)式计算$F2$和$F1$统计量:

(1) 首先分别计算3种形式的模型：变参数模型、变截距模型和不变参数模型，在每个模型的回归统计量里可以得到相应的残差平方和$S1$_zhong=20571.8、$S2$_zhong =30004.2 和$S3$_zhong = 69605.61。

(2) 按(10.2.7)式和(10.2.8)式计算F统计量，其中N=8、k=1、T=4，得到的两个F统计量分别为:

$F1$_zhong=(($S2$_zhong-$S1$_zhong)/7)/($S1$_zhong /16) = 1.048

$F2$_zhong=(($S3$_zhong-$S1$_zhong)/14)/($S1$_zhong /16) = 2.724

利用函数 @qfdist(d,k1,k2) 得到F分布的临界值，其中d 是临界点，k1和k2是自由度。在给定5%的显著性水平下(d=0.95)，得到相应的临界值为:

$$F_{\alpha 2}(14, 16) = 2.37 \qquad F_{\alpha 1}(7, 16) = 2.66$$

由于 $F2> 2.37$ ，所以拒绝H_2；而 $F1< 2.66$ ，所以接受H_1。因此，中部地区模型应采用变截距的形式。同样方法可以计算东部和西部地区的模型形式.

表10.4　东部、中部和西部城镇居民消费模型形式设定检验结果

检验统计量	东部	中部	西部
F_2	3.67* （2.07）	2.72* （2.37）	2.04** （2.81）
F_1	1.90 （2.30）	1.05 （2.66）	0.52 （2.39）
Hausman检验 （w统计量）	2.51 （3.84）	0.29 （3.84）	0.001 （3.84）

注：括号内为t 统计量临界值，“*”，“**”分别表示在5%、10%的显著性水平下拒绝原假设。

从表10.4中可以看出，东部、中部和西部城镇居民消费模型的F_2 均在5%或10%的显著性水平下显著，而各模型的 F_1却均小于相应临界值。

可见，对三个模型进行模型设定检验时均拒绝H_2且接受H_1，因此，东、中和西部城镇居民消费模型均应采用变截距形式。

同时，从三个模型的进一步Hausman检验结果中可以看出，三个模型的W统计量均小于临界值，这说明各模型均无法拒绝个体影响与解释变量不相关的原假设，因此应该将三个区域的城镇居民消费模型中的个体影响确定为随机影响形式，即分别建立东部、中部和西部城镇居民消费的随机影响变截距模型。模型形式为：

$$CS_{j,it} = \alpha_j + \beta_j \cdot YD_{j,it} + v_{j,i} + u_{j,it}$$

其中，j =1,2,3 分别代表东部、中部和西部地区；j =1,2,…，N_j，N_j分别表示东、中、西部包含的地区个数，N_1=11, N_2=8, N_3=10；t 表示时间期间。α_j 为东、中、西部地区的平均自发消费水平，为随机变量，β_j 代表j地区中i 城市或省份的随机影响，用来反映地区内部不同省市间的消费特征差异，为各地区的平均边际消费倾向，反映着不同地区城镇居民消费行为的差异。

中部地区随机影响模型：

Dependent Variable: ?CS
Method: Pooled EGLS (Cross-section random effects)
Date: 08/07/07 Time: 08:18
Sample: 1991 1994
Included observations: 4
Cross-sections included: 8
Total pool (balanced) observations: 32
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	101.2438	26.61983	3.803322	0.0007
?YD	0.777074	0.010963	70.87900	0.0000
Random Effects (Cross)				
AH_--C	26.98698			
HLJ_--C	27.35269			
HEN_--C	24.10450			
HUB_--C	23.99242			
HUN_--C	-40.36641			
JL_--C	-1.632951			
JX_--C	-20.75515			
SX_--C	-39.68208			

Effects Specification

	S.D.	Rho
Cross-section random	35.46421	0.4909
Idiosyncratic random	36.11829	0.5091

Weighted Statistics

R-squared	0.994203	Mean dependent var	772.7402
Adjusted R-squared	0.994010	S.D. dependent var	461.1331
S.E. of regression	35.68868	Sum squared resid	38210.46
F-statistic	5145.510	Durbin-Watson stat	1.817838
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.990799	Mean dependent var	1702.913
Sum squared resid	69797.33	Durbin-Watson stat	0.995173

使用Swamy-Arora方法、估计成分方差，对三个随机影响变截距模型模型进行估计，估计结果由表10.5给出（其中各模型中随机影响的估计结果略）。

表10.5　各地区边际消费倾向(β_j)的估计结果

	东部	中部	西部
α_j	175.77 (5.24)*	101.24 (3.80)*	112.19 (3.02)*
β_j	0.766 (80.99)*	0.777 (70.88)*	0.784 (53.83)*
R^2	0.993	0.994	0.987
样本容量	44	32	40

注：括号内数值为估计值对应的t统计量，“*”表示在5%的显著性水平下拒绝原假设。

10.4 变系数模型

前面所介绍的变截距模型中，横截面成员的个体影响是用变化的截距来反映的，即用变化的截距来反映模型中忽略的反映个体差异的变量的影响。

然而现实中变化的经济结构或不同的社会经济背景等因素，有时会导致反映经济结构的参数、随着横截面个体的变化而变化。因此，当现实数据不支持变截距模型时，便需要考虑这种系数随横截面个体的变化而改变的变系数模型。

变系数模型的基本形式如下：

$$y_i = \alpha_i + x_i \beta_i + u_i$$

$$i = 1, 2, \ldots, N \quad (10.4.1)$$

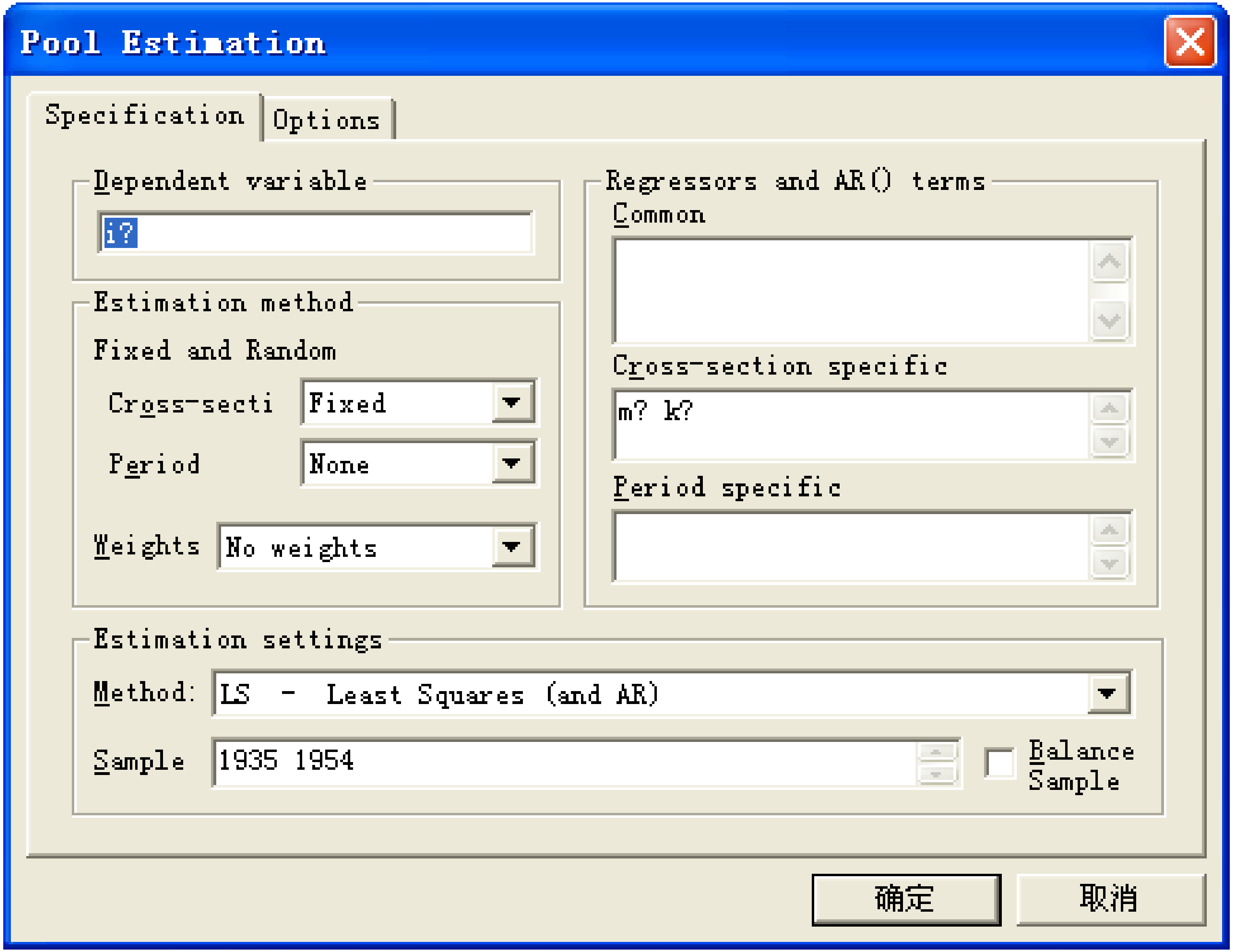
其中：y_i 为因变量向量，x_i 为 $T \times k$ 维解释变量矩阵，参数 α_i 表示模型的常数项，β_i 为对应于解释变量矩阵 x_i 的系数向量。随机误差项 u_i 相互独立，且满足零均值、等方差的假设。

在式(10.4.1)所表示的变系数模型中，常数项 α_i 和系数向量 β_i 都是随着横截面个体的改变而变化的，因此可以将变系数模型改写成如下形式：

$$y_i = \tilde{x}_i \delta_i + u_i \, , \qquad i = 1, 2, \ldots, N \qquad (10.4.2)$$

其中：$\tilde{x}_i = (1, x_i)$ ，$\delta_i = (\alpha_i, \beta_i')'$ 。

类似于变截距模型，变系数模型也分为固定影响变系数模型和随机影响变系数模型两种类型。


Pool Estimation
Specification
Options
Dependent variable
i?
Regressors and AR() terms
Common
Estimation method
Fixed and Random
Cross-secti
Fixed
Period
None
Weights
No weights
Cross-section specific
m? k?
Period specific
Estimation settings
Method:
LS - Least Squares (and AR)
Sample
1935 1954
Balance Sample
确定
取消

Dependent Variable: I?
Method: Pooled Least Squares
Date: 11/27/06 Time: 11:18
Sample: 1935 1954
Included observations: 20
Cross-sections included: 5
Total pool (balanced) observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-39.36133	32.46156	-1.212552	0.2287
_GM--M_GM	0.119281	0.017779	6.709095	0.0000
_CH--M_CH	0.077948	0.095009	0.820422	0.4143
_GE--M_GE	0.026551	0.035262	0.752959	0.4536
_WE--M_WE	0.052894	0.097138	0.544526	0.5875
_US--M_US	0.156571	0.048704	3.214716	0.0018
_GM--K_GM	0.371445	0.025513	14.55883	0.0000
_CH--K_CH	0.315718	0.137059	2.303518	0.0237
_GE--K_GE	0.151694	0.058228	2.605151	0.0108
_WE--K_WE	0.092406	0.346948	0.266341	0.7906
_US--K_US	0.423866	0.095831	4.423042	0.0000
Fixed Effects (Cross)				
_GM--C	-110.4211			
_CH--C	33.17137			
_GE--C	29.40502			
_WE--C	38.85194			
_US--C	8.992796			

Effects Specification

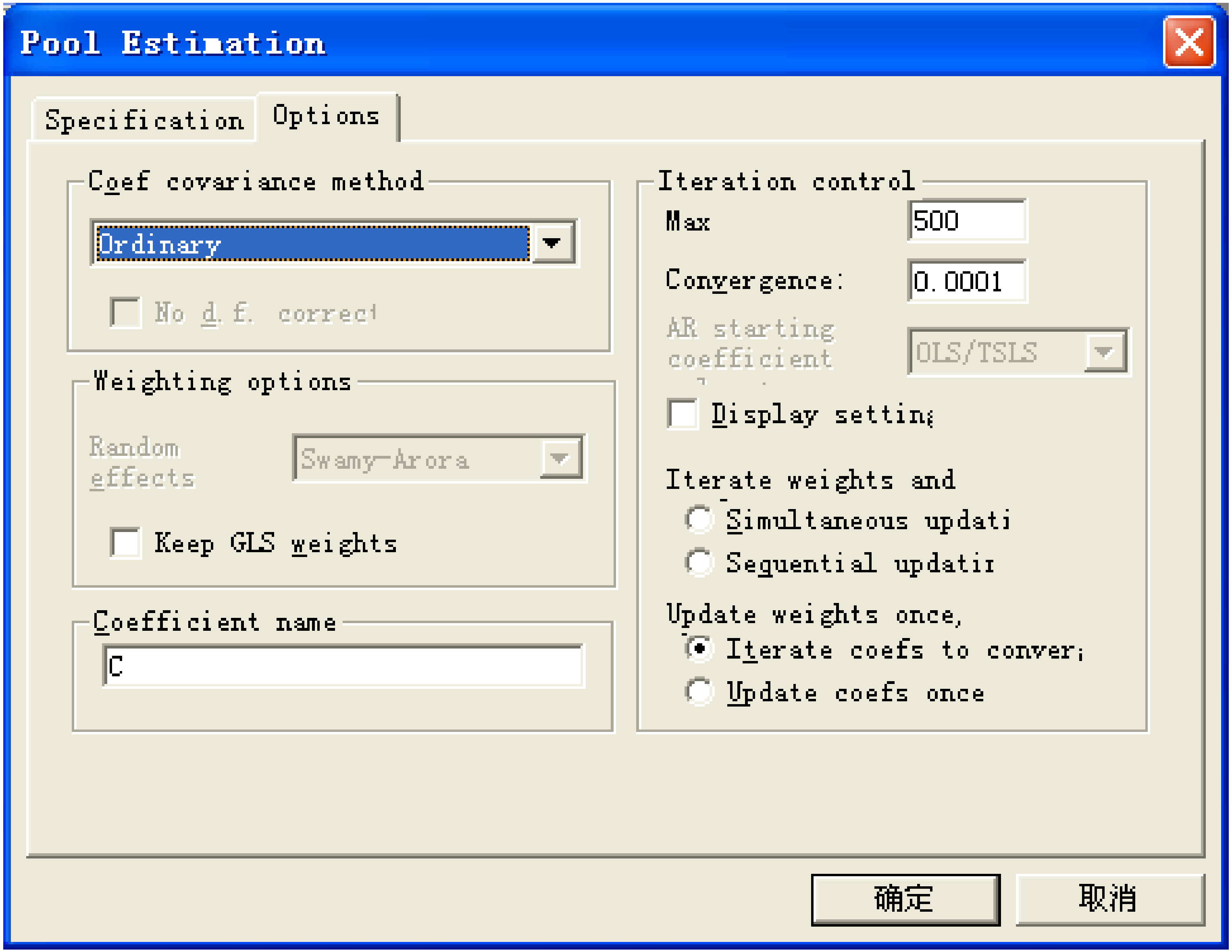
Cross-section fixed (dummy variables)

R-squared	0.952260	Mean dependent var	248.9570
Adjusted R-squared	0.944396	S.D. dependent var	267.8654
S.E. of regression	63.16379	Akaike info criterion	11.26682
Sum squared resid	339121.5	Schwarz criterion	11.65760
Log likelihood	-548.3410	F-statistic	121.1043
Durbin-Watson stat	0.978693	Prob(F-statistic)	0.000000

估计Pool方程的其他选项(Options)

EViews不能估计这样的模型：很少的时期或者庞大的截面成员。所用的时期数应至少不小于解释变量个数。即使有足够的观测值，估计的残差相关矩阵还必须是非奇异的。如果有一条不满足EViews的要求，EViews会显示错误信息："Near Singular Matrix"。

当选择加权时，复选框Iterate to convergence控制可行GLS程序。如果选择，EViews就一直迭代权重和系数直到收敛。如果模型中包括AR项，这个选择就没有意义，因为在AR估计中，EViews会一直迭代直至收敛。

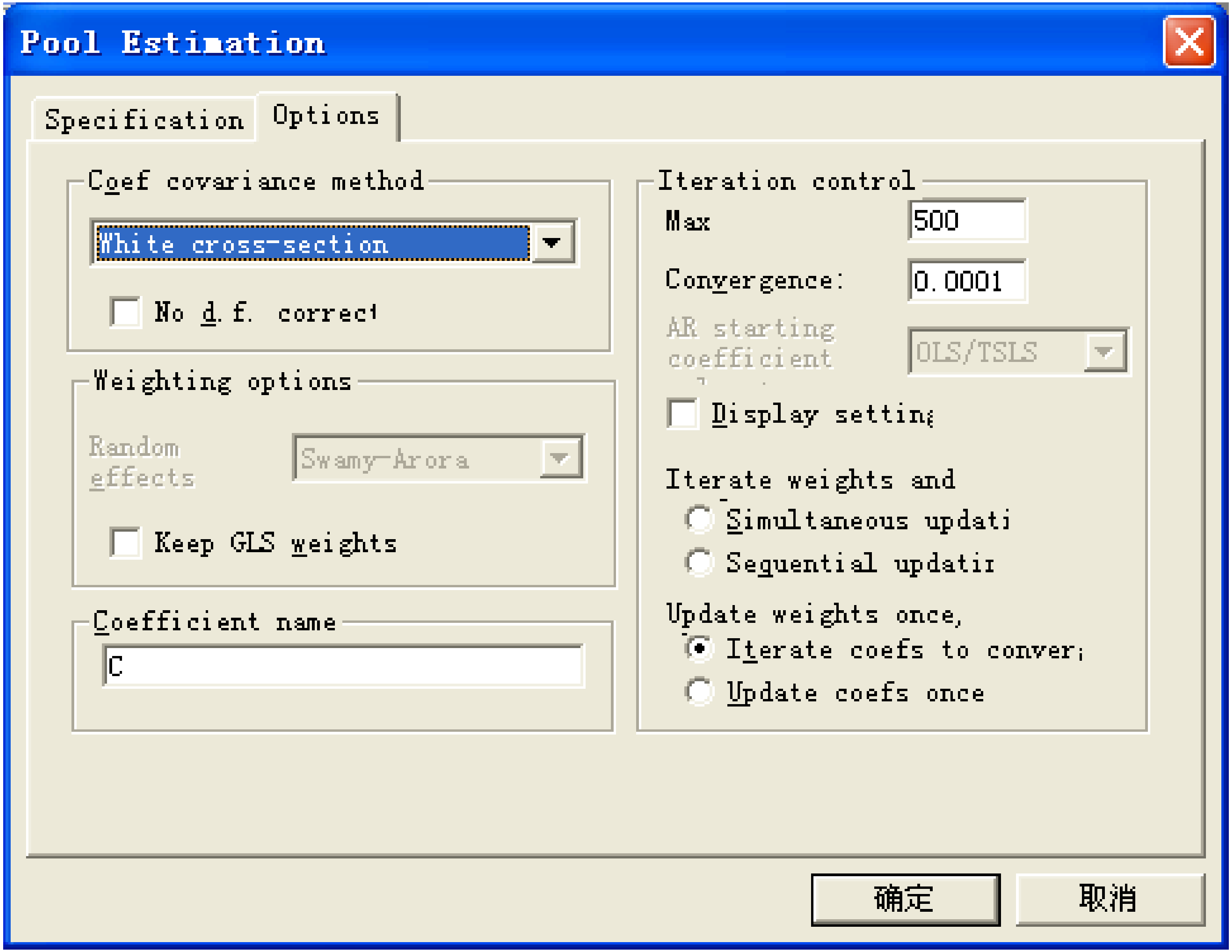

Pool Estimation
Specification
Options
Coef covariance method
Ordinary
No d.f. correct
Weighting options
Random effects
Swamy-Arora
Keep GLS weights
Coefficient name
C
Iteration control
Max
500
Convergence:
0.0001
AR starting coefficient
OLS/TSLS
Display setting
Iterate weights and
Simultaneous updati
Sequential updatir
Update weights once,
Iterate coefs to conver
Update coefs once
确定
取消

10.5 White 异方差协方差
(White Heteroskedasticity Covariance)

EViews能估计那些广义异方差性强的协方差。这种形式的异方差性比上面介绍的截面异方差性更普遍，因为一个截面成员内的方差可以随时间不同。

要得到怀特标准差和协方差，点Options按钮，选择Coef covariance method。EViews给出了一个下拉列表，列表中包含8种选项。默认的是最上方的Ordinary项，对应式(10.3.7) 和式(10.3.8)给出的系数协方差形式。在此下拉列表中的另外7种系数协方差形式参见10.5节的10.5.1~ 10.5.5式。

注意此选项不适用于随机影响估计。


Pool Estimation
Specification
Options
Coef covariance method
White cross-section
No d.f. correct
Weighting options
Random effects
Swamy-Arora
Keep GLS weights
Coefficient name
C
Iteration control
Max
500
Convergence:
0.0001
AR starting coefficient
OLS/TSLS
Display setting
Iterate weights and
Simultaneous updati
Sequential updatir
Update weights once,
Iterate coefs to conver
Update coefs once
确定
取消

10.6 面板数据的单位根检验和协整检验

10.6.1 面板数据的单位根检验

在单序列单位根检验方法的基础上，本节将简要地介绍五种面板数据的单位根检验方法。

对面板数据考虑下面的AR(1)过程：

$$y_{it} = \rho_i y_{it-1} + x_{it}'\delta_i + u_{it}$$

$$i = 1, 2, \ldots, N \quad t = 1, 2, \ldots, Ti \qquad (10.6.1)$$

其中：x_{it} 表示模型中的外生变量向量，包括各个体截面的固定影响和时间趋势。N 表示个体截面成员的个数，T_i 表示第 i 个截面成员的观测时期数，参数 ρ_i 为自回归的系数，随机误差项 u_{it} 相互满足独立同分布假设。可见，对于式（10.6.1）所表示的AR(1)过程，如果$|\rho_i|<1$，则对应的序列 y_i 为平稳序列；如果$|\rho_i|=1$，则对应的序列y_i为非平稳序列。

根据对式（**10.6.1**）中参数ρ_i的不同限制，可以将面板数据的单位根检验方法划分为两大类**:**

一类为相同根情形下的单位根检验，这类检验方法假设面板数据中的各截面序列具有相同的单位根过程（**common unit root process**），即假设式（**10.6.1**）中的参数ρ_i 满足 $\rho_i = \rho$（ $i = 1, 2, \ldots, N$ ）；

另一类为不同根情形下的单位根检验，这类检验方法允许面板数据中的各截面序列具有不同的单位根过程（**individual unit root process**），即允许参数ρ_i 跨截面变化。

1．相同根情形下的单位根检验

（1）LLC检验

（2）Breitung检验

（3）Hadri检验

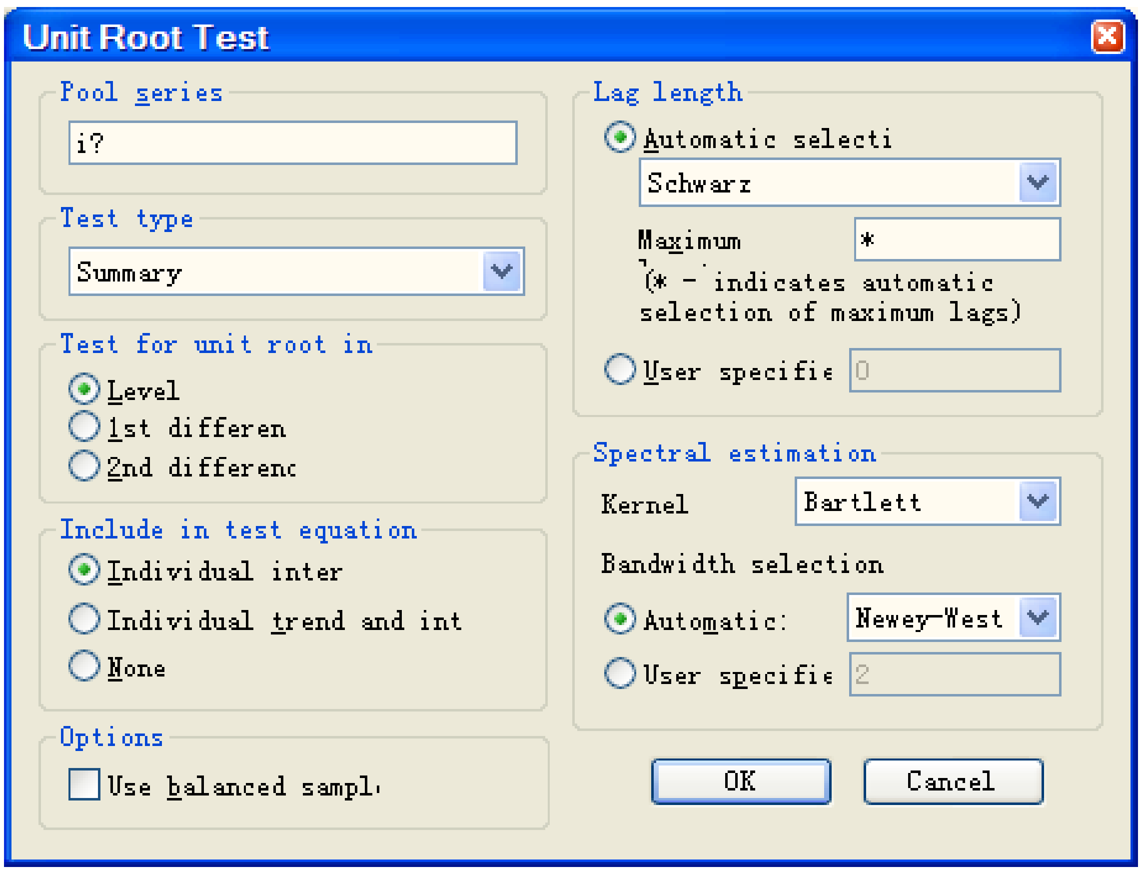
2．不同根情形下的单位根检验

（1）Im-Pesaran-Skin检验

（2）Fisher-ADF检验和Fisher-PP检验

Pool序列的单位根检验

在Pool对象中，对ADF、PP等单位根检验方法均可以实现。在Pool对象的工具栏中，选择View/Unit Root Test，并输入相应的Pool序列名，可以实现ADF、PP等多种方法下的Pool序列的单位根检验。



例10.4中I?的水平变量的所有方法的单位根检验结果:

Pool Unit Root Test on I?

Pool unit root test: Summary
Date: 08/07/07 Time: 10:25
Sample: 1935 1954
Series: I_GM, I_CH, I_GE, I_WE, I_US
Exogenous variables: Individual effects
Automatic selection of maximum lags
Automatic selection of lags based on SIC: 0 to 4
Newey-West bandwidth selection using Bartlett kernel

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	2.76785	0.9972	5	87
Breitung t-stat	-1.77320	0.0381	5	82
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	2.89169	0.9981	5	87
ADF - Fisher Chi-square	4.00150	0.9473	5	87
PP - Fisher Chi-square	7.05649	0.7201	5	95
Null: No unit root (assumes common unit root process)				
Hadri Z-stat	6.62821	0.0000	5	100

** Probabilities for Fisher tests are computed using an asympotic Chi -square distribution. All other tests assume asymptotic normality.

各种方法的结果(除Breitung检验 外)都接受原假设， I?存在单位根，是非平稳的。

例10.4中I?的一阶差分变量的所有方法的单位根检验结果:

Pool Unit Root Test on D(I?)

Pool unit root test: Summary
Date: 08/07/07 Time: 10:28
Sample: 1935 1954
Series: I_GM, I_CH, I_GE, I_WE, I_US
Exogenous variables: Individual effects
Automatic selection of maximum lags
Automatic selection of lags based on SIC: 0 to 4
Newey-West bandwidth selection using Bartlett kernel

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-4.07224	0.0000	5	82
Breitung t-stat	-3.46305	0.0003	5	77
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-5.37837	0.0000	5	82
ADF - Fisher Chi-square	47.9544	0.0000	5	82
PP - Fisher Chi-square	42.5813	0.0000	5	90
Null: No unit root (assumes common unit root process)				
Hadri Z-stat	2.92246	0.0017	5	95

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

各种方法的结果都拒绝原假设，所以可以得出结论：I?是I(1)的。

10.6.2 面板数据的协整检验

面板数据的协整检验方法可以分为两大类，一类是建立在Engle and Granger二步法检验基础上的面板协整检验，具体方法主要有Pedroni检验和Kao检验；另一类是建立在Johansen协整检验基础上的面板协整检验。

1．Pedroni检验

2．Kao检验

3．Johansen面板协整检验

例10.6 城镇居民消费和收入的面板数据的协整关系检验

对例10.5中城镇居民消费和收入的面板数据（样本区间为1991-2003年）进行协整关系检验，各检验方法的检验结果由表10.8和表10.9给出。

Pool序列的协整检验

对于面板数据的协整检验，可以使用pool 对象，或者在panel 工作文件的组中完成。在此我们主要介绍如何在pool 对象中实现面板数据的协整检验。

在EViews中打开pool对象，选择Views/ Cointegration Test…，则显示协整检验的对话框。

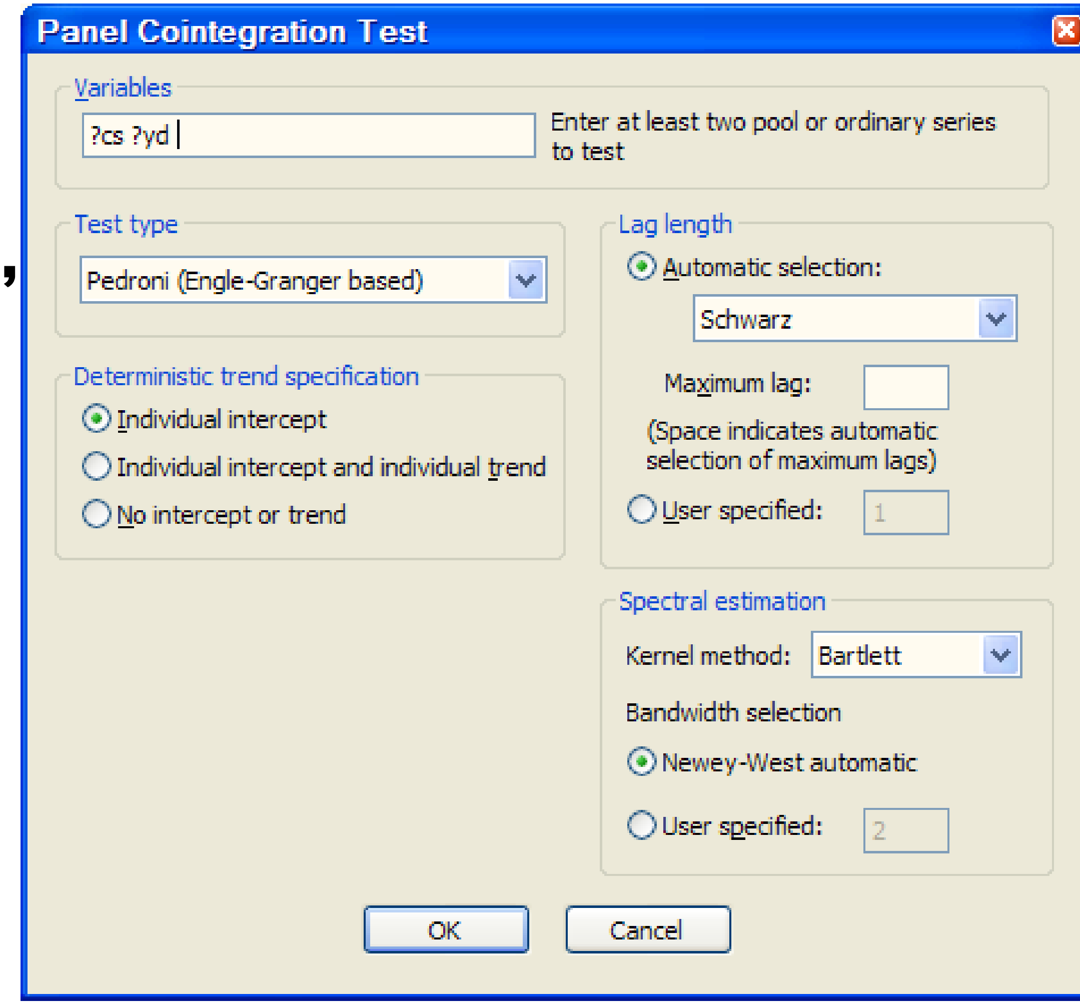


图10.6　面板数据的协整检验的对话框

在对话框的左上角可以输入要检验的变量的名称，同时通过对话框中左下角的下拉菜单可以在Pedroni (Engle-Granger based), Kao (Engle-Granger based), Fisher (combined Johansen). 三种检验方法之间切换。当选择不同的检验方法时，对话框中其余的部分会发生相应的变化。

Pedroni检验：

Pedroni Residual Cointegration Test
Series: ?CS ?YD
Date: 08/07/07 Time: 11:29
Sample: 1991 2003
Included observations: 13
Cross-sections included: 29
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
Lag selection: Automatic SIC with a max lag of 2
Newey-West bandwidth selection with Bartlett kernel

Alternative hypothesis: common AR coefs. (within-dimension)

	Statistic	Prob.	Weighted Statistic	Prob.
Panel v-Statistic	2.099652	0.0440	0.359988	0.3739
Panel rho-Statistic	-3.415758	0.0012	-4.315135	0.0000
Panel PP-Statistic	-5.991403	0.0000	-8.196673	0.0000
Panel ADF-Statistic	-7.835311	0.0000	-9.223079	0.0000

Alternative hypothesis: individual AR coefs. (between-dimension)

	Statistic	Prob.
Group rho-Statistic	-0.837712	0.2809
Group PP-Statistic	-6.990581	0.0000
Group ADF-Statistic	-7.194068	0.0000

表10.8　Kao检验和Pedroni检验结果 (滞后阶数由SIC准则确定)

检验方法	检验假设	统计量名	统计量值（P值）
Kao检验	H_0：$\rho = 1$	ADF	-6.787326（0.0000）*
Pedroni检验	H_0：$\rho = 1$ H_1：$(\rho_i = \rho) < 1$	Panel v-Statistic	2.099652（0.044）*
		Panel rho-Statistic	-3.415758（0.0012）*
		Panel PP-Statistic	-5.991403（0.0000）*
		Panel ADF-Statistic	-7.835311（0.0000）*
	H_0：$\rho = 1$ H_1：$(\rho_i = \rho) < 1$	Group-rho-Statistic	-0.837712（0.2809）
		Group PP-Statistic	-6.990581（0.0000）*
		Group ADF-Statistic	-7.194068（0.0000）*

表10.8 Johansen面板协整检验结果

(选择序列有确定性趋势而协整方程只有截距的情况)

原假设	Fisher联合迹统计量(*p*值)	Fisher联合*λ-max*统计量(*p*值)
0个协整向量	133.4 (0.0000)*	128.7 (0.0000)*
至少1个协整向量	65.74 (0.2266)	65.74 (0.2266)

注：加“*”表示在5%的显著性水平下拒绝原假设而接受备择假设。

上述检验结果检验的样本区间为1991-2003年，从表10.8和表10.9的检验结果可以看出，我国29个省市的城镇居民消费和收入的面板数据之间存在协整关系。

10.7 Pool方程视图和过程

估计出Pool方程后，可以按下述方法检验输出结果：

1. 表达式

选择View/Representations检查输出。EViews把Pool估计成一个方程的系统，每个截面成员一个方程。

Estimation Equations:
=========================

I_GM = C(16) + C(1)*F_GM + C(6)*S_GM + C(11)*I_GM(-1)

I_CH = C(17) + C(2)*F_CH + C(7)*S_CH + C(12)*I_CH(-1)

I_GE = C(18) + C(3)*F_GE + C(8)*S_GE + C(13)*I_GE(-1)

I_WE = C(19) + C(4)*F_WE + C(9)*S_WE + C(14)*I_WE(-1)

I_US = C(20) + C(5)*F_US + C(10)*S_US + C(15)*I_US(-1)

2. 估计输出

View/Estimation Output选项会改变合并估计结果的输出形式。

像其他估计对象一样，可通过在Pool中选择View/Coef Covariance Matrix来检查系数协方差矩阵的估计。

3. 检验

EViews可以进行Pool方程估计参数的系数检验。选择View/Wald Coefficient tests…并输入要检验的限制条件。

4. 残差

选择View/Residuals/Table或View/Residuals/Graph可把残差表示成表格形式或图形形式。EViews会显示每个截面方程的残差。残差命名形式为基本名RES后跟截面识别名。如果想用这些名称存储残差序列，选择Procs/Make Resids。

5. 残差协方差/相关性

可以检查估计残差的同步协方差矩阵和相关矩阵。选择View/Residual，然后选择Correlation Matrix或Covariance Matrix查看矩阵。

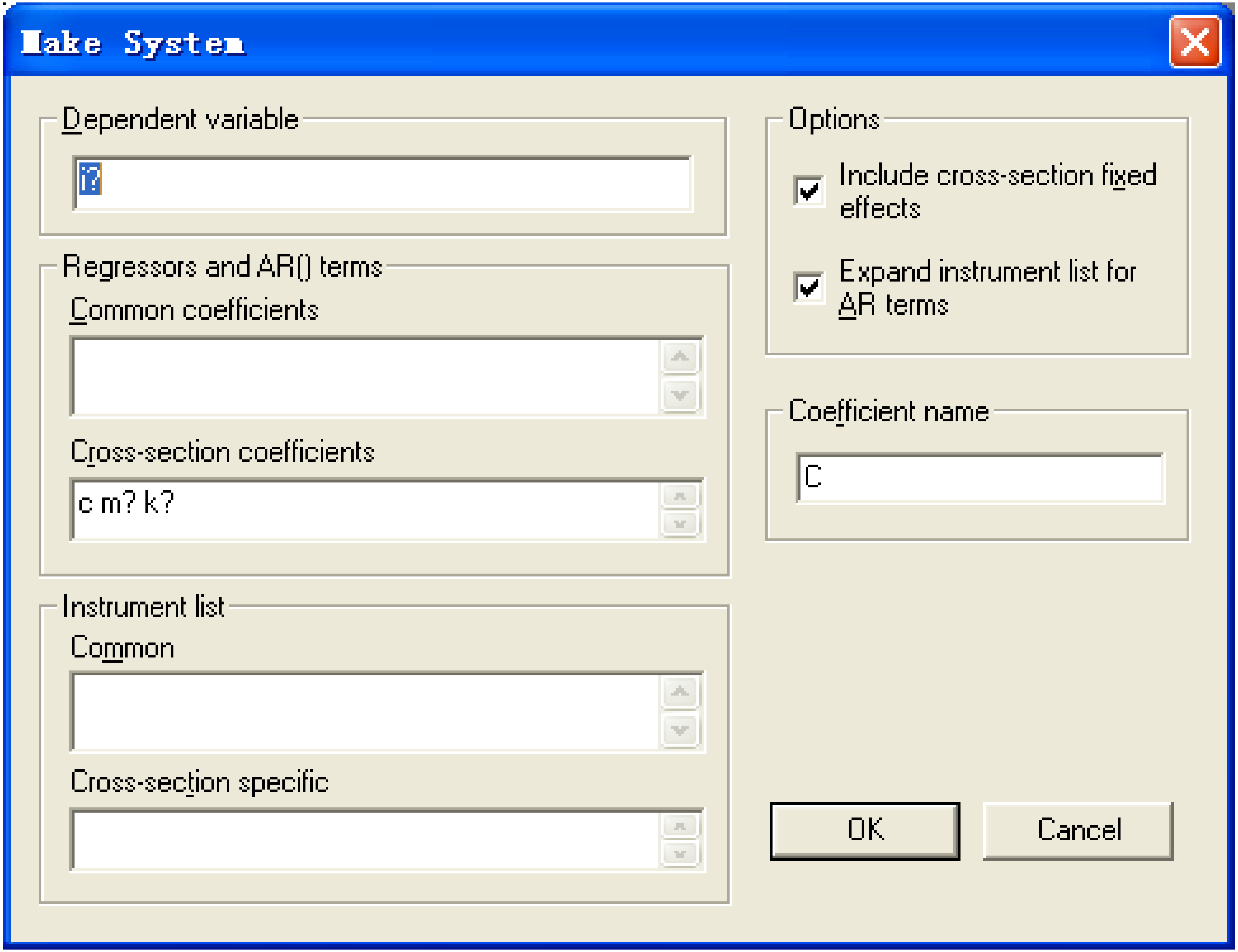
6. 预测

要使用Pool方程进行预测必须先建立一个模型。选择Procs/Make Model建立一个包括所有估计系数的未命名模型对象。模型可以根据需要进行编辑。求解模型能对每个截面成员的因变量进行预测。

7. 用Pool创建系统

可能有些复杂的Panel Data方程不能用Pool对象进行估计。要使用更多的估计方法，如二阶段最小二乘法，三阶段最小二乘法，GMM，或使用任意系数限制，需要用Pool对象创建一个系统对象。可以用一个已估计的Pool创建系统，也可以提供信息从Pool生成系统。系统对象可进一步使用高级技术进行估计。

选择Procs/Make System…，填写对话框。像前面一样输入因变量，规定截距，填写共同系数和截面特定系数变量。还可以填入工具变量。在每个方程中普通变量可作为工具变量，Pool变量会在对应截面方程中把截面特定变量作为工具变量：


Make System
Dependent variable
i?
Regressors and AR() terms
Common coefficients
Cross-section coefficients
c m? k?
Instrument list
Common
Cross-section specific
Options
Include cross-section fixed effects
Expand instrument list for AR terms
Coefficient name
C
OK
Cancel

System: UNTITLED Workfile: 10_5\21_1

View | Proc | Object | Print | Name | Freeze | MergeText | Estimate | Spec

```
I_GM = C(11) + C(1)*M_GM + C(6)*K_GM

I_CH = C(12) + C(2)*M_CH + C(7)*K_CH

I_GE = C(13) + C(3)*M_GE + C(8)*K_GE

I_WE = C(14) + C(4)*M_WE + C(9)*K_WE

I_US = C(15) + C(5)*M_US + C(10)*K_US
```

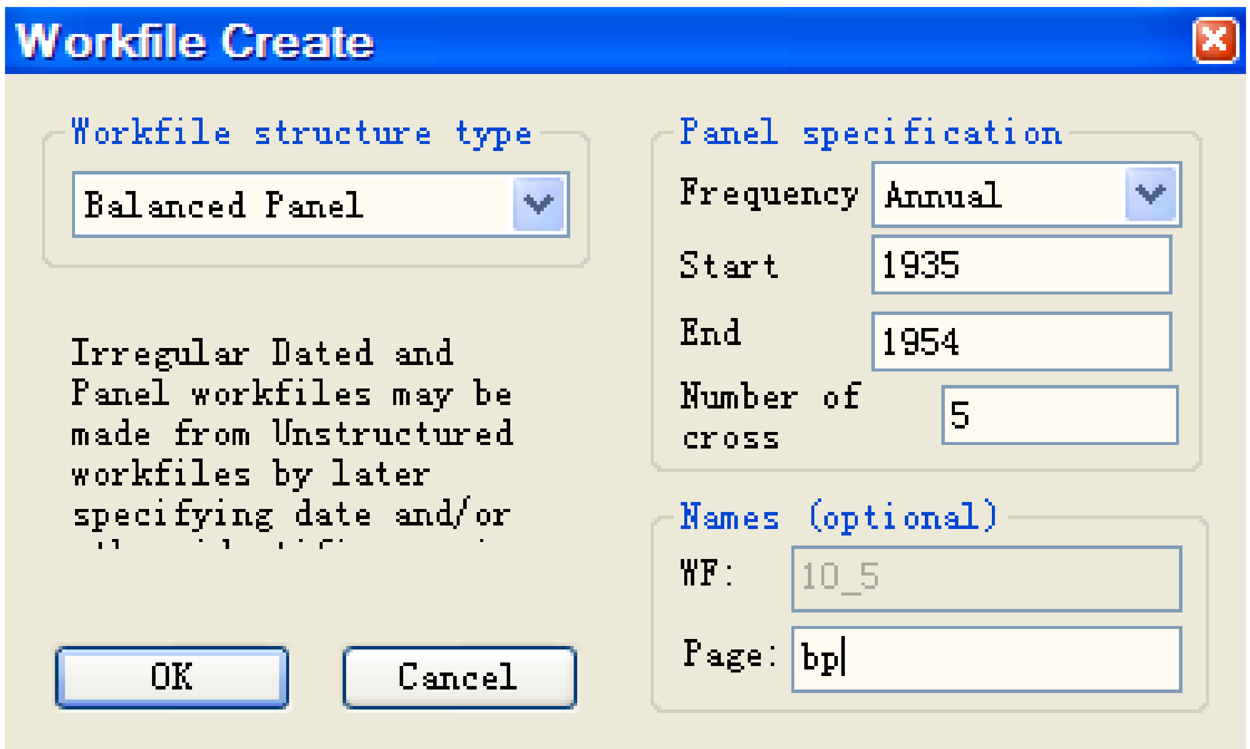
10.8 面板结构的工作文件

对于处理和分析面板数据，除了Pool对象外，EViews还提供了一种特殊结构的工作文件——面板结构的工作文件（Panel Workfile）。当面板数据的截面成员较多但时期较少时，一般都是侧重进行截面分析。可以通过EViews中的面板工作文件对这种“宽而短”的数据进行处理和建模分析。面板结构的工作文件的操作比较复杂，下面仅简要地介绍它的特点。

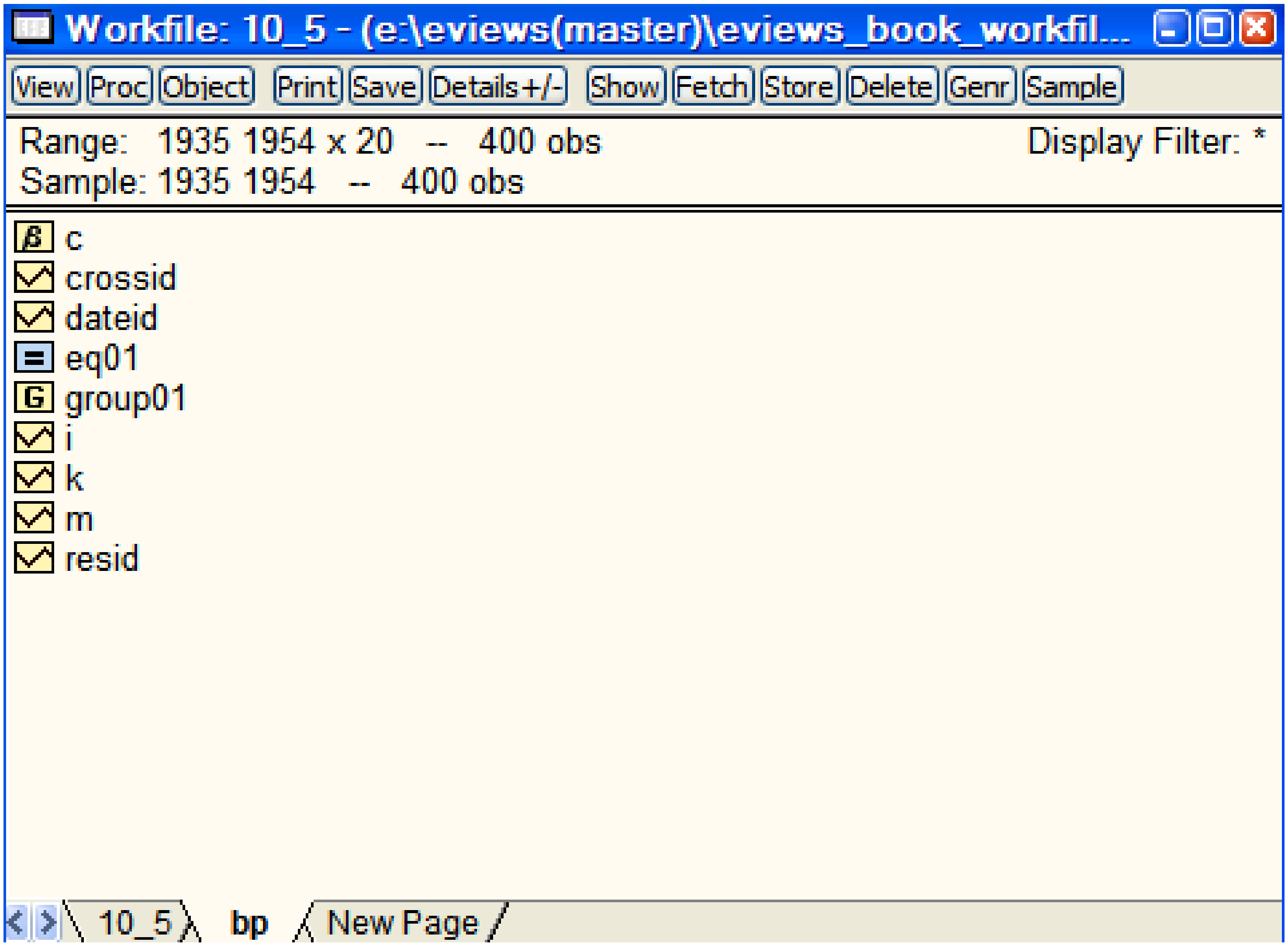
建立平衡面板工作文件

创建一个平衡面板结构，在下拉列表中选择Balanced Panel，选定频率（Frequency），输入起始日期（Start date）和终止日期（End date）以及截面成员的个数（Number of cross）。可以命名工作文件和命名工作文件页，单击OK按钮。EViews将创建一个给定频率的平衡面板工作文件，使用特定的起始和终止日期以及截面成员的个数。

下面的例子中，EViews创建了一个5个截面成员、固定频率、年度面板工作文件，观测值起始于1935年，终止于1954年。



例10.5建立的平衡面板工作文件:

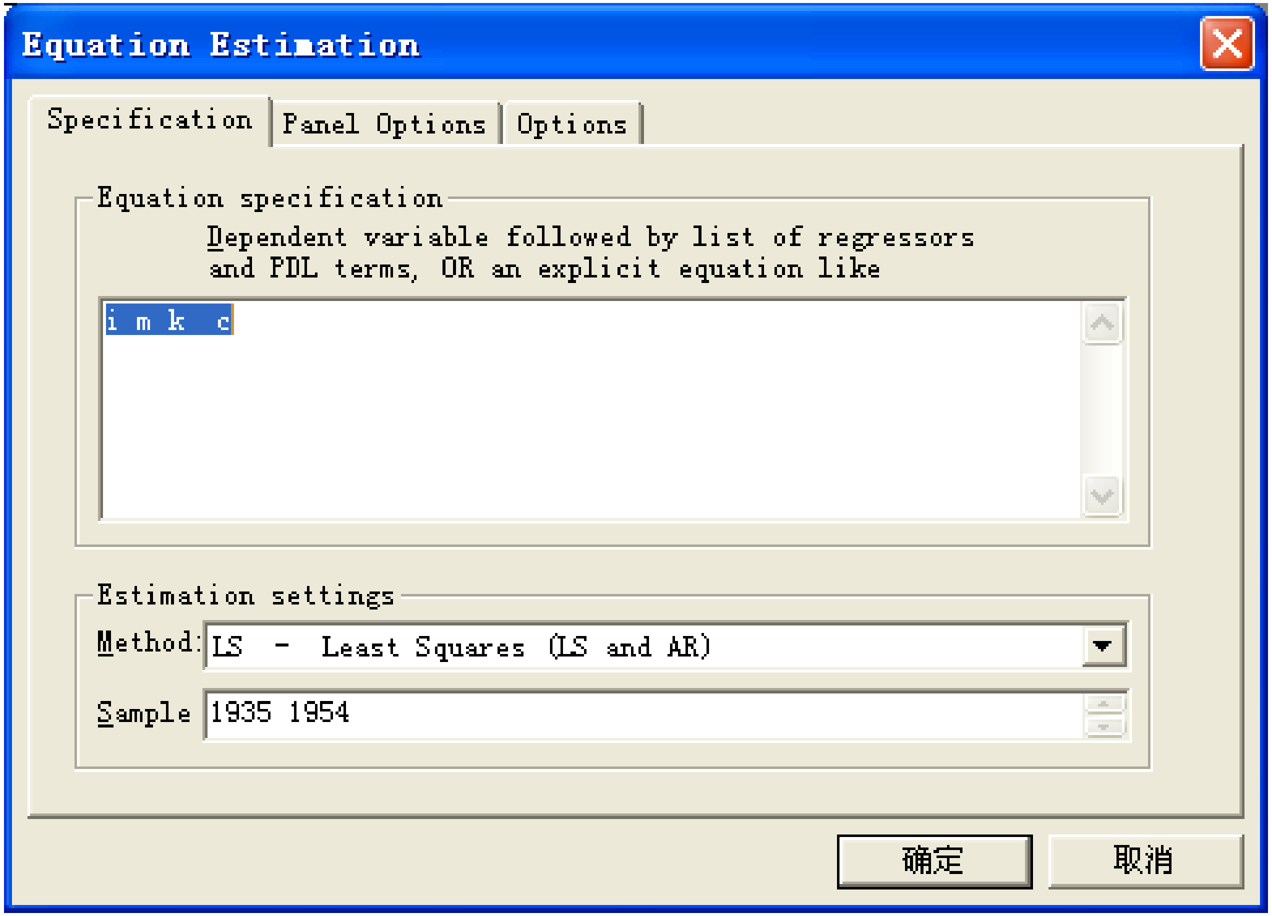


在面板工作文件中，数据是以堆积的形式存放的，故称其为<u>堆积面板数据</u>。工作文件中每一个序列的各期观测值都具有二维信息，即每个序列的观测值标签都由两部分构成，一部分反映观测值的截面个体信息，另一部分反映观测值的时期信息。

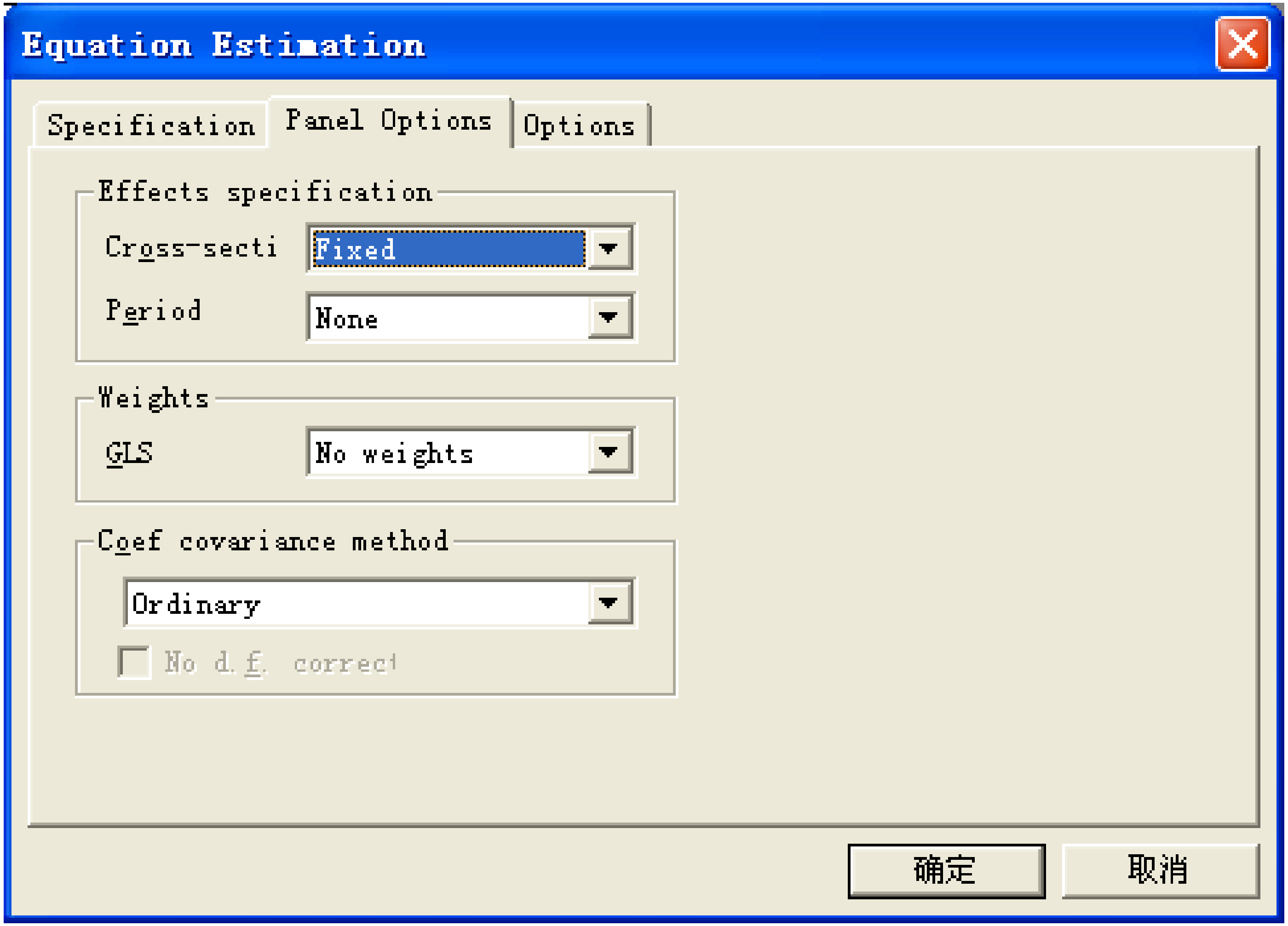
例10.5平衡面板工作文件中的序列 I, M, K 是以堆积形式存放的:

obs	I	M	K
obs	I	M	K
CH - 35	40.29000	417.5000	10.50000
CH - 36	72.76000	837.8000	10.20000
CH - 37	66.26000	883.9000	34.70000
CH - 38	51.60000	437.9000	51.80000
CH - 39	52.41000	679.7000	64.30000
CH - 40	69.41000	727.8000	67.10000
CH - 41	68.35000	643.6000	75.20000
CH - 42	46.80000	410.9000	71.40000
CH - 43	47.40000	588.4000	67.10000
CH - 44	59.57000	698.4000	60.50000
CH - 45	88.78000	846.4000	54.60000
CH - 46	74.12000	893.8000	84.80000
CH - 47	62.68000	579.0000	96.80000
CH - 48	89.36000	694.6000	110.2000
CH - 49	78.98000	590.3000	147.4000
CH - 50	100.6600	693.5000	163.2000
CH - 51	160.6200	809.0000	203.5000
CH - 52	145.0000	727.0000	290.6000
CH - 53	174.9300	1001.500	346.1000
CH - 54	172.4900	703.2000	414.9000
GE - 35	33.10000	1170.600	97.80000
GE - 36	45.00000	2015.800	104.4000
GE - 37	77.20000	2803.300	118.0000

面板工作文件其方程建立过程，与普通工作文件中的方程建立过程相同，只是在方程说明窗口中增加了与Pool对象中方程说明项类似的方程面板结构说明页，同时在估计方法中增加了Pool对象所没有的广义矩估计（GMM）法，用来估计模型解释变量中含有滞后因变量的变截距模型。



只能求解变截距模型:


Equation Estimation
Specification
Panel Options
Options
Effects specification
Cross-secti
Fixed
Period
None
Weights
GLS
No weights
Coef covariance method
Ordinary
No d.f. correct
确定
取消

Dependent Variable: I
Method: Panel Least Squares
Date: 11/29/06 Time: 16:09
Sample: 1935 1954
Cross-sections included: 5
Total panel (balanced) observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M	0.105980	0.015891	6.669182	0.0000
K	0.346660	0.024161	14.34781	0.0000
C	-62.59439	29.44191	-2.126030	0.0361

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.937454	Mean dependent var	248.9570
Adjusted R-squared	0.933419	S.D. dependent var	267.8654
S.E. of regression	69.11798	Akaike info criterion	11.37694
Sum squared resid	444288.4	Schwarz criterion	11.55930
Log likelihood	-561.8468	F-statistic	232.3194
Durbin-Watson stat	0.806789	Prob(F-statistic)	0.000000

固定影响模型的变截距的系数要在视图（view）里的Fixed/Random effects中看到。

Cross-section Fixed Effects

	ID01	Effect
1	_CH	33.22081
2	_GE	-179.5764
3	_GM	-13.47235
4	_US	155.1329
5	_WE	4.694980

一般情况下，面板工作文件中数据的分析和处理与其他文件中数据的分析和处理过程是一致的。在面板工作文件中能够实现对堆积数据的显示图、单位根检验等操作，其单位根检验过程和检验结果同组序列的单位根检验过程和结果基本类似。

虽然利用EViews中的面板工作文件可以对堆积数据进行多种分析和处理，但是对堆积形式的序列的处理仍具有一定的局限性。目前还不能对堆积形式的序列进行季节调整也不能利用该形式的序列进行VAR模型和VEC模型的估计（当前已可以）。