计量分析与 STATA 应用

钟经樊 连玉君

关于作者： 钟经樊 台湾中央研究院 经济研究所

连玉君 中山大学 岭南学院 金融系

中文版本： 版本 2.0，二〇一〇年六月

这份文档是我们即将出版的书稿，目前免费提供给中山大学岭南学院的师生使用。
发布这份文档的目的有二:
其一，用做授课讲义，帮助岭南学院的同学们学习 STATA;
其二，恳请大家对书稿提出修改意见，包括书稿的结构安排、表述错误，以及错别字等细节。

目录

第八章

面板模型及 STATA 应用

8.1 简介

面板数据 (Panel Data)，简言之，是时间序列和截面数据的混合。 严格地讲是指对一组个体 (如居民、国家、公司等) 连续追踪观察多期得到的资料。 所以很多时候也称其为“追踪资料”。相对于单纯的截面资料和时序资料， 这种特殊的资料结构，使得我们可以建立更为符合实际的计量模型。 当然，由于资料结构的复杂性，也对模型的估计和分析提出了更高的要求。例如，由于面板资料是对特定的个体追踪多年得到的，此时，观察值之间彼此独立的假设可能不再成立。 这会在很大程度上增加分析的难度，在非线性模型或动态模型中更是如此。

近年来，由于面板数据获得变得相对容易，使得其应用范围也不断扩大。 而关于面板数据模型的计量理论也几乎涉及到了以往截面分析和时间序列分析中所有可能出现的主题， 如近年来发展出的面板数据向量自回归模型 (Panel VAR)、面板数据单位根检验 (Panel Unit Root test)、面板数据协整分析 (Panel Cointegeration)、面板数据门槛模型 (Panel Threshold) 等， 都是在现有截面分析和时间序列分析中的热点主题的基础上发展起来的。

使用面板数据主要有以下几方面的优点：

- 便于控制个体的异质性。比如，我们在研究全国 30 个省份居民人均消费青岛啤酒的数量时， 可以选取居民的收入、当地的啤酒价格、上一年的啤酒消费量等变量作为解释变量。 但同时我们也会认为 民族习惯、[1] 风俗文化、[2] 广告投放等因素也会显著地影响居民的啤酒消费量。 对于特定的个体而言，前两种因素不会随时间的推移而有明显的变化，通常称为个体效应。 而广告的投放往往通过电视或广播，我们可以认为在特定的年份所有省份所接受的广告投放量是相同的， 通常称为“时间效应”。 这些因素往往因为难以

[1] 如宁夏属于回族自治区，那里的回民因为信仰伊斯兰教，所以不允许饮酒的， 而生活 在宁夏的许多汉民也往往因为自己的回民朋友无法饮酒而无形中减少了啤酒的消费量。

[2] 如中国南部地区啤酒的消费量比较大，而北方很多地区只有在夏天才会饮用较多的啤酒， 冬天他们一般只喝白酒。

获得数据或不易衡量而无法进入我们的模型， 在截面分析中者往往会引起遗漏变量的问题。 而面板数据模型的主要用途之一就在于处理这些不可观测的个体效应或时间效应。

- 包含的信息量更大，降低了变量间共线性的可能性，增加了自由度和估计的有效性。
- 便于分析动态调整。

本章主要介绍目前文献中常用的面板数据模型。 第 8.2 节介绍两种基本的静态面板模型：固定效应模型和随机效应模型。 第 8.4 节介绍异方差和序列相关稳健性估计量。 第 8.6 节介绍动态面板模型，包括 FD-GMM 和 SYS-GMM 两种估计方法。 有关面板数据模型的更为详尽的介绍，请参考 Baltagi (2001), Wooldridge (2002), Hsiao (2003) 以及 Arellano (2003).

8.2 静态面板数据模型

我们一般所说的静态面板数据模型，是指解释变量中不包含被解释变量的滞后项 (通常为一阶滞后项) 的情形。但严格地讲，随机干扰项服从某种序列相关 (如 AR(1)，AR(2)，MA(1) 等) 的模型也不是静态模型。动态模型和静态模型在处理方法上往往有较大的差异。 本节中我们重点介绍两种最为常用的静态模型 — 固定效应模型 (Fixed Effect Model) 和随机效应模型 (Random Effect Model)。

对于面板数据，在模型设定过程中，我们通常采用 i 表示个体[3] ($i = 1, 2, \cdots, N$)， 用 t 表示时间 ($t = 1, 2, \cdots, T$)。最直接的想法可能是设定如下线性模型：

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

其中，$\boldsymbol{\beta}_{it}$ 用于衡量个体 i 在第 t 时点，$\mathbf{x}_{it}$ 对 y_{it} 的边际影响。 显然，这个模型的设定过于一般化，因为我们假设 α_{it} 和 $\boldsymbol{\beta}_{it}$ 都会随着个体 i 和时点 t 发生变化。 为此，需要对上述模型做进一步限定，例如，假设 $\boldsymbol{\beta}_{it}$ 为常数，即 $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}$， 但常数项可以随着个体的不同而有所差异，可以表示如下：

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{8-1}$$

$\mathbf{x}_{it}$ 为 $K \times 1$ 列向量 (不包含常数项)， K 为解释变量的个数，$\boldsymbol{\beta}$ 为 $K \times 1$ 系数列向量。 这意味着，对于所有的个体和时点，x 的边际效果都相同，但个体 i 的平均水平不同于个体 j。 对于特定的个体 i 而言，a_i 表示那些不随时间改变的影响因素， 而这些因素在多数情况下都是无法直接观测或难以量化的， 如个人的消费习惯、企业文化和经营风格、国家的社会制度等， 我们一般称其为“个体效应”(individual effects)。 一般情况下，我们假设 ε_{it} 具有独立同分布的特征，[4] 均值为 0，方差为 σ_ε^2。 由于模型 (8-1) 中，α_i 可以看做随个体变化的截距项， 那么也

[3]这里的个体可以是公司，国家，行业，或个人。

[4]也就是说，不同个体之间，以及同一个个体的不同时间点上，干扰项 ϵ_{it} 都是不相关的。

就可以进一步将 α_i 视为 N 个未知参数。也正因为如此，模型 (8-1) 通常被称为“固定效应模型”(Fixed effects model)。

与固定效应模型相对应的另一种设定方式是所谓的“随机效应模型”(Random effects model)。该模型假设个体的截距项虽然有差异，但不是固定的，而是从一个服从均值为 μ，方差为 σ_μ^2 的分布中随机抽取的。模型设定如下：

$$y_{it} = \mu + \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \tag{8-2}$$

该模型的干扰项包含两个部分：不随时间改变的的干扰项 α_i 和通常意义上的 (可以随时间改变的) 干扰项 ε_{it}。[5] 需要说明的是，由于我们在模型 (8-2) 中增加了截距项 μ，此时 α_i 的均值为 0。

对比二者的设定方式可知，两种模型的差异主要反映在对“个体效应”的处理上。固定效应模型假设个体效应在组内是固定不变的，个体间的差异反映在每个个体都有一个特定的截距项上；随机效应模型则假设所有的个体具有相同的截距项，个体间的差异是随机的，这些差异主要反应在随机干扰项的设定上。基于此，一种常见的观点认为，当我们的样本来自一个较小的母体时，我们应该使用固定效应模型，而当样本来自一个很大的母体时，应当采用随机效应模型。比如在研究中国地区经济增长的过程中，我们以全国 28 个省区为研究对象，可以认为这 28 个省区几乎代表了整个母体。同时也可以假设在样本区间内，各省区间的经济结构、人口素质等不可观测的特质性因素是固定不变的，因此采用固定效应模型是比较合适的。而当我们研究西安市居民的消费行为时，即使样本数为 10000 人，相对于西安市 600 万人口的母体而言仍然是个很小的样本。此时，可以认为不同的居民在个人能力、消费习惯等方面的差异是随机的，此时采用随机效应模型较为合适。

遗憾的是，很多情况下，我们并不能明确地区分我们的样本来自一个较大母体还是较小的母体。因此有些学者认为，区分固定效应模型和随机效应模型应当看使用二者的假设条件是否满足。由于随机效应模型把个体效应 α_i 设定为干扰项的一部分，所以就要求解释变量与个体效应不相关。而在固定效应模型中，个体效应 α_i 被视为 N 个待估参数，并不受这个假设条件的限制。因此，如果我们的检验结果表明该假设满足，那么就应该采用随机效应模型，因为它更为有效 (所需估计的参数较少)，反之，就需要采用固定效应模型。

另外，有些学者认为具体采用哪一种模型主要决定于我们的分析目的。如果主要目的在于估计模型的参数，而模型中个体的数目又不是很大的情况下，采用固定效应模型是个不错的选择，因为它非常容易估计。但当我们需要对模型的误差成分进行分析时 (通常分解为长期效果和短期效果)，就只能采用随机效应模型。在这种情况下，即使模型中的部分解释变量与个体效应相关，我们仍然可以通过工具变量法对模型进行估计。

简言之，两种模型有各自的优缺点和适用范围，在实证分析的过程中，我们一方面要根据分析的目的选择合适的模型，同时也要以 8.2.3 节中介绍的假设检验方法为基础进行模型筛选。

[5] 也正因为如此，该模型也被称为“误差成分模型”(error components model)。

8.2.1 固定效应模型

模型的假设条件

在估计模型 (8-1) 时，通常需要设定如下两个基本假设：[6]

假设 1 :

$$\mathrm{E}(\boldsymbol{\varepsilon}_{it}|\mathbf{x}_{it},\alpha_i)=0$$

假设 2 :

$$\mathrm{Var}\left(\boldsymbol{\varepsilon}_{it}|\mathbf{x}_{it},\alpha_i\right)=\sigma^2$$

假设 1 表明干扰项 $\boldsymbol{\varepsilon}$ 与解释变量 $\mathbf{x}$ 的当期观察值、前期观察值以及未来的观察值均不相关，也就是说模型中所有的解释变量都是严格外生的。假设 2 就是一般的同方差假设，在此假设下模型 (8-1) 的 OLS 估计是 BLUE (Best Linear Unbiased Estimator) 的。当此假设无法满足时，我们就需要处理异方差或序列相关以便得到稳健性估计量。

最小二乘虚拟变量估计量

在假设 1 和假设 2 同时成立的情况下，我们可以采用虚拟变量的方式将模型 (8-1) 重新表述如下：

$$y_{it}=\sum_{j=1}^{N}\alpha_i d_{ij}+\mathbf{x}_{it}'\boldsymbol{\beta}+\varepsilon_{it} \tag{8-3}$$

其中，若 $i=j$ 时，$d_{ij}=1$，否则为 0，即模型中包含了 N 个反应个体特征的虚拟变量。参数 $\alpha_1,\alpha_2,\cdots,\alpha_N$ 以及 $\boldsymbol{\beta}$ 可以采用普通最小二乘法 (OLS) 估计得到。由此得到的 $\boldsymbol{\beta}$ 系数称为“最小二乘虚拟变量估计量”(least squares dummy variable (LSDV) estimator)。当 N 较小时，采用这种方法非常简便，所有能执行 OLS 估计的计量软件都可以完成固定效应模型的估计。然而，当 N 比较大时，模型中将包含 $N+K$ 个解释变量，计算的工作量往往很大，对于 N 相当大的情况 (如 $N=100,000$)，一般的计算机都无法胜任。因此，有必要先进行一些变换以消除固定效应，进而对简化的模型进行估计，随后三个小节介绍的法都是基于此目的进行的。

从模型 (8-3) 的设定形式可知，对于固定效应模型而言，由于 α_i 不随时间变化，所以 $\mathbf{x}_{it}$ 中不能包含不随时间改变的变量，如性别、种族、出生地等，因为这些变量都会与 α_i 存在完全共线性。[7]

组内估计量

1. 基本思想

[6]一般应用中，我们也常采用如下两个相对较弱的假设。假设 1′: $\mathrm{E}\left(\boldsymbol{\varepsilon}_i|\mathbf{x}_i\right)=0$ 和 假设 2′: $\mathrm{Var}\left(\boldsymbol{\varepsilon}_i|\mathbf{x}_i\right)=\sigma^2\mathbf{I}_T$。

[7]若 $\mathbf{x}_{it}$ 包含了任何不随时间变化的变量，STATA 会自动将这些变量删除。

给定 (8-1) 式，我们可以进一步得到如下模型：

$$\bar{y}_i = \alpha_i + \bar{\mathbf{x}}_i\boldsymbol{\beta} + \bar{\varepsilon}_i \tag{8-4}$$

其中，$\bar{y}_i = (1/T_i)\sum_{t=1}^{T_i} y_{it}$，$T_i$ 表示第 i 个个体的观察区间 (如 $T_1 = 5$ 年，$T_2 = 3$ 年)。$\bar{\mathbf{x}}_i$ 和 $\bar{\varepsilon}_i$ 的定义方式与此相同。换言之，模型 (8-4) 表示个体 i 在样本观察区间内的平均值之间的关系。

模型 (8-1) 与模型 (8-4) 相减可以去除个体效应 α_i：[8]

$$(y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i) \tag{8-5}$$

若设定 $\dot{y}_{it} = (y_{it} - \bar{y}_i), \dot{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$，以及 $\dot{\varepsilon}_{it} = (\varepsilon_{it} - \bar{\varepsilon}_i)$ 则我们只需对如下模型执行 OLS 估计即可得到 $\boldsymbol{\beta}$ 的估计值：

$$\dot{y}_{it} = \dot{\mathbf{x}}_{it}'\boldsymbol{\beta} + \dot{\varepsilon}_{it} \tag{8-6}$$

简言之，要得到固定效应模型 (8-1) 的估计系数，只需要从原始数据中间去其组内平均值，进而对变换后的组内差分模型 (8-6) 执行 OLS 估计即可。为此，该估计量也成为“组内估计量”(within group estimator)，记为 $\hat{\boldsymbol{\beta}}_{WG}$。

2. 更为严格的推导过程[9]

模型 (8-1) 可以采用向量的形式表示为：

$$\mathbf{y}_i = \alpha_i \mathbf{1}_T + \mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \tag{8-7}$$

其中，$\mathbf{y}_i = (y_{i1}y_{i2}, \cdots y_{iT})'$，$\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \cdots, \mathbf{x}_{iT})'$，$\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$，$\mathbf{1}_T$ 是一个所有元素都为 1 的 $T \times 1$ 列向量。将所有观察值进行堆叠，模型 (8-1) 可用矩阵形式表示为：

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{8-8}$$

其中，$\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \cdots, \mathbf{y}_N')'$，$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \cdots, \boldsymbol{\varepsilon}_N')'$，均为 $NT \times 1$ 向量，$\mathbf{D} = \mathbf{I}_N \otimes \mathbf{1}_T$，$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots a_N)'$。

需要注意的是，在模型 (8-8) 中，$\mathbf{D}$ 项实际上对应着 N 个虚拟变量，因此，模型 (8-8) 等价于在模型 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 中加入 N 个虚拟变量。为了避免共线性问题，解释变量 $\mathbf{X}$ 中不应再包含常数项。[10]

[8] 多数面板模型都具有大 N 小 T 结构，此时我们重点关心的仍然是系数 $\boldsymbol{\beta}$。在 N 较小的面板模型中，α_i 的估计值可能成为分析的重点。此时可以采用 LSDV 进行估计，或采用 (8-13) 式获得 α_i 的估计值。

[9] 对于矩阵运算不熟悉的读者可以跳过此节，这并不影响你对固定效应模型估计方法的理解。

[10] 当然，我们也可以在 $\mathbf{X}$ 中加入常数项，但此时要同时加入约束条件：$\sum_{i=1}^{N} a_i = 0$。这样我们估计出的个体效应 $\hat{a}_i$ 就应当解释为个体 i 的相对截距项，而不是前面得到的绝对截距项。STATA8.0 就采取了在 $\mathbf{X}$ 中包含常数项的处理方式。

在正式估计模型之前，我们先定义一些有用的矩阵运算，它们将在后面的分析中反复使用。定义 $\mathbf{DD}' = \mathbf{I}_N \otimes \mathbf{J}_T$，其中，$\mathbf{J}_T = \mathbf{1}_T\mathbf{1}_T'$ 为 $T \times T$ 维矩阵，每个元素均为 1。同时，定义 $\mathbf{P} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \mathbf{I}_N \otimes \bar{\mathbf{J}}_T$，$\bar{\mathbf{J}}_T = (1/T)\mathbf{J}_T$ 是 $T \times T$ 维矩阵，每个元素均为 $1/T$；$\mathbf{Q} = \mathbf{I}_{NT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \mathbf{I}_{NT} - \mathbf{P}$。矩阵 $\mathbf{P}$ 和 $\mathbf{Q}$ 都具有如下性质：

(1) 对称、幂等性: $\mathbf{P}' = \mathbf{P}$，且 $\mathbf{P}^2 = \mathbf{P}$；

(2) 正交性: $\mathbf{PQ} = \mathbf{0}$；

(3) 和为单位矩阵: $\mathbf{P} + \mathbf{Q} = \mathbf{I}_{NT}$。

我们可以从上述三个性质中的任意两个推导出第三个。易于证明，$\mathbf{QD} = \mathbf{0}$，因此，我们可以通过在等式 (8-8) 两边同时左乘 $\mathbf{Q}$ 以消除固定效应：

$$\mathbf{Qy} = \mathbf{QX}\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\varepsilon} \tag{8-9}$$

易于证明，(8-9) 式变换后的结果其实就是前文提到的 (8-5) 式。变换后的模型的 OLS 估计量为：

$$\hat{\boldsymbol{\beta}}_{WG} = (\mathbf{X}'\mathbf{QX})^{-1}\mathbf{X}'\mathbf{Qy} \tag{8-10}$$

方差估计量为：

$$\mathrm{Var}\,(\hat{\boldsymbol{\beta}}_{WG}) = \sigma^2(\mathbf{X}'\mathbf{QX})^{-1} \tag{8-11}$$

显然，σ^2 的一致估计量为：

$$\hat{\sigma}^2 = \frac{1}{NT-N-K}(\mathbf{Qy} - \mathbf{QX}\hat{\boldsymbol{\beta}}_{WG})'(\mathbf{Qy} - \mathbf{QX}\hat{\boldsymbol{\beta}}_{WG}) \tag{8-12}$$

个体效应的估计值为：

$$\hat{a}_i = \bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{WG} \tag{8-13}$$

一阶差分估计量

除了上述通过“组内去心”的办法消除固定效应外，还可以通过一阶差分的方式去除固定效应。对 (8-1) 式取一阶差分，得到

$$\begin{aligned} \triangle y_{i2} &= \triangle \mathbf{x}_{i2}\boldsymbol{\beta} + \triangle \boldsymbol{\varepsilon}_{i2} \\ &\vdots \\ \triangle y_{iT} &= \triangle \mathbf{x}_{iT}\boldsymbol{\beta} + \triangle \boldsymbol{\varepsilon}_{iT} \end{aligned} \tag{8-14}$$

采用矩阵形式可表示为

$$\mathbf{By}_i = \mathbf{Bx}_i\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\varepsilon}_i \tag{8-15}$$

其中，

$$\mathbf{B}=\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}_{(T-1)\times T} \tag{8-16}$$

对所有观察值进行堆叠，得到

$$(\mathbf{I}_N\otimes\mathbf{B})\mathbf{y}=(\mathbf{I}_N\otimes\mathbf{B})\mathbf{X}+(\mathbf{I}_N\otimes\mathbf{B})\boldsymbol{\varepsilon} \tag{8-17}$$

设 $\mathbf{Q}_B=\mathbf{I}_N\otimes\mathbf{B}$，则相应的 OLS 的估计量为：

$$\hat{\boldsymbol{\beta}}_{OLS}=(\mathbf{X}'\mathbf{Q}_B\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_B\mathbf{y} \tag{8-18}$$

根据假设 1 可知，$\mathrm{E}(\boldsymbol{\varepsilon}\mathbf{X})=0$，所以 $\hat{\boldsymbol{\beta}}_{OLS}$ 是 $\hat{\beta}$ 的无偏 估计量，在 N 较大的情况下，$\hat{\boldsymbol{\beta}}_{OLS}$ 也是一致的。由假设 2 可知，$\boldsymbol{\varepsilon}$ 满足同方差假设，且不存在序列相关。但变换后的干扰项 $\mathbf{B}\boldsymbol{\varepsilon}$ 却并不满足同方差的假设，

$$\mathrm{Var}\,(\mathbf{Q}_B\boldsymbol{\varepsilon})=\sigma^2\mathbf{Q}_B\mathbf{Q}_B' \tag{8-19}$$

根据第四章中介绍的 GLS 理论可知，模型 (8-17) 的 GLS 估计量是 BLUE 的，

$$\hat{\boldsymbol{\beta}}_{FD}=[\mathbf{X}\mathbf{Q}_B(\mathbf{Q}_B\mathbf{Q}_B')^{-1}\mathbf{Q}_B\mathbf{X}]^{-1}\mathbf{X}\mathbf{Q}_B(\mathbf{Q}_B\mathbf{Q}_B')^{-1}\mathbf{Q}_B\mathbf{y}. \tag{8-20}$$

易于证明 $\mathbf{Q}_B(\mathbf{Q}_B\mathbf{Q}_B')^{-1}\mathbf{Q}_B=\mathbf{Q}$。[11] 因此，

$$\hat{\boldsymbol{\beta}}_{GLS}\sim\hat{\boldsymbol{\beta}}_{WG}$$

也就是说，采用一阶差分去除“固定效应”后， 再用 GLS 估计差分后的模型得到的 GLS 估计量与我们前面介绍的组内估计是等价的。 由于二者都满足经典回归模型的基本假设，所以都是 BLUE 的。

---

[11]利用矩阵直乘的性质：$(\mathbf{A}\otimes\mathbf{F})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC})\otimes(\mathbf{FD})$， 可以得到 $\mathbf{Q}_B(\mathbf{Q}_B\mathbf{Q}_B')^{-1}\mathbf{Q}_B=\mathbf{I}_N\otimes\mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{B}$。进一步，可以证明 $\mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{B}=\mathbf{I}_T-\bar{J}_T$：由于矩阵

$$\mathcal{H}=\begin{bmatrix} T^{-1/2}\mathbf{1}_T' \\ (\mathbf{BB}')^{-1/2}\mathbf{B} \end{bmatrix}$$

满足 $\mathcal{H}\mathcal{H}'=\mathbf{I}_T$，所以 $\mathcal{H}'\mathcal{H}=\mathbf{I}_T$，即

$$\mathbf{1}_T'\mathbf{1}_T/T+\mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{B}=\mathbf{I}_T$$

因此，$\mathbf{Q}_B(\mathbf{Q}_B\mathbf{Q}_B')^{-1}\mathbf{Q}_B=\mathbf{I}_N\otimes(\mathbf{I}_T-\bar{J}_T)=\mathbf{I}_{NT}-\mathbf{P}=\mathbf{Q}$。

前向正交分解

“前向正交分解”(forward orthogonal debiations, FOD) 法由 Arellano and Bover (1995) 提出。类似于上面介绍的一阶差分法。它也可以去除个体效果，但却不会在变换后的干扰项中引入序列相关问题。虽然在处理静态模型时这种方法略显繁复，但在动态面板数据模型的分析中，该方法显得格外重要。

正交变换基于如下 $(T-1)\times T$ 矩阵

$$\mathbf{A}=(\mathbf{B}\mathbf{B}')^{-1/2}\mathbf{B}$$

如果将 $(\mathbf{B}\mathbf{B}')^{-1/2}$ 视为裘拉斯基 (Cholesky) 分解的上三角阵，那么 A 矩阵可表示为

$$\mathbf{A}=\mathrm{diag}[(T-1)/T(T-2)/(T-1),\cdots 1/2]^{-1/2}\mathbf{A}^{\dagger}$$

其中，

$$\mathbf{A}^{\dagger}=\begin{bmatrix} 1 & (1-T)^{-1} & (1-T)^{-1} & \cdots & (1-T)^{-1} & (1-T)^{-1} & (1-T)^{-1} \\ 0 & 1 & (2-T)^{-1} & \cdots & (2-T)^{-1} & (2-T)^{-1} & (2-T)^{-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \tag{8-21}$$

因此，干扰项 $\boldsymbol{\varepsilon}_i$ 经矩阵 $\mathbf{A}$ 转换后得到的 $\boldsymbol{\varepsilon}_i^*=\mathbf{A}\boldsymbol{\varepsilon}_i$ 将具有如下 $T-1$ 个元素：

$$\varepsilon_{it}^*=c_t\big[\varepsilon_{it}-\frac{1}{T-t}(\varepsilon_{it+1})+\cdots+\varepsilon_{iT})\big] \tag{8-22}$$

其中，$c_t^2=(T-t)/(T-t+1)$。显然，$\mathbf{A}'\mathbf{A}=\mathbf{I}_T-\bar{J}_T$，$\mathbf{A}\mathbf{A}'=\mathbf{I}_{T-1}$。进一步，我们可以得到 $\mathbf{Q}=\mathbf{I}_N\otimes\mathbf{A}'\mathbf{A}$。采用 OLS 估计经过这种“前向正交分解”变换后的模型同样可以得到组内估计量 $\hat{\boldsymbol{\beta}}_{WG}$。因此，正交分解可以视为一种类似于“一阶差分”的处理方式，其优点在于不会在转换后的干扰项中引入序列相关问题。

简言之，无论采用“组内去心”、“一阶差分”还是“正交分解”，我们都可以得到组内估计量。在第 8.6 节的动态面板数据模型中我们将主要应用这种转换方法来去除个体效应。

时间效应

前面介绍的固定效应模型着重在于考虑不可观测的个体效应，按照同样的思路，我们还可以在某些分析中考虑不可观测的时间效应。如在研究区域经济增长的过程中，全球石油价格的上涨、金融危机的爆发都会对所有研究对象在特定年份的产出有所影响。我们注意到，这些因素在特定的年份会对经济体中的所有个体产生影响，这启发我们可以通过设定时间虚拟变量来反映这些时间效应的影响。

1. 直觉的解释和估计方法

我们可以在模型第 4 页中介绍的 LSDV 模型 (8-3) 的基础上进一步增加 $T-1$ 个时间虚拟变量 $\mathbf{s}_{t\tau}$ 来反映时间效应的影响：

$$y_{it}=\sum_{j=1}^{N}\alpha_i d_{ij}+\sum_{\tau=2}^{T}\lambda_t s_{t\tau}+\mathbf{x}_{it}'\boldsymbol{\beta}+\varepsilon_{it} \tag{8-23}$$

其中，若 $t=p$ 时，$s_{t\tau}=1$，否则为 0。采用 OLS 即可获得所有系数的无偏估计量。即使在 N 较大的情况下，我们仍然可以在经过组内去心的模型 (8-5) 中加入 $T-1$ 个时间虚拟变量来控制时间效应。简言之，对于时间效应，我们完全可以将其视为 $\mathbf{x}_{it}$ 的一部分。

2. 更为严格的推导过程

模型的基本设定为：

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta}+u_{it} \\ u_{it} &= a_i+\lambda_t+\varepsilon_{it} \end{aligned} \tag{8-24}$$

其中，$i=1,2,\cdots,N$；$t=1,2,\cdots,T$。相应的向量形式为：

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta}+\mathbf{u} \\ \mathbf{u} &= (\mathbf{I}_N\otimes\mathbf{1}_T)\mathbf{a}+(\mathbf{1}_N\otimes\mathbf{I}_T)\boldsymbol{\lambda}+\boldsymbol{\varepsilon} \end{aligned} \tag{8-25}$$

其中，$\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_T)'$，$\mathbf{a}$ 和 $\boldsymbol{\varepsilon}$ 的定义同前。 假设对于任何 i 和 t 而言，$\mathbf{x}_{it}$ 均不与 ε_{it} 相关。 为了分析的方便， 令 $\mathbf{D}_a=\mathbf{I}_N\otimes\mathbf{1}_T$，$\mathbf{D}_\lambda=\mathbf{1}_N\otimes\mathbf{I}_T$。那么模型的矩阵形式可表示为：

$$\mathbf{y}=\mathbf{D}_a\mathbf{a}+\mathbf{D}_\lambda\boldsymbol{\lambda}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{8-26}$$

我们注意到，$\mathbf{D}_a$ 和 $\mathbf{D}_\lambda$ 分别为 $(NT\times N)$ 和 $(NT\times T)$ 维矩阵， 当 N 或 T 较大时，运算量都会很大。 因此，我们需要事先进行一些简单的运算以去除个体效应和时间效应。 类似于前面 $\mathbf{J}_T$ 的定义方式，设 $\mathbf{J}_N=\mathbf{1}_N\mathbf{1}_N'$， 于是，$\mathbf{D}_\lambda\mathbf{D}_\lambda'=\mathbf{J}_N\otimes\mathbf{I}_T$。 同时， 定义 $\mathrm{E}_N=\mathbf{I}_N-\bar{\mathbf{J}}_N$，其中 $\bar{\mathbf{J}}_N=(1/N)\mathbf{J}_N$。 进一步， 定义转换矩阵：

$$\mathbf{Q}=\mathrm{E}_N\otimes\mathrm{E}_T=\mathbf{I}_N\otimes\mathbf{I}_T+\mathbf{I}_N\otimes\bar{\mathbf{J}}_T-\bar{\mathbf{J}}_N\otimes\mathbf{I}_T+\bar{\mathbf{J}}_N\otimes\bar{\mathbf{J}}_T \tag{8-27}$$

该转换矩阵可以去除个体效应 a_i 和时间效应 λ_t。如，$\tilde{\mathbf{y}}=\mathbf{Q}\mathbf{y}$ 中的特定元素为： $\tilde{y}_{it}=y_{it}-\bar{y}_i-\bar{y}_t+\bar{y}$， 其中，$\bar{y}_i=(1/N)\sum_{i=1}^{N}y_{it}$，$\bar{y}_t=(1/T)\sum_{t=1}^{T}y_{it}$， $\bar{y}=(1/NT)\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it}$。因此，我们可以用 $\tilde{\mathbf{y}}=\mathbf{Q}\mathbf{y}$ 对 $\tilde{\mathbf{X}}=\mathbf{Q}\mathbf{X}$ 进行 OLS 回归， 得到模型 (8-24) 的组内估计量为：

$$\tilde{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{y} \tag{8-28}$$

个体效应和时间效应的估计量分别为：

$$\hat{a}_i = (\bar{y}_i-\bar{y})-\tilde{\boldsymbol{\beta}}(\bar{x}_i-\bar{x}) \tag{8-29}$$

$$\hat{\lambda}_t = (\bar{y}_t-\bar{y})-\tilde{\boldsymbol{\beta}}(\bar{x}_t-\bar{x}) \tag{8-30}$$

这里有两点需要注意： 其一，模型 (8-24) 中不能包含不随时间或不随个体变化的解释变量，因为这些变量在转换过程中都被消除了； 其二，我们没有特意强调模型中是否包含常数项，事实上只要保证不出现完全共线性问题即可， 即，如果要加入常数项，那么就必须同时约束 $\sum_{i=1}^{N} a_i = 0$ 和 $\sum_{t=1}^{T} \lambda_t = 0$； 如果不加常数项，那么就无需作任何约束了。 但加入常数项与否将影响到 $\hat{a}_i$ 和 $\hat{\lambda}_t$ 的含义， 在解释系数的经济含义时需要注意。

8.2.2 随机效应模型

模型的基本设定

当 N 很大时，采用固定效应模型往往会使参数的数目迅速增加，自由度的损失往往较大。前文已经提到，在固定小模型的设定中，$\mathbf{x}_{it}$ 中不能包含不随时间改变的变量，如性别、种族等。然而，在有些研究中，我们研究的重点可能恰恰是这些变量。此时，随机效应模型可能更为适用。模型的基本设定同 (8-2):[12]

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + u_{it} \\ u_{it} &= \alpha_i + \varepsilon_{it} \end{aligned} \tag{8-31}$$

随机效应模型可以视为固定效应模型的一个扩展， 这需要我们在上一节中假设 1 和假设 2 的基础上再增加如下假设：

假设 3 :

$$\alpha_i \sim i.i.d\,(0, \sigma_\alpha^2)$$

假设 4 :

$$\mathrm{Cov}(\alpha_i, \mathbf{x}_{it}) = 0$$

假设 5 :

$$\mathbf{u}_i|\mathbf{x}_i \sim i.i.d\,(0, \sigma_\varepsilon^2\mathbf{I}_T + \sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T')$$

其中， 假设 3 将个体效应 α_i 设定为服从均值为 0，方差为 σ_a^2 的随机变数， 而我们在固定效应模型的设定中 α_i 只是一个普通的解释变量，因此无需对它作任何限制；假设 4 非常显然，因为此时我们将 α_i 视为随机干扰项的一部分，所以它不能与解释变量相关；假设 5 表明 α_i 与 ε_{it} 相互独立。

序列相关性

易于证明：

$$\mathrm{Cov}\,(u_{it}, u_{js}) = \begin{cases} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \text{for} \quad i = j, \ \ t = s \\ \sigma_\alpha^2 & \text{for} \quad i = j, \ \ t \neq s \\ 0 & \text{for} \quad i \neq j, \ \ t \neq s \end{cases} \tag{8-32}$$

[12]这里，为了表述的方便，我们把模型 (8-2) 中的常数项 μ 放在了 $\mathbf{x}_{it}$ 中。

和

$$\rho = \mathrm{Corr}\,(u_{it}, u_{js}) = \begin{cases} 1 & \text{for} \quad i = j, \ t = s \\ \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2) & \text{for} \quad i = j, \ t \neq s \\ 0 & \text{for} \quad i \neq j, \ t \neq s \end{cases} \tag{8-33}$$

从 (8-33) 式可以看出，由于随机效应的引入使得组内不同时期的观察值之间存在固定不变的自相关关系，相关系数为 $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$。这很容易理解，因为尽管个体效应是随机的，但在组内并不随时间改变，组内不同期间固定的相关性也就必然存在。从另一个角度来看，ρ 的含义在于，模型中的总方差中有 $\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$ 来自于不随时间改变的干扰，而余下的部分则归因于随个体和时间改变的干扰。当然，在某些情况下这个假设显得过于严格。如在研究投资或消费时，我们往往会假设组内不同期间的相关性是随时间逐渐减弱的。

GLS 估计

基于以上设定，可以写出干扰项的方差-协方差矩阵：

$$\mathbf{\Omega} = \mathrm{E}(\mathbf{uu}') = \mathbf{I}_N \otimes (\sigma_\varepsilon^2 \mathbf{I}_T + \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T') = \mathbf{I}_N \otimes \mathbf{\Sigma} \tag{8-34}$$

其中，$\mathbf{\Sigma} = \sigma_\varepsilon^2 \mathbf{I}_T + \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T'$，具体形式为：

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{bmatrix} \tag{8-35}$$

那么，$\boldsymbol{\beta}$ 的 GLS 估计量为：

$$\hat{\boldsymbol{\beta}}_{GLS} = [\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} \tag{8-36}$$

方差估计量为：

$$\mathrm{Var}\,(\hat{\boldsymbol{\beta}}_{GLS}) = [\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}]^{-1} \tag{8-37}$$

为了进一步说明上述 GLS 估计量与前面介绍的组内估计量之间的关系，我们可以对 $\mathbf{\Sigma}$ 矩阵进行分解，得到 $\mathbf{G} = \mathbf{\Omega}^{-1/2} = [\mathbf{I}_n \otimes \mathbf{\Sigma}]^{-1/2}$，进而采用 $\mathbf{G}$ 矩阵对模型 (8-31) 进行转换。显然，我们只需要求出 $\mathbf{\Sigma}^{-1/2}$ 即可，

$$\mathbf{\Sigma}^{-1/2} = \frac{1}{\sigma_\varepsilon}\left[\mathbf{I} - \frac{\theta}{T}\mathbf{1}_T\mathbf{1}_T'\right]$$

其中，

$$\theta = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_a^2}}$$

于是我们可以对原始数据作如下转换：

$$\boldsymbol{\Sigma}^{-1/2}\mathbf{y}_i = \frac{1}{\sigma_\varepsilon}\begin{bmatrix} y_{i1} - \theta\bar{y}_i \\ y_{i2} - \theta\bar{y}_i \\ \vdots \\ y_{iT} - \theta\bar{y}_i \end{bmatrix} \tag{8-38}$$

按照同样的方法我们可以对 $\mathbf{x}_i$ 进行转换，对模型 (8-2) 转换后可得：

$$(y_{it} - \theta\bar{y}_i) = (\mathbf{x}_{it} - \theta\bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (1-\theta)\alpha_i + (\varepsilon_{it} - \theta\bar{\varepsilon}_i) \tag{8-39}$$

对模型 (8-39) 执行 OLS 估计即可得到与 (8-36) 式相同的结果。[13] 我们注意到，如果 (8-38) 式中的 $\theta = 1$，则上述变换就是我们前面讲到的“组内去心”，得到的就是固定效应模型对应的组内估计量 (8-10)。事实上，我们可以证明 $\hat{\boldsymbol{\beta}}_{GLS}$ 可以表示为组内估计量和组间估计量的加权平均，详细过程请参考 Greene (2000, pp.295-296)。

FGLS 估计

上面介绍的 GLS 估计是在假设方差成分已知的前提下进行了， 但多数情况下我们并不知道 σ_ε^2 和 σ_α^2， 因此需要先估计这两个未知参数，继而用它们去代替 (8-35) 式中的真实值并采用 GLS 估计即可。基本思路是：先估计固定效应模型，得到 σ_ε^2 的估计值 $\hat{\sigma}_\varepsilon^2$，继而估计混合 OLS 模型， 利用其残差和第一步得到的 $\hat{\sigma}_\varepsilon^2$ 即可估计出 $\hat{\sigma}_u^2$。

由于组内估计量是无偏且一致的，所以我们可以利用固定效应模型的残差来估计 σ_ε^2，因为在估计固定效应模型的过程中我们已经去除了个体效应。 设 $e_{it} = (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\hat{\boldsymbol{\beta}}_{WG}$ 为固定效应模型的残差，则

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2}{nT - n - K} \tag{8-40}$$

接下来需要估计 σ_α^2。模型 (8-31) 的 OLS 估计仍然无偏且一致的。 设 $\tilde{e}_{it}$ 为模型 (8-31) 的 OLS 残差，则

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T \tilde{e}_{it}^2}{nT - K - 1} = \hat{\sigma}_\varepsilon^2 + \hat{\sigma}_a^2 \tag{8-41}$$

由此，我们可以得到：

$$\hat{\sigma}_a^2 = \hat{\sigma}_u^2 - \hat{\sigma}_\varepsilon^2$$

由于该估计量可能为负值，所以我们可以略去 (8-40) 式和 (8-41) 式中对自由度的调整。这样就可以保证 $\hat{\sigma}_u^2$ 一定是大于 $\hat{\sigma}_\varepsilon^2$ 的，因为前者是后者在附加约束条件下的估计量。 这种处理方法的依据在于我们只需要 σ_ε^2 和 σ_α^2 的一致估计即可，至于是否无偏并不影响大样本性质。

[13]当然，在 θ 未知的情况下，这一看似简单的估计方法是无法执行的。通过下一小节的介绍可以发现，事实上，θ 中的参数 σ_ε^2 和 σ_ε^2 可以通过组内估计量和 Pooled OLS 估计量获得。因此，从实际操作的角度来讲，任何能执行 OLS 操作的软件，都可以用来估计随机效应模型。

上述估计方法虽然简单易行，但是当随机效应模型中包含不随时间改变的变量，如性别、种族等，我们就无法通过估计固定效应模型来估计 σ_ε 了。不过此时我们可以沿袭上面的思路，利用组间估计和混合 OLS 估计的残差来估计 σ_ε^2 和 σ_α^2。采用 OLS 估计模型 (8-2) 的组内平均模型：

$$\bar{\mathbf{y}}_i = \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\boldsymbol{\mu}}_i \tag{8-42}$$

可以得到一致估计量 $m^* = \hat{\sigma}_a^2 + (\hat{\sigma}_\varepsilon^2/T)$，结合 m^* 和 $\hat{\sigma}_u^2$ 我们可以得到：

$$\hat{\sigma}_\varepsilon^2 = \frac{T}{T-1}(\hat{\sigma}_u^2 - m^*)$$

$$\hat{\sigma}_a^2 = \frac{T}{T-1}m^* - \frac{1}{T-1}\hat{\sigma}_u^2$$

那么以上介绍的各种 FGLS 估计量哪个更为有效呢？我们知道，对于随机效应模型而言，针对方差成分的真实值进行 GLS 估计将得到 BLUE 估计量。而以上介绍的 FGLS 估计量在 $N \to \infty$ 或 $T \to \infty$ 或二者 都成立 的情况下，都是渐进有效的。Maddala 和 Mount(1973) 采用蒙特卡罗模拟方法对各种 FGLS 估计量的比较表明，在小 样本下各种估计方法难分仲伯，所以建议采用简单易行的方法进行估计。 Taylor (1980) 比较了小样本下随机效应的 FGLS 估计和固定效应的 LSDV 估计，结果表明：

(1) 相对于 LSDV，FGLS 更具有效性，且具有较小的自由度；

(2) FGLS 的方差不会大于 Cramer-Rao 下限的 17%。

(3) 选择相对有效的方差成分估计量并不必然能够提高 FGLS 估计量的有效性。

8.2.3 假设检验

根据前面的介绍，我们大体可以采用三种方法估计面板数据模型：混合 OLS、固定效应模型和随机效应模型。 那么如何对这三种模型进行区分和筛选呢？ 这就需要进行假设检验。 显然，如果个体效应 (固定效应或随机效应) 显著异于零，那么就需要采用固定效应或随机效应模型。 对于随机效应模型，它要求 $\mathrm{Cov}(\alpha_i, \mathbf{x}_i) = 0$，而固定效应模型则没有这一限制， 所以如果这一假设无法满足，我们就只能采用固定效应模型，或采用工具变量法来估计随机效应模型。

固定效应的检验

由 8.2.1 小节的分析可知，固定效应模型的本质是通过个体间截距项的差异来捕捉不可观测的个体效果。 但是，如果个体间 (组间) 并不存在统计意义上的显著差异，我们只需对混合数据执行 OLS 估计即可。[14] 检验的基本思路为，在个体效应不显著的原假设下， 应当有如下关系成立：

$$H_0: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_n$$

[14] 此时，模型设定为 $y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$。文献中也将此模型成为“混合数据模型”，相应的估计量称为混合 OLS (Pooled OLS) 估计量。

我们可以采用 F 统计量来检验上述假设是否成立，

$$F=\frac{(R_u^2-R_r^2)/(n-1)}{(1-R_u^2)/(nT-n-K)}\sim F(n-1,nT-n-K) \tag{8-43}$$

其中，u 表示不受约束的模型，即我们的固定效应模型； r 表示受约束的模型，即混合数据模型， 仅有一个公共的常数项。

同理，我们可以构造相应的 F 统计量来检验时间效应的显著性，以及个体效应和时间效应的联合显著性。

检验随机效应

Breusch and Pagan (1980) 建议基于模型 (8-2) OLS 估计的残差构造 LM 统计量，针对如下原假设来检验随机效应，

$$H_0:\quad \sigma_\alpha^2=0 \quad v.s. \quad H_1:\sigma_\alpha^2\neq 0$$

相应的检验统计量为：

$$\mathrm{LM}=\frac{nT}{2(T-1)}\left[\frac{\sum_{i=1}^n\left[\sum_{t=1}^T e_{it}\right]^2}{\sum_{i=1}^n\sum_{t=1}^T e_{it}^2}-1\right]^2 \tag{8-44}$$

在原假设下，LM 统计量服从一个自由度为 1 的卡方分布。如果拒绝原假设则表明存在随机效应。 如果采用矩阵的形式，该 LM 统计量可以表示为：

$$\mathrm{LM}=\frac{nT}{2(T-1)}\left[\frac{\mathbf{e}'\mathbf{DD}'\mathbf{e}}{\mathbf{e}'\mathbf{e}}-1\right]^2 \tag{8-45}$$

需要说明的是，该检验假设模型的设定是正确的，即 α_i 与解释变量不相关，而这一假设是否正确 还需要作进一步的检验，这是我们下面要分析的内容。

固定效应还是随机效应？Hausman 检验

在前面的分析中，我们从不同角度比较了固定效应模型和随机效应模型的差别， 那么在实际分析中应该使用哪个模型呢？ 某些学者指出，试图区分固定效应和随机效应本身就是错误的，二者似乎不具可比性。 Mundlak (1978) 指出，一般情况下，我们都应当把个体效应视为随机的。如果从单纯的实际操作角度来 考虑，固定效应模型往往会耗费很大的自由度，尤其是对于截面数目很大的面板数据，随机效应模型似 乎更合适。但另一方面，固定效应模型有一个独特的优势，我们无须做个体效应与其它解释变数不相关 的假设，而在随机效应模型中，这个假设是必须的， 否则就会导致内生性问题，并进而导致参数估计的非一致性。

因此，我们可以通过检验固定效应 α_i 与其它解释变量是否相关作为进行固定效应和随机效应模型筛选的依据。此时，我们可以采用 Hausman 检验。其基本思想是，在 α_i 与其他解释变量不相关的原假设下，我们采用 OLS 估计固定效应模型和采用 GLS 估计随机效应模型得到的

参数估计都是无偏且一致的，只是前者不具有效性。若原假设不成立，则固定效应模型的参数估计仍然是一致的，但随机效应模型却不是。因此，在原假设下，二者的参数估计应该不会有显著的差异， 我们可以基于二者参数估计的差异构造统计检验量 。

假设 $\mathbf{b}$ 和 $\hat{\boldsymbol{\beta}}$ 分别为固定效应模型的 OLS 估计和随机效应模型的 GLS 估计，则

$$\mathrm{Var}(\mathbf{b}-\hat{\boldsymbol{\beta}})=\mathrm{Var}(\mathbf{b})+\mathrm{Var}(\hat{\boldsymbol{\beta}})-\mathrm{Cov}(\mathbf{b},\hat{\boldsymbol{\beta}})-\mathrm{Cov}(\mathbf{b},\hat{\boldsymbol{\beta}})' \tag{8-46}$$

基于上述 Hausman 检验的思想，有效估计量与它和非有效估计量之差的协方差应当为零，即

$$\mathrm{Cov}[(b-\hat{\boldsymbol{\beta}}),\hat{\boldsymbol{\beta}}]=\mathrm{Cov}(\mathbf{b},\hat{\boldsymbol{\beta}})-\mathrm{Var}(\hat{\boldsymbol{\beta}})=\mathbf{0} \tag{8-47}$$

由此我们可以得到：

$$\mathrm{Cov}(\mathbf{b},\hat{\boldsymbol{\beta}})=\mathrm{Var}(\hat{\boldsymbol{\beta}}) \tag{8-48}$$

将 (8-48) 式代入 (8-46) 式得到：

$$\mathrm{Var}(b-\hat{\boldsymbol{\beta}})=\mathrm{Var}(\mathbf{b})-\mathrm{Var}(\hat{\boldsymbol{\beta}})=\boldsymbol{\Psi} \tag{8-49}$$

Hausman 检验基于如下 Wald 统计量：

$$W=[b-\hat{\boldsymbol{\beta}}]'\hat{\boldsymbol{\Psi}}^{-1}[b-\hat{\boldsymbol{\beta}}]\sim\chi^2(K-1) \tag{8-50}$$

其中，$\hat{\boldsymbol{\Psi}}$ 采用固定效应和随机效应模型的协方差矩阵进行计算. 如果拒绝了原假设，就表明个体效应 α_i 和解释变量 $\mathbf{x}_{it}$ 是相关的，此时我们有两种处理办法： 一是采用固定效应模型，某些情况下这是一种无奈的选择；[15] 二是采用工具变量法来处理内生问题。[16]

[15]因为有时我们通过 B-P 检验发现存在随机效应，但 Hausman 检验又表明使用随机效应模型的前提假设得不到满足，而我们又往往很难找到合适的工具变量，所以只能采用固定效应模型。

[16]在 STATA 中可以采用 xthtaylor 和 xtivreg 命令来完成相应的估计。 但这两个命令的侧重点还是有所差别的，前者重点处理的是模型 (8-31) 中 α_i 与 $\mathbf{x}_i$ 之间的相关性， 而后者则重点处理通常意义上的内生性问题， 即 $\boldsymbol{\varepsilon}_i$ 与 $\mathbf{x}_i$ 之间的相关性。

8.3 STATA 实现 I：静态面板模型

8.3.1 简介

在目前比较流行的计量软件中，STATA 在面板数据处理方面的优势比较明显。这一方面得益于 STATA 自身强大的数据处理功能，另一方面则归因于其快捷的更新速度。目前，STATA 不但能估计此前介绍的两种基本的静态面板模型 (`xtreg` 命令)，还能估计随后将要介绍的多种动态面板模型 (如 `xtabond`, `xtdpdsys`, `xtdpd` 命令)。对于内生性问题的处理也是 STATA 的一个强项 (如 `xtivreg`, `xtivreg2` 命令)。在 STATA11 中，我们还可以很方便的完成面板单位根检验 (`xtunitroot` 命令)、面板协整分析 (`nharvey`, `xtwest` 命令)，以及面板误差修正模型的分析 (`xtpmg` 命令)。与此同时，全球大量的 STATA 用户还提供了新近发展的面板门槛模型 (`xtthres`, `xtptm` 命令)，面板 VAR 模型 (`pvar`, `xtvar`) 等命令的估计程序。[17]

8.3.2 基本设定

截面变量和时间变量的设定

我们首先通过一份简单的数据来说明如何在 STATA 中设定面板数据的结构。

```
. use http://www.stata-press.com/data/r11/invest2.dta, clear

. rename company id

. rename time year

. order id year

. replace year = year + 1990
(100 real changes made)

. list if (id<=3 & year<=1995), noobs
```

id	year	invest	market	stock
1	1991	317.6	3078.5	2.8
1	1992	391.8	4661.7	52.6
1	1993	410.6	5387.1	156.9
1	1994	257.7	2792.2	209.2
1	1995	330.8	4313.2	203.4
2	1991	40.29	417.5	10.5
2	1992	72.76	837.8	10.2
2	1993	66.26	883.9	34.7
2	1994	51.6	437.9	51.8
2	1995	52.41	679.7	64.3
3	1991	33.1	1170.6	97.8
3	1992	45	2015.8	104.4
3	1993	77.2	2803.3	118

[17] 若想全面了解 STATA 中有关面板数据的命令，可输入 `help xt` 命令。若想搜索有关面板数据的外部命令，可输入 `findit panel data` 命令。

```
  3    1994      44.6    2039.7    156.2
  3    1995      48.1    2256.2    172.6
```

这份数据共包含五个变量，其中，id 和 year 分别为截面变量和时间变量，分别对应于模型 (8-1) 中的下标 i 和 t。显然，通过这两个变量我们可以非常清楚地确定 panel data 的数据存储格式。因此，在使用 STATA 估计模型之前，我们必须告诉它截面变量和时间变量分别是什么，所用的命令为 xtset：[18]

```
. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1991 to 2010
                delta:  1 unit
```

这里，STATA 确认了我们所设定的截面变量 (id) 和时间变量 (year)，并提示说我们的数据结构为“strongly balanced”。这表示，在样本中，每个公司都有相同的年度观察值 (1991-2010)。在多数情况下，我们所收集的数据都是非平行数据 (unbalanced panel data)。[19]

若想面板数据的结构有更详细的了解，可以输入 xtdes 命令：

```
. xtdes

      id:  1, 2, ..., 5                                     n =          5
    year:  1991, 1992, ..., 2010                            T =         20
           Delta(year) = 1 unit
           Span(year)  = 20 periods
           (id*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                        20      20      20        20        20      20      20

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+----------------------
        5    100.00  100.00 |  11111111111111111111
 ---------------------------+----------------------
        5    100.00         |  XXXXXXXXXXXXXXXXXXXX
```

在我们的样本中，共包含 $n = 5$ 家公司，每家公司有 $T = 20$ 年的观察值。同时，由于该样本是平行面板，所以 T_i (每个公司对应的观察年数) 的分布在各个分位上都为 20。

统计描述

在正式进行模型的估计之前，我们必须对样本的基本分布特性有一个总体的了解。也要大体了解主要变量的均值、标准差、最大值、最小值等情况。显然，STATA 中有关统计描述

[18] 另一个与该命令功能相似的命令是 tsset。需要注意的是，若同一家公司有两个以上相同年度的观察值，则截面变量和时间变量将无法 (联合起来) 唯一标示样本中的每一个观察值，此时，执行 xtset 命令时，STATA 将报告错误信息“repeated time values within panel”。解决办法是在执行 xtset 命令前，先删除重复的观察值，命令为“duplicates drop id year, force”。

[19] 对于上市公司而言，有些公司上市较晚，有些公司中途退市，都可能导致个体间时间跨度的差异，从而使我们的数据是非平行的。随后我们会介绍这种两种数据结构对面板分析的影响，以及二者的转换。

的命令 (summarize, tabstat, histogram, kdensity 等) 仍然适用于面板数据的分析。同时，STATA 也专门为面板数据定制了一些进行描述性统计分析的命令，如 xtsum、xttab，以及 xttrans 等。[20]

xtsum 命令事实上是我们经常使用的命令 summarize 的扩展，各个统计量都分别在样本总体、组内和组间三个层次上进行计算。二者的对比如下：

```
. sum   invest

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      invest |       100     248.957    267.8654      12.93     1486.7

.
. xtsum invest

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
invest   overall |   248.957   267.8654      12.93     1486.7 |     N =     100
         between |             246.9354    42.8915     608.02 |     n =       5
         within  |             149.9249   -101.363   1127.637 |     T =      20
```

相比于 summarize 命令，xtsum 提供了更为详细的信息。它将变量 x_{it} 分解成 "组间" ($\bar{x}_i$) 和 "组内" ($x_{it} - \bar{x}_i - \bar{\bar{x}}$) 两个部分。[21]

8.3.3 面板数据的处理

1. 产生滞后项和差分项

由于 Panel Data 兼具截面数据和时间序列二者的特性，所以对时间序列进行操作的运算同样可以应用于 Panel Data。这使得某些数据的处理变得非常方便。例如，对于上述数据， 我们想产生一个新的变量 invest$_{it-1}$，也就是变量 invest 的一阶滞后项，那么我们可以采用如下命令：

```
gen Lag_invest = L.invest
```

按照这样的思路，还可以产生某个变量的移动平均、差分等。总之，凡是可以应用到时间序列上的操作，基本上都可以应用到 Panel Data 中来，例如：

```
gen  F_invest =  F.invest   // 超前项
gen  D_invest =  D.invest   // 一阶差分项
gen D2_invest = D2.invest   // 二阶差分项
```

2. 产生组内均值

```
bysort   id: egen mi_stock = mean(stock)  // 每个公司的平均值
bysort year: egen mt_stock = mean(stock)  // 每个年度的平均值
```

[20] 另外一些用于面板数据统计性描述的命令可以从网上下载 (使用 findit 命令)，包括：xtcount, countby, xtpattern, panels 等。

[21] 这里所谓的 "组间" 其实就是个体的平均值。此外，真正意义上的 "组内" 统计量应该是 $x_{it} - \bar{x}_i$，即 (8-5) 式的变换。这里之所以再加上总样本的平均值 $\bar{\bar{x}}$，是为了保证上述各统计量之间的可比性。由下文的分析可知，STATA 在估计固定效应模型时，采用也是这一变换，而不是理论推导公式 (8-5) 或 (8-9)。

3. 产生观察期末变量

```
bysort id: gen end_year = year[_n+1]
bysort id: replace end_year = year if _n==_N
```

或

```
bysort id: egen end_year2 = max(year)
```

4. 将非平行面板数据转换为平行面板数据

有时候我们的数据在经过初步处理后并非是平行面板数据，即每个截面的观察期数可能不同，可是许多情况下我们又必须使用平行数据，这就需要把非平行数据“削平”后转化为平行数据。STATA 官方发布的命令并不能快捷地处理这个看似简单问题。为此，笔者自行编写 `xtbalance` 命令来处理这一问题。在使用之前，读者需要输入如下命令安装 `xtbalance`：

```
ssc install xtbalance, replace
```

完成安装后，可输入 `help xtbalance` 命令查看其帮助文件。为了说明 `xtbalance` 的使用方法，我们先调入一份非平行面板数据：

```
. use http://www.stata-press.com/data/r11/abdata.dta, clear

. qui xtset id year

. xtdes

      id:  1, 2, ..., 140                                    n =        140
    year:  1976, 1977, ..., 1984                             T =          9
           Delta(year) = 1 unit
           Span(year)  = 9 periods
           (id*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         7       7       7         7         8       9       9

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-----------
       62     44.29   44.29 |  1111111..
       39     27.86   72.14 |  .1111111.
       19     13.57   85.71 |  .11111111
       14     10.00   95.71 |  111111111
        4      2.86   98.57 |  11111111.
        2      1.43  100.00 |  ..1111111
 ---------------------------+-----------
      140    100.00         |  XXXXXXXXX
```

假设我们想保留 1977-1983 年样本区间内的平行面板，便可输入如下命令：

```
. xtbalance, range(1977 1983)
(115 observations deleted due to out of range)
(384 observations deleted due to discontinues)
```

可见，有 115 个观察值由于在 1977-1983 年区间以外而被删除，另有 384 个观察值则因为观察年份不连续而被删除。读者可以输入 `xtdes` 命令验证上述命令的效果。需要注意的是，虽然从表明上看，上述处理似乎使数据变成“banlanced”，但由于部分变量仍包含缺漏值 (可以

输入 sum 命令查验)，致使其实际上并非平行数据。更为稳妥的处理方法是附加 miss(_all) 选项，以便在删除所有变量的缺漏值后，再执行平行面板转换。[22]

8.3.4 面板模型的估计

STATA 面板模型命令概览

STATA 11 主要提供了多种面板模型的估计方法，如表 8-1 所示 (参见 [XT] **xtreg**)。其中多数模型的估计方法我们都会在随后的章节中陆续讲到。

表 8-1: STATA 11.0 中用于估计 Panel Data 模型的主要命令一览

命令	模型
xtreg	Fixed-， between- and random-effects， and population-averaged linear models
xtregar	Fixed- and random-effects linear models with an AR(1) disturbance
xtgls	Panel-data models using GLS
xtpcse	OLS or Prais-Winsten models with panel-corrected standard errors
xtrc	Random coefficients models
xtivreg	Instrumental variables and two-stage least squares for panel-data models
xtivreg2	IV/2SLS, GMM and AC/HAC, LIML regression for panel data models (需要下载)
xtabond	Arellano-Bond linear dynamic panel-data estimator
xtdpdsys	Arellano-Bond/Blundell-Bond estimation
xtdpd	Linear dynamic panel-data estimation
xtabond2	Arellano-Bond system dynamic panel data estimator (需要下载)
xttobit	Random-effects tobit models
xtintreg	Random-effects interval data regression models
xtlogit	Fixed-effects， random-effects， population-averaged logit models
xtprobit	Random-effects and population-averaged probit models
xtcloglog	Random-effects and population-averaged cloglog models
xtpoisson	Fixed-effects， random-effects， population-averaged Poisson models
xtmixed	Multilevel mixed-effects linear regression
xtnbreg	Fixed-effects， random-effects， population-averaged negative binomial models
xtfrontier	Stochastic frontier models for panel-data
xthtylor	Hausman-Taylor estimator for error-components models
xtunitroot	Panel-data unit-root tests
xtwest	Westerlund error correction based panel cointegration tests
xtpmg	Pooled mean-group, mean-group, and dynamic fixed-effects models (需要下载)

[22] 当然，也可以在 miss() 选项中指定变量的名称，xtbalance 将只删除这些变量中包含的缺漏值。

固定效应模型和随机效应模型的估计

第 8.2 节介绍的固定效应模型和随机效应模型 (以下分别简称 FE 和 RE)，可以采用 xtreg 命令估计，基本语法格式如下：

xtreg *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *model_type other_options*]

其中，model_type 选项用于指定需要估计的模型，对应关系如表 8-2 所示。这里有三点需要说明：其一，如果不填 model_type 选项，则 STATA 默认采用第 8.2.2 小节介绍 GLS 方法估计 RE 模型；其二，若设定 mle 选项，则 STATA 会采用 MLE 估计 RE 模型，相应的似然函数参见 [XT] **xtreg** (pp.466)；其三，上述命令格式只是一个基本形式，对于不同模型，还有一些相当灵活的控制选项，读者可以参考相应的帮助。

表 8-2: xtreg 命令中选项的含义

model_type	模型
be	Between-effects estimator
fe	Fixed-effects estimator
re	GLS Random-effects estimator
pa	GEE population-averaged estimator
mle	Maximum-likelihood Random-effects estimator

下面，我们通过一个具体实例来说明上述命令的使用方法。我们仍然采用第 8.3.2 小节的 invest2.dta 数据，除了第 17 页中提到的 id 和 year 变量，另外三个变量分别是：invest 表示投资支出，market 表示市场价值，stock 表示资本存量。我们的目的是研究公司的投资额和资本存量如何影响其市场价值，为此建立了如下实证模型：

$$market_{it} = \mu + invest_{it} + stock_{it} + \alpha_i + \varepsilon_{it}$$

1. FE 模型

若假设 α_i 为固定效应，则上述模型就是一个典型的 FE 模型，估计结果如下：

```
. xtreg market invest stock, fe

Fixed-effects (within) regression               Number of obs      =       100
Group variable: id                              Number of groups   =         5

R-sq:  within  = 0.4168                         Obs per group: min =        20
       between = 0.6960                                        avg =      20.0
       overall = 0.6324                                        max =        20

                                                F(2,93)            =     33.23
corr(u_i, Xb)  = 0.5256                         Prob > F           =    0.0000

------------------------------------------------------------------------------
      market |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
      invest |   3.05273   .4577368     6.67   0.000     2.143756    3.961705
       stock | -.6763434   .2216246    -3.05   0.003    -1.116446    -.236241
       _cons |  1372.613   76.96444    17.83   0.000     1219.776    1525.449
-------------+----------------------------------------------------------------
     sigma_u | 1023.5914
     sigma_e |  370.9569
         rho | .88390837   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(4, 93) =    97.68              Prob > F = 0.0000
```

其中，选项 fe 表明我们采用的是固定效应模型。表头部分的前两行呈现了模型的估计方法 (Fixed-effects (within) regression)、截面变量的名称 (id)、以及估计中使用的样本数目和个体的数目。第 3 行到第 5 行列示了模型的拟合优度 R^2，分为组内、组间和样本总体三个层次 (详见第 8.3.4 页的解释)。第 6 行和第 7 行分别列示了针对模型中所有非常数变量执行联合检验得到的 F 统计量以及相应的 P 值，本例中分别为 33.23 和 0.0000，表明参数整体上 相当显著。[23]第 8-11 行列示了解释变量的估计系数、标准误、t 统计量和相应的 P 值，以及 95% 置信区间，这和我们在进行截面回归时得到的结果是一样的。最后四行列示了固定效应模型中个体效应 (α_i) 和随机干扰项 (ε_{it}) 的方差估计值、[24]以及二者在总方差中的相对比例，即 $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2) = 0.8839$。[25]

需要注意的是，表中最后一行列示了检验固定效应是否显著的 F 统计量 (参见 (8-43) 式) 和相应的 P 值。显然，本例中固定效应非常显著。

细心的读者可能会产生如下疑问：通过 (8-5) 或 (8-9) 式进行组内变换后的模型都不再包含常数项，但为何上面的估计结果中还有常数项呢？这是因为，STATA 中的“组内变换”与 (8-5) 式略有差异：

$$(y_{it} - \bar{y}_i + \bar{\bar{y}}) = \alpha + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i + \bar{\bar{\mathbf{x}}})'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i + \bar{\bar{\varepsilon}}) \tag{8-51}$$

其中，$\bar{\bar{y}} = (1/N)\sum_{i=1}^N \bar{y}_i$，即 y_{it} 的样本平均值，$\bar{\bar{\mathbf{x}}}$ 和 $\bar{\bar{\varepsilon}}$ 的定义与此相似。相对于 (8-5) 式，该模型中增加了一个常数项，它表示个体效应 α_i 的样本平均值。显然，$\boldsymbol{\beta}$ 系数的估计值和标准误在两种变换下并不存在任何差异。[26]

前文已经提到，除了采用组内变换的方式，亦可采用最小二乘虚拟变量法估计固定效应模型，即 (8-3) 式。STATA 中有多种方法可以执行这一分析。最为简单的方法莫过于在 Pooled OLS 中增加 N 个反映个体效应的虚拟变量：

---

[23]F 统计量服从自由度分别为 2 和 93 的 F 分布。其中，2 表示除常数项外，模型中有两个解释变量，$93 = NT - N - k = 5 \times 20 - 5 - 2$。

[24]STATA 列出的是二者的标准差，分别为 sigma_u 和 sigma_e。其中，u 和 e 分别表示个体效应 α_{it} 和干扰项 ε_{it}。

[25]这里的 σ_α^2 经由个体效应的估计值 $\hat{\alpha}_i$ (由 (8-13) 式估得) 的标准差计算而得。显然，它与 (8-33) 式中呈现的随机效应模型中的 σ_α^2 具有不同的含义。

[26]STATA 之所以采用 (8-51) 式的变换，一方面是为了保证不同模型之间的可比性，另一方面则因为在模型中附加了常数项而保证了 R^2 仍然是有意义的。

```
. qui tabulate id, gen(dum_a)

. reg market invest stock dum_a*, noconstant

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  7,    93) =  577.77
       Model |   556540367     7  79505766.7           Prob > F      =  0.0000
    Residual |    12797639    93  137609.021           R-squared     =  0.9775
-------------+------------------------------           Adj R-squared =  0.9758
       Total |   569338006   100  5693380.06           Root MSE      =  370.96

------------------------------------------------------------------------------
      market |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |    3.05273   .4577368     6.67   0.000     2.143756    3.961705
       stock |  -.6763434   .2216246    -3.05   0.003    -1.116446    -.236241
      dum_a1 |   2916.289   194.5067    14.99   0.000     2530.037     3302.54
      dum_a2 |   512.3015   85.89466     5.96   0.000     341.7317    682.8712
      dum_a3 |   1899.707   99.82705    19.03   0.000      1701.47    2097.944
      dum_a4 |   597.8959   83.66869     7.15   0.000     431.7464    764.0453
      dum_a5 |     936.87   158.2156     5.92   0.000     622.6851    1251.055
------------------------------------------------------------------------------
```

这里，我们首先采用 tabulate 命令附加 gen() 选项，生成了 5 个虚拟变量，[27]进而采用 regress 命令执行 OLS 估计。与此前的理论分析一致，此时得到的 invest 和 stock 变量的系数估计值与 xtreg, fe 的结果完全相同。然而，两种方法下得到的 R^2 存在显著差异：LSDV 下得到的 R^2 (0.9775) 明显高于 xtreg, fe 命令得到的 R^2 (within R-sq=0.4168)，这是因为后者在进行组内变换过程中去除个体效应 α_i，致使 within R-sq 中并未包含 α_i 对方差的贡献。

我们亦可用 areg 命令，或在 regress 命令前附加 xi：前缀的方式估计 LSDV 模型：

```
. qui areg market invest stock, absorb(id)  // areg

. est store fe_areg

. qui xi: reg market invest stock i.id      // xi: reg

. est store fe_xi_reg

. local m "fe_areg fe_xi_reg"

. esttab `m', mtitle(`m') nogap scalar(N r2 r2_a) star(* 0.1 ** 0.05 *** 0.01)
>

--------------------------------------------
                      (1)             (2)
                  fe_areg       fe_xi_reg
--------------------------------------------
invest              3.053***        3.053***
                   (6.67)          (6.67)
stock              -0.676***       -0.676***
                  (-3.05)         (-3.05)
_Iid_2                            -2404.0***
--------------------------------------------
```

[27]为了防止共线性问题，这里附加了 noconstant 选项。当然，也可以仅放入 $N-1$ 个虚拟变量，保留一个公共的常数项，此时无需再附加 noconstant 选项。

```
                              (-12.40)
_Iid_3                        -1016.6***
                               (-4.59)
_Iid_4                        -2318.4***
                              (-11.30)
_Iid_5                        -1979.4***
                              (-15.50)
_cons            1372.6***     2916.3***
                 (17.83)       (14.99)
--------------------------------------------
N                   100           100
r2                0.936         0.936
r2_a              0.932         0.932
--------------------------------------------
t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01
```

可以看出，两种方式得到的估计值和相同的统计量并不存在任何差异。需要说明的是，对于 N 较小的面板而言，采用第二种方法比较方便，能够直接得到个体效应的估计值 $\hat{\alpha}_i$。但当 N 较大时，这种方法将不再奏效，而 areg 和 xtreg,fe 命令则较好。[28]相比于 xtreg,fe，虽然 areg 也同样未呈现个体效应的估计值，但在计算 R^2 时，它却考虑了个体效应对整个模型的方差贡献，因此得到的 R^2 相对较高。

2. RE 模型

若假设本例的样本公司是从一个很大的母体中随机抽取的，且 α_i 与解释变量 invest 和 stock 均不相关，则我们可以将 α_i 视为随机干扰项的一部分。此时，设定随机效应模型 (8-31) 更为合适。估计过程相当简单，仅需把上例中的 fe 选项去掉或附加 re 选项即可，例如：

```
. xtreg market invest stock, re

Random-effects GLS regression                   Number of obs      =       100
Group variable: id                              Number of groups   =         5

R-sq:  within  = 0.4163                         Obs per group: min =        20
       between = 0.7054                                        avg =      20.0
       overall = 0.6380                                        max =        20

Random effects u_i ~ Gaussian                   Wald chi2(2)       =     95.98
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      market |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |   3.847014   .4834565     7.96   0.000     2.899457    4.794572
       stock |  -.7981618    .256522    -3.11   0.002    -1.300936   -.2953879
       _cons |   1212.764   154.6209     7.84   0.000     909.7122    1515.815
-------------+----------------------------------------------------------------
     sigma_u |  223.80826
     sigma_e |   370.9569
         rho |  .26686395   (fraction of variance due to u_i)
```

[28]在 STATA S.E. 版本中，矩阵的最大维度为 11000，因此，当 $N > 11000$ 时，在 regress 命令中附加虚拟变量来估计 LSDV 模型的方法就不再适用了。

从列表形式上来看，此时得到的结果与 FE 模型并无大异。细心的读者可能注意到表头中呈现了如下信息“corr(u_i,X)=0(assumed)”，这其实就是第 8.2.2 节中的假设 4，我们曾反复强调，该假设是保证 RE 模型估计结果无偏的基本前提。至于其他方面的差异，留待读者自行品味。

3. 模型的筛选和检验

这是模型设定过程中最为关键同时也是最难的一步，主要涉及使用混合 OLS 模型、FE 模型还是 RE 模型，更进一步还可能包括序列相关和异方差的检验等问题。 在这方面功力的提高需要大量的实践经验和对理论的深入理解。

(1) 检验个体效应

对于固定效应模型而言，回归结果中最后一行汇报的 F 统计量便在于检验所有的个体效应整体上是否显著。在我们的例子中，上面的检验结果表明固定效应模型优于混合 OLS 模型。

(2) 检验随机效应

我们可以采用 (8-44) 式的 LM 统计量来检验随机效应是否显著，相应的命令为 xttest0：

```
. qui xtreg market invest stock, re

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        market[id,t] = Xb + u[id] + e[id,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                  market |    2018625       1420.783
                       e |     137609       370.9569
                       u |   50090.14       223.8083

        Test:   Var(u) = 0
                             chi2(1) =   325.74
                         Prob > chi2 =     0.0000
```

这里，qui 命令的作用在于不把估计结果输出到屏幕上。LM 检验得到的 P 值为 0.0000，表明随机效应非常显著。 可见，随机效应模型也优于混合 OLS 模型。

(3) Hausman 检验

虽然通过上面的分析，我们可以确认在模型中加入个体效应 α_i，将显著优于 α_i 为常数假设下的混合 OLS 模型，但还无法明确区分 FE 和 RE 的优劣。此时需要执行第 14 页中介绍的 Hausman 检验，具体步骤为：

- step1：估计固定效应模型，存储估计结果；
- step2：估计随机效应模型，存储估计结果；
- step3：进行Hausman检验；

相应的 STATA 命令为：

```
. qui xtreg market invest stock, fe
. est store fe
. qui xtreg market invest stock, re
. est store re
. hausman fe re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
      invest |    3.05273     3.847014        -.794284               .
       stock |  -.6763434    -.7981618        .1218184               .
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =   -47.57    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
```

这里我们仍然采用 `qui` 命令屏蔽了结果的输出，进而采用 `est store` 命令分别把 FE 和 RE 的估计结果存储到名称为 `fe` 和 `re` 的临时性文件中，并最终用 `hausman` 命令调用二者的结果得到 (8-50) 式中的 Wald 统计量和相应的 P 值。

我们注意到，`sqrt(diag(V_b-V_B))` 全为缺失值，而更为令人疑惑的是，Hausman 检验得到的统计量 `chi2(2)=-47.57`，是一个小于零的数值。理论上来讲，`chi2` 统计量一定为正数，这是在进行 Hausman 检验过程中经常遇到的问题。产生这些情况的原因可能有多种，但我认为一个主要的原因是模型设定有问题，导致 Hausman 检验的基本假设得不到满足。[29]这时，最好先重新审视一下模型设定是否合理，看看是否遗漏了重要的解释变量，或者某些变量是非平稳的等等。在确定模型的设定没有问题的情况下再进行 Hausman 检验，如果仍然拒绝原假设或是出现上面的问题，那么我们就认为随机效应模型的基本假设 (个体效应与解释变量不相关) 得不到满足。此时，需要采用工具变量法或是使用固定效应模型。

对于采用 STATA 9.0 或以上版本的读者而言，使用 `hausman` 命令中新增的 `sigmaless` 和 `sigmamore` 两个选项可以大大缓解 `chi2` 值为负的问题。看下面的例子：

```
. hausman fe re, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
```

[29]不过，STATA 手册的观点恰好相反，认为当 `chi2` 统计量为负时，意味着无法拒绝原假设。参见 [U] **hausman**, pp.642。

```
-------------+----------------------------------------------------------------
      invest |     3.05273     3.847014       -.794284        .3300321
       stock |   -.6763434    -.7981618       .1218184        .1205094
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       38.91
                Prob>chi2 =      0.0000

. hausman fe re, sigmaless

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
      invest |     3.05273     3.847014       -.794284        .2580749
       stock |   -.6763434    -.7981618       .1218184        .0942346
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       63.63
                Prob>chi2 =      0.0000
```

我们注意到，虽然通过设定 `sigmaless` 或 `sigmamore` 选项可以保证得到的 `chi2` 统计量为正数，但相应的 P 值却表明，FE 和 RE 不存在显著差异的原假设被高度拒绝了。这印证了我们的观点，即 `chi2` 为负是 Hausman 检验的原假设被拒绝的征兆。

时间固定效应

如果希望进一步在上述模型中加入时间效应，以便估计模型 (8-23)，那么可以采用时间虚拟变量来实现。首先，我们需要定义 $T-1$ 个时间虚拟变量：

```
tab year,  gen(dumt)
drop dumt1
```

这里，`tab` 命令用于列示变量 `year` 的组类别，选项 `gen(dumt)` 用于产生 T 个以 `dumt` 开头的年度虚拟变量。第二条命令的作用在于去掉第一个虚拟变量以避免完全共线性。若在固定效应模型中加入时间虚拟变量，则估计模型 (8-23) 的命令为：

```
xtreg market invest stock dumt*, fe
```

若估计随机效应模型 (8-31) 中进一步控制时间效应，则只需将上述命令中的 `fe` 选项修改为 `re` 即可。

另一个我们非常关心的问题可能是时间效应的联合显著性，即采用类似于 (8-43) 式的 F 统计量来执行 Wald 检验。这可以采用 `test` 命令来完成。假设我们想检验 1992-1995 年的时间效应整体上是否显著，则可以执行如下命令：

```
.   qui xtreg market invest stock dumt*, fe

.   test dumt2 = dumt3 = dumt4 = dumt5 = 0

 ( 1)  dumt2 - dumt3 = 0
 ( 2)  dumt2 - dumt4 = 0
 ( 3)  dumt2 - dumt5 = 0
 ( 4)  dumt2 = 0

       F(  4,    74) =    9.20
            Prob > F =    0.0000
```

拟合值和残差的获取

完成上述模型的估计后，可以采用 predict 命令获取被解释变量的拟合值，以及模型的残差估计值。使用 xtreg 命令完成模型的估计后，predict 命令的语法格式如下：

predict *newvar* [*if*] [*in*] [, *statistic*]

表 8-3 呈现了 statistic 选项中各个统计量的代码及其含义。

表 8-3: predict 命令中 statistic 选项的含义

统计量	含义
xb	$\mathbf{x}_{it}'\hat{\boldsymbol{\beta}}$，$y_{it}$ 的拟合值，默认选项
ue	$\hat{\alpha}_i + e_{it}$，复合残差项
*xbu	$\mathbf{x}_{it}'\hat{\boldsymbol{\beta}} + \hat{\alpha}_i$，包含个体效应的拟合值
*u	$\hat{\alpha}_i$，固定效应估计值或随机效应的误差成分
*e	e_{it}，残差

注：不带星号的统计量会针对所有样本进行计算，若只需计算参与回归的样本对应的统计量，可以采用 predict...if e(sample)... 命令；带星号的统计量仅针对参与回归的样本进行计算。

在下面的例子中，我们首先采用 1997-2010 年期间的样本估计了固定效应模型 ([L1])，进而采用 predict 命令计算了被解释变量 market 的拟合值 ([L2])。由于在 [L2] 行的命令中，我们附加了 if e(sample) 副指令，计算出的拟合值 market_hat 仅包含在 [L1] 行中参与回归分析的样本对应的观察值，即 1997-2010 年样本区间。因此，market_hat 变量中会包含 30 个缺漏值 (5 家公司在 1991-1996 年期间的观察值)。在 [L3] 和 [L4] 行，我们采用两种方式计算了 $\hat{\alpha}_i + e_{it}$，从最后一行 sum 命令呈现的结果来看，这两种计算方法完全等价。第 [L5] 和 [L6] 行的命令，分别通过附加 e 和 u 选项，计算出了个体效应 $\hat{\alpha}_i$ 和残差 e_{it}。

```
.   qui xtreg market invest stock if year>1996, fe  //                [L1]

.   predict market_hat if e(sample), xb    // y_hat = Xb              [L2]
(30 missing values generated)

.   gen ae_my = market - market_hat        // y - y_hat = a_i + e_it [L3]
(30 missing values generated)

.   predict ae_stata if e(sample), ue      // a_i + e_it              [L4]
```

```
(30 missing values generated)

.   predict e, e                      // e_it                    [L5]
(30 missing values generated)

.   predict a, u                      // a_i                     [L6]
(30 missing values generated)

.   sum ae_my ae_stata e a

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       ae_my |        70    1.30e-13    890.2898  -997.8857   1991.987
    ae_stata |        70    1.30e-13    890.2898  -997.8857   1991.987
           e |        70   -2.92e-14    290.6643  -646.4621   863.0408
           a |        70    1.59e-13    841.5047   -783.008   1407.327
```

最后需要说明的是，对于固定效应模型而言，由于原始数据经过了组内变换，所以最终得到的残差序列不但在总样本中均值为零 ($\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it} = 0$)，而且在每个截面内部，其均值同样为零 ($\sum_{t=1}^{T} e_{it} = 0$)。这一看似非常明了的结论在有些分析中显得尤为重要。例如，在研究资本结构文献中，学者们经常采用模型的残差来衡量公司的实际负债率与目标负债率之间的偏离程度；在现金持有文献中，残差通常用来衡量所谓的“超额现金持有 (excess cash)”；而在企业投资行为相关的文献中，残差则用以衡量投资不足或过度投资。[30]在这种情况下，若采用固定效应模型进行估计，在计算残差时，应采用表 8-3 中的“复合残差”，即在 `predict` 命令中附加 `ue` 选项。换言之，个体效应 $\hat{\alpha}_i$ 也应考虑在内。

拟合优度——R^2

虽然在传统的线性回归模型中，大家经常采用 R^2 来比较不同模型的优劣，然而，对于面板模型而言，R^2 的这一功能大为减弱。一方面，由于面板数据兼顾了截面资料和时序资料的特征，在有些研究中人们更关注组间 (between) 方差，而有些研究中则更为关注组内方差，而二者之间往往不具可比性。另一方面，只有在采用 OLS 进行估计，且模型中包含常数项的情况下，R^2 才能作为模型比较的标准，而对于多数面板模型，基本上都是采用 GLS、MLE 或 GMM 来估计的。尤其是在后两种情况下，R^2 几乎没有任何意义。

除了多数教科书讲到方差分解的方式获取 R^2 (参见第三章)，R^2 亦可定义为被解释变量的实际观察值与拟合值之间相关系数的平方。这里所言的拟合值意指上一小节中的 $\mathbf{x}'_{it}\hat{\boldsymbol{\beta}}$，即忽略 $\hat{\alpha}_i$。在上述例子中，`xtreg` 命令都会报告三个 R^2，下面解释其含义。

设 $\hat{\beta}$ 为经由 `xtreg` 命令得到的系数估计值 (可以附加 `be`、`fe` 或 `re` 选项)，同时，用 $\rho^2(x, y)$ 表示 x 和 y 之间相关系数的平方，则

[30]代表性的文献包括：资本结构 (Flannery and Rangan, 2006; Harford et al., 2009)、现金持有 (Opler et al., 1999; Mikkelson and Partch, 2003; Frésard and Salva, 2010)、投资支出 (Richardson, 2006)。

$$
\begin{aligned}
\text{Within } R^2 &: \rho^2\left\{(y_{it}-\bar{y}_i),(\mathbf{x}_{it}'\hat{\boldsymbol{\beta}}-\bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}})\right\} \\
\text{Between } R^2 &: \rho^2(\bar{y}_i,\bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}) \\
\text{Overall } R^2 &: \rho^2(y_{it},\mathbf{x}_{it}'\boldsymbol{\beta})
\end{aligned}
$$

因此，对于固定效应模型而言，只有 Within R^2 是真正意义上的 R^2，而对于组间效应模型而言，只有 Between R^2 是真正意义上的 R^2，对于随机效应模型而言，并不存在真正意义上的 R^2。[31]

8.4 非均齐方差

在 8.2.1 和 8.2.2 小节的模型设定中，我们假设干扰项 ε_{it} 具有独立同分布的特征 (假设 2)。在很多情况下，这一假设显得过于严格。例如，对于大 N 小 T 型面板 (多见于微观个体或企业资料)，由于它主要表现出截面资料的特征，异方差便是一个需要重点考虑的问题；而对于大 T 小 N 型的面板 (多见于宏观资料)，截面内的时序特征往往是分析的重点，此时需要谨慎处理序列相关问题。除此之外，在两种资料形态中，截面相关也是一个不可忽略的问题。 当模型中存在异方差、序列相关或截面相关时，在同方差假设下得到的估计量虽然仍旧是无偏且一致的，但不具有效性。

在公司金融和资产定价领域，面板模型的应用越来越广泛，虽然多数学者多采用 White (1980) 的方法计算了异方差稳健型标准误，但对于序列相关和截面相关却并未给予足够的重视。Petersen (2009) 收集了在 2001-2004 年发表于 JF，JFE 和 RFS [32]中的 207 篇文献，发现其中有 42% 并未针对可能存在的序列相关或组间相关调整其标准误。根据 Petersen 的理论分析和模拟分析，这可能导致严重的统计推断偏误。

8.4.1 异方差

在 FE 模型中，异方差主要源于 ε_{it}，即 $Var(\varepsilon_{it})=\sigma_i^2$；而对于 RE 模型而言，由于其干扰项包含了 ε_{it} 和 α_i 两个部分，二者都可能导致异方差。本节中仅介绍第一种情况，至于后两种情况，文献中应用有限，有兴趣的读者可以参考 Baltagi (2001)。

组间异方差：FGLS 估计

模型的基本设定同 8.2.1 小节，考虑模型 (8-8)，

$$
\mathbf{y}=\mathbf{D}\mathbf{a}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{8-52}
$$

[31]有关这一主题的更多介绍，请参见 STATA11 Manual [xt], p.448，以及 Verbeek (2004, section 10.2.4)。

[32]这是金融领域最顶尖的三本期刊：Journal of Finance (JF), Journal of Financial Economics (JFE), Review of Financial Studies (RFS)。

其中，$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \cdots, \boldsymbol{\varepsilon}_n')'$。这里我们将第 4 页中的假设 2 放松为：

$$\mathrm{Var}(\boldsymbol{\varepsilon}_i|\mathbf{x}_i, \alpha_i) = \sigma_i^2 \mathbf{I}_T \tag{8-53}$$

令 $\boldsymbol{\Sigma} = \mathrm{diag}[\sigma_i^2]$，为 $N \times N$ 矩阵，则

$$\boldsymbol{\Omega}_0 = \mathrm{Var}(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma} \otimes \mathbf{I}_T. \tag{8-54}$$

在 (8-52) 两边同乘 $\mathbf{Q}$ 以消除个体效应，得到

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \tag{8-55}$$

其中，$\mathbf{y}^* = \mathbf{Q}\mathbf{y}$，$\mathbf{X}^* = \mathbf{Q}\mathbf{X}$，$\boldsymbol{\varepsilon}^* = \mathbf{Q}\boldsymbol{\varepsilon}$。干扰项的方差-协方差矩阵可以表示为：

$$\boldsymbol{\Omega} = \mathrm{Var}(\boldsymbol{\varepsilon}^*) = E[\mathbf{Q}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{Q}'] = \mathbf{Q}\boldsymbol{\Omega}_0\mathbf{Q}' \tag{8-56}$$

于是，模型 (8-55) 的 GLS 估计量为：

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{GLS} &= [\mathbf{X}^{*'}\boldsymbol{\Omega}^{-1}\mathbf{X}^*]^{-1}\mathbf{X}^{*'}\boldsymbol{\Omega}^{-1}\mathbf{y}^* \\
&= [\mathbf{X}'\boldsymbol{\Omega}_0^{-1}\mathbf{X}]^{-1}\mathbf{X}'\boldsymbol{\Omega}_0^{-1}\mathbf{y} \\
&= \left[\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\mathbf{X}_i'\mathbf{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\mathbf{X}_i'\mathbf{y}_i\right]
\end{aligned} \tag{8-57}$$

由此可以看出，$\hat{\boldsymbol{\beta}}_{GLS}$ 之所以比 $\hat{\boldsymbol{\beta}}_{WG}$ 更为有效，主要归因为采用 $1/\sigma_i^2$ 作为权重，从而为波动性较大的个体分配较小的权重。

进一步，可以得到 $\hat{\boldsymbol{\beta}}_{GLS}$ 的方差估计量为：

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_{GLS}) = [\mathbf{X}^{*'}\boldsymbol{\Omega}^{-1}\mathbf{X}^*]^{-1} = [\mathbf{X}'\boldsymbol{\Omega}_0^{-1}\mathbf{X}]^{-1} \tag{8-58}$$

要获得相应的 FGLS 估计量，需要首先估计出 $\boldsymbol{\Sigma}$ 中的未知参数 σ_i^2。由于组内估计量在异方差设定下仍然是无偏且一致的，所以我们可以基于组内估计的残差来估计 σ_i^2。令 $e_{it} = y_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}_{WG}$，其中，$\hat{\boldsymbol{\beta}}_{WG}$ 为模型 (8-52) 在同方差设定下的组内估计量，即 (8-10)。由此可以得到 σ_i^2 的一致估计量：

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} e_{it}^2 \tag{8-59}$$

进而可以得到 $\hat{\boldsymbol{\Sigma}} = \mathrm{diag}[\hat{\sigma}_i^2]$，以及 $\hat{\boldsymbol{\Omega}}_0 = \hat{\boldsymbol{\Sigma}} \otimes \mathbf{I}_T$。用 $\hat{\boldsymbol{\Omega}}_0$ 代替 (8-57) 和 (8-58) 式中的 $\boldsymbol{\Omega}_0$ 即可得到相应的 FGLS 估计量。

组间异方差：Wald 检验

Greene (2000, pp.598) 建议采用如下修正后的 Wald 统计量来检验组间异方差：

$$W' = \sum_{i=1}^{N} \frac{(\hat{\sigma}_i^2 - \hat{\sigma}^2)^2}{V_i} \tag{8-60}$$

其中，

$$V_i = \frac{1}{T}\frac{1}{T-1}\sum_{t=1}^{T}(e_{it}^2 - \hat{\sigma}_i^2)^2 \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2$$

在原假设 ($H_0 : \sigma_i^2 = \sigma^2$) 成立的情况下，该统计量的渐进分布为：

$$W' \xrightarrow{d} \chi^2(N)$$

▷ Example

易于证明，给定假设条件 (8-53)，模型 (8-52) 的 GLS 估计量 (8-55) 与如下不考虑个体效应，但假设组间存在异方差的混合模型的 GLS 估计量完全相同：

$$\mathbf{y}_i = \mathbf{X}_i'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \qquad \text{Var}(\boldsymbol{\varepsilon}_i|\mathbf{x}_i, \alpha_i) = \sigma_i^2\mathbf{I}_T$$

首先，需要采用 (8-60) 式的 W' 统计量来检验模型中是否存在组间异方差。Baum(2001) 编写的 `xttest3` 命令可以很方便地完成这一任务：

```
.  qui xtreg market invest stock, fe

.  xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (5)   =      862.08
Prob>chi2 =      0.0000
```

显然，原假设被拒绝了。此时，需要采用 (8-57) 获得 $\boldsymbol{\beta}$ 的 GLS 估计量，命令为 `xtgls`：

```
.  xtgls market invest stock, panels(heteroskedastic)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   no autocorrelation

Estimated covariances      =         5          Number of obs      =        100
Estimated autocorrelations =         0          Number of groups   =          5
Estimated coefficients     =         3          Time periods       =         20
                                                Wald chi2(2)       =     159.43
```

```
                                                Prob > chi2          =     0.0000

      market        Coef.   Std. Err.       z     P>|z|      [95% Conf. Interval]

      invest     3.616278    .5102399     7.09    0.000      2.616226     4.61633
       stock     .7711491     .378626     2.04    0.042      .0290557    1.513243
       _cons      543.555    78.41464     6.93    0.000      389.8652    697.2449
```

其中，组间异方差通过 `panels()` 选项来设定。有兴趣的读者可以对比一下 GLS 估计值与第 21 页中不考虑异方差时的估计结果。

上述结果是采用两步法获得，即，先采用 OLS 估计不考虑异方差的模型 (8-52)，进而利用其残差计算 (8-59) 式中的 $\hat{\sigma}_i^2$，并最终得到 FGLS 估计量。

当然，在完成上述 FGLS 估计后，我们可以采用其残差重新计算 $\hat{\sigma}_i^2$，并带入 (8-57) 式，计算新一轮的估计系数。这一过程可以反复执行，直到两次估计值之间的差异小于某一预设的临界值为止。这一过程称为迭代 GLS 法，只需在上述命令中进一步附加 `igls` 选项即可得到相应的结果：

```
.   xtgls market invest stock, p(het) igls
Iteration 1: tolerance = .15864396
  (output omitted)
Iteration 45: tolerance = 8.265e-08

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   no autocorrelation

Estimated covariances      =          5     Number of obs        =       100
Estimated autocorrelations =          0     Number of groups     =         5
Estimated coefficients     =          3     Time periods         =        20
                                            Wald chi2(2)         =    236.61
Log likelihood             = -759.5837      Prob > chi2          =    0.0000

      market        Coef.   Std. Err.       z     P>|z|      [95% Conf. Interval]

      invest     3.214628    .3395728     9.47    0.000      2.549078    3.880179
       stock     .5016944     .305698     1.64    0.101     -.0974626    1.100851
       _cons     467.0456    36.60579    12.76    0.000      395.2995    538.7916
```

可见，经过 45 轮迭代，模型最终达到了收敛，系数估计值和标准误都有所变化，尤其是 `stock` 变量的标准误明显增大了，此时其系数估计值并不显著。

◁

8.4.2 序列相关

在此前的分析中，我们假设 FE 或 RE 模型中的个体效应 α_i 可以有效捕捉截面内的跨期相关性，[33]并进而假设干扰项 ε_{it} 是 $i.i.d$ 的，即，对于不同的个体，以及同一个个体的不同时点上，ε_{it} 均不相关。然而，对于 T 较大的面板而言，α_i 往往无法完全反映时序相关性，此时 ε_{it} 便可能存在序列相关，在多数情况下被设定为 $AR(1)$ 过程。

这里，我们重点介绍 RE 模型中序列相关的处理方法，对于 FE 模型而言，处理方法相对简单，有兴趣的读者可以参考 Bhargava, Franzini and Narendranathan (1982)，以及 Stata 11 手册 [XT] **xtregar** 中的相关说明。

这里，我们仅介绍 ε_{it} 服从 $AR(1)$ 过程时的估计方法，[34]模型的基本设定如下：

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it} \tag{8-61a}$$

$$u_{it} = \alpha_i + \varepsilon_{it} \tag{8-61b}$$

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + v_{it} \tag{8-61c}$$

其中，$\alpha_i \sim i.i.d\,(0, \sigma_u^2)$，$v_{it} \sim i.i.d\,(0, \sigma_v^2)$，同时我们假设 α_i 与所有 ε_{it} 均不相关，且满足稳定性条件 $|\rho| < 1$。处理的基本思路很简单，我们首先采用传统的处理时间序列模型的方法消除序列相关，进而采用 GLS 估计变换后的模型。对于第一期观察值，同样有两种处理方法：一是采用 Cochrane-Orcutt (1949) 建议的方法，舍弃第一期观察值；二是依据 Prais-Winsten (1956) 的方法，对第一期观察值进行特别处理。

Cochrane-Orcutt 估计

我们对模型 (8-61) 进行准差分处理，得到

$$(y_{it} - \rho y_{it-1}) = (\mathbf{x}_{it} - \rho\mathbf{x}_{it-1})'\boldsymbol{\beta} + (1-\rho)\alpha_i + v_{it}$$

对应的向量形式为：

$$\mathbf{C}\mathbf{y}_i = \mathbf{C}\mathbf{x}_i + \mathbf{C_1}\alpha_i + \mathbf{C}\mathbf{v}_i \tag{8-62}$$

其中，$\mathbf{C_1} = (1-\rho)\mathbf{1}_{T-1}$ 为 $(T-1)\times 1$ 列向量，

$$\mathbf{C} = \begin{bmatrix} -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}_{(T-1)\times T} \tag{8-63}$$

[33]例如，在 8.2.2 小节中已经提到，在 RE 模型中，由于 α_i 不随时间变化，所以组内观察值会存在不随时间改变的序列相关性 (参见 (8-32) 式)。当然，在有些情况下，这显然是一个过于严格的假设条件。

[34]至于 ε_{it} 服从 $AR(2)$、$AR(4)$ 或 $MA(1)$ 过程时的估计方法，有兴趣的读者可以参考 Baltagi (2001) 第五章相关内容。

我们可进一步写出 (8-62) 的矩阵形式：

$$(\mathbf{I}_N \otimes \mathbf{C})\mathbf{y} = (\mathbf{I}_N \otimes \mathbf{C})\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N \otimes \mathbf{C}_1)\mathbf{a} + (\mathbf{I}_N \otimes \mathbf{C})\mathbf{v} \tag{8-64}$$

其中，$\mathbf{a} = (\alpha_1, \alpha_2, \cdots, \alpha_n)'$。令 $\mathbf{u}^* = (\mathbf{I}_N \otimes \mathbf{C}_1)\mathbf{a} + (\mathbf{I}_N \otimes \mathbf{C})\mathbf{v}$，则模型 的方差-协方差矩阵为：[35]

$$\boldsymbol{\Omega} = E[\mathbf{u}^*\mathbf{u}^{*'}) = (1-\rho)^2\sigma_\alpha^2(\mathbf{I}_N \otimes \mathbf{J}_{T-1}) + \sigma_v^2(\mathbf{I}_N \otimes \mathbf{C}\mathbf{C}') \tag{8-65}$$

其中，$\mathbf{J}_{T-1}$ 为 $(T-1)\times(T-1)$ 方阵，所有元素都为 1。 于是，模型 (8-64) 的 GLS 估计为：

$$\hat{\boldsymbol{\beta}}_{GLS} = [\mathbf{X}^{*'}\boldsymbol{\Omega}^{-1}\mathbf{X}^*]^{-1}\mathbf{X}^{*'}\boldsymbol{\Omega}^{-1}\mathbf{y}^* \tag{8-66}$$

系数的方差-协方差矩阵为：

$$\text{Var}(\hat{\boldsymbol{\beta}}_{GLS}) = [\mathbf{X}^{*'}\boldsymbol{\Omega}^{-1}\mathbf{X}^*]^{-1} \tag{8-67}$$

其中，$\mathbf{y}^* = (\mathbf{I}_N \otimes \mathbf{C})\mathbf{y}$，$\mathbf{X}^* = (\mathbf{I}_N \otimes \mathbf{C})\mathbf{X}$。

由于 $\boldsymbol{\Omega}$ 中含有未知参数 ρ，σ_u^2 和 σ_v^2，所以要进行 FGLS 估计就需要先获得这三个参数的一致估计量。由于模型 (8-61) 的组内估计量是无偏且一致的，所以我们可以利用其残差 e_{it} 来估计 ρ：

$$\hat{\rho} = \sum_{i=1}^{n}\hat{\rho}_i \qquad \text{其中，} \quad \hat{\rho}_i = \frac{\sum_{t=2}^{T} e_{it}e_{it-1}}{\sum_{t=2}^{T} e_{it-1}^2} \tag{8-68}$$

同时，由 (8-61b) 式可知

$$\text{Var}[u_{it}] = \sigma_\alpha^2 + \sigma_v^2/(1-\rho^2) \tag{8-69}$$

而由 (8-61c) 式可知

$$\text{Var}[\varepsilon_{it}] = \sigma_v^2/(1-\rho^2) \tag{8-70}$$

这提示我们可以采用模型 (8-61) 的混合最小二乘估计残差和组内估计残差来估计 σ_u^2 和 σ_v^2。

设 $\hat{e}_{it}$ 为对 (8-61a) 采用 OLS 估计得到的残差，[36]那么依据 (8-69) 有如下关系成立：

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2 = \hat{\sigma}_a^2 + \frac{\hat{\sigma}_v^2}{1-\hat{\rho}^2} \tag{8-71}$$

同时，依据 (8-70) 有如下关系成立：

$$\frac{1}{NT-N-K}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2 = \frac{\hat{\sigma}_v^2}{1-\hat{\rho}^2} \tag{8-72}$$

[35]由于前面假设 α_i 与所有 ε_{it} 均不相关，所以二者的交乘项均为零。同时，计算中要利用矩阵直乘的性质 $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$。

[36]注意，这里我们忽略模型中的个体效应和序列相关的设定。

因此，联立 (8-68)、(8-71) 和 (8-70) 三式，我们可以得到 $\hat{\rho}$、$\hat{\sigma}_v^2$ 和 $\hat{\sigma}_a^2$，并进而得到

$$\hat{\mathbf{\Omega}} = (1-\hat{\rho})^2\hat{\sigma}_a^2(\mathbf{I}_N \otimes \mathbf{J}_{T-1}) + \hat{\sigma}_v^2(\mathbf{I}_N \otimes \mathbf{CC}') \tag{8-73}$$

最终我们得到模型的 FGLS 估计量为：

$$\hat{\boldsymbol{\beta}}_{FGLS} = [\mathbf{X}^{*'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X}^*]^{-1}\mathbf{X}^{*'}\hat{\mathbf{\Omega}}^{-1}\mathbf{y}^* \tag{8-74}$$

和

$$\text{Var}[\hat{\boldsymbol{\beta}}_{FGLS}) = [\mathbf{X}^{*'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X}^*]^{-1} \tag{8-75}$$

Prais-Winsten 估计

当 T 较小而 N 较大时，采用 Cochrane-Orcutt 的方法舍弃第一期观察值会在很大程度上影响估计结果。[37] lillard 和 willis (1978) 采用 Prais 和 Winsten (1956) 处理时间序列模型的方法将我们在 8.2.2 小节 中介绍的随机效应模型扩展为允许干扰项服从自相关的情形。 由于此时需要考虑第一期观察值，所以我们增加对干扰项初始值的假设 $v_{i0} \sim (0, \sigma_\varepsilon^2)$，即 要求干扰项是从一个平稳的均衡态开始的。

首先，我们采用 Prais-Winsten (PW) 转换矩阵

$$\mathbf{C} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}_{T\times T} \tag{8-76}$$

去除模型中的存在的序列相关性。 对于面板数据而言，该转换需要应用到 N 个截面中。转换后干扰项的向量形式为：[38]

$$\mathbf{u}^* = (\mathbf{I}_N \otimes \mathbf{C})\mathbf{u} = [\mathbf{I}_N \otimes (\mathbf{C}\mathbf{1}_T)]\mathbf{a} + (\mathbf{I}_N \otimes \mathbf{C})\boldsymbol{\varepsilon} \tag{8-77}$$

我们注意到 $\mathbf{C1}_T = (1-\rho)\mathbf{1}_T^a$，其中，$\mathbf{1}_T^a = (\delta, \mathbf{1}_{T-1}')'$， 而 $\delta = \sqrt{(1+\rho)/(1-\rho)}$。 因此，(8-77) 式可表示为：

$$\mathbf{u}^* = (1-\rho)(\mathbf{I}_N \otimes \mathbf{1}_T^a)\mathbf{a} + (\mathbf{I}_N \otimes \mathbf{C})\boldsymbol{\varepsilon} \tag{8-78}$$

于是变换后干扰项的方差-协方差矩阵为：[39]

$$\mathbf{\Omega}^* = E[\mathbf{u}^*\mathbf{u}^{*'}) = \sigma_\alpha^2(1-\rho)^2[\mathbf{I}_N \otimes (\mathbf{1}_T^a\mathbf{1}_T^{a\,\prime})] + \sigma_v^2\mathbf{I}_{NT} \tag{8-79}$$

---

[37] 虽然 Cochrane-Orcutt 估计在大样本下能够得到一致性的估计量，但一般的面板数据 (T 较小，N较大) 显然 无法满足获得渐进性估计量的条件。我们一般采用 Cochrane-Orcutt 估计多出于计算的方便。

[38] 推导过程中需要注意，$(\mathbf{I}_N \otimes \mathbf{C})(\mathbf{I}_T \otimes \mathbf{1}_T) = \mathbf{I}_T \otimes \mathbf{C1}_T$。

[39] 我们可以证明 $E[(\mathbf{I}_N \otimes \mathbf{C})\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\mathbf{I}_N \otimes \mathbf{C})') = \sigma_v^2\mathbf{I}_{NT}$。 因为，$\mathbf{C}\boldsymbol{\varepsilon}_i = (\sqrt{1-\rho^2}\varepsilon_{i1}v_{i1}v_{i2}, \cdots v_{iT})'$， 而 由 (8-70) 式可知 $\text{Var}[\varepsilon_{i1}) = \sigma_v^2/(1-\rho^2)$。 所以，$E[\mathbf{C}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{C}') = \sigma_v^2\mathbf{I}_T$。由此，我们就可以很轻易地得到上述结果。

得到 $\mathbf{\Omega}^*$ 后，我们就可以进一步得到类似于 (8-66) 式和 (8-67) 式的 GLS 估计量，并可以 利用前面得到的 $\hat{\rho}$、$\hat{\sigma}_v^2$ 和 $\hat{\sigma}_a^2$ 来获得相应的 FGLS 估计量。

下面我们说明上述变换的方差分解过程，并借此说明序列相关模型与前面提到的一般性随机效应模型之间的关系。

令 $d^2 = \mathbf{1}_T'\mathbf{1}_T$，$\mathbf{J}_T^a = \mathbf{1}_T\mathbf{1}_T'$ 及 $\bar{\mathbf{J}}_T^a = \mathbf{J}_T^a/d^2$。同时我们定义 $\mathbf{I}_T = \mathrm{E}_T^a + \bar{\mathbf{J}}_T^a$， 其中，$\mathrm{E}_T^a = \mathbf{I}_T - \bar{\mathbf{J}}_T^a$。那么 (8-79) 可重新表示为：

$$\mathbf{\Omega}^* = \sigma_\tau^2(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T^a) + \sigma_v^2(\mathbf{I}_N \otimes \mathrm{E}_T^a) \tag{8-80}$$

其中，$\sigma_\tau^2 = d^2\sigma_\alpha^2(1-\rho)^2 + \sigma_v^2$。因此，

$$\sigma_v^2\mathbf{\Omega}^{*-1/2} = (\sigma_v/\sigma_\tau)(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T^a) + (\mathbf{I}_N \otimes \mathrm{E}_T^a) = \mathbf{I}_{NT} - \theta_\tau(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T^a) \tag{8-81}$$

其中，$\theta_\tau = 1 - (\sigma_v/\sigma_\tau)$。

对经过 PW 转换的观察值 $\mathbf{y}^* = (\mathbf{I}_N \otimes \mathbf{C})\mathbf{y}$ 左乘 $\sigma_v^2\mathbf{\Omega}^{*-1/2}$ 得到 $\mathbf{y}^{**} = \sigma_v^2\mathbf{\Omega}^{*-1/2}\mathbf{y}^*$，其特定元素为：

$$\mathbf{y}_i^{**} = (y_{i1}^* - \theta_\tau\delta b_i, y_{i2}^*\theta_\tau b_i, \cdots, y_{iT}^*\theta_\tau b_i)' \tag{8-82}$$

其中，$b_i = [(\delta y_{i1}^* + \sum_2^T y_{it}^*)/d^2]$ $(i = 1, 2, \cdots, n)$。可以看出，我们在 AR(1) 随机效应模型中对第一期观察值进行了非常特殊的处理。同时我们注意到：

1. 当 $\rho = 0$，即不存在序列相关时， $\mathbf{\Omega}^* = T\sigma_\alpha^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_v^2(\mathbf{I}_N \otimes \mathbf{I}_T)$，等价于 (8-34) 式，这 表明此时模型转化为普通的随机效应模型。
2. 当 $\sigma_\alpha^2 = 0$，即不存在随机效应，我们将得到 $\sigma_\tau^2 = \sigma_v^2$ 及 $\theta_\tau = 0$， 这表明 y_{it}^{**} 将转化为仅作 PW 转换得到的 y_{it}^*。

作上述分解的另一个好处是使我们可以基于对 $\mathbf{\Omega}^*$ 的分解来估计方差成分 σ_α^2 和 σ_v^2。 可以证明：[40]

$$(\mathbf{I}_N \otimes \mathrm{E}_T^a)\mathbf{u}^* \sim (\mathbf{0}, \sigma_v^2[\mathbf{I}_N \otimes \mathrm{E}_T^a]) \tag{8-83}$$

和

$$(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T^a)\mathbf{u}^* \sim (\mathbf{0}, \sigma_\tau^2[\mathbf{I}_N \otimes \bar{\mathbf{J}}_T^a]) \tag{8-84}$$

由此，我们可以得到 σ_v^2 和 σ_τ^2 的如下一致估计量：

$$\hat{\sigma}_v^2 = \frac{\mathbf{u}^{*'}(\mathbf{I}_N \otimes \mathrm{E}_T^a)\mathbf{u}^*}{N(T-1)} \quad \text{和} \quad \hat{\sigma}_\tau^2 = \frac{\mathbf{u}^{*'}(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T^a)\mathbf{u}^*}{N} \tag{8-85}$$

[40] $\mathrm{Var}[(\mathbf{I}_N \otimes \mathrm{E}_T^a)\mathbf{u}^*) = (\mathbf{I}_N \otimes \mathrm{E}_T^a)\mathbf{\Omega}^*(\mathbf{I}_N \otimes \mathrm{E}_T^a)' = \sigma_\alpha^2(1-\rho)^2[\mathbf{I}_N \otimes (\mathrm{E}_T^a\bar{\mathbf{J}}_T^a\mathrm{E}_T^a)] + \sigma_v^2[\mathbf{I}_N \otimes (\mathrm{E}_T^a\mathbf{I}_T\mathrm{E}_T^a)]$。由于 E_T^a、$\bar{\mathbf{J}}_T^a$ 和 $\mathbf{J}_T^a$ 均为幂等矩阵，且 $\mathrm{E}_T^a\bar{\mathbf{J}}_T^a = (\mathbf{I}_T - \bar{\mathbf{J}}_T^a)\mathbf{J}_T^a = \mathbf{0}$，所以 $\mathrm{Var}[(\mathbf{I}_N \otimes \mathrm{E}_T^a)\mathbf{u}^*) = \sigma_v^2[\mathbf{I}_N \otimes \mathrm{E}_T^a]$。同理，可得 (8-84) 式。

这里，我们先可以采用 ρ 的一致估计量 (如 $\tilde{\rho}$) 得到 PW 转换矩阵的一致估计 $\hat{\mathbf{C}}$，然后 用 $(\mathbf{I}_N \otimes \hat{\mathbf{C}})\mathbf{y}$ 对 $(\mathbf{I}_N \otimes \hat{\mathbf{C}})\mathbf{X}$ 作 OLS 回归，得到的残差 $\hat{\mathbf{u}}^*$ 便是 (8-85) 式中 $\mathbf{u}^*$ 的一致估计量。这样我们便可得到 $\hat{\sigma}_v^2$ 和 $\hat{\sigma}_\tau^2$ 的估计值。

因此，AR(1) 随机效应的模型的估计过程可总结如下：[41]

1. 对原始模型进行 PW 转换，这类似于我们处理一般时间序列模型的方法；
2. 从经过 PW 转换后的观察值中减去“伪均值”，如 (8-82) 式的设定；
3. 对经过第二步转换后的模型进行 OLS 估计。

序列相关检验

1. Wooldridge 检验

对于固定效应模型 (8-1)，其一阶差分的形式为：

$$\triangle y_{it} = \triangle \mathbf{x}_{it}\boldsymbol{\beta} + \triangle \varepsilon_{it} \tag{8-86}$$

若设定 $\varepsilon_{it} = \rho\varepsilon_{it-1} + u_{it}$，则 $\Delta\varepsilon_{it} = \rho\Delta\varepsilon_{it-1} + \Delta u_{it}$。那么序列相关的原假设为：

$$H_0 : \rho = 0 \quad v.s. \quad \rho \neq 0$$

在原假设 H_0 成立的情况下，易于证明有如下关系成立：

$$Corr(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-1}) = -0.5 \tag{8-87}$$

即使存在序列相关，(8-86) 式的 OLS 估计量仍然是一致的，假设其残差估计值为 $\tilde{e}_{it}$，设用 $\tilde{e}_{it}$ 对 $\tilde{e}_{it-1}$ 进行 OLS 回归得到的系数估计值为 $\hat{\theta}$，那么上述序列相关 检验就转化为检验 $\hat{\theta}$ 是否显著异于 -0.5，这采用一般的 t 检验即可完成。[42]

至于随机效应模型设定下的序列相关检验就要相对复杂一些，有兴趣的读者可以参考 Baltagi (2001, Section 5.2)，以及 Sosa-Escudero and Bera (2008)。

2. Durbin-Waston 检验

Bhargava, Franzini and Narendranathan(1982) 将时间序列中广泛应用的 Durbin-Waston 检验扩展到了面板情形下。原假设为：

$$H_0 : \rho = 0 \quad v.s. \quad \rho > 0 \quad \text{or} \quad \rho < 0$$

检验统计量为：

$$DW_p = \frac{\sum_{i=1}^N \sum_{t=2}^T (e_{it} - e_{it-1})^2}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} \tag{8-88}$$

[41] STATA 11 手册 [XT] **xtregar** (pp.485-486)，对这一转换过程进行了非常简洁明了的说明。

[42] 对于这部分内容的详细介绍，请参考 Wooldridge (2002, pp.282)。Drukker(2003) 详细介绍了这一检验的基本思路，并编写一个 STATA 命令 (`xtserial`) 来执行这一检验。

其中，e_{it} 为固定效应模型 (8-3) 或 (8-5) 的残差。Bhargava et al. (1982) 沿袭 Durbin and Waston 的思路，进一步推导出在不同 N, T 和 K 取值下，DW_p 的临界值的上限和下限，见表 8-4。

在时间序列分析中，D-W 统计量可能落入所谓的“不确定区域 (inconclusive region)”，这是该统计量的主要局限。然而，Bhargava et al. (1982) 发现，基于面板数据得到的 DW_p 统计量的不确定区域要窄得多，在 N 较大的面板中尤其如此，这从见表 8-4 中的结果可以得到清晰的印证。

对于一个 $N = 100, T = 6$，且包含三个解释变量 ($K = 3$) 的模型而言，若 $DW_p < 1.859$，则我们可以在 5% 显著水平上拒绝 $H_0 : \rho = 0$。对于 N 非常大的面板而言，Bhargava et al. (1982) 建议了一个简单的判断准则：若 DW_p 小于 2，则认为存在显著为正的序列相关。虽然 Bhargava et al. (1982) 仅针对 FE 模型推导了 DW_p 统计量，但由于组内估计量也是 RE 模型的一致估计量，所以对于 RE 模型而言，DW_p 统计量仍然适用。

表 8-4: 面板 Durbin-Waston 统计量的 5% 上限和下限 (Bhargava et al., 1982)

		$N = 100$		$N = 500$		$N = 1000$	
		d_L	d_U	d_L	d_U	d_L	d_U
$T = 6$	$K = 3$	1.859	1.880	1.939	1.943	1.957	1.959
	$K = 9$	1.839	1.902	1.935	1.947	1.954	1.961
$T = 10$	$K = 3$	1.891	1.904	1.952	1.954	1.967	1.968
	$K = 9$	1.878	1.916	1.949	1.957	1.965	1.970

Bhargava et al. (1982) 提出的 DW_p 统计量仅适用于平行面板数据，Baltagi and Wu (1999) 进一步将其扩展至非平行面板，并提出另一种统计量。在随后将要介绍的 `xtregar` 命令中附加 `lbi` 选项可以很方便地计算出这两个统计量。

除了上述两种检验方法外，Arellano and Bond (1991) 还提出了一个基于 GMM 估计的更为一般化的序列相关检验方法。Roodman (2009, pp.119-120) 对这一检验方法进行较为详细的论述，并编写了 `abar` 命令来执行这一检验。由于该统计量是在 GMM 估计的基础上提出来的，而 OLS 又是 GMM 的特例，因此，该统计量不但适用于一般的线性回归模型，是适用于 IV 估计活 GMM 估计。至于其用法，我们会在后续章节中介绍。

对于 RE 模型，序列相关检验相对复杂一些。Sosa-Escudero and Bera (2008) 非常详细地探讨了四种相关的检验方法，并编写了 `xttest1` 命令来执行这些检验。

▷ **Example**

1. FE 模型的序列相关检验

对于固定效应模型，可以采用第 38 页中介绍的 Wooldridge 检验法，命令为 `xtserial`：

```
. xtserial market invest stock
```

```
Wooldridge test for autocorrelation in panel data
H0: no first order autocorrelation
    F(  1,       4) =      4.442
           Prob > F =      0.1028
```

这里 $p = 0.0448$，我们可以在 5% 的显著水平上拒绝不存在序列相关的原假设。考虑到本例中样本的时间跨度为 20 年，这个结论应该在预料之中。该检验的最后一步是用 $\tilde{e}_{it}$ 对 $\tilde{e}_{it-1}$ 进行 OLS 回归，因此，输入 `mat list e(b)` 命令可以得到 $\hat{\theta} = -0.319$。

我们尚可采用 Bhargava et al. (1982) 提出的 DW_p 统计量，即 (8-88) 式，来检验序列相关是否存在，以便得到更为稳健的结论。命令如下：

```
.   qui xtreg market invest stock dumt*, fe  // FE estimation
.   predict e, e                             //   e_it
.   qui gen de = D.e                         // D.e_it
.   egen sum_e_sq  = sum(e^2)                // DW_p 分母
.   egen sum_de_sq = sum(de^2)               // DW_p 分子
.   gen dw = sum_de_sq/sum_e_sq              // DW_p
.   dis "D-W = " dw[1]
D-W = 1.4965359
```

这里得到的 $DW_p = 1.497$，明显小于表 8-4 中相应的临界值 1.878。因此，基于 DW_p 统计量的检验同样在 5% 水平上拒绝了原假设。我们亦可采用 `xtregar` 并附加 `lbi` 选项，更为快捷地计算出 DW_p 统计量：

```
.   qui xtregar market invest stock dumt*, fe lbi   // 注意附加 [lbi] 选项
.   dis "D-W = "  e(d1)
D-W = 1.4965359
```

2. RE 模型的序列相关检验

对于 RE 模型，可以采用 `xttest1` 命令来执行 Sosa-Escudero and Bera (2008) 检验：[43]

```
.   qui xtreg market invest stock dumt*, re
.   xttest1
Tests for the error component model:

        market[id,t] = Xb + u[id] + v[id,t]
           v[id,t] = lambda v[id,(t-1)] + e[id,t]

        Estimated results:
                 |       Var     sd = sqrt(Var)
        ---------+-----------------------------
          market |    2018625       1420.783
               e |   88403.93      297.32798
```

[43] 该命令需要下载，可以在命令窗口中输入 `net install sg164_1.pkg` 直接下载，亦可输入 `findit xttest1` 命令，然后依据弹出页面的提示下载之。

```
                    u |    52550.39       229.23873

        Tests:
           Random Effects, Two Sided:
           ALM(Var(u)=0)          =  163.26 Pr>chi2(1)  =  0.0000

           Random Effects, One Sided:
           ALM(Var(u)=0)          =   12.78 Pr>N(0,1)   =  0.0000

           Serial Correlation:
           ALM(lambda=0)          =   15.94 Pr>chi2(1)  =  0.0001

           Joint Test:
           LM(Var(u)=0,lambda=0)  =  234.95 Pr>chi2(2)  =  0.0000
```

这里汇报了 4 个统计量，分别用于检验 RE 模型中随机效应 (单尾和双尾)、序列相关以及二者的联合显著性。检验结果表明存在随机效应和序列相关，而且对随机效应和序列相关的联合检验也非常显著。

3. 估计

上述结果表明，无论是 FE 还是 RE 模型，干扰项中都存在显著的序列相关。为此，我们进一步采用 `xtregar` 命令来估计模型 (8-61)。首先考虑固定效应模型：

```
.  xtregar market invest stock dumt*, fe lbi

FE (within) regression with AR(1) disturbances  Number of obs      =        95
Group variable: id                              Number of groups   =         5

R-sq:  within  = 0.6588                         Obs per group: min =        19
       between = 0.6379                                        avg =      19.0
       overall = 0.5554                                        max =        19

                                                F(20,70)           =      6.76
corr(u_i, Xb)  = 0.4674                         Prob > F           =    0.0000

------------------------------------------------------------------------------
      market |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |   2.547768   .4751971     5.36   0.000     1.600017    3.495519
       stock |  -.7422344   .2788278    -2.66   0.010    -1.298339   -.1861299
  (output omitted)
       _cons |    1916.31    158.156    12.12   0.000     1600.878    2231.742
-------------+----------------------------------------------------------------
      rho_ar |  .26817919
     sigma_u |  1145.2561
     sigma_e |  289.77582
     rho_fov |  .93983152   (fraction of variance because of u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(4,70) =    74.38               Prob > F = 0.0000
modified Bhargava et al. Durbin-Watson = 1.4965359
Baltagi-Wu LBI = 1.6364756
```

这里，我们通过附加 `fe` 选项来估计固定效应模型，此时 STATA 会采用 Cochrane-Orcutt 转换来消除序列相关。由于样本中有 5 家公司，在完成转换后，观察值从最初的 100 减少为 95。由于附加了 `lbi` 选项，最后两行报告了 Bhargava et al. (1982) 提出的 DW_p 统计量，同时还报

告了 Baltagi and Wu (1999) 的修正统计量 `Baltagi-Wu LBI = 1.636`，略高于前者。遗憾的是，Baltagi and Wu (1999) 并未提供该统计量的临界值，虽然他们提供了标准化后的统计量，但并不适用于 N 较小的面板数据。同时，我们也注意到，这里的估计结果与此前不考虑序列相关时得到的结果存在一定的差异。

若需估计包含 AR(1) 干扰项的 RE 模型，只需在 `xtregar` 命令后附加 `re` 选项即可。此时，STATA 会采用相对有效的 Prais-Winsten 转换来消除序列相关，命令如下：

```
.   xtregar market invest stock dumt*, re lbi

RE GLS regression with AR(1) disturbances       Number of obs      =       100
Group variable: id                              Number of groups   =         5

R-sq:  within  = 0.6906                         Obs per group: min =        20
       between = 0.7060                                        avg =      20.0
       overall = 0.6374                                        max =        20

                                                Wald chi2(22)      =    141.60
corr(u_i, Xb)       = 0 (assumed)               Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      market |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |   2.990253   .5047607     5.92   0.000      2.00094    3.979565
       stock |  -.6175912   .3100871    -1.99   0.046    -1.225351   -.0098317
       (output omitted)
       _cons |   897.5951   250.6101     3.58   0.000     406.4084    1388.782
-------------+----------------------------------------------------------------
      rho_ar |  .26817919   (estimated autocorrelation coefficient)
     sigma_u |  514.27681
     sigma_e |  424.74901
     rho_fov |  .59448231   (fraction of variance due to u_i)
       theta |   .7594229
------------------------------------------------------------------------------
modified Bhargava et al. Durbin-Watson = 1.4965359
Baltagi-Wu LBI = 1.6364756
```

由于 Prais-Winsten 转换保留第一期观察值，因此这里参与回归的样本是 100。我们也注意到，此时得到的 DW_p 和 `LBI` 统计量与 `fe` 模型中完全相同，这是因为，这两个统计量的计算公式并不依赖于我们所估计的模型 (FE 或 RE)。

◁

8.4.3 方差形式未知时的稳健性估计

在上述分析中，我们通过特定的假设设定了异方差和序列相关的具体形式。然而，在实证分析过程中，经常需要作出如下权衡：若上述估计方法的前提假设是正确的，则相应的估计量较为有效； 然而，当这些假设条件不满足，或曰我们错误地设定了干扰项的协方差矩阵，则上述估计量反而不及普通的 FE 或 RE 估计量。

由于在多数情况下，我们都不知道真实的数据生成过程 (GDP)，这就使得出现后一种偏误的可能性较大。为此，在多数文献中，学者们都沿袭了 White (1980) 和 Newey and West (1987)

的思路，仍然采用 Pooled OLS, FE 或 RE 得到系数估计值，但需要根据可能存在的异方差、序列相关或截面相关来调整标准误。

根据不同的面板数据结构，STATA 提供了多种获取稳健型标准误的估计命令，见表 8-5。总体而言，对于“大 T 小 N”型面板而言，`xtgls`, `xtpsce`, `xtsur` 都是不错的选择，而对于“大 N 小 T”型面板而言，则应使用 `xtreg` 命令附加 `vce(robust)` 或 `vce(cluster)` 选项，或使用 `xtscc` 命令。在公司金融领域，Fama and MacBeth (1973) 提出的两步估计法也得到了广泛的应用，使用 `xtfmb` 命令可以很方便地实现这一估计。此外，近十年来，得益于计算机技术的发展，以重复抽样为基础的 `bootstrap` 法也获得了快速的发展。采用该方法计算标准误的好处在于，无需预先假设干扰项的分布特征，而完全基于样本的特征，通过重复抽样重现其经验分布 (emperical distribution)。对于样本较大的面板数据而言，这是个不错的选择。

表 8-5: 面板模型中计算稳健型标准误的 STATA 命令

Command	Option	SE estimates are robust to disturbances that are	Notes
`reg`, `xtreg`	`robust`	heteroscedastic	
`reg`, `xtreg`	`cluster()`	heteroscedastic and autocorrelated	
`xtregar`		autocorrelated with AR(1)[a]	
`newey2`		heteroscedastic and autocorrelated of type MA(q)[b]	can perform IV regression
`xtgls`	`panels()`, `corr()`	heteroscedastic, contemporaneously cross-sectionally correlated, and autocorrelated of type AR(1)	$N < T$ required for feasibility; tends to produce optimistic SE estimates
`xtpcse`	`correlation()`	heteroscedastic, contemporaneously cross-sectionally correlated, and autocorrelated of type AR(1)	large-scale panel regressions with `xtpcse` take a lot of time
`xtscc`		heteroscedastic, autocorrelated with MA(q), and cross-sectionally dependent	

[a] AR(1) refers to first-order autoregression

[b] MA(q) denotes autocorrelation of the moving average type with lag length q.

* Source: Hoechle (2007), Table 1.

“异方差-序列相关”稳健型标准误

当第 8.2.1 小节中的假设 1 成立，但假设 2 不成立时，[44]基于组内估计量的方差矩阵 (参见

[44]若使用 (8-6) 式中的符号，可表示为 $E(\dot{\boldsymbol{\varepsilon}}_i \mid \mathbf{x}_i) = 0$，但 $Var(\dot{\boldsymbol{\varepsilon}}_i \mid \mathbf{x}_i) \neq \sigma^2 1_T$。

(8-11) 式) 得到的系数标准误将是非一致的。我们可以将这一估计式重新表述为：

$$\widehat{Var}\,(\hat{\boldsymbol{\beta}}_{WG}) = \hat{\sigma}^2(\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1} \tag{8-89}$$

其中，$\dot{\mathbf{X}} = \mathbf{QX}$，$\dot{\mathbf{y}} = \mathbf{Qy}$，$\hat{\sigma}^2$ 由 (8-12) 式给出。然而，由于 $E(\dot{\mathbf{X}}_i'\dot{\varepsilon}_i) = 0$，且

$$\left(\frac{1}{N}\dot{\mathbf{X}}'\dot{\mathbf{X}}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}_{WG} - \boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{X}}_i'\dot{\varepsilon}_i$$

上式右侧可以视为以均值为 0 的随机数 ($\dot{\varepsilon}_i$) 为权重的样本加权平均，当 T 固定，而 N 趋于无穷时，利用中央极限定理可得：

$$\frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{X}}_i'\dot{\varepsilon}_i \xrightarrow{d} N[\mathbf{0}, E(\dot{\mathbf{X}}_i'\dot{\varepsilon}_i\dot{\varepsilon}_i'\dot{\mathbf{X}}_i)]$$

因此，对于大 N 小 T 型面板，组内估计量的“异方差-序列相关”稳健型方差-协方差渐进估计量为：

$$\begin{aligned}
\widetilde{Var}(\hat{\boldsymbol{\beta}}_{WG}) &= (\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1}\left(\sum_{i=1}^{N}\dot{\mathbf{X}}_i'\dot{\mathbf{e}}_i\dot{\mathbf{e}}_i'\dot{\mathbf{X}}_i\right)(\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1} \\
&= \left[\sum_{i=1}^{N}\dot{\mathbf{X}}_i'\dot{\mathbf{X}}_i\right]^{-1}\left(\sum_{i=1}^{N}\dot{\mathbf{X}}_i'\dot{\mathbf{e}}_i\dot{\mathbf{e}}_i'\dot{\mathbf{X}}_i\right)\left[\sum_{i=1}^{N}\dot{\mathbf{X}}_i'\dot{\mathbf{X}}_i\right]^{-1} \\
&= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\mathbf{x}}_{it}'\dot{\mathbf{x}}_{it}\right]^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\dot{e}_{it}\dot{e}_{is}\dot{\mathbf{x}}_{it}\dot{\mathbf{x}}_{it}'\right)\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\mathbf{x}}_{it}'\dot{\mathbf{x}}_{it}\right]^{-1}
\end{aligned} \tag{8-90}$$

其中，$\dot{\mathbf{e}}_i = \dot{\mathbf{y}}_i - \dot{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_{WG}$ (Arellano, 1987)，即组内估计量的残差。然而，对于大 T 小 N 型面板而言，该估计量不再具有一致性，此时要采用下一小节介绍的 Driscoll and Kraay (1998) 稳健型估计量。

对于随机效应模型而言，若 α_i 与 $\mathbf{X}_{it}$ 不相关，则只需采用 (8-39) 式转换原始变量，进而用转换后的变量替代上述公式中相应的变量即可。

▷ Example

要获得“异方差-序列相关”稳健型标准误，只需在 `xtreg` 命令中附加 `vce(robust)` 或 `vce(cluster)` 选项即可。例如，对于 FE 模型，我们可以执行如下命令：

```
.  xtreg market invest stock, fe vce(robust)

Fixed-effects (within) regression               Number of obs      =       100
Group variable: id                              Number of groups   =         5

R-sq:  within  = 0.4168                         Obs per group: min =        20
```

```
      between = 0.6960                                        avg =      20.0
      overall = 0.6324                                        max =        20
                                               F(2,4)             =     38.64
corr(u_i, Xb)  = 0.5256                        Prob > F           =    0.0024

                                    (Std. Err. adjusted for 5 clusters in id)
```

market	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
invest	3.05273	1.13323	2.69	0.054	-.0936203	6.199081
stock	-.6763434	.501297	-1.35	0.249	-2.068167	.7154801
_cons	1372.613	130.4248	10.52	0.000	1010.495	1734.73
sigma_u	1023.5914					
sigma_e	370.9569					
rho	.88390837	(fraction of variance due to u_i)				

相比于此前未附加 vce(robust) 选项时的结果 (第 21 页)，虽然系数的估计值未发生变化，但此时得到的标准误明显增大了，致使最终的统计推断也更为保守 (invest 仅在 10% 水平上显著，而 stock 则不在显著)。同时，表头最后一行增加了一条信息“Std. Err. adjusted for 5 clusters in id”，这意味着，对于面板模型而言，STATA 在计算所谓的“robust”标准误时，是以个体为单位调整标准误的。因此，我们得到的“robust”标准误其实是同时调整了异方差和序列相关后的标准误。换言之，上述结果与设定 vce(cluster) 选项的结果完全相同：

```
. xtreg market invest stock, fe vce(cluster id)
```

(*output omitted*)

```
                                    (Std. Err. adjusted for 5 clusters in id)
```

market	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
invest	3.05273	1.13323	2.69	0.054	-.0936203	6.199081
stock	-.6763434	.501297	-1.35	0.249	-2.068167	.7154801
_cons	1372.613	130.4248	10.52	0.000	1010.495	1734.73

(*output omitted*)

◁

“异方差-序列相关-截面相关”稳健型标准误

虽然上述估计方法在估计方差-协方差矩阵时考虑了异方差和序列相关的影响，但都未考虑截面之间的相关性。在早期研究中，Parks (1967) 提出了一个同时考虑异方差和截面相关性的可行性广义最小二乘 (FGLS) 估计量。[45]然而，该模型存在两个主要的局限：其一，虽然多数面板

[45]Kmenta (1986) 在其基础上进行了扩展。Greene (2000, Section 15.2.2, pp.599-601) 对这一方法的推导过程和相关的检验方法进行了较为详细的介绍。在 STATA 中，要估计该模型，可以使用 xtgls 命令并附加 panels(correlated) 选项。

数据都是“大 N 小 T”型结构，但该模型却仅适用于“大 T 小 N”型面板；其二，Beck and Katz (1995) 研究表明，该方法估得的标准误严重偏低。

鉴于 Park-Kmenta 方法的局限，Beck and Katz (1995) 建议仍然采用 Pooled OLS 估计系数，但标准误则采用经过截面相关调整后的标准误 (panel-corrected standard errors, PCSEs)。在 STATA 中，可以采用 `xtpcse` 命令估计这一模型。由于这一模型需要以 $N \times N$ 维协方差矩阵的估计值为基础，因此，对于 N 较大的面板数据而言，其参数估计值将变得非常不不准确。

由于多数面板数据都是“大 T 小 N”型的，截面相关系数的数目会以 N^2 的速度增长，而观察值的数目的增长速度则仅为 N。因此，为了保证模型可以估计，多数研究都会假设样本中的所有个体之间都具有相同的截面相关性，此时，在模型中附加 $T-1$ 个时间虚拟变量即可控制截面相关性。然而，这一处理方法的假设条件在多数情况下都显得过于严格了。

为此，Driscoll and Kraay (1998) 在第五章中介绍的 Newey and West (1987) 估计量的基础上，提出了一个基于非参数方法的协方差估计量，在考虑序列相关的基础上，进一步控制了截面相关的影响。[46]通过这种方法获取稳健型标准误的好处在于，所得到的干扰项的方差-协方差矩阵的估计量并不依赖于截面数目 N。这使得该方法能够有效克服上文提到的 Parks (1967)、Kmenta (1986)、Beck and Katz (1995) 估计量的缺陷 (只有在 T 远大于 N 的情况下，这些估计量才是一致的)。

Driscoll and Kraay (1998) 最初提出的估计量是仅适用于平行面板，Hoechle (2007) 进一步将其扩展为非平行面板。当第 8.2.1 小节中的假设 1 成立，但假设 2 不成立时，组内估计量 $\hat{\boldsymbol{\beta}}_{WG}$ 仍然是 $\boldsymbol{\beta}$ 的无偏估计量，但 Driscoll and Kraay (1998) 建议采用如下“异方差-序列相关-截面相关”稳健型协方差矩阵来估计系数的标准误：

$$\widetilde{Var}(\hat{\boldsymbol{\beta}}_{WG}) = (\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1}\hat{\mathbf{S}}_T(\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1} \tag{8-91}$$

其中，$\hat{\mathbf{S}}_T$ 具有与 Newey and West (1987) 相似的形式：

$$\hat{\mathbf{S}}_T = \hat{\Omega}_0 + \sum_{j=1}^{m(T)} w(j,m)[\hat{\Omega}_j + \hat{\Omega}'_j]\,. \tag{8-92}$$

其中，$m(T)$ 表示残差可能存在序列相关的滞后阶数，[47]$w(j,m)$ 则表示修正后的 Bartlett 权重，定义为 $w(j,m) = 1 - j/(m(T)+1)$，其作用在于保证 $\hat{\mathbf{S}}_T$ 为半正定矩阵，同时对高阶滞后项赋以较小的权重。$(K+1)\times(K+1)$ 维矩阵 $\hat{\Omega}_j$ 定义为 (K 表示 x_it 中非常数项解释变量的个数)：

$$\hat{\Omega}_j = \sum_{t=j+1}^{T} \mathbf{h}_t(\hat{\boldsymbol{\beta}}_{WG})\mathbf{h}_{t-j}(\hat{\boldsymbol{\beta}}_{WG})' \quad \text{其中} \quad \mathbf{h}_t(\hat{\boldsymbol{\beta}}_{WG}) = \sum_{i=1}^{N(t)} \mathbf{h}_{it}(\hat{\boldsymbol{\beta}}_{WG}) \tag{8-93}$$

[46]Arellano (2003, pp.19) 也对这一估计方法进行了较为详细的介绍。

[47]在随后要介绍的 `xtscc` 命令中，$m(T)$ 的默认值为 $m(T) = \text{floor}[4(T/100)^{2/9}]$。更为详细的介绍请参见 Hoechle (2007, p.259)。

其中，$\mathbf{h}_{it}(\hat{\boldsymbol{\beta}}_{WG})$ 表示第 t 年的矩条件，而 $\mathbf{h}_t(\hat{\boldsymbol{\beta}}_{WG})$ 则表示第 t 年所有个体 (从 1 到 $N(t)$) 的矩条件之加总，$N(t)$ 表示第 t 年的公司数目。这意味着 (8-93) 适用于非平行面板。采用 组内估计量获得 $\boldsymbol{\beta}$ 的估计值后，(8-93) 式中的矩条件 $\mathbf{h}_{it}(\hat{\boldsymbol{\beta}}_{WG})$ 可基于组内估计的残差 $(\hat{e}_{it})$ 表示为如下 $(K+1)\times 1$ 维矩条件：

$$\mathbf{h}_{it}(\hat{\boldsymbol{\beta}}_{WG}) = \mathbf{x}_{it}\dot{e}_{it} = \mathbf{x}_{it}(y_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}_{WG})$$

由 (8-92) 和 (8-93) 可以看出，Driscoll and Kraay (1998) 将同一个年度 (t) 内的所有个体的矩条件 $\mathbf{h}_{it}(\hat{\boldsymbol{\beta}})$ 进行平均化处理，从而将 Newey and West (1987) 基于时间序列提出的异方差-序列相关稳健型方差矩阵估计量适用于面板数据。由于 $\mathbf{h}_t(\hat{\boldsymbol{\beta}})$ 定义为所有个体的平均值，经由上述方法得到的标准误始终是一致的，与 N 的大小无关。更为重要的是，通过这种方法得到的标准误对于一般形式的截面相关都是稳健的。

显然，在上述推导过程中，若不对模型进行组内变换，我们便可获得针对 Pooled OLS 估计量的 Driscoll and Kraay (1998) “异方差-序列相关-截面相关”稳健型标准误。

有关截面相关检验的介绍，请参见 Sarafidis and De Hoyos (2006)，以及 Greene(2000，pp.598-601)。

▷ Example

1. 截面相关检验

对于 FE 模型，可以利用 Baum(2001) 编写的 `xttest2` 命令来检验截面相关性：[48]

```
.   qui xtreg market invest stock, fe
.   xttest2

Correlation matrix of residuals:

          __e1     __e2     __e3     __e4     __e5
__e1    1.0000
__e2    0.7939   1.0000
__e3    0.6092   0.5348   1.0000
__e4    0.2504   0.4066   0.7326   1.0000
__e5    0.3103   0.1165   0.3728  -0.1097   1.0000

Breusch-Pagan LM test of independence: chi2(10) =    46.258, Pr = 0.0000
Based on 20 complete observations
```

显然，截面之间不存在相关性的原假设被高度拒绝了，为此，在计算系数标准误时，我们需要考虑这一问题，其处理方式将在下一小节中介绍。

对于 RE 模型而言，我们可以采用 Sarafidis and De Hoyos (2006) 编写的 `xtcsd` 命令来检验截面相关性是否显著：[49]

[48] 下载方式为：`ssc install xttest2, replace`。由于该方法基于 SUR 估计进行检验，因此要求面板资料为 $N < T$ 型结构。

[49] 下载方式为：`ssc install xtcsd, replace`。

```
.   qui xtreg market invest stock, re  // or `fe´ for FE model
.   xtcsd, pesaran  // Pesaran´s Test

Pesaran´s test of cross sectional independence =     4.385, Pr = 0.0000
.   xtcsd, frees    // Frees´s Test

  Frees´ test of cross sectional independence =     0.508
|--------------------------------------------------------|
  Critical values from Frees´ Q distribution
                    alpha = 0.10 :    0.1294
                    alpha = 0.05 :    0.1695
                    alpha = 0.01 :    0.2468
```

这里，我们首先估计了 RE 模型，进而在 `xtcsd` 命令中分别附加 `pesaran` 和 `frees` 选项，采用两种不同的方法检验了截面相关性。

2. 获取稳健型标准误

对于 FE 模型，在确认存在截面相关的情况下，我们可以采用 Hoechle (2007) 编写的 `xtscc` 命令获取 Driscoll and Kraay (1998) 提出的“异方差-序列相关-截面相关”稳健型标准误：

```
.   xtscc market invest stock, fe

Regression with Driscoll-Kraay standard errors   Number of obs     =       100
Method: Fixed-effects regression                 Number of groups  =         5
Group variable (i): id                           F(  2,       4)   =     51.52
maximum lag: 2                                   Prob > F          =    0.0014
                                                 within R-squared  =    0.4168

------------------------------------------------------------------------------
             |             Drisc/Kraay
      market |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |    3.05273   .5832634     5.23   0.006     1.433331    4.672129
       stock |  -.6763434   .3666318    -1.84   0.139    -1.694276    .3415896
       _cons |   1372.613   102.5325    13.39   0.000     1087.937    1657.289
------------------------------------------------------------------------------
```

这里，`xtscc` 根据脚注 47 中的公式自动选择的滞后阶数为 2。系数估计值和 Within-R^2 与 `xtreg,fe` 的结果完全相同，但标准误存在较大差异。可见，在本例中，截面相关对统计推断有较大的影响。

若读者有更好的方法来确定自相关的滞后阶数，则可以通过 `lag()` 选项设定之。当然，在多数情况下，这很难做到。不过，我们可以通过附加 `lag(0)` 选项，来估计仅考虑异方差和截面相关的稳健型标准误，命令如下：

```
xtscc market invest stock, fe lag(0)
```

对于 RE 模型而言，若希望计算 Driscoll and Kraay (1998) 稳健型标准误，可以先采用

xtdata, re 命令进行 (8-39) 式的 RE 转换，进而对转换后的数据执行 xtscc 命令即可。[50]

◁

采用 Bootstrap 获取标准误

Bootstrap 由 Efron (1979) 提出，[51]其基本思想在于，通过从原始样本中执行多次可重复抽样来计算统计量的标准误。相对于传统的统计推断方法 (通常会假设干扰项的分布特征)，该方法只需要假设观测样本是从母体中随机抽取的即可。

为了说明 Bootstrap 的基本原理，考虑如下简单线性回归模型：

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad (i = 1, 2, \cdots, N)$$

只要 x_i 是严格外生的，β 的 OLS 估计量 $\hat{\beta}$ 就是无偏且一致的。然而，在无法确知干扰项 ε_i 的分布特征或样本数较小的情况下，基于渐进分布得到的标准误往往缺乏有效性，致使传统的统计推断方法可能变得非常不准确。[52]

此时，我们可以采用如下三个步骤计算出 $\hat{\beta}$ 的 Bootstrap 标准误：

(1) 从原始样本中可重复地 (with replacement) 抽取 N 个观察值，称之为经验样本 (empirical sample)。由于是可重复抽样，有些观察值会被抽中 1 次，有些被抽中 2 次，而有些则可能一次都未被抽中；

(2) 采用第一步得到的经验样本估计上述模型，得到 β 的 OLS 估计值 $\hat{\beta}_1^{bs}$；

(3) 将上述两个步骤进行 k 次，可以得到 $\hat{\beta}_1^{bs}, \hat{\beta}_2^{bs}, \cdots, \hat{\beta}_k^{bs}$。

完成上述估计后，我们得到了 k 个 β 的经验估计，它们都是无偏且一致估计量。在计算 β 的标准误之前，我们需要简单回顾一下“标准误”这个看似简单的概念。对于随机数 z 而言，我们往往采用标准差 (standard deviation) 来描述其波动性，而对于统计量 (如这里的 $\hat{\beta}$)，我们同样可以采用标准差来衡量其准确程度，并称之为标准误 (standard error)。换言之，标准误其实就是统计量的标准差。在传统的统计推断中，由于我们只能观察到从母体中随机抽取的 N 个特定的观察值，由此只能计算出一个统计量，这使得我们无法计算“统计量的标准差 (这需要多个统计量的观察值)”，而只能依据特定假设条件推导出该统计量的分布特征。[53]然而，在 Bootstrap 中，由于可以通过重复抽样来获得 k 个经验样本，并进而得到 k 个 β 的经验估计值 $\hat{\beta}_1^{bs}, \hat{\beta}_2^{bs}, \cdots, \hat{\beta}_k^{bs}$，从而使得我们能够计算出“统计量的标准差”，即所谓的标准误。显然，采用这种方法，

[50]需要说明的是，这种处理方法得到的估计结果存在微小的偏差。详情请参阅 [XT] **xtdata**, pp.63。

[51]Efron and Tibshirani (1993) 非常详细地介绍了 Bootstrap 的相关理论和应用，Cameron and Trivedi (2009, chap. 13) 介绍了 STATA 中的相关操作方法。

[52]回顾一下我们在 OLS 部分的介绍中是如何获得 s.e.($\hat{\beta}$) 的。在小样本下，我们往往假设 $\varepsilon_i \sim N(0, \sigma^2)$，进而得到 $\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$，由此计算出 s.e.($\hat{\beta}$) 后，进一步计算出 $t = \hat{\beta}/\mathrm{s.e.}(\hat{\beta})$。在 $\varepsilon_i \sim N(0, \sigma^2)$ 的假设下，该统计量服从 t 分布。在大样本下，可以采用中央极限定理，因此上述推断无需假设 ε_i 服从正态分布。显然，当 ε_i 的分布特征与上述假设不一致时，$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ 不再是 $\mathrm{Var}(\hat{\beta})$ 的有效估计量。

[53]如，在 OLS 中，假设 $\varepsilon_i \sim N(0, \sigma^2)$，我们便可推导出 $\hat{\beta} \sim N(\beta, \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1})$。

仅需假设 x_i 是严格外生的，以保证 $\hat{\beta}_j^{bs}$，$(j=1,2,\cdots,k)$ 是 β 的无偏估计量即可。换言之，采用这种方法计算出的标准误将是“异方差-序列相关”稳健型估计量。

基于上述分析，我们可以计算出 $\{\hat{\beta}_1^{bs},\hat{\beta}_2^{bs},\cdots,\hat{\beta}_k^{bs}\}$ 的样本标准差，从而得到 $\hat{\beta}$ 的 Bootstrap 标准误：

$$\widehat{\mathrm{se}}_{bs}=\left\{\frac{1}{k-1}\sum_{j=1}^{k}\left(\hat{\beta}_j^{bs}-\bar{\beta}^{bs}\right)^2\right\}^{1/2} \tag{8-94}$$

其中，$\bar{\beta}^{bs}$ 是 $\hat{\beta}_j^{bs}$ 的样本平均值。

`STATA` 中的多数回归分析命令都支持 `vce(bootstrap)` 选项，以便计算出 Bootstrap 稳健型标准误。[54]对于面板数据模型而言，此前介绍的多数命令 (`xtreg`, `xtregar`, `xtgls` 等) 都支持这一选项。需要说明的是，在针对面板数据进行抽样时，为了保持每个截面内的时序特征，Bootstrap 是以个体 (公司、国家或地区) 为单位进行抽样的。换言之，我们可以把面板中的每家公司视为一个“数据块 (block)”，而 Bootstrap 的抽样单位是这些数据块，而非单个的观察值。[55]

Bootstrap 的精度决定于抽样的次数 (k) 和原始样本 (N) 的大小。STATA 默认的抽样次数为 $k=50$，多数情况下都能够提供比较可信的统计推断。在模型筛选阶段，这有助于节省抽样时间。Efron and Tibshirani (1993) 建议，若采用 Bootstrap 获取标准误，抽样 50-200 次已经可以达到比较稳定的效果。但在样本较大的情况下，尤其是在报告论文的最终结果时，选择 500 或 1000 次抽样可能更为稳妥，这也是多数文献中的做法。

▷ Example

对于 FE 模型，我们可以执行如下命令来获取 Bootstrap 稳健型标准误：

```
.   xtreg market invest stock, fe vce(bootstrap, reps(200) seed(123))
(running xtreg on estimation sample)

Bootstrap replications (200)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200

Fixed-effects (within) regression               Number of obs      =       100
Group variable: id                              Number of groups   =         5

R-sq:  within  = 0.4168                         Obs per group: min =        20
       between = 0.6960                                        avg =      20.0
       overall = 0.6324                                        max =        20

                                                Wald chi2(2)       =      6.09
```

[54]需要说明的是，设定 `vce(bootstrap)` 选项只会改变标准误，而系数估计值则仍然采用相应模型的估计公式。从直觉上我们可能认为 $\bar{\beta}^{bs}$ 是一个优于 $\hat{\beta}_{OLS}$ 的估计量，但事实并非如此。详情请参阅 [U] **bootstrap**, pp.202。

[55]参见 Cameron and Trivedi(2005, section 11.6.2)。

```
corr(u_i, Xb)  = 0.5256                    Prob > chi2       =    0.0475
                                 (Replications based on 5 clusters in id)
```

market	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
invest	3.05273	1.265414	2.41	0.016	.5725642	5.532896
stock	-.6763434	.4211295	-1.61	0.108	-1.501742	.1490553
_cons	1372.613	364.8994	3.76	0.000	657.4229	2087.802
sigma_u	1023.5914					
sigma_e	370.9569					
rho	.88390837	(fraction of variance due to u_i)				

在上述命令中，我们通过 `reps()` 选项将抽样的次数设定为 200。为了使估计过程不至于太枯燥，[56]每完成一次抽样和相应的估计，STATA 会在屏幕上打印一个点，为此，本例中共打印出 200 个点。设定 `nodots` 选项可以避免在屏幕上打点。

这里，我们还通过附加 `seed()` 选项将 Bootstrap 的种子值 (seed) 设定为 135，其作用在于保证每次执行上述命令得到的结果都相同。[57]在撰写学术论文过程中，为了保证日后修改时结果不因随机抽样而发生变化，或保证其他学者可以验证我们的结论，设定这一选项显得尤为重要。

相比于不附加 `vce(bootstrap)` 选项的结果而言，此时得到的标准误较大，从而使统计推断更为保守 (t 值较小)。表头下方的"(Replications based on 5 clusters in id)"信息表明，对于面板模型而言，STATA 是以公司为单位进行抽样的，这有助于保持截面内的时序特征。[58]

◁

2. 采用 Bootstrap 执行 Hausman 检验[59]

传统的 Hausman 检验 (第 8.2.3 小节) 要求在原假设下，RE 模型是完全有效的估计量 (fully efficient)，采用 Bootstrap 可以在该假设不满足的情况下执行 Hausman 检验。

▷ Example

首先，我们需要编写一个简单的程序，以便得到每一轮抽样过程中重点关注的统计量：FE

[56]在有些情况下，Bootstrap 会非常耗时，这种看似"单调"的娱乐方式的确有助于缓解你的急躁心情。

[57]Bootstrap 抽样是一个随机的过程，若不附加 `seed()` 选项，则每次得到的标准误都不会完全相同。在每一轮 Bootstrap 抽样过程中，STATA 都会调用一个所谓的随机数发生器函数，并根据这些随机数从原始样本中抽样。种子值其实就是这个随机数发生器的初始值。详情请参阅 STATA 11 手册 [D] **functions** (pp.223-224)。种子值可以设定为任意正整数，选取不同的种子值仅会导致标准误的微小变动，但统计推断的结论通常不会发生实质性的变化。

[58]显然，采用这种方法得到的标准误将是"异方差-截面相关"稳健型估计量。若希望进一步考虑序列相关，则可以以单个观察值为抽样单位，但同时考虑个体效应，可以采用第 23 页中介绍的 LSDV 法或 `areg` 命令，并附加 `vce(bootstrap)` 选项。

[59]这一小节的分析深受 Cameron (2009, p.430) 的启发。有关这一部分的理论分析，请参见 Cameron (2005, section 21.4.3, pp.717-718)。

和 RE 估计系数的差异：

```
.    * Program to return (bl-b2) for Hausman test of FE v.s. RE
.    cap program drop Hausman_FE_RE

.    program define Hausman_FE_RE, eclass
  1.    version 10
  2.      tempname b bfe bre
  3.      xtreg market invest stock, fe
  4.      matrix `bfe´ = e(b)              // FE coefficient
  5.      xtreg market invest stock, re
  6.      matrix `bre´ = e(b)              // RE coefficient
  7.      matrix `b´ = `bfe´ - `bre´       // Difference between FE and RE
  8.      ereturn post `b´
  9.    end
```

我们将该程序命名为 Hausman_FE_RE，并定义为 eclass 类型，以便将该程序的计算结果传递给后续程序。这主要通过最后一行的 ereturn post 命令来实现的。

接下来，就可以执行 Bootstrap 估计了：[60]

```
.    * Bootstrap estimates for robust Hausman test
.    bootstrap _b, reps(500) seed(135) nodots nowarn: Hausman_FE_RE

Bootstrap results                               Number of obs      =       100
                                                Replications       =       500
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
invest	-.794284	.4802484	-1.65	0.098	-1.735554	.1469854
stock	.1218184	.2707035	0.45	0.653	-.4087508	.6523875
_cons	159.8489	165.7438	0.96	0.335	-165.0031	484.7008

这里，reps() 选项用于设定 Bootstrap 的次数，nodots 选项用于告知 STATA 无需在屏幕上打点，而 nowarn 则可以屏蔽所有警示信息。需要注意的是，根据 Hausman_FE_RE 的定义，这里报告的系数估计值其实是 FE 和 RE 估计的系数差异，显然，除了 invest 变量的系数差异在 10% 水平上显著异于零外，其他两个变量并不存在显著差异。

我们可以进一步执行如下 Wald 检验，以便确认所有变量是否在 FE 和 RE 估计之间存在显著差异，这其实就是稳健型 Hausman 检验：

```
.    * Perform robust Hausman test for FE v.s. RE
.    test invest = stock = _cons = 0

 ( 1)   invest - stock = 0
 ( 2)   invest - _cons = 0
 ( 3)   invest = 0
```

[60]在执行 Bootstrap 命令之前，需要先将上述 Hausman_FE_RE 程序另存为一个同名的 ado 文件，并保存到 personal 文件夹下。另一种方式是，选中上面定义的 Hausman_FE_RE 程序代码，执行快捷键 Ctrl+R，以便将 Hausman_FE_RE 程序定义到内存中。

```
       chi2(  3) =     2.90
     Prob > chi2 =     0.4080
```

可见，若依据 Bootstrap 的检验结果，我们无法拒绝 Hausman 检验的原假设，应该选择相对更为有效的 RE 模型。

◁

Fama-MacBeth 两步法

在公司财务领域，Fama and MacBeth (1973) 两步估计法具有非常广泛的应用。[61]

Fama-MacBeth 估计法分为两步：(1) 在各个年度上分别针对所有样本公司执行 OLS (截面) 回归，得到分年度系数估计值 $\hat{\beta}_t$ $(t=1,2,\cdots,T)$；(2) 计算上述 T 次回归的平均系数，得到全样本的系数估计值，即

$$\begin{aligned}\hat{\beta}_{FM} &= \sum_{t=1}^{T}\frac{\hat{\beta}_t}{T} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sum_{i=1}^{N}X_{it}y_{it}}{\sum_{i=1}^{N}X_{it}^2}\right) = \beta + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sum_{i=1}^{N}X_{it}\varepsilon_{it}}{\sum_{i=1}^{N}X_{it}^2}\right) \qquad (8\text{-}95)\end{aligned}$$

进一步可以计算出其标准误：

$$s.e.(\hat{\beta}_{FM}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{(\hat{\beta}_t-\hat{\beta}_{FM})^2}{T-1}} \qquad (8\text{-}96)$$

由 (8-95) 式可知，在 X_{it} 严格外生的假设下，$E(X_{it}\varepsilon_{it})=0$，此时 $\hat{\beta}_{FM}$ 是其真实值的无偏估计量。对于其标准误而言，(8-96) 成立的前提是 β_t 彼此独立，二者要要求 Corr $(X_{it}\varepsilon_{it}, X_{is}\varepsilon_{is}) \neq 0$ $(t \neq s)$。由于 Fama-MacBeth 两步法并未考虑个体效应，当个体效应对 y_{it} 有显著影响时，将主要反映在干扰项 ε_{it} 中。此时，上述独立性假设不再成立，这会导致 $s.e.(\hat{\beta}_{FM})$ 严重偏低 (Petersen, 2009, pp.446)，进而导致过度拒绝原假设 (因为此时计算出来的 t 值偏大)。

▷ **Example**

我们可以采用 STATA 用户编写的 `xtfmb` 命令实现 Fama-MacBeth 两步法：

```
.  xtfmb market invest stock

Fama-MacBeth (1973) Two-Step procedure          Number of obs     =       100
                                                Num. time periods =        20
```

[61] 虽然，Fama and MacBeth (1973) 提出这一方法的初衷是为了克服干扰项的序列相关问题，但 Petersen (2009) 在对比了该方法与此前介绍的 Pooled OLS、FE，以及考虑异方差、序列相关和截面相关的稳健型估计量的统计性质后，发现在多数情况下，该方法都存在比较严重的偏误。因此，就笔者的观点来看，使用该方法只能作为与前期相关研究进行对比分析时使用。

```
                                              F(  2,     19)     =     28.59
                                              Prob > F           =    0.0000
                                              avg. R-squared     =    0.8954
------------------------------------------------------------------------------
             |            Fama-MacBeth
      market |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |    3.61214   .6539737     5.52   0.000     2.243357    4.980922
       stock |    4.91288   1.334844     3.68   0.002     2.119019     7.70674
       _cons |   185.2234   36.44294     5.08   0.000     108.9475    261.4994
------------------------------------------------------------------------------
```

对比此前的 FE 估计，可以发现此时 stock 变量的系数显著为负，这是因为，Fama-MacBeth 两步法主要反映了数据的横截面特征，而 FE 模型则重点关注组内差异。从这个意义上来讲，Fama-MacBeth 两步法的估计结果与组间效应模型的估计结果更加接近。[62]同时，我们也发现，采用 Fama-MacBeth 两步法得到的 R^2 远高于 FE 或 RE 估计，甚至明显高于 Pooled OLS。当然，在本例中，每个年度上只有 $N = 5$ 个观察值，这可能导致 Fama-MacBeth 两步法的估计结果很不准确。

为了使大家更为深入地理解 Fama-MacBeth 两步法的估计过程，我们可以通过如下简单程序来估计其系数：

```
. qui tsset id year
. egen t = group(year)   // t=1,2,3,...,T
. qui tsset id t
. qui sum t
. local T = r(max)
. qui gen b_cons   = .   // 用于存储各年度的系数估计值
. qui gen b_invest = .
. qui gen b_stock  = .
. qui gen r2       = .   // 用于存储各年度的 R-sq
. forvalues t = 1/`T'{   // 分年度估计
  2.    qui{
  3.          reg market invest stock if (t==`t')
  4.          replace b_cons   = _b[_cons]  in `t'
  5.          replace b_invest = _b[invest] in `t'
  6.          replace b_stock  = _b[stock]  in `t'
  7.          replace r2       = e(r2)      in `t'
  8.        }
  9.  }
. sum b_* r2             // 呈现估计结果

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      b_cons |        20    185.2234    162.9778  -15.99471   682.8146
```

[62]STATA 命令为：xtreg market invest stock, be。

```
    b_invest |        20     3.61214    2.924659  -.6602821   9.543836
     b_stock |        20     4.91288    5.969603  -.2354539   26.91374
          r2 |        20    .8954157    .1047199   .6821356   .9991304
```

当然，我们也可以采用更为简洁的 statsby 命令来完成上述操作：

```
.   preserve

.     qui statsby _b e(r2), by(year) clear: reg market invest stock

.     sum _b* _eq2

    Variable |       Obs        Mean    Std. Dev.        Min        Max
-------------+--------------------------------------------------------
   _b_invest |        20     3.61214    2.924659  -.6602821   9.543836
    _b_stock |        20     4.91288    5.969603  -.2354539   26.91374
     _b_cons |        20    185.2234    162.9778  -15.99471   682.8146
 _eq2_stat_1 |        20    .8954157    .1047198   .6821356   .9991304

.   restore
```

执行 statsby 命令后，内存中现有的数据会被自动清除，代之以各个年度的平均估计系数估计值。为了避免当前数据被篡改，我们首先使用了 preserve 命令，以便在执行 statsby 命令之前备份当前数据，待执行完 statsby 命令，并采用 sum 命令在屏幕上呈现出所需的估计结果后，我们进一步采用 restore 命令重新调入此前执行 preserve 命令时自动备份的数据。

◁

8.5 内生性问题与 IV/GMM 估计

参见 Cameron (2005)

Hausman-Taylor estimator [Sun, 2004, pp.37]

Sun(2004) An Introduction to Panel Data

to be finished

8.6 动态面板模型

动态面板数据模型的典型特征是解释变量中包含被解释变量的滞后项。Nerlove (1971) 和 Nickell (1981) 研究表明，对于 T 很小的数据而言，常用的针对固定效应模型的最小平方虚拟变量估计量 (LSDV) 是有偏的，即使 N 很大甚至趋于无穷大，我们仅能通过增加 T 来降低偏误。

对于 T 很小这一类的动态面板数据而言，文献中提供了多种获得一致估计量的方法，这其中的多数方法 都是在 GMM 框架下进行的。包括 Anderson and Hsiao (1981), Holtz-Eakin et al. (1988), Arellano and Bond (1991), 以及 Ahn and Schmidt (1995, 1997)。

8.6.1 简介

考虑如下动态面板模型：

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \tag{8-97}$$

假设 $\varepsilon_{it} \sim i.i.d.(0, \sigma_\varepsilon^2)$。相比于此前介绍的模型，该模型的解释变量中包含了被解释变量一阶滞后项 $y_{i,t-1}$，在很多情况下，其系数 γ 是分析的重点。

为了便于说明，我们先不考虑外生变量 $\mathbf{x}_{it}$，重点分析如下简化模型：

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}, \quad |\gamma| < 1 \tag{8-98}$$

依据该模型的设定，我们可以将 $y_{i,t-1}$ 表示为：

$$y_{i,t-1} = \gamma y_{i,t-2} + \alpha_i + \varepsilon_{i,t-1} \tag{8-99}$$

这一表述对于理解动态面板模型很有帮助。

下面要说明的是，模型 (8-98) 的 Pooled OLS、组内估计量，以及一阶差分估计量都是有偏且非一致的。

若采用 Pooled OLS 估计模型 (8-98)，则可将模型表述为：

$$y_{it} = \gamma y_{i,t-1} + u_{it}, \quad u_{it} = \alpha_i + \varepsilon_{it}$$

显然，$\text{Corr}(y_{i,t-1}, u_{it}) \neq 0$，因为二者都包含 α_i，这意味着 $y_{i,t-1}$ 是一个内生变量，Pooled OLS 估计量是有偏的。

事实上，即使采用组内变换或差分变换去除个体效应 α_i，仍然无法解决动态面板模型 (8-98) 的内生性问题。先考虑组内估计量，

$$y_{it} - \bar{y}_i = \gamma(y_{i,t-1} - \bar{y}_{i,-1}) + (\varepsilon_{it} - \bar{\varepsilon}_i) \tag{8-100}$$

其中，$\bar{y}_i = (1/T)\sum_{t=1}^{T} y_{it}$，$\bar{y}_{i,-1} = [1/(T-1)]\sum_{t=2}^{T} y_{i,t-1}$。由于 $\bar{\varepsilon}_i$ 中包含 $\varepsilon_{i,t-1}$，而根据 (8-99)，它与 $y_{i,t-1}$ 是相关的，这意味着 $\text{Corr}\,(y_{i,t-1} - \bar{y}_{i,-1}, \varepsilon_{it} - \bar{\varepsilon}_i) \neq 0$，即，$\gamma$ 的组内估计量也是有偏的。[63]

接下来，我们考虑一阶差分估计量，对模型 (8-98) 执行一阶差分变换可得：

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta\varepsilon_{it}, \quad t = 2, 3, \cdots, T \tag{8-101}$$

[63] 只有在 $N \to \infty$ 和 $T \to \infty$ 的情况下，组内估计量才是无偏的。对于多数大 N 小 T 型面板而言，组内估计量都存在严重的下偏偏误。例如，Verbeek (2004, pp.361) 发现，若 γ 的真实值为 0.5，即使 $N \to \infty$，当 $T = 2$ 时，$\text{plim}\ \hat{\gamma}_{WG} = -0.25$；当 $T = 3$ 时，$\text{plim}\ \hat{\gamma}_{WG} = -0.04$；当 $T = 10$ 时，$\text{plim}\ \hat{\gamma}_{WG} = 0.33$。Hisao (2003, Section 4.2) 以及 Canova (2007) 也对这一问题进行了非常细致的探讨。

虽然去除了个体效应 α_i，但 $\text{Corr}(\Delta y_{i,t-1}, \Delta\varepsilon_{it}) \neq 0$，[64]因此，一阶差分估计量也是有偏的。

事实上，我们也不难证明，RE 估计量同样面临上述问题。

这里需要说明的是，虽然在模型 (8-98) 中，γ 的 Pooled OLS 估计量和组内估计量分别上偏 (upwoard bias) 和下偏 (downwoard bias) 于其真实值，但二者却构成了 γ 估计值的上限和下限 (Roodman, 2009)。因此，在后续分析中，我们可以用 OLS 和 FE 估计值作为评估估计结果好坏的一个粗略标准。

8.6.2 IV 估计

通过上述分析可以看出，对于动态面板模型而言，此前在静态面板模型使用的估计方法都不再适用。然而，我们注意到，问题的关键在于 $y_{i,t-1}$ 作为解释变量导致的内生性问题。为此，可以采用 IV 或 GMM 估计得到 γ 的一致估计量。

IV 估计量由 Anderson and Hisao (1981) 提出。对于模型 (8-98) 而言，若 ε_{it} 不存在序列相关，则可以先通过一阶差分去除个体效应 α_i。在模型 (8-101) 中，$y_{i,t-2}$ 可以作为 $\Delta y_{i,t-1}$ 的工具变量，因为 $y_{i,t-2}$ 与 $\Delta y_{i,t-1}$ 相关，但与 $\Delta\varepsilon_{it}$ 不相关。基于此，我们可以得到如下 IV 估计量：

$$\hat{\gamma}_{IV}^1 = \frac{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-2}\Delta y_{it}}{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-2}\Delta y_{i,t-1}} \tag{8-102}$$

采用矩阵的形式可表述为：

$$\hat{\gamma}_{IV}^1 = (\mathbf{y}'_{-2}\Delta\mathbf{y}_{-1})^{-1}(\mathbf{y}'_{-2}\Delta\mathbf{y}) \tag{8-103}$$

上述估计量具有一致性的前提条件 (矩条件) 是：

$$E(y_{i,t-2}\Delta\varepsilon_{it}) = 0 \tag{8-104}$$

按照相似的思路，Anderson and Hisao (1981) 建议亦可采用 $\Delta y_{i,t-2}$ 作为 $\Delta y_{i,t-1}$ 的工具变量，相应的 IV 估计量为：

$$\hat{\gamma}_{IV}^2 = \frac{\sum_{i=1}^N \sum_{t=2}^T \Delta y_{i,t-2}\Delta y_{it}}{\sum_{i=1}^N \sum_{t=2}^T \Delta y_{i,t-2}\Delta y_{i,t-1}} \tag{8-105}$$

采用矩阵的形式可表述为：

$$\hat{\gamma}_{IV}^2 = (\Delta\mathbf{y}'_{-2}\Delta\mathbf{y}_{-1})^{-1}(\Delta\mathbf{y}'_{-2}\Delta\mathbf{y}) \tag{8-106}$$

对应的矩条件为：

$$E(\Delta y_{i,t-2}\Delta\varepsilon_{it}) = 0 \tag{8-107}$$

对于上述两个 IV 估计量的优劣，可以从两个角度来看。其一，相比于 $\hat{\gamma}_{IV}^1$ 估计量，在 $\hat{\gamma}_{IV}^2$ 中，为了构造工具变量 $\Delta y_{i,t-2}$，不得不多损失一年的资料。换言之，要获得 $\hat{\gamma}_{IV}^1$ 至少需要 3 年的观察值，而获得 $\hat{\gamma}_{IV}^2$ 则至少需要 4 年的观察值。其二，Arellano (1989) 发现，若使用

[64]这是因为，$\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$，$\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$，而根据 (8-99)，$y_{i,t-1}$ 与 $\varepsilon_{i,t-1}$ 是相关的。

$\Delta y_{i,t-2}$ 作为工具变量，会导致估计量的方差异常增大，从而使统计推断变得相当不准确。因此，Arellano 建议采用 $y_{i,t-2}$ 做工具变量来执行上述 IV 估计，即推荐使用 $\hat{\gamma}_{IV}^1$。

不过，自从 Hansen (1982) 提出广义矩估计 (GMM) 方法后，上述 IV 估计量很少用于估计动态面板模型，致使以上比较其实上不再是个重要的议题。例如，Ahn and Schmidt (1995) 认为，虽然上述 IV 估计量是一致的，但并非最有效的，因为二者都未充分利用所有可用的矩条件，同时，IV 估计量也未充分考虑差分后的干扰项 ($\Delta\varepsilon_{it}$) 的方差结构。[65]

8.6.3 一阶差分 GMM (FD-GMM) 估计量

基本思路

我们可以从 GMM 的角度重新解读 Anderson and Hisao (1981) 提出的两个 IV 估计量 $\hat{\gamma}_{IV}^1$ 和 $\hat{\gamma}_{IV}^2$，事实上，二者分别对应了矩条件 (8-104) 和 (8-107)。沿袭这一思路，若我们可以找到更多合理的矩条件，[66]相应估计量将更为有效。Arellano and Bond (1991) 正是沿着这一思路提出了在文献中广泛应用的一阶差分 GMM 估计量 (以下简称 FD-GMM)。例如，对于差分方程 (8-101)，当 $t=3$ 时，我们能够观察到的第一期观察值为：

$$y_{i3}-y_{i2}=\gamma(y_{i2}-y_{i1})+(\varepsilon_{i3}-\varepsilon_{i2}) \tag{8-108}$$

显然，此时 y_{i1} 是合理的工具变量。因为，只要 ε_{it} 不存在序列相关，y_{i1} 便与 $(\varepsilon_{i3}-\varepsilon_{i2})$ 不相关，而 y_{i1} 与 $(y_{i2}-y_{i1})$ 则是高度相关的。我们可以将上述分析归结为如下矩条件：

$$E[y_{i1}(\varepsilon_{i3}-\varepsilon_{i2})]=0$$

进一步分析 $t=4$ 的情形，在 (8-101) 中，我们能够进一步观察到第二期观察值：

$$y_{i4}-y_{i3}=\gamma(y_{i3}-y_{i2})+(\varepsilon_{i4}-\varepsilon_{i3}) \tag{8-109}$$

此时，y_{i2} 和 y_{i1} 都是 $(y_{i3}-y_{i2})$ 的合理工具变量，因为二者与 $(\varepsilon_{i4}-\varepsilon_{i3})$ 的相关系数均为零。换言之，此时可以进一步得到如下两个矩条件：

$$E[y_{i1}(\varepsilon_{i4}-\varepsilon_{i3})]=0$$

$$E[y_{i2}(\varepsilon_{i4}-\varepsilon_{i3})]=0$$

显然，上述分析思路可以一直延续到样本中的所有观察区间，随着 t 的增加，工具变量的数目也会逐渐增加，当 $t=T$ 时，合理的工具变量集为 $(y_{i1}, y_{i2}, \cdots, y_{i,T-2})$。

[65]若假设 $\varepsilon_{it}\sim i.i.d.(0,\sigma_\varepsilon^2)$，则 $\Delta\varepsilon_{it}$ 不再服从 $i.i.d.$ 分布。详见 (8-121) 式。

[66]其实，一个非常直接的 GMM 估计量就是把上述两个矩条件联合起来构造相应的 GMM 估计量。

相比于上一节介绍的 IV 估计量，GMM 的主要差异在于可以为不同的观察期设定不同的工具变量集。[67]因此，对于个体 i 而言，各期的工具变量集合可汇总为如下矩阵：

$$Z_i = \begin{bmatrix} [y_{i1}] & 0 & \cdots & 0 \\ 0 & [y_{i1}, y_{i2}] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [y_{i1}, y_{i2}, \cdots, y_{i,T-2}] \end{bmatrix} \tag{8-110}$$

由于模型 (8-98) 中仅包含了 y_{it} 的一阶滞后项，设 $p=1$，则 Z_i 矩阵包含 $T-p-1$ 行，$\sum_{m=p}^{T-2} m$ 列 (这也是矩条件的数目)。样本中所有个体工具变量构成的矩阵为

$$\mathbf{Z} = [Z_1', Z_2', \cdots, Z_N']' \tag{8-111}$$

同时，上述矩条件简写如下：

$$E(Z_i' \Delta\varepsilon_i) = \mathbf{0} \tag{8-112}$$

为了推导出 GMM 估计量，可将该矩条件重新表述为对应的样本矩条件：

$$E[Z_i'(\Delta y_i - \gamma \Delta y_{i,-1})] = \mathbf{0} \tag{8-113}$$

由于工具变量的数目可能远远多于未知参数的个数，我们无法保证上述矩条件严格等于零，此时需要极小化如下目标函数 (参见第 7 章)：

$$\min_{\gamma} \left[\frac{1}{N}\sum_{i=1}^{N}(Z_i'(\Delta y_i - \gamma \Delta y_{i,-1}))\right]' \mathbf{W}_N \left[\frac{1}{N}\sum_{i=1}^{N}(Z_i'(\Delta y_i - \gamma \Delta y_{i,-1}))\right] \tag{8-114}$$

其中，$\mathbf{W}_N$ 是一个对称且正定的权重矩阵。[68] 对 γ 求一阶偏导数，可以得到 γ 的 GMM 估计量：

$$\begin{aligned} \hat{\gamma}_{GMM} = & \left[\left(\sum_{i=1}^{N} \Delta\mathbf{y}_{i,-1}' Z_i\right) \mathbf{W}_N \left(\sum_{i=1}^{N} Z_i' \Delta\mathbf{y}_{i,-1}\right)\right]^{-1} \times \\ & \left[\left(\sum_{i=1}^{N} \Delta\mathbf{y}_{i,-1}' Z_i\right) \mathbf{W}_N \left(\sum_{i=1}^{N} Z_i' \Delta\mathbf{y}_{i}\right)\right] \end{aligned} \tag{8-115}$$

显然，$\hat{\gamma}_{GMM}$ 的性质决定于权重矩阵 $\mathbf{W}_N$ 的选择。例如，若选择 $\mathbf{W}_N = \mathbf{I}$，则上述 GMM 估计量便转化为普通的 IV 或 2SLS 估计量 $\hat{\gamma}_{2SLS}$。[69]虽然当 $N \to \infty, T \to \infty$ 时，$\hat{\gamma}_{IV}$ 是 $\hat{\gamma}$ 的渐进一致

[67]有关面板 GMM 的详细介绍，请参见 Cameron and Trivedi(2005, Chp 22)。

[68]需要说明的是，$\mathbf{W}_N$ 的下标 N 仅表明该权重矩阵依赖于样本数 N，并非用于标示矩阵的维度。

[69]其估计式为：$\hat{\gamma}_{2SLS} = \left[\Delta\mathbf{y}_{-1}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\Delta\mathbf{y}_{-1}\right]^{-1}\left[\Delta\mathbf{y}_{-1}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\Delta\mathbf{y}\right]$。

性估计量，但如同我们在第 58 页脚注 65 中所强调的，该估计量并未考虑干扰项的方差结构，因此并不是 $\hat{\gamma}$ 的有效估计量。

对于小样本而言，我们必须慎重选择 $\mathbf{W}_N$ 以便获得最为有效 (Var $(\hat{\gamma}_{GMM})$ 最小) 的估计量。根据第 7 章中有关 GMM 理论的介绍，我们知道，最优权重矩阵与样本矩 (sample moments) 的协方差矩阵的逆矩阵成正比。这意味着，最优权重矩阵应该满足如下条件：

$$\operatorname*{plim}_{N\to\infty} \mathbf{W}_N = V(Z_i'\Delta\varepsilon_i)^{-1} = E(Z_i'\Delta\varepsilon_i\Delta\varepsilon_i' Z_i)^{-1} \tag{8-116}$$

这意味着，只要能获得 γ 的一致估计量，进而得到 $\hat{\varepsilon}_{it}$，即可用样本均值代替 (8-116) 的期望运算，以便获得最优权重的估计值：

$$\widehat{\mathbf{W}}_N^{opt} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\Delta\hat{\varepsilon}_i\Delta\hat{\varepsilon}_i' Z_i\right)^{-1} \tag{8-117}$$

例如，我们可以选择 $\mathbf{W}_N = \mathbf{I}$，经由 (8-115) 估得 $\hat{\gamma}_{GMM}$ 后，进一步得到 $\Delta\hat{\varepsilon}_i$，由此计算出 $\widehat{\mathbf{W}}_N^{opt}$ 后重新代入 (8-115) 式，即可得到 $\hat{\gamma}_{GMM}$ 的有效估计量。然而，这一处理方法并未考虑 $\Delta\varepsilon_i$ 的方差结构，因此，并非最佳选择。

下面，我们首先利用工具变量集 $\mathbf{Z}$，并采用 GLS 估计差分模型 (8-101)，从而得到 γ 的一步估计量 (one-step estimation) $\hat{\gamma}_1$，并将其残差代入 (8-117) 式，得到最优权重矩阵 $\widehat{\mathbf{W}}_N^{opt}$，进而设 $\mathbf{W}_N = \widehat{\mathbf{W}}_N^{opt}$，由 (8-115) 式估计出 γ 的两步估计量 (two-step estimator) $\hat{\gamma}_2$。

一步 GMM 估计量

为了便于推导，采用向量形式将 (8-98) 式重新表示如下：

$$\mathbf{y}_i = \gamma\mathbf{y}_{i,-1} + \mathbf{1}_T\alpha_i + \varepsilon_i \tag{8-118}$$

采用差分矩阵 $\widetilde{\mathbf{B}}$ 左乘 (8-118) 以去除个体效应 α_i：

$$\widetilde{\mathbf{B}}\mathbf{y}_i = \gamma\widetilde{\mathbf{B}}\mathbf{y}_{i,-1} + \widetilde{\mathbf{B}}\boldsymbol{\varepsilon}_i \tag{8-119}$$

其中，[70]

$$\widetilde{\mathbf{B}} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}_{(T-2)\times T} \tag{8-120}$$

干扰项 $\widetilde{\mathbf{B}}\boldsymbol{\varepsilon}_i$ 的方差-协方差矩阵为：

$$\text{Var}\,(\widetilde{\mathbf{B}}\boldsymbol{\varepsilon}_i) = \sigma_\varepsilon^2\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}' = \sigma_\varepsilon^2\mathbf{G} \tag{8-121}$$

[70]细心的读者会发现，这里的 $\widetilde{\mathbf{B}}$ 矩阵与 (8-16) 式的 $\mathbf{B}$ 矩阵非常相似，只是后者的维度为 $(T-1)\times T$。这是因为，对于模型 (8-101) 而言，只有在 $t=3$ 时，我们才能观察到第一期观察值。

其中，

$$\mathbf{G}=\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}'=\begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix}_{(T-2)\times(T-2)} \tag{8-122}$$

将所有观察值累叠后可将 (8-119) 表示为：

$$(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})\mathbf{y}=\gamma(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})\mathbf{y}_{-1}+(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})\boldsymbol{\varepsilon} \tag{8-123}$$

亦可表示为：

$$\Delta\mathbf{y}=\gamma\Delta\mathbf{y}_{-1}+\Delta\boldsymbol{\varepsilon} \tag{8-124}$$

其中，$\Delta\mathbf{y}_{-1}=(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})\mathbf{y}_{-1}$，相对而言，这种表述方法更容易理解。两边同时左乘 $\mathbf{Z}'$ ((8-111) 式) 可得：

$$\mathbf{Z}'\Delta\mathbf{y}=\gamma\mathbf{Z}'\Delta\mathbf{y}_{-1}+\mathbf{Z}'\Delta\boldsymbol{\varepsilon} \tag{8-125}$$

干扰项的方差-协方差矩阵为：[71]

$$\mathrm{Var}\,(\mathbf{Z}'\Delta\boldsymbol{\varepsilon})=\sigma_{\varepsilon}^2\mathbf{A}_1 \tag{8-126}$$

其中，

$$\mathbf{A}_1=\mathbf{Z}'(\mathbf{I}_N\otimes\mathbf{G})\mathbf{Z}=\sum_{i=1}^{N}Z_i'\mathbf{G}Z_i$$

采用 GLS 估计模型 (8-126) 即可得到 Aellano and Bond (1991) 提出的一步 GMM 估计量：

$$\hat{\gamma}_1=\left[\Delta\mathbf{y}_{-1}'\mathbf{Z}\mathbf{A}_1^{-1}\mathbf{Z}'\Delta\mathbf{y}_{-1}\right]^{-1}\left[\Delta\mathbf{y}_{-1}'\mathbf{Z}\mathbf{A}_1^{-1}\mathbf{Z}'\Delta\mathbf{y}\right] \tag{8-127}$$

若设 $\mathbf{W}_N=\mathbf{A}_1^{-1}$，则可用 (8-115) 式的形式将 (8-127) 式重新表示如下：

$$\hat{\gamma}_1=\mathbf{Q}_1^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{y}_{i,-1}'Z_i\right)\mathbf{A}_1^{-1}\left(\sum_{i=1}^{N}Z_i'\Delta\mathbf{y}_i\right) \tag{8-128}$$

其中，

$$\mathbf{Q}_1=\left(\sum_{i=1}^{N}\Delta\mathbf{y}_{i,-1}'Z_i\right)\mathbf{A}_1^{-1}\left(\sum_{i=1}^{N}Z_i'\Delta\mathbf{y}_{i,-1}\right)$$

[71]推导过程为：$\mathrm{Var}\,(\mathbf{Z}'\Delta\boldsymbol{\varepsilon})=\mathrm{Var}\,(\mathbf{Z}'(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})\boldsymbol{\varepsilon})=\mathbf{Z}'(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})\mathrm{Var}\,(\boldsymbol{\varepsilon})(\mathbf{I}_N\otimes\widetilde{\mathbf{B}}')\mathbf{Z}=\mathbf{Z}'(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})(\sigma_{\varepsilon}^2\mathbf{I}_{NT})(\mathbf{I}_N\otimes\widetilde{\mathbf{B}}')\mathbf{Z}=\sigma_{\varepsilon}^2\mathbf{Z}'(\mathbf{I}_N\otimes\widetilde{\mathbf{B}})(\mathbf{I}_N\otimes\widetilde{\mathbf{B}}')\mathbf{Z}=\sigma_{\varepsilon}^2\mathbf{Z}'(\mathbf{I}_N\otimes\mathbf{G})\mathbf{Z}=\sigma_{\varepsilon}^2\mathbf{A}_1$。可利用如下矩阵运算：$(\mathbf{A}\otimes\mathbf{B})'=\mathbf{A}\otimes\mathbf{B}'$，$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC})\otimes(\mathbf{BD})$。

显然，(8-128) 式中未包含任何未知参数。此外，从 (8-126) 式的变换过程来看，$\hat{\gamma}_1$ 与 Anderson and Hisao 提出的 IV 估计量非常相似，区别在于，它使用了更多的工具变量，更为重要的式，它考虑了 $\Delta\varepsilon_i$ 的方差-协方差结构，因此更为有效。

估计出 $\hat{\gamma}_1$ 后，可以进步得到一步估计量的残差：

$$\Delta\hat{\varepsilon}_i = \Delta\mathbf{y}_i - \hat{\gamma}_1\Delta\mathbf{y}_{i,-1} \tag{8-129}$$

若假设 ε_{it} 是同方差的，则 $\hat{\gamma}_1$ 的方差-协方差矩阵为：

$$\hat{V}_1 = \hat{\sigma}_1^2\mathbf{Q}_1^{-1} \tag{8-130}$$

其中，

$$\hat{\sigma}_1^2 = \frac{1}{NT-K}\sum_{i=1}^{N}\Delta\hat{\varepsilon}_i'\Delta\hat{\varepsilon}_i$$

稳健型估计量为：

$$\hat{V}_{1r} = \mathbf{Q}_1^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{y}_{i,-1}'Z_i\right)\mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{A}_1^{-1}\left(\sum_{i=1}^{N}Z_i'\Delta\mathbf{y}_{i,-1}\right)\mathbf{Q}_1^{-1} \tag{8-131}$$

其中，

$$\mathbf{A}_2 = \sum_{i=1}^{N}Z_i'\Delta\hat{\varepsilon}_i'\Delta\hat{\varepsilon}_i Z_i$$

两步 GMM 估计量

将 (8-129) 式中由一步估计量得到的残差 $\Delta\hat{\varepsilon}_i$ 代入 (8-117) 式，即可得到 GMM 估计的最优权重，$\widehat{\mathbf{W}}_N^{opt} = \mathbf{A}_2$。因此，Arellano and Bond (1991) 提出的两步估计量 (一个真正意义上的 GMM 估计量) 可表示为：

$$\hat{\gamma}_2 = \mathbf{Q}_2^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{y}_{i,-1}'Z_i\right)\mathbf{A}_2^{-1}\left(\sum_{i=1}^{N}Z_i'\Delta\mathbf{y}_i\right) \tag{8-132}$$

其中，

$$\mathbf{Q}_2 = \left(\sum_{i=1}^{N}\Delta\mathbf{y}_{i,-1}'Z_i\right)\mathbf{A}_2^{-1}\left(\sum_{i=1}^{N}Z_i'\Delta\mathbf{y}_{i,-1}\right)$$

此时，$\hat{\gamma}_2$ 的方差-协方差矩阵为：

$$\hat{V}_2 = \mathbf{Q}_2^{-1} \tag{8-133}$$

Arellano and Bond (1991) 的 Monte Carlo 模拟分析表明，对于小样本而言，两步估计量的标准误存在严重的下偏偏误 (downward bias)。因此，他们建议在进行系数统计推断时，仍然采用一步 GMM 估计量，因为一步估计量的权重矩阵与待估参数无关，而两步估计量的权重矩阵则依赖于一步估计得到的残差。在最近的一篇文章中，Windmeijer (2005) 提出了一种在小样本下修正这种偏误方法，其 Monte Carlo 模拟分析表明，修正后的两步估计量的标准误具有很好的统计性质。在 STATA 中，两步估计量的稳健型标准误便是采用 Windmeijer (2005) 的修正公式计算的。

8.6.4 假设检验

序列相关检验

前面已经提到，无论是 IV 估计还是 GMM 估计，工具变量的合理性都依赖于一个重要的前提假设：ε_{it} 不存在序列相关。为此，Arellano and Bond (1991) 提出了如下统计量，以检验差分方程的干扰项 $\Delta\varepsilon_i$ 的 m 阶序列相关是否显著：

$$AR_m = \frac{\sum_{i=1}^{N} \Delta\hat{\varepsilon}'_{mi} \Delta\hat{\varepsilon}_i}{B_1^{1/2}} \tag{8-134}$$

其中，

$$\Delta\hat{\varepsilon}_{mi} = L_m(\Delta\hat{\varepsilon}_i)$$

这里，L_m 表示 m 阶滞后算子。同时，

$$\begin{aligned} B_1 = \sum_{i=1}^{N} \Delta\hat{\varepsilon}'_{mi}\mathbf{G}\Delta\hat{\varepsilon}_{mi} & -2\left(\sum_{i=1}^{N} \Delta\hat{\varepsilon}'_{mi}\Delta\mathbf{y}_{i,-1}\right)\mathbf{Q}_1^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{y}'_{i,-1}Z_i\right)\mathbf{A}_1^{-1}\left(\sum_{i=1}^{N} Z'_i\mathbf{G}\Delta\hat{\varepsilon}_{mi}\right) \\ & +\left(\sum_{i=1}^{N} \Delta\hat{\varepsilon}'_{mi}\Delta\mathbf{y}_{i,-1}\right)\hat{V}_1\left(\sum_{i=1}^{N}\Delta\mathbf{y}'_{i,-1}\Delta\hat{\varepsilon}_{mi}\right) \end{aligned}$$

需要说明的是，即使 ε_{it} 不存在序列相关，其差分项 $\Delta\varepsilon_{it}$ 也必然存在一阶序列相关，[72]但不存在二阶序列相关。因此，在实证分析过程中，我们重点关注的是 $\Delta\varepsilon_{it}$ 是否存在二阶序列相关，即 AR_2 是否显著。

上述 AR_m 统计量是在 ε_{it} 为同方差假设下得到的。若考虑异方差并采用 (8-131) 式计算稳健型标准误，在计算 AR_m 统计量时，需要将 (8-135) 式 B_1 中的 $\mathbf{G}$ 替换为 $\Delta\hat{\varepsilon}'_i\Delta\hat{\varepsilon}_i$，$A_1$ 替换为 A_2，$\hat{V}_1$ 替换为 $\hat{V}_{1r}$。对于两步 GMM 估计量而言，在计算 AR_m 统计量时，要将 (8-135) 式 B_1 中的 $\mathbf{Q}_1$ 替换为 $\mathbf{Q}_2$，A_1 替换为 A_2，$\hat{V}_1$ 替换为 $\hat{V}_2$。同时，还需将 $\Delta\hat{\varepsilon}_{mi}$ 替换为 $\Delta\hat{\hat{\varepsilon}}_{mi} = \Delta\mathbf{y}_i - \hat{\gamma}_2\Delta\mathbf{y}_{i,-1}$，并将 $\mathbf{G}$ 替换为 $\Delta\hat{\hat{\varepsilon}}'_{mi}\Delta\hat{\hat{\varepsilon}}_{mi}$。详见 Stata 11 Manual，[XT] **xtabond**。

过度识别检验：Sargan 检验

在上述 GMM 估计过程中，工具变量的数目 (r) 通常远远大于未知参数的数目 (k)，这意味着 GMM 估计中包含了 $(r-k)$ 个过度识别约束 (over-identifying restrictions)。Arellano and Bond (1991) 建议采用 Sargan (1958) 提出的检验统计量。对于一步 GMM 估计量而言，相应的 Sargan 统计量为：

$$S_1 = \left(\sum_{i=1}^{N} \Delta\hat{\varepsilon}'_i Z_i\right) A_1 \left(\sum_{i=1}^{N} Z'_i\Delta\hat{\varepsilon}_i\right)\left(\frac{1}{\hat{\sigma}_1^2}\right) \sim \chi^2_{r-k} \tag{8-135}$$

[72] 因为 $\Delta\varepsilon_{it}$ 和 $\Delta\varepsilon_{i,t-1}$ 中都包含 $\varepsilon_{i,t-1}$。

对于两步 GMM 估计量而言，相应的 Sargan 统计量为：

$$S_2 = \left(\sum_{i=1}^{N} \Delta\hat{\hat{\varepsilon}}_i' Z_i\right) A_2 \left(\sum_{i=1}^{N} Z_i' \Delta\hat{\hat{\varepsilon}}_i\right) \sim \chi^2_{r-k} \tag{8-136}$$

其中，$\Delta\hat{\hat{\varepsilon}}_{mi} = \Delta\mathbf{y}_i - \hat{\gamma}_2 \Delta\mathbf{y}_{i,-1}$。如果 S_1 或 S_2 大于给定置信水平 (如 5%) 上的临界值，则过度识别的矩条件被拒绝，这意味着部分 Z_i 中的工具变量与干扰项相关，并非合理的工具变量。此时需要重新设定模型或选择新的工具变量。

需要特别说明的是，虽然我们同时列出了 S_1 或 S_2 两个统计量，但 Arellano and Bond (1991) 以及后续大量研究都表明，应该使用 S_2 进行 Sargan 检验。因为，S_1 并未考虑异方差问题，因此存在严重的偏误。

8.6.5 包含其它解释变量的动态面板模型

考虑如下一般化的动态面板模型：

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \tag{8-137}$$

首先，需要通过一阶差分去除个体效应，

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta\mathbf{x}_{it}'\boldsymbol{\beta} + \Delta\varepsilon_{it} \tag{8-138}$$

我们仍然可以采用前面介绍的 GMM 进行估计，差别仅在于 (8-110) 式的工具变量集会有所变化。在正式分析之前，有必要对 $\mathbf{x}_{it}$ 中包含的变量类型进行区分：

(1) 外生变量 (exgenous variables)。若对于任意的 s 和 t，都有 $E(x_{it}\varepsilon_{is}) = 0$ 成立，则称 x_{it} 为严格外生变量。换言之，任何时点的外来干扰都不会影响 x_{it}。

(2) 内生变量 (endogenous variables)。若当 $s \le t$ 时，$E(x_{it}\varepsilon_{is}) \ne 0$，则称 x_{it} 为内生变量。

(3) 先决变量 (predetermined variables)。若当 $s < t$ 时，$E(x_{it}\varepsilon_{is}) \ne 0$，但当 $s \ge t$ 时，$E(x_{it}\varepsilon_{is}) = 0$，则称 x_{it} 为先决变量。对于此类变量而言，虽然当期干扰 ε_{it} 不会影响当期 x_{it}，但会影响其随后各期的观察值，即 $x_{i,t+p}$ $(p > 0)$。

完成上述界定后，我们可以依据如下原则来设定模型 (8-138) 中的工具变量：

(1) 若 x_{it} 为严格外生变量，则 Δx_{it} 可以作为自身的工具变量；

(2) 若 x_{it} 为内生变量，其处理方法与 $y_{i,t-1}$ 完全相同，可以采用滞后二阶以上的水平变量 $(x_{it-s}, s \ge 2)$ 作为 Δx_{it} 的工具变量；

(3) 若 x_{it} 为先决变量，则 $x_{i,t-1}$ 也是 Δx_{it} 的合理工具变量，为此，可以采用滞后一阶以上的水平变量 $(x_{it-s}, s \ge 1)$ 作为 Δx_{it} 的工具变量。

例如，若假设 $\mathbf{x}_{it}$ 中包含的所有变量都是严格外生的，则工具变量集可以表述为：

$$Z_i = \begin{bmatrix} [y_{i1}] & 0 & \cdots & 0 & \Delta\mathbf{x}_{i3}' \\ 0 & [y_{i1}, y_{i2}] & \cdots & 0 & \Delta\mathbf{x}_{i4}' \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & [y_{i1}, y_{i2}, \cdots, y_{i,T-2}] & \Delta\mathbf{x}_{iT}' \end{bmatrix} \tag{8-139}$$

此时，Z_i 的行数与 (8-110) 相同，仍然是 $T-p-1$，而其列数则增加为 $\sum_{m=p}^{T-2} m+k$，其中，k 表示 $\mathbf{x}$ 中包含的变量的个数。

在实际分析过程中，$\mathbf{x}$ 中可能同时包含外生变量、内生变量和先决变量，我们只需按照上述原则设定工具变量矩阵 Z_i 即可。当然，若能够获得其它的工具变量，亦可将其加入 Z_i 矩阵。

滞后阶数的选择 (弱工具变量问题)

显然，当 $\mathbf{x}$ 中包含一个或数个内生变量或先决变量时，工具变量的数目随着样本区间 T 的增大会迅速增加。相对于 Anderson and Hisao (1981) 提出的 IV 估计量，Arellano and Bond (1991) 的 FD-GMM 估计量，以及随后将要介绍的 Blundell and Bover (1998) 的 SYS-GMM 估计量都采用了大量的工具变量 (工具变量的数目随着样本区间 T 的增大会迅速增加)，以便提高 GMM 估计的有效性。虽然从理论上讲，对于内生变量 Δw_{it}，滞后两阶以上的水平变量 $(w_{it-s}, s\geq 2)$ 都是合理的工具变量，但当 $s>4$ 时，w_{it-s} 与 Δw_{it} 之间的相关性可能很低，此时便可能出现所谓的“弱工具变量 (weak instruments)”问题。虽然文献中并未明确给出弱工具变量的定义，但基本的共识是，当工具变量与内生变量 (或先决变量) 之间的相关性较低时，便会出现弱工具变量问题。此时，GMM 估计量将不再具有一致性。近期的研究发现 (Ziliak, 1997; Judson and Owen, 1999)，使用过度的工具变量会导致 GMM 估计量的小样本性质非常糟糕。

工具变量过多还会对假设检验产生重要的影响，[73]例如，Anderson and Sorenson (1996) 以及 Bowsher (2002) 均发现，这可能大幅弱化 Hansen 检验的检定力，表现为 Hansen J 统计量对应的 p 值为 1。Sargan (1958) 在提出 Sargan 统计量时，也特别指出该统计量的偏误与工具变量的个数成正比，因此，若使用 Sargan 统计量的渐进分布进行统计推断，则工具变量的个数不能过大。[74] 因此，对于一个 $T=10$ 的面板而言，选择 $s\leq 3$ 或 $s\leq 4$ 比选择 $s\leq T-2$ 要更合理一些。

估计和检验方法

在 8.6.3 小节介绍的估计和检验方法同样适用于模型 (8-138)。设 $\Delta\mathbf{x}_i^*=[\Delta\mathbf{y}_{i,-1},\mathbf{x}_i{}']'$，则只需用 $\Delta\mathbf{x}_i^*$ 替换 8.6.3 小节中的 $\Delta\mathbf{y}_{i,-1}$，并根据 x_{it} 是内生变量、外生变量还是先决变量调整 Z_i 矩阵的定义，便可利用 8.6.3 小节中介绍的公式得到模型 (8-138) 的各个估计量。

8.6.6 系统 GMM (SYS-GMM) 估计量

to be finished

[73]当然，对于何谓“过多”这一问题，文献中并未形成一致观点。或许是因为即使在工具变量数目很少的情况下，GMM 估计量仍然可能存在偏误 (Roodman, 2009, pp.99)。

[74]“They were found to be proportional to the number of instrumental variables, so that, if the asymptotic approximations are to be used, this number must be small.” 参见 Sargan (1958, pp.393)。

8.7 面板门槛模型

8.7.1 简介

自从 Tong (1978) 提出门限自回归模型 (Threshold Auto-regression, 简称 TAR) 后， 这种非线性时间序列模型在经济和金融领域得到了广泛的应用。 虽然 TAR 模型大部分被应用于时间序列资料， 但 Tiao and Tsay (1994)、Potter (1995)、Marterns, Kofman and Vorst (1998) 也利用此方法分析横截面资料或面板资料。 门限自回归模型在计量方法上有较客观地研究方式， 利用门限变量 (Threshold variable) 来决定不同的分界点， 进而利用门限变量的观察值估计出适合的门限值， 这可以有效避免了一般研究者所使用的主观判定分界点法所造成的偏误。

在估计门限自回归模型时， 必须首先检验是否存在门限效应。 由于未知参数 (nuisance) 的存在将导致检验统计量的分布是非标准的， 即出现了所谓的“Davies Problem”。 因此，Hansen (1999) 建议采用“自体抽样法” (Bootstrap，也称为“拔靴法”) 来计算检验统计量的渐进分布， 以便检验门限效应的显著性。 在拒绝原假设， 即存在门限效应的情况下，Chan (1993) 研究表明， 门限自回归模型的 OLS 估计量具有超一致性， Chan 进而推导出了 OLS 估计量的渐进分布。 不幸的是，未知参数的存在会导致该分布呈现非标准态。 Hansen (1999) 通过似然比检验 (Likelihood Ratio test) 构造“非拒绝域”的方法解决了这个问题。

Hansen (1999) 建议采用两阶段最小二乘法来估计门限面板数据模型。 第一步，对于给定的门限值 (γ)，计算相应的残差平方和 (SSR)， 进而所有 SSR 中的最小值所对应的 γ 值 $\hat{\gamma}$。 第二步，利用 $\hat{\gamma}$ 值来估计模型中不同区间 (Regime) 的系数并作相关的分析。

8.7.2 单一门槛模型

模型的基本设定

基本模型设定如下：

$$y_{it} = u_i + x'_{it}\boldsymbol{\beta}_1 \cdot I(q_{it} \le \gamma) + x'_{it}\boldsymbol{\beta}_2 \cdot I(q_{it} > \gamma) + \varepsilon_{it}, \tag{8-140}$$

其中，$i = 1, 2, \cdots, N$ 表示不同的个体， $t = 1, 2, \cdots, T$ 表示时间，q_{it} 为门槛变量，y_{it} 和 X_{it} 分别为被解释变量和解释变量， $I(\cdot)$ 为一个指标函数，相应的条件成立时取值为 1，否则取值为0。 用下面的方式表示 (8-140) 式可能更为清晰：

$$y_{it} = \begin{cases} \mu_i + x_{it}\boldsymbol{\beta}'_1 + \varepsilon_{it} & q_{it} \le \gamma \\ \mu_i + x_{it}\boldsymbol{\beta}'_1 + \varepsilon_{it} & q_{it} > \gamma \end{cases} \tag{8-141}$$

为了下面分析的方便，我们采用一种更为紧凑的方式来表示 (8-140) 式，设

$$x_{it}(\gamma) = \begin{cases} x_{it} I(q_{it} \le \gamma) \\ x_{it} I(q_{it} > \gamma) \end{cases} \tag{8-142}$$

且 $\boldsymbol{\beta} = (\boldsymbol{\beta}_1' \ \ \boldsymbol{\beta}_2')'$，于是 (8-140) 式等价于：

$$y_{it} = \mu_i + x_{it}'(\gamma)\boldsymbol{\beta} + \varepsilon_{it} \tag{8-143}$$

依据门限变量 q_{it} 与门限值 γ 的相对大小，我们可以将样本观察值分成两个区间。区间的差异表现在回归系数 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$ 的不同上。为了保证 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$ 可以识别，x_{it} 中不能包含诸如性别、国籍等不随时间改变的变量。同时，我们也要求门限变量 q_{it} 是不随时间改变的。假设 ε_{it} 服从均值为 0，方差 σ_ε^2 有限的独立、同分布，即 $\varepsilon_{it} \sim i.i.d\ N(0, \sigma_\varepsilon^2)$。$iid$ 的假设使得我们的解释变量中不能包含被解释变量的滞后期，也就是说，我们的模型是一个静态模型。当然，如何将此模型扩展到动态模型或包含异方差的模型是一个有待进一步研究的问题。

模型的估计

我们需要先去除个体效应 μ_i，采用的方法就是去除组内平均值，这与我们处理一般的固定效应模型所采用的方法是相同的。对 (8-140) 式中所有截面取组内平均得到：

$$\bar{y}_i = \mu_i + \bar{\mathbf{x}}_i(\gamma) + \bar{\boldsymbol{\varepsilon}}_i \tag{8-144}$$

其中，

$$\bar{\mathbf{y}}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{\boldsymbol{\varepsilon}}_i = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}, \quad \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}(\gamma) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} x_{it} I(q_{it} \le \gamma) \\ \sum_{t=1}^{T} x_{it} I(q_{it} > \gamma) \end{pmatrix}.$$

用 (8-143) 式减去 (8-144) 式，得到

$$y_{it}^* = x_{it}^*(\gamma)\boldsymbol{\beta} + \varepsilon_{it}^*. \tag{8-145}$$

其中，$y_{it}^* = y_{it} - \bar{y}_i$，$x_{it}^*(\gamma) = x_{it} - \bar{x}_i$，$\varepsilon_{it}^* = \varepsilon_{it} - \bar{\boldsymbol{\varepsilon}}_i$. 进一步将个体观察值进行累叠[75]

$$y_i^* = \begin{pmatrix} y_{i2}^* \\ \vdots \\ y_{iT}^* \end{pmatrix}, \quad x_i^*(\gamma) = \begin{pmatrix} x_{i2}^*(\gamma) \\ \vdots \\ x_{iT}^*(\gamma) \end{pmatrix}, \quad \varepsilon_i^* = \begin{pmatrix} \varepsilon_{i2}^* \\ \vdots \\ \varepsilon_{iT}^* \end{pmatrix}$$

我们进而可以对所有个体的观察值进行累叠，分别表示为 $\mathbf{Y}^*$、$\mathbf{X}^*(\gamma)$ 和 $\boldsymbol{\varepsilon}^*$，例如，

$$\mathbf{X}^*(\gamma) = \begin{pmatrix} \mathbf{x}_1^*(\gamma) \\ \vdots \\ \mathbf{x}_i^*(\gamma) \\ \vdots \\ \mathbf{x}_n^*(\gamma) \end{pmatrix}$$

[75] 由于在得到 (8-145) 式的过程中，我们去除了组内平均，所以第一期的观察值已经无用，不过如果不去除第一期的观察值，结果也不会受到影响。笔者曾写信给 Hansen 教授确认过这个问题。

采用这种表示方式，(8-145) 式等价于

$$\mathbf{Y}^* = \mathbf{X}^*(\gamma)\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \tag{8-146}$$

对于给定的 γ 值，我们可以采用普通最小二乘法 (OLS) 得到参数 $\boldsymbol{\beta}$ 的一致估计量，即

$$\hat{\boldsymbol{\beta}}(\gamma) = \left[\mathbf{X}^*(\gamma)'\mathbf{X}^*(\gamma)\right]^{-1}\mathbf{X}^*(\gamma)'\mathbf{Y}^*, \tag{8-147}$$

相应的残差向量为，

$$\hat{\mathbf{e}}^*(\gamma) = \mathbf{Y}^* - \mathbf{X}^*(\gamma)\boldsymbol{\beta}^*(\gamma), \tag{8-148}$$

残差平方和为，

$$S_1(\gamma) = \hat{\mathbf{e}}^*(\gamma)'\hat{\mathbf{e}}^*(\gamma) \tag{8-149}$$

Chan (1993) 和 Hansen (1997) 建议采用最小二乘法来估计 γ。我们可以通过最小化 (8-149) 式对应的 $S_1(\gamma)$ 来获得 γ 的估计值，即

$$\hat{\gamma} = \arg\min_{\gamma} S_1(\gamma). \tag{8-150}$$

一旦我们得到了 $\hat{\gamma}$，便可进而得到 $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\gamma})$、残差向量 $\hat{\mathbf{e}}^* = \hat{\mathbf{e}}^*(\hat{\gamma})$ 以及残差的方差

$$\hat{\sigma} = \hat{\sigma}^2(\hat{\gamma}) = \frac{1}{n(T-1)}\hat{\mathbf{e}}^{*'}\hat{\mathbf{e}}^* = \frac{1}{n(T-1)}S_1(\hat{\sigma}) \tag{8-151}$$

其中，n 表示公司数目，T 表示时间跨度。

计算相关问题

在我们对 (8-150) 式极小化从而得到门限值 γ 之 OLS 估计量的过程中，由于残差平方和 $S_1(\gamma)$ 仅决定于 γ，所以它是 γ 的一个非连续函数，其中至多有 nT 个间断点 (因为样本容量为 nT)。但事实上，我们在极小化 (8-150) 式的过程中，仅需要对那些非重复的门槛值 q_{it} 进行搜索即可。因此，一般情况下搜索次数会小于 nT 次。我们可以采用以下步骤对 (8-150) 式进行求解:

- 第一步，对门限变量 q_{it} 中非重复的观察值进行排序，去掉其中最大和最小的观察值各 $\eta\%$ 个 ($\eta > 0$)。
- 第二步，用余下的 N^* 个观察值作为估计样本，将 q_{it} 值从小到大依次代入 (8-146) 式进行估计，得到相应的残差平方和 (8-149)，后者的最小值便对应着 γ 的估计值 $\hat{\gamma}$。

在实际操作过程中，N 可能是一个很大的数值，上面介绍的方法可能相当费时。我们可以采用一种更为简洁的等价方式，使得需要搜索的 γ 值的个数大幅减少。我们不必搜索所有介于第 $\eta\%$ 和 $(1-\eta)\%$ 百分位之间的 q_{it} 值，而是仅仅搜索特定百分位上的数值。采用这种方法在多数实证分析中都能达到我们所期望的精度。在笔者的一篇文章中 (连玉君和程建, 2006)，我们按照 $\{1.00\%, 1.25\%, 1.50\%, 1.75\%, 2\%, \cdots, 99.0\%\}$ 栅格进行搜索，共包括 393 个分位值。

假设检验

1. 门槛效应的检验

虽然在模型的设定中我们假设存在门限效应，但是其是否具有统计上的显著性，还需要做进一步的检验。原假设为不存在门限效应，可以表示为：

$$H_0 : \beta_1 = \beta_2$$

相应的备择假设为：

$$H_1 : \beta_1 \neq \beta_2$$

在原假设 H_0 下，门限值 γ 是无法识别的，[76] 此时传统检验统计量的分布是非标准的，这就是所谓的“Davies Problem”。[77] Hansen (1996) 建议采用“自体抽样法”(Bootstrap) 来模拟似然比检验的渐进分布。在不存在门限效应的原假设下，模型 (8-146) 转化为，

$$y_{it} = \mu_i + x_{it}'\boldsymbol{\beta}_1 + \varepsilon_{it}, \tag{8-152}$$

去除个体效应后，可得，

$$y_{it}^* = x_{it}^*\boldsymbol{\beta}_1^* + \varepsilon_{it}^* \tag{8-153}$$

我们可以采用 OLS 得到 $\boldsymbol{\beta}_1$ 的估计量，表示为 $\tilde{\boldsymbol{\beta}}_1$，相应的残差为 $\tilde{e}_{it}^*$，残差平方和为 $S_0 = \tilde{e}_{it}^{*\prime}\tilde{e}_{it}^*$。似然比检验 (LR test) 基于如下统计量，

$$F_1 = \frac{S_0 - S_1(\hat{\gamma})}{\sigma_\varepsilon^2} = \frac{S_0 - S_1(\hat{\gamma})}{S_1(\hat{\gamma})/n(T-1)} \tag{8-154}$$

F_1 的渐进分布是非标准的，完全不同于卡方分布。而且，一般而言其分布依赖于样本的矩 (如均值、方差、峰度和偏度等)，所以临界值无法查表得到。Hansen (1996) 研究表明，采用“自体抽样法”可以获得其一阶渐进分布，基于此构造的 p 值也将是渐进有效的 (asymptotically valid)。考虑到面板模型的数据结构特征，我们建议采用如下步骤进行自体抽样。

- 第一步，在反复抽样的过程中，假设解释变量 x_{it} 和门槛变量 q_{it} 都是固定不变的。将估计备择假设对应的模型 (8-145) 得到的残差 $\hat{\mathbf{e}}^*$ 按个体分组：$\hat{\mathbf{e}}_i^* = (\hat{\mathbf{e}}_{i1}^*, \hat{\mathbf{e}}_{i2}^*, \cdots, \hat{\mathbf{e}}_{iT}^*)$，把由此得到的样本观察值 $\hat{\mathbf{e}}_1^*, \hat{\mathbf{e}}_2^*, \cdots, \hat{\mathbf{e}}_n^*$ 视为“自体抽样”的实证分布 (empirical distribution)。
- 第二步，从实证分布中随机抽取 (可重复) n 个样本观察值，构造出原假设 H_0 下的“自体抽样”样本，即，$\mathbf{Y}_{bs} = \mathbf{X}^*(\gamma)\boldsymbol{\beta} + \mathbf{e}_{bs}$。[78]

[76] 因为在原假设 H_0 下，我们事实上对 (8-140) 式施加了线性约束 $\beta_1 - \beta_2 = 0$，所以不存在一个唯一的 γ 值，使 (8-140) 式成立。

[77] Davies Problem 是指由于未知参数的存在使得检验统计量服从非标准分布的问题 (Davies, 1977, 1987)。在近期的研究中，Andrews and Ploberger(1994) 以及 Hansen (1996) 又重新审视了这个问题，并在处理方法上做了一些改进。

[78] 需要说明的是，在原假设 H_0 下，F_1 统计量与参数 β_1 无关，所以任何的 β_1 值都可以使用。

- 第三步，利用第二步构造的“自体抽样”样本分别估计原假设对应的模型 (8-153) 和备择假设对应的模型 (8-145), 计算由 (8-154) 式决定的似然比统计量。
- 第四步，重复以上过程多次 (如 500 次)， 计算模拟值大于真实值的概率， 这便是我们采用“自体抽样”法得到的原假设 H_0 下 F_1 统计量的渐进 P 值。

如果得到的 P 值小于我们设定的临界值 (如 5%)，那么就拒绝原假设，从而认为存在门槛效应。

2. 门槛估计值的渐进分布特征

在已经确认存在门槛效应的情况下 ($\beta_1 \neq \beta_2$)，Chan (1993) 以及 Hansen (1997) 研究表明 $\hat{\gamma}$ 是 γ_0 (γ 的真实值) 的一致估计量，然而其渐进分布是高度非标准的。 Hansen (1997) 指出， 构造 γ 的置信区间的最佳方法是利用似然比统计量构造出“非拒绝域”。 对于原假设 $H_0 : \gamma = \gamma_0$ 而言，似然比统计量为：

$$LR_1(\gamma) = \frac{S_1(\gamma) - S_1(\hat{\gamma})}{\hat{\sigma}^2} \tag{8-155}$$

如果 $LR_1(\gamma_0)$ 的值足够大，那么我们就拒绝原假设。 需要注意的是，(8-155) 式的统计量和 (8-154) 式的统计量所检验的原假设是不同的。 $LR_1(\gamma_0)$ 用于检验 $H_0 : \gamma = \gamma_0$，而 F_1 用于检验 $H_0 : \beta_1 = \beta_2$。

在满足一系列假设条件及 $H_0 : \gamma = \gamma_0$ 的情况下， Hansen (1999) 导出 $LR_1(\gamma)$ 的渐进分布满足定理 1：

$$LR_1(\gamma) \to_d \xi \quad (n \to \infty) \tag{8-156}$$

其中，ξ 是一个具有如下分布函数的随机变量，

$$p(\xi \leq x) = (1 - \exp(-x/2))^2 \tag{8-157}$$

这表明似然比统计量具有非标准分布，但不存在“未知参数”问题 (free of nuisance parameters)。 由于定理 1 中的渐进分布是枢轴的 (pivotal)， 因此可用来构造一个有效渐进的置信区间。进一步地，分配函数 (8-157) 的反函数可表示为，

$$c(\alpha) = -2\ln(1 - \sqrt{1-\alpha}) \tag{8-158}$$

据此我们可以很方便的计算出临界值。 比如，10% 显著水平下的临界值为 6.53，5% 临界值为 7.35，1% 临界值为 10.59。 如果 $LR_1(\gamma_0)$ 的值大于 $c(\alpha)$，那么我们就在 α 显著水平上拒绝原假设 $H_0 : \gamma = \gamma_0$。

为了得到 γ 的渐进置信区间，我们可以引入“非拒绝域”的概念。 在 $1-\alpha$ 置信水平上的“非拒绝域”是指一系列满足 $LR_1(\gamma) \leq c(\alpha)$的 γ 值， 二者分别由 (8-155) 式和 (8-158) 式确定。 作为一种简易的直观判断方法，我们可以绘制出以 $LR_1(\gamma)$ 为纵坐标， γ 为横坐标的二维图，同时在 $c(\alpha)$ 处画一条水平线， 以确定置信区间。

3. 系数估计值的渐进分布特征

估计值 $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\gamma})$ 与门槛估计值 $\hat{\gamma}$ 相关联，这似乎使得对 $\boldsymbol{\beta}$ 的统计推论变得相当复杂。Chan (1993) 和 Hansen (1997) 指出这种相关性并不会显著影响一阶渐进性质，所以对 $\boldsymbol{\beta}$ 的统计推论可以在假设门槛估计值 $\hat{\gamma}$ 就是其真实值的情况下进行。因此，$\hat{\boldsymbol{\beta}}$ 的渐进分布将是正态分布，其协方差矩阵可以用下式估计得到：

$$\hat{V} = \hat{\sigma}^2 \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})' \right)^{-1}. \tag{8-159}$$

虽然在构造门槛 γ 的置信区间时，我们需要假设残差项为独立同分布的 (iid)，但是在此处构造估计系数的置信区间时，我们完全可以放松这个假设。在允许残差项存在条件异方差的情况下，$\hat{\boldsymbol{\beta}}$ 的方差-协方差矩阵的估计式为：

$$\hat{V}_h = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})' \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})' (\hat{e}_{it}^*)^2 \right) \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})' \right)^{-1}. \tag{8-160}$$

这事实上是 White 估计式。

8.7.3 多重门槛模型

模型 (8-140) 中仅有一个门槛，但在许多情况下门槛的个数可能不止一个。如双重门槛模型的设定如下:

$$y_{it} = u_i + x_{it}'\boldsymbol{\beta}_1 \cdot I(q_{it} \le \gamma_1) + x_{it}'\boldsymbol{\beta}_2 \cdot I(\gamma_1 < q_{it} \le \gamma_2) + x_{it}'\boldsymbol{\beta}_3 \cdot I(q_{it} > \gamma_2) + \varepsilon_{it}, \tag{8-161}$$

其中 $\gamma_1 < \gamma_2$。这里我们集中讨论双重门槛模型，因为由此可以很方便地扩展到跟多门槛的情形。下面我们重点讨论三个方面的问题：(1) 双重门槛模型的估计; (2) 检验双重门槛效应的显著性; (3) 构造门槛参数 γ_1 和 γ_2 的置信区间。

估计

对于给定的 (γ_1, γ_2)，(8-161) 式是关于 $(\beta_1, \beta_2, \beta_3)$ 的线性模型，所以我们可以用 OLS 进行估计。因此，对于每一组给定的 (γ_1, γ_2)，我们可以首先计算出相应的残差平方和 $S(\gamma_1, \gamma_2)$，这与我们对单一门槛模型的处理相似。(γ_1, γ_2) 的联合 OLS 估计值就是使得 $S(\gamma_1, \gamma_2)$ 最小时对应的值 $(\hat{\gamma}_1, \hat{\gamma}_2)$。虽然从理论上讲，同时估计 $(\hat{\gamma}_1, \hat{\gamma}_2)$ 是可行的，但是实际操作的过程中这种做法是非常费时的。我们对 (γ_1, γ_2) 进行逐个搜索大约需要进行 $N^2 = (nT)^2$ 次回归。

为了减小运算量，我们采用“循环法”进行估计。在含有多个结构突变点的模型中，这种方法可以得到参数的一致估计量，如 Chong (1994)、Bai (1997) 以及 Bai and Perron (1998)。在我们估计多重门槛模型时，也可以采用同样的方法。第一步，设 $S_1(\gamma)$ 为由 (8-149) 式所定义的单一门槛设定下的残差平方和，γ_1 为使得 $S_1(\gamma)$ 最小时对应的门槛估计值。Chong (1994) 以及 Bai (1997) 的研究表明无论对于 γ_1 还是 γ_2 而言，$\hat{\gamma}_1$ 都是 γ_1 的一致估计量 固定第一步得到的 $\hat{\gamma}_1$，估计 (8-161) 式，则第二步的筛选标准为，

$$S_2^{\gamma}(\gamma_2) = \begin{cases} S(\hat{\gamma}_1, \gamma_2) & 若\ \hat{\gamma}_1 < \gamma_2 \\ S(\gamma_2, \hat{\gamma}_1) & 若\ \gamma_2 < \hat{\gamma}_1 \end{cases} \tag{8-162}$$

进而得到第二步的门槛估计值为，

$$\hat{\gamma}_2^{\gamma} = \arg\min_{\gamma_2} S_2^{\gamma}(\gamma_2). \tag{8-163}$$

由于若个别区间内的观察值过少我们无法进行有效的估计， 所以我们可以在搜索 (8-162) 式的过程中限制每个区间内的最小观察值个数， 即，若某个以门槛值划分的一个区间内的观察值少于 30 个， 我们就跳过此门槛值进行下一个门槛值的搜索。

Bai (1997) 的研究表明 $\hat{\gamma}_2^{\gamma}$ 是渐进有效的， 但 $\hat{\gamma}_1$ 的估计却不具有此性质。 这是因为在估计 $\hat{\gamma}_1$ 的过程中， 残差平方和中包含了我们所忽略的区间造成的影响。 但由于 $\hat{\gamma}_2^{\gamma}$ 是渐进有效的，所以我们可以固定 $\hat{\gamma}_2^{\gamma}$， 然后重新估计 ，此时参数的筛选标准为，

$$S_1^{\gamma}(\gamma_1) = \begin{cases} S(\gamma_1, \hat{\gamma}_2^{\gamma}) & 若 \quad \gamma_1 < \hat{\gamma}_2 \\ S(\hat{\gamma}_2, \gamma_1) & 若 \quad \hat{\gamma}_2 < \gamma_1 \end{cases} \tag{8-164}$$

得到更新后的 $\hat{\gamma}_1$的估计值为，

$$\hat{\gamma}_1 = \arg\min_{\gamma_1} S_1^{\gamma}(\gamma_1). \tag{8-165}$$

Bai (1997) 指出在结构突变模型中， 更新后的估计值 $\hat{\gamma}_1^{\gamma}$ 是渐进有效的， 所以我们认为在门槛模型中这个结论仍然成立。

门槛个数的确定

在模型 (8-161) 的设定中，可能不存在门槛， 也可能存在一个或两个门槛值，因此我们需要对此进行检验。 在上一节中，我们采用 F_1 统计量来检验单一门槛效应的显著性， 并采用“自体抽样”法获得了该统计量的置信区间。 如果 F_1 拒绝了原假设，即存在一个门槛，那么在模型 (8-161) 的设定中， 我们就需要作进一步的检验以便区分单一门槛和双重门槛。

我们从第二步中估计出的最小残差平方和为 $\hat{\sigma}^2 = S_2^{\gamma}/n(T-1)$， 因此我们可以基于下面的统计量来检验单一门槛与双重门槛那个更为显著:

$$F_2 = \frac{S_1(\hat{\gamma}_1) - S_2^{\gamma}(\hat{\gamma}_2^{\gamma})}{\hat{\sigma}^2} \tag{8-166}$$

该检验量事实上是一个近似的似然比检验。 如果 F_2 的值较大，那么我们就拒绝仅存在一个门槛值的原假设。 由于在原假设下，似然比统计检验量的渐进分布是非枢轴的 (non-pivotal), Hansen (1999) 建议采用“自体抽样法”来近似其样本分布。 抽样方法和检验统计量的构造方法与我们在上一节中介绍的 F_1 统计量的构造方法相似。 区别在于，此时的我们的被择假设对应的模型为 (8-161) 式， 而原假设对应的模型为 (8-140)。因此，我们产生的自抽样样本为，

$$y_{it}^{\#} = x_{it}'\hat{\beta}_1 \cdot I(q_{it} \le \hat{\gamma}) + x_{it}'\hat{\beta}_2 \cdot I(q_{it} > \hat{\gamma}) + \varepsilon_{it}^{\#} \tag{8-167}$$

从 (8-167) 式我们可以看出，F_2 的抽样分布依赖于回归参数 $\beta_1 - \beta_2$ 和 γ。 这不同于 LR_1 的抽样分布，LR_1 并不依赖于任何回归参数。 所以，F_2 的抽样分布是非枢轴的， 使得我们无法保证“自体抽样法”能产生很好的近似效果。

置信区间的构造

最后，我们来构造两个门槛参数 (γ_1,γ_2) 的置信区间。 Bai (1997) 研究表明， 我们在本节第一部分得到的更新后的 $\hat{\gamma}_1$ 的估计值与在单一门槛模型中得到的估计值具有相同的渐进分布。这启发我们可以采用与上一节相同的方法来构造置信区间。 设

$$LR_2^{\gamma}(\gamma)=\frac{S_2^{\gamma}-S_2^{\gamma}(\hat{\gamma}_2^{\gamma})}{\hat{\sigma}^2} \tag{8-168}$$

及

$$LR_2^{\gamma}(\gamma)=\frac{S_1^{\gamma}-S_1^{\gamma}(\hat{\gamma}_1^{\gamma})}{\hat{\sigma}^2} \tag{8-169}$$

其中，$S_2^{\gamma}(\gamma)$ 和 $S_1^{\gamma}(\gamma)$ 分别由 (8-162) 和 (8-164) 式定义。 γ_2 和 γ_1 在 $(1-\alpha)\%$ 显著水平上的置信区间分别为一系列满足 $LR_2^{\gamma}(\gamma)\le c(\alpha)$ 和 $LR_1^{\gamma}(\gamma)\le c(\alpha)$ 的 γ 值。

8.7.4 STATA 实现

Hansen (1999) 在其个人网站上发布了估计面板门槛模型的 Gauss 程序， 附带有程序的说明和其实证分析中所用的数据。 笔者采用 STATA8.0 软件包完成了该程序的估计、检验和绘图程序， 主要包括两个部分，分别为 xtthres.ado 和 xttr_graph.ado。 由于程序代码较长，所以这里仅呈现程序的帮助文件。 需要程序源文件的读者可以向笔者索取 (arlionn@163.comn)。

作为这一模型的应用，国外已有数十位学者采用它进行实证分析， 国内使用该模型的文献十分有限， 就我所知，目前仅有两篇：魏锋和孔煜 (2005)、连玉君和程建 (2006)。

参考文献

[1] Arellano, M. 1987. ”Computing Robust Standard Errors for within-Groups Estimators.” Oxford Bulletin of Economics and Statistics, 49(4), pp. 431-34.

[2] Arellano, M. 2003. Panel Data Econometrics. New York: Oxford University Press.

[3] Arellano, M and S Bond. 1991. ”Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” The Review of Economic Studies, 58(2), pp. 277-97.

[4] Arellano, M and O Bover. 1995. ”Another Look at the Instrumental Variable Estimation of Error-Components Models.” Journal of Econometrics, 68(1), pp. 29-51.

[5] Arellano, M and B Honoré. 2001. ”Panel Data Models: Some Recent Developments.” Handbook of Econometrics, 5, pp. 3229-96.

[6] Baltagi, BH. 2001. Econometric Analysis of Panel Data. Chichester: John Wiley & Sons.

[7] Baum, CF. 2001. ”Residual Diagnostics for Cross-Section Time Series Regression Models.” STATA JOURNAL, 1(1), pp. 101-04.

[8] Baum, CF; ME Schaffer and S Stillman. 2003. ”Instrumental Variables and Gmm: Estimation and Testing.” STATA JOURNAL, 3(1), pp. 1-31.

[9] Bhargava, A; L Franzini and W Narendranathan. 1982. ”Serial Correlation and the Fixed Effects Model.” The Review of Economic Studies, 49(4), pp. 533-49.

[10] Bruno, Giovanni S. F. 2005. ”Approximating the Bias of the Lsdv Estimator for Dynamic Unbalanced Panel Data Models.” Economics Letters, 87(3), pp. 361-66.

[11] Cameron, AC and PK Trivedi. 2009. Microeconometrics Using Stata. Stata Press.

[12] Cameron, AC and PK Trivedi. 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

[13] Canova, F. 2007. Methods for Applied Macroeconomic Research. Princeton University Press.

[14] Cornelissen, T. 2008. ”The Stata Command Felsdvreg to Fit a Linear Model with Two High-Dimensional Fixed Effects.” STATA JOURNAL, 8(2), pp. 170-89.

[15] Drukker, DM. 2003. ”Testing for Serial Correlation in Linear Panel-Data Models.” STATA JOURNAL, 3(2), pp. 168-77.

[16] Efron, B. 1979. ”Bootstrap Methods: Another Look at the Jackknife.” The annals of statistics, 7(1), pp. 1-26.

[17] Efron, B and RJ Tibshirani. 1993. An Introduction to the Bootstrap. New York: Chapmann & Hall.

[18] Flannery, Mark J. and Kasturi P. Rangan. 2006. ”Partial Adjustment toward Target Capital Structures.” Journal of Financial Economics, 79(3), pp. 469-506.

[19] Frésard, Laurent and Carolina Salva. 2010. ”The Value of Excess Cash and Corporate Governance: Evidence from U.S. Cross-Listings.” Journal of Financial Economics, Forthcoming.

[20] Greene, W H. 2000. Econometric Analysis, 4th Edition. New Jersey: Prentice Hall.

[21] Hansen, LP. 1982. ”Large Sample Properties of Generalized Method of Moments Estimators.” Econometrica, 50(4), pp. 1029-54.

[22] Harford, Jarrad; Sandy Klasa and Nathan Walcott. 2009. ”Do Firms Have Leverage Targets? Evidence from Acquisitions.” Journal of Financial Economics, 93(1), pp. 1-14.

[23] Hoechle, D. 2007. ”Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence.” STATA JOURNAL, 7(3), pp. 281-312.

[24] Hsiao, C. 2003. Analysis of Panel Data. Cambridge University Press.

[25] Judson, RA and AL Owen. 1999. ”Estimating Dynamic Panel Data Models: A Guide for Macroeconomists.” Economics Letters, 65(1), pp. 9-15.

[26] Mikkelson, WH and MM Partch. 2003. ”Do Persistent Large Cash Reserves Hinder Performance?” Journal of Financial and Quantitative Analysis, 38(2), pp. 275-94.

[27] Newey, WK and KD West. 1987. ”A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” Econometrica, 55(3), pp. 703-08.

[28] Opler, T; L Pinkowitz; R Stulz and R. Williamson. 1999. ”The Determinants and Implications of Corporate Cash Holdings.” Journal of Financial Economics, 52(1), pp. 3-46.

[29] Petersen, Mitchell A. 2009. ”Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches.” Review of Financial Studies, 22(1), pp. 435-80.

[30] Richardson, S. 2006. ”Over-Investment of Free Cash Flow.” Review of Accounting Studies, 11(2), pp. 159-89.

[31] Roodman, David. 2009. ”How to Do Xtabond2: An Introduction to Difference and System Gmm in Stata.” STATA JOURNAL, 9(1), pp. 86-136.

[32] Sarafidis, V and RE De Hoyos. 2006. ”Testing for Cross-Sectional Dependence in Panel-Data Models.” STATA JOURNAL, 6, pp. 482-96.

[33] Schaffer, M.E. 2010. ”Xtivreg2: Stata Module to Perform Extended Iv/2sls, Gmm and Ac/Hac, Liml and K-Class Regression for Panel Data Models.” http://ideas.repec.org/c/boc/bocode/s456501.html.

[34] Sosa-Escudero, W and AK Bera. 2008. ”Tests for Unbalanced Error-Components Models under Local Misspecification.” STATA JOURNAL, 8(1), pp. 68-78.

[35] Verbeek, M. 2004. A Guide to Modern Econometrics. Wiley.

[36] Wooldridge, JM. 2002. Econometric Analysis of Cross Section and Panel Data. MIT press.

[37] Ziliak, JP. 1997. ”Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators.” Journal of Business & Economic Statistics, 15(4), pp. 419-31.

一份不太长的 Stata 简介

连玉君

中山大学 岭南学院

arlionn@163.com

2010-7-14

目 录

1　Stata 概貌

自从 2003 年开始使用 Stata 以来，我一直把“Stata”读为“Stay-ta”。有一次和一个从日本回来的朋友聊天，她把 Stata 读为“Star-ta”，让我甚感不适。经查阅，方才发现，原来“Stata”并非数个单词的缩写（因此其正确拼写为 Stata 而非 STATA），而是由“statistics”和“data”合成的一个新词，Stata 公司的员工都将其读做“Stay-ta”。从这个小小的趣闻中，可以看出 Stata 在问世之初（1985 年）的主要功能在于统计分析和数据处理。经历了二十余年的发展，Stata 已经升级到第 11.1 版（表 1），在不断强化上述功能的同时，Stata 在矩阵运算、绘图、编程等方面的功能也在不断加强。

表 1　Stata 发展历程

1.0	January 1985	6.0	January 1999
1.1	February 1985	7.0	December 2000
1.2	March 1985	8.0	January 2003
1.4	August 1986	8.1	July 2003
1.5	February 1987	8.2	October 2003
2.0	June 1988	9.0	April 2005
2.05	June 1989	9.1	September 2005
2.1	September 1990	9.2	April 2006
3.0	March 1992	10.0	June 2007
3.1	August 1993	10.1	August 2008
4.0	January 1995	11.0	July 2009
5.0	October 1996	11.1	June 2010

Source: http://www.Stata.com/support/faqs/res/history.html

Stata 擅长数据处理、面板数据分析、时间序列分析、生存分析，以及调查数据分析，但其它方面的功能也并不逊色（表 2）。

表 2　Stata的功能一览

数据处理和绘图			
Data management	Graphics		
统计分析和检验			
Basic statistics	Nonparametric methods	Exact statistics	
ANOVA/MANOVA	其它检验方法和函数		
回归分析			
Linear models	GLM	MLE	GMM
Multilevel mixed models	Panel data	Probit/Logit/Count	Time series
多变量模型（多元统计）		抽样和模拟分析	
Multivariate methods	Cluster analysis	Resampling and simulation	
调查分析和生存分析			
Survey methods	Survival analysis	Epidemiologists	
编程			
Programming language	Mata	User-written commands	

2　为何选择 Stata?

这是个不太容易回答的问题。Stata网站列举了数条可能的原因。Edwards(2005)曾经非常细致地对比了Stata, SPSS和SAS的优劣。Princeton大学的Torres-Reyna博士则将四种常用软件的特征总结为表 3。整体而言，Stata具有较强的优势。

表 3　四款统计软件的对比分析

Features	**Stata**	SPSS	SAS	R
Learning curve	Steep/gradual	Gradual/flat	Pretty steep	Pretty steep
User interface	Programming/ point-and-click	Mostly point-and-click	Programming	Programming
Data manipulation	Very strong	Moderate	Very strong	Very strong
Data analysis	Powerful	Powerful	Powerful/versatile	Powerful/versatile
Graphics	Very good	Very good	Good	Good

Source：http://dss.princeton.edu/training/StataTutorial.pdf，p.3.

我为何钟情于 Stata?

就我个人的经历而言，如下几个原因使我自 2003 年以来一直钟情于 Stata。

Stata的数据处理功能很强大。由于将数据导入内存后进行运算，其速度非常快。在多个数据文件的合并和追加，以及文字资料、时序资料，以及调查资料的处理方面，Stata总能以极为简洁的命令完成分析。虽然Stata管方命令仅能支持txt和xml格式数据文件的导入和导出，但借助Stat/Transfer软件，我们可以非常方便地实现不同软件数据格式的转换，如Excel, Access, SPSS, SAS, Eviews, Gauss, Limdep, S-Plus, R等。我是做公司财务的，每年 5 月，在GTA、CCER、Wind等数据库提供商提供了最新的数据后，我也需要更新自己的Stata数据库(我把这些数据库提供的几十个子库合并为一个名为"Arlion_data.dta"的Stata数据文件，并与我的合作者们分享)。借助Stata的数据处理功能，我只需在上一年度已经完成的do-files中稍作修改即可完成数据的更新工作。整个过程仅需 2 天的时间。我无法想象，如果没有Stata提供的merge、append、forvalues等命令，这个数据更新的过程将会有多么痛苦。

Stata 的 do-files 带来的便利。我很少点击 Stata 的菜单，也很少在命令窗口中输入命令，我使用 do-files（当然，每天要在这个窗口中敲入几十次 `help` 命令）。简单而言，Stata 的 do-files 只是一个包含了多行 Stata 命令的文本文件而已（[U]16 Do-files, Long(2009)）。有些时候，要完成一篇文章的数据处理过程需要数周的时间，do-files 就显得格外重要，它使得我们很容易对此前的处理过程进行修改。更为重要的是，后续文章都可以在这个 do-files 的基础上扩展。我与搭档合作时，每天只需通过电子邮件发送只有几 k 大小的 do-file 即可；而我的学生们则可以通过 do-files 重现我上课时讲解的每一个估计命令；很多学生的第一篇实证分析的论文都是在我已经完成的 do-files 基础上完成的。

Stata 绘制的图形非常精美。这也为回归分析提供了一种可视化的分析工具，自

Stata10 发布以来，Stata 增加了图形编辑、多种字体支持，以及数学符号支持等功能。Stata 可以输出十余种图片格式，可以非常方便地插入 Word、LaTeX 等文字排版软件。即使采用点击鼠标的方式绘制图形，Stata 也会自动生成命令代码，为图形的修改提供了极大的便利。

Stata在编程方面提供了良好的平台。比如，做非线性最小二乘(NLS)、最大似然估计(MLE)、广义矩估计(GMM)，只需要设定函数形式，编写一些简单的程序即可完成，至于数值算法等比较复杂的技术问题，Stata都已帮你做好了。例如，我完成的第一篇实证分析的论文便是以NLS为基础的（连玉君 and 钟经樊(2007)），随后，我又采用MLE完成了异质性随机边界模型（连玉君 and 苏治(2009)）和双边随机边界模型（Lian and Chung(2008)；连玉君(2009)）的估计。自Stata11 发布以来，GMM的实现也变得非常简单了，你只需设定残差方程、指定工具变量，并选择何时得权重矩阵即可完成估计。

Stata具有良好的扩展性。Stata具有自己的编程语言，其所有命令都对应着一个以“.ado”为后缀的同名程序文件。对于Stata用户而言，我们可以使用viewsource或doedit命令查看这些程序的代码。更为重要的是，我们可以非常方便地自行编写命令，以实现对Stata官方命令的补充和扩展。这种特殊的扩展功能赋予了Stata用户极大的灵活性，我们可以用findit命令下载到大量的外部命令，以便适时跟进最新的统计方法。这同时也推动了Stata自身的发展，例如，Stata用户开发出的可绘制地图的命令tamp，spmap，china_map等就是一个很好的例证；由David Roodman编写的xtabond2命令则被Stata11设定为估计动态面板模型的官方命令（xtdpd，xtdpdsys）；同样，由F. Bornhorst and C.F. Baum编写的ipshin、levinlin命令，C.F. Baum编写的hadrilm命令，以及S. Merryman编写的xtfisher等用于执行面板单位根检验的命令都被Stata11 设定为官方命令xtunitroot。饮水思源，我自己也贡献了xtbalance等命令。若想发布自己编写的Stata命令，只需发邮件给波士顿大学的C.F. Baum教授即可。

最后，从我身边这些老师和朋友的经验来看，Stata受到了越来越多的关爱。我的导师使用Gauss十年有余，在 2001 年接触Stata后，毅然改用Stata。还有很多国外的朋友，基本上都在使用Stata。当越来越多的人开始使用Stata时，我们的交流成本会迅速下降。

当然，软件本身并无好还之分，只是一个习惯的问题。关键的问题还是对统计和计量理论的掌握，这是决定你是否能驾驭软件的关键。

正在消弭的 Stata 缺陷

Stata并不完美，但她正在趋近完美。Evan Stark博士非常精辟的概括了这一特征：“You get the sense that at Stata they thought of everything, and when they or a user points out that they didn’t, they quickly provide a fix or new functionality. Although it did take me a while to understand its syntax [switching from SPSS], I did master it and statistical life became thereafter very *enjoyable*.”

诚如MacStats网站的评价，Stata结果似乎无法像SPSS或Eviews那样非常美观地输出（或粘贴）到Word/Excel文档中。然而，得益于广大Stata用户的努力，这不再是个问题，我们可以使用`tabout` (Watson(2007)), `esttab` (Jann and Long(2010)), `logout`, `outreg2` (Jann(2005), Jann(2007)), `xml_tab` (Lokshin and Sajaia(2008)) 等命令非常方便的把Stata结果输出到Excel, Word, LaTeX和HTML (Gini and Pasquini(2006)) 等文件中。连玉君博士制作的视频文件*Stata与Word、Excel、LaTeX的亲密接触* 非常细致地介绍了这一主题。他的另一份文档*Stata与LaTeX的完美结合* 则较为全面的介绍了如何将Stata结果输出到LaTeX。

在早期版本中，Stata的do-files编辑器过于简单。Stata11 发布后，其do-files编辑器已然从灰姑娘变成了白雪公主，具有了语法高亮显示、结构代码折叠、书签设定等功能，而且，对于书写大型do-files的用户而言，命令的行数也不再受到任何限制。对于中文用户而言，只需稍作调整，即可获得很好的显示效果。

Stata9 以前的版本无法对图形进行二次编辑，且图形中的可供选择的字体也非常有限。自从Stata10 和Stata11 发布以来，这两个问题得到了很好的解决。图形中的文字可以是粗体、斜体，亦可包含多种数学符号；在用户手动编辑图形时，相应的命令会自动显示在屏幕上，进而用于处理其他类似的图形。

不同于SAS等从硬盘上读取数据的统计软件，Stata是将数据调入内存后执行运算的，这使得其运算速度非常快。然而，对于经常处理高频数据和大型调查数据的用户而言，Stata的这种运算机制反而成了其缺陷——它能够处理的数据量受限于计算机的内存容量。虽然在既有的多个Stata版本中，Stata11 家族中进一步增加了Stata/MP，使其在配有多核处理器的计算机中运算速度进一步得到提升，但数据容量的限制问题仍然未能得到实质性的改进。

3 如何学习 Stata？

我经常会被问到“Stata 好学吗”、“我多长时间能学会 Stata”，诸如此类的问题。诚然，相比于 SPSS 和 Eviews 等软件，Stata 的门槛的确要高一些。然而，问题的关键并不在于 Stata 本身有多么难学，而在于你在统计和计量方面花费了多少时间，这与学习 Stata 所需的时间显著负相关。因此，我的回答往往会是：“哦，这个不好说，如果……，其实很简单……”。

相比于十年前，现在学习 Stata 的资料已经非常丰富了。虽说殊途同归，但不同的学习路径却存在着巨大的效率差异。对于初学者而言，我的建议是，首要的问题是知道“Stata 能做什么”，继而才是“Stata 如何做什么”。

第一个问题之所以重要，是因为从本质上讲，Stata只是我们完成统计分析的工具而已，因此，其基本平台是否宽广、是否有扩展潜力，以及它提供的分析工具是否能满足你的专业需求，都是你在选择Stata之前需要深入了解的。Stata User’s Guide（400 页，中

文）对这些问题做出了很好的解答，是一幅绝佳的导航图，能帮助你在短时间内了解Stata的基本架构、语法特征和核心功能。对于第二个问题，则有众多的资料可供参考：

（1）网络资源

在附录A中，我精选了一些链接。值得一提的有如下几个：

- **Stata官方网站。**Stata公司提供的Web resources，涵盖了大量相关网络资源；其FAQ则提供了各种常见问题的解答；Statalist则是一个类似于人大经济论坛的免费的讨论区。加入Statalist的方法很简单，你只需要发送邮件至*majordomo@hsphsun2.harvard.edu*，邮件内容无需任何称谓，只需写上"**subscribe Statalist**"的字样即可。接到确认信息后，你便成为一名Statalist的成员了。当然，即使不加入，你仍然可以浏览，但不能提问。
- **UCLA(加州大学洛杉矶分校)提供的网络教程。**该网站提供的Data Management、Graphics、Regression、Logistic Regression、Multilevel Modeling、Survey Data Analysis等模块都非常出色；其Web Books、Textbook Examples模块则非常细致地呈现了几十本非常流行的统计和计量教材的Stata实例；对于LaTeX感兴趣的朋友，则可以通过Stata Tools for LaTeX模块获得诸多有用的信息；在Graph examples模块中，则列举了四十余种图形的绘制方法；最后，在Classes and Seminars模块中，你可以在线观看数十个Stata教学视频。
- **人大经济论坛。**若从人数上来讲，人大经济论坛或许是全球最大的经济类论坛了。目前，其计量经济学板块又细分出多个计量软件专题讨论区。在Stata专版已发布了 4000 余个讨论主题（18000 余条回复），而Stata上传下载区则汇集了大量学习资料。在统计软件培训班VIP答疑区中，Stata培训班的学员所提出的问题，可以在 24 小时内得到详尽的回复。

（2）相关的书籍

自从Hamilton（1990）出版*Statistics with Stata*后，一系列将计量理论与软件操作结合起来的书籍开始相继面世，而在此之前，人们似乎都认为软件操作是件非常简单的事情。也正因为如此，很多学生在修习完了一个学年的计量经济学课程后，仍然不知道该如何完成OLS估计。为此，我列举的书籍多附有Stata实例（* 表示我的推荐程度），多数书中的范例数据都可通过Stata官方网站下载。

- **一份详细的书单。**UCLA提供了的书单：Statistics Books for Loan。
- **入门教材：**Baum（2006）*、Newton and Cox（2009）、Chen et al.（2005）、Adkins and Hill（2008）*；Wooldridge（2009）*，波士顿大学的网站上提供了该书所有章节的Stata范例，是一套非常好的学习资料。
- **综合性教材：**Cameron and Trivedi（2005）撰写的 *Microeconometrics: Methods and applications* 一书全面介绍了微观计量中的基本分析工具，其中不乏最近十年中得到广泛应用的 Bootstrap、Monte Carlo 模拟，以及非参数估计法。二人于 2009

年出版的另一力作（Cameron and Trivedi（2009）*）是这本书的姊妹篇，重点介绍了常用计量模型的 Stata 实现方法。

- **Stata手册**。我一直非常佩服撰写Stata手册的那些人（我的导师也有相似的感觉），他们总能以最简洁的语言说清楚纠结我很久的问题。Stata11 附有 16 本电子手册，仅需统一放置于D:\stata11\utilities目录下，即可从Stata内部的帮助文件中的Also see部分直接链接到相应的PDF说明书中。作为初学者，我强烈建议你将 **[U]** 和 **[D]** 打印出来，反复研读。
- **统计方法**：Rabe-Hesketh and Everitt（2006）。
- **Stata 绘图**：Mitchell（2008），非常细致地介绍了各种图形的绘制方法。
- **Stata 数据处理**：Kohler and Kreuter（2005）*、Long（2009）*、杨菊华（2008）。
- **Stata 编程**：Baum（2009），当然，该书中有关数据处理的介绍也非常精彩。
- **Logit/Probit模型**：Hosmer and Lemeshow(2000)*对相关的理论进行非常细致的介绍，是我学习Logit模型的入门教材；Long and Freese(2001)*、Long and Freese(2006)、Hilbe(2009)则涉及了大量的Stata实例，对解读Logit/Probit模型的结果很有帮助；Rabe-Hesketh et al.(2004)提供了在GLLAMM架构下估计xtlogit, xtprobit, xtmelogit, 以及xtmepoisson模型的方法。
- **Panel Data 和多层次模型**：Stata11 手册[XT]*，简洁明了，附有大量实例；Cameron and Trivedi（2009）*、王志刚（2008）、Rabe-Hesketh and Skrondal（2008）。
- **Mata**：Schmidheiny（2008）*，简洁明了介绍了 Mata 的基本用法；详情则可参考 Stata11 手册 **[M]**。
- **GLLAMM**：Rabe-Hesketh et al.（2004）（下载）。
- **Meta**：Sterne（2009）。
- **GLM**：Hardin et al.（2007）。
- **MLE**：Harrison（2008）（Lectures）、Gould et al.（2006）。
- **生存分析**：Cleves et al.（2008）。

（3）**Stata 视频**

相比于网络教程和纸本教材，通过视频学习Stata可能是最快捷的方式了。坊间流传有两套Stata视频教程：**一套是**UCLA免费发布的视频教程，内容涉及Stata入门、数据处理和绘图等。该视频教程采用英文讲解，思路清晰。局限在于所涉及内容不够系统，但对于想快速入门的学生则是一份不错的参考资料。同时，藉由这份资料也可以练习一下英语听力。**另一套是**由中山大学岭南学院的连玉君博士制作的Stata视频教程。该教程分为初级（36 学时）、高级（48 学时）和Panel data专题（12 学时）三个部分。该视频教程涵盖了Stata简介、数据处理、矩阵、绘图、编程等基本操作，同时还包含了OLS、GLS、MLE、GMM、Bootstrap、Monte Carlo模拟、时间序列分析、面板数据模型等分析工具。详见附录C。

4 最后的话

(1) **好脑瓜不如烂笔头。**这是一个适用于学习任何新知识的“秘诀”，对于功能强大，以敲命令为基础的Stata软件而言尤其如此。因此，你要时刻记录新学到的命令、方法和技巧，并定期整理。若能将这些手记与其他Stata用户分享，你会有更多的收获。我的博客中便提供了不少这样的笔记。

(2) **学以致用。**在了解了 Stata 的基本功能和架构后，想要进一步提升自己的最佳途径就是动手写一篇实证分析的论文，并自始至终用 Stata 解决所有问题。这项工作的起点是一份以 txt 或 Excel 格式存储的原始数据文件，中间过程完整地记录于一个 do-files 文档中，最终的分析结果要自动输出到 Word, Excel 或 LaTeX 文档中。

(3) **不耻下问。**这个不用多言了，你只需克服“不耻”，进而多花些精力考虑考虑该如何提问即可（注：很多人不会提问）。

参考文献

Adkins, L., R. Hill. Using stata for principles of econometrics[M]. Wiley, 2008.

Baum, C. An Introduction to Modern Econometrics using Stata[M]. Stata Corp, 2006.

Baum, C. An Introduction to Stata Programming[M]. Stata Press, 2009.

Cameron, A., P. Trivedi. Microeconometrics: methods and applications[M]. Cambridge University Press, 2005.

Cameron, A., P. Trivedi. Microeconometrics Using Stata[M]. Stata Press, 2009.

Chen, X., P. Ender, M. Mitchell, C. Wells, 2005, Stata Web Books: Regression with Stata (http://www.ats.ucla.edu/stat/stata/webbooks/reg/default.htm.

Cleves, M., W. Gould, R. Gutierrez, Y. Marchenko. An introduction to survival analysis using Stata[M]. Stata Press, 2008.

Edwards, M., **2005**, SPSS, STATA, and SAS: Flavours of Statistical Software, *URI: http://hdl.handle.net/1873/250*.

Gini, R., J. Pasquini, **2006**, Automatic generation of documents, *STATA JOURNAL*, 6 (1): 22-39.

Gould, W., J. Pitblado, W. Sribney. Maximum likelihood estimation with Stata[M]. Stata Press, 2006.

Hamilton, L. Statistics with Stata[M]. Brooks/Cole, 1990.

Hardin, J., J. Hilbe, J. Hilbe. Generalized linear models and extensions[M]. Stata Press, 2007.

Harrison, G. Maximum Likelihood Estimation of Utility Functions Using Stata[M]. University of Central Florida, http://web.bus.ucf.edu/documents/economics/workingpapers/2006-12.pdf, 2008.

Hilbe, J. Logistic regression models[M]. Chapman & Hall/CRC Press, 2009.

Hosmer, D., S. Lemeshow. Applied Logistic Regression[M]. New York: John Wiley & Sons, Inc, 2000.

Jann, B., **2005**, Making regression tables from stored estimates, *STATA JOURNAL*, 5 (3): 288-308.

Jann, B., **2007**, Making regression tables simplified, *STATA JOURNAL*, 7 (2): 227-244.

Jann, B., J. Long, **2010**, Tabulating SPost results using estout and esttab, *STATA JOURNAL*, 10 (1): 46-60.

Kohler, U., F. Kreuter. Data Analysis Using Stata[M]. Stata Press, 2005.

Lian, Y., C.-F. Chung, **2008**, Are Chinese Listed Firms Over-Investing?, *SSRN working paper, Available at SSRN:* *http://ssrn.com/abstract=1296462*.

Lokshin, M., Z. Sajaia, **2008**, Creating print-ready tables in Stata, *STATA JOURNAL*, 8 (3): 374-389.

Long, J. The workflow of data analysis using Stata[M]. Stata Press, 2009.

Long, J., J. Freese. Regression models for categorical dependent variables using Stata[M]. Stata press, 2001.

Long, J., J. Freese. Regression Models for Categorical Dependent Variables using Stata[M]. Stata press, 2006.

Mitchell, M. A visual guide to Stata graphics[M]. Stata Corp, 2008.

Newton, H., N. Cox. Seventy-six Stata Tips[M]. Stata Press, 2009.

Rabe-Hesketh, S., B. Everitt. A Handbook of Statistical Analyses Using Stata[M]. Chapman & Hall/CRC, 2006.

Rabe-Hesketh, S., A. Skrondal. Multilevel and Longitudinal Modelling Using Stata (Second Edition)[M]. Stata Press, 2008.

Rabe-Hesketh, S., A. Skrondal, A. Pickles, **2004**, GLLAMM manual, *UC Berkeley Division of Biostatistics working paper series 160*, *http://www.bepress.com/ucbbiostat/paper160/*.

Schmidheiny, K., **2008**, Coding with Mata in Stata, *Lectures in Universitat Pompeu Fabra*, *http://kurt.schmidheiny.name/teaching/statamata.pdf*.

Sterne, J. Meta-analysis in stata: An updated collection from the stata[M]. Stata Press, 2009.

Watson, I., **2007**, Publications quality tables in Stata: a tutorial for the tabout program, *Working Paper*, *http://fmwww.bc.edu/repec/bocode/t/tabout_tutorial.pdf*.

Wooldridge, J. Introductory econometrics: A modern approach[M]. South Western Cengage Learning, 2009.

王志刚. 面板数据模型及其在经济分析中的应用[M]. 北京：经济科学出版社, 2008.

杨菊华. 社会统计分析与数据处理技术——STATA 软件的应用[M]. 北京：中国人民大学出版社, 2008.

连玉君. 中国上市公司投资效率研究[M]. 北京：经济管理出版社, 2009.

连玉君，苏治, **2009**，融资约束、不确定性与上市公司投资效率，*管理评论*，(01): 19-26.

连玉君，钟经樊, **2007**，中国上市公司资本结构动态调整机制研究，*南方经济*，(01): 23-38.

附录 A：一些有用的 Stata 链接

I. Websites of Stata CP

Stata website: http://www.Stata.com 导航图
Sata resources: http://www.Stata.com/links/resources1.html (大量网络教程链接)
Stata journal: http://www.Stata.com/support/faqs/res/sj.html 中
Stata library: http://www.ats.ucla.edu/stat/Stata/library/
Statalist archive: http://www.hsph.harvard.edu/cgi-bin/lwgate/STATALIST/archives/
Stata FAQs: http://www.Stata.com/support/faqs/
Stata statistics FAQs: http://www.Stata.com/support/faqs/stat/
Stata listserver: http://www.Stata.com/support/Statalist/
Stata discussion list: Statalist@hsphsun2.harvard.edu
Stata bookstore: http://www.Stata.com/bookstore/ Example Datasets 中
Stata Manual: http://www.Stata-press.com/manuals/ Example Datasets 中

II. Websites in China

- 人大经济论坛 （国内最大的经济论坛）
  - 人大经济论坛Stata专版：http://www.pinggu.org/bbs/forum-67-1.html
  - 人大经济论坛Stata上传下载区：
    http://www.pinggu.org/bbs/forum-121-1.html 汇
  - 人大经济论坛统计软件培训班VIP答疑区（针对Stata视频教程学员）：
    http://www.pinggu.org/bbs/forum-114-1.html (所有Stata问题 24 小时内回复)

III. UCLA Academic Technology Services (极力推荐)

- Classes and Seminars
  - Introduction to Stata 10 *with movies*
  - Introduction to Stata (for Stata version 8 and 9) *with movies*
  - Regression with Stata
  - Logistic Regression with Stata *with movies*
  - Beyond Binary Logistic Regression with Stata *with movies*
  - Factor Variables and Interactions in Stata 11
  - Main Effects and Interactions for Logit Models in Stata *with movies*
  - Multiple Imputation in Stata, Part 1
  - Multiple Imputation in Stata, Part 2
  - Survey Data Analysis with Stata 8 *with movies*
  - Introduction to Survey Data Analysis with Stata 9
  - Applied Survey Data Analysis with Stata 9 with movie and mp3
  - Survival Analysis Using Stata
  - Graphics using Stata 8 *with movies*
  - Introduction to Programming in Stata
  - What's New in Stata 8
  - What's New in Stata 9
  - What's New in Stata 10

- Links by Topic
  - Data Management
  - Graphics
  - ANOVA
  - Regression
  - Logistic (and Categorical) Regression
  - Count Models
  - Multilevel Modeling
  - Survival Analysis
  - Survey Data Analysis
- Frequently Asked Questions (FAQ)
- Statistical Analysis
  - Data Analysis Examples (绝佳的数据处理专题)
  - Annotated Output (详细解读Stata输出结果)
  - Textbook Examples (包含十余本教科书的Stata实例)
  - Web Books (两本Stata网络教程)
  - What statistical analysis should I use? (常用统计分析的Stata实例)

IV. Stata Portal (a comprehensive links)

- 不错的入门资料：
  - Getting Started with Stata.
  - Introduction to STATA with Econometrics in Mind
- Stata and Related Resources
  - Stata Web Site (including)
    - Links to other Stat Resourcess, NetCourses, Stata Press, Product Information, sample Stata Session, Capabilities, The Stata Journal
  - Frequently Asked Questions from Stata Corporation
  - StataList, the Stata listserv
    - Information from Stata Corp
    - Weblog: http://Statalist.blogspot.com
    - RSS Feed: http://feeds.feedburner.com/Statalist
  - Stata Tips (几十个Stata应用的小贴士)
- Course Notes
  - ED 230A: Introduction to Research Design and Statistics from Phil Ender
  - ED 230BC: Linear Statistical Models from Phil Ender
  - ED 231A: Multivariate Analysis from Phil Ender
  - ED 231C: Applied Categorical & Nonnormal Data Analysis from Phil Ender
  - e-Tutorial on Stata for Econ 508 from University of Illinois
- Textbook Examples
  - Stata Textbook Examples: Introductory Econometrics by Jeffrey Wooldridge from Boston College
- Beginning Stata Tutorials
  - Introduction to Stata, Jeroen Weesie, Utrecht University, Netherlands

- Getting Started with Stata for MS Windows: A Brief Introduction, Robert Yaffee, New York University, USA
- Online Help for Stata from DSS at Princeton University
- Introduction to Stata from the Department of Economics at Princeton University
- Publications on Statistical Software, from University of Wisconsin-Madison Social Science Computing Cooperative
- Analysis of Survey Data for Social Science Research (introduces Stata), a collaborative project of the University of Cape Town and the University of Michigan.
- Stata Programming: Data Management from The Carolina Population Center at UNC Chapel Hill

- Data Files
    - Data in Stata format from the Center for International Development, Harvard University, USA
- Statistical
    - Survival Analysis with Stata: Course EC968, Institute for Social and Economic Research, University of Essex, UK
    - Generalized Linear Models from Princeton University.
    - S-Post: Post Estimation Commands in Stata from J. Scott Long and J. Freese
    - gllamm for complex problems by Stas Kolenikov
    - Statistics and Social Science support from NYU Information Technology Services
- Stata Programs
    - SSC Stata Program Archive Boston College Department of Economics
    - Statistical Software from Gary King
    - Stata Code by Christopher Ferrall, Queen's University, Canada.
    - Stata program by Tony Brady, Sealed Envelope Ltd
    - Generalized Linear Latent and Mixed Models (GLLAMM) by Sophia Rabe-Hesketh, Kings College London
    - Stata Software by Nicola Orsini, Institute of Environmental Medicine, Karolinska Institutet
- Other
    - Some notes on text editors for Stata users, Statalist members
    - StyleRules - Suggestions on Programming Style by Nicholas J. Cox
    - A SAS User's Guide to Stata courtesy of The Carolina Population Center at UNC Chapel Hill
    - A review of random effects modelling in Stata from the Centre for Multilevel Modeling

附录 B：43 个不可不知的 Stata 命令

虽然Stata已经历了二十余年的发展，命令不断增加，但牢记如下 43 个基本命令却是作为一个Stata用户的立身之本（Source：Stata 11 Manual，[U]27, p.375）：

Getting online help	
`help`, `hsearch`,	
`net search`, `search`	[U] 4 Stata's help and search facilities
Keeping Stata up to date	
`ado`, `net`, `update`	[U] 28 Using the Internet to keep up to date
`adoupdate`	[R] adoupdate
Operating system interface	
`pwd`, `cd`	[D] cd
Using and saving data from disk	
`save`	[D] save
`use`	[D] use
`append`, `merge`	[U] 22 Combining datasets
`compress`	[D] compress
Inputting data into Stata	[U] 21 Inputting data
`input`	[D] input
`edit`	[D] edit
`infile`	[D] infile (free format); [D] infile (fixed format)
`infix`	[D] infix (fixed format)
`insheet`	[D] insheet
Basic data reporting	
`describe`	[D] describe
`codebook`	[D] codebook
`list`	[D] list
`browse`	[D] edit
`count`	[D] count
`inspect`	[D] inspect
`table`	[R] table
`tabulate`	[R] tabulate oneway and [R] tabulate twoway
Data manipulation	[U] 13 Functions and expressions
`generate`, `replace`	[D] generate
`egen`	[D] egen
`rename`	[D] rename
`drop`, `keep`	[D] drop
`sort`	[D] sort
`encode`, `decode`	[D] encode
`order`	[D] order
`by`	[U] 11.5 by varlist: construct
`reshape`	[D] reshape
Keeping track of your work	
`log`	[U] 15 Saving and printing output—log files
`notes`	[D] notes
Convenience	
`display`	[R] display

附录 C：Stata 视频教程

自 2007 年以来，人大经济论坛陆续推出了SAS、Eviews、Stata等软件的视频教程。相比于传统的教科书和课堂授课方式，视频教学大大降低了学习统计软件的门槛，因而受到了广大学员的一致好评。

Stata视频教程由中山大学岭南学院的连玉君博士制作，分为初级、高级和Panel data专题三个部分，是一套学习计量经济学和Stata应用的绝佳教程。

（a）Stata 初级视频教程。主要介绍 Stata 的操作方法，包括 Stata 入门、数据处理、绘图、矩阵和编程初步五个部分，共计 36 个学时，全面介绍了 Stata 的基本操作方法。

相关链接：说明书 http://baoming.pinggu.org/Default.aspx?id=16（大纲和试听视频）

（b）Stata 高级视频教程。主要介绍各种计量模型的基本思想及其在 Stata 中的实现方法，包括 OLS、GLS、MLE、IV-GMM、时间序列分析、面板模型、Stata 高级编程、Bootstrap 和 Monte Carlo 模拟等内容，共计 48 个学时，全面的覆盖了计量经济学的核心内容。该视频的特点是以实 Stata 证分析为导向，视频中介绍了大量的应用实例。

相关链接：说明书 http://baoming.pinggu.org/Default.aspx?id=25（大纲和试听视频）

（c）Panel Data 专题。重点介绍近十年应用非常广泛的各种面板模型，包括固定效应模型、随机效应模型、IV-GMM 估计、异方差和序列相关、动态面板模型、随机边界面板模型等，并采用模拟的方式非常细致地呈现了各种模型的小样本性质，对于深入理解面板模型的理论基础颇有裨益。该视频包含 12 个视频文件，每个文件 40-60 分钟。

相关链接：目录 http://baoming.pinggu.org/Default.aspx?id=26（大纲和试听视频）