第 15 章 短 面 板

15.1 面板数据的特点

面板数据(panel data 或 longitudinal data)，指的是在一段时间内跟踪同一组个体(individual)的数据。

它既有横截面的维度(n 个个体)，又有时间维度(T 个时期)。

一个$T=3$的面板数据结构如表 15.1。

表 15.1　面板数据的结构

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

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

反之，如果 T 较大，而 n 较小，则被称为“长面板”(long panel)。

在面板模型中，如果解释变量包含被解释变量的滞后值，则称为“动态面板”(dynamic panel)；反之，则称为“静态面板”(static panel)。

如果在面板数据中，每个时期在样本中的个体完全一样，则称为“平衡面板数据”(balanced panel)；反之，则称为“非平衡面板数据”(unbalanced panel)。

面板数据的优点：

(1) 解决遗漏变量问题：

遗漏变量常由不可观测的个体差异或“异质性”(heterogeneity)造成。

如果个体差异“不随时间而改变”(time invariant)，则面板数据可解决遗漏变量问题。

(2) 提供个体动态行为的信息：

例：考虑区分规模效应与技术进步对企业生产效率的影响。对于截面数据，没有时间维度，无法观测到技术进步。对于时间序列，无法区分生产效率的提高究竟有多少由于规模扩大，有多少

由于技术进步。

例：对于失业问题，截面数据能告诉在某个时点上哪些人失业，时间序列数据能告诉某个人就业与失业的历史，但这两种数据均无法告诉是否失业的总是同一批人(低流转率)，还是失业的人群总在变动(高流转率)。

(3) 样本容量较大：同时有截面维度与时间维度，面板数据的样本容量更大，可提高估计精度。

面板数据也会带来问题，比如，数据通常不满足独立同分布的假定，因为同一个体在不同期的扰动项一般存在自相关。

面板数据的收集成本通常较高，不易获得。

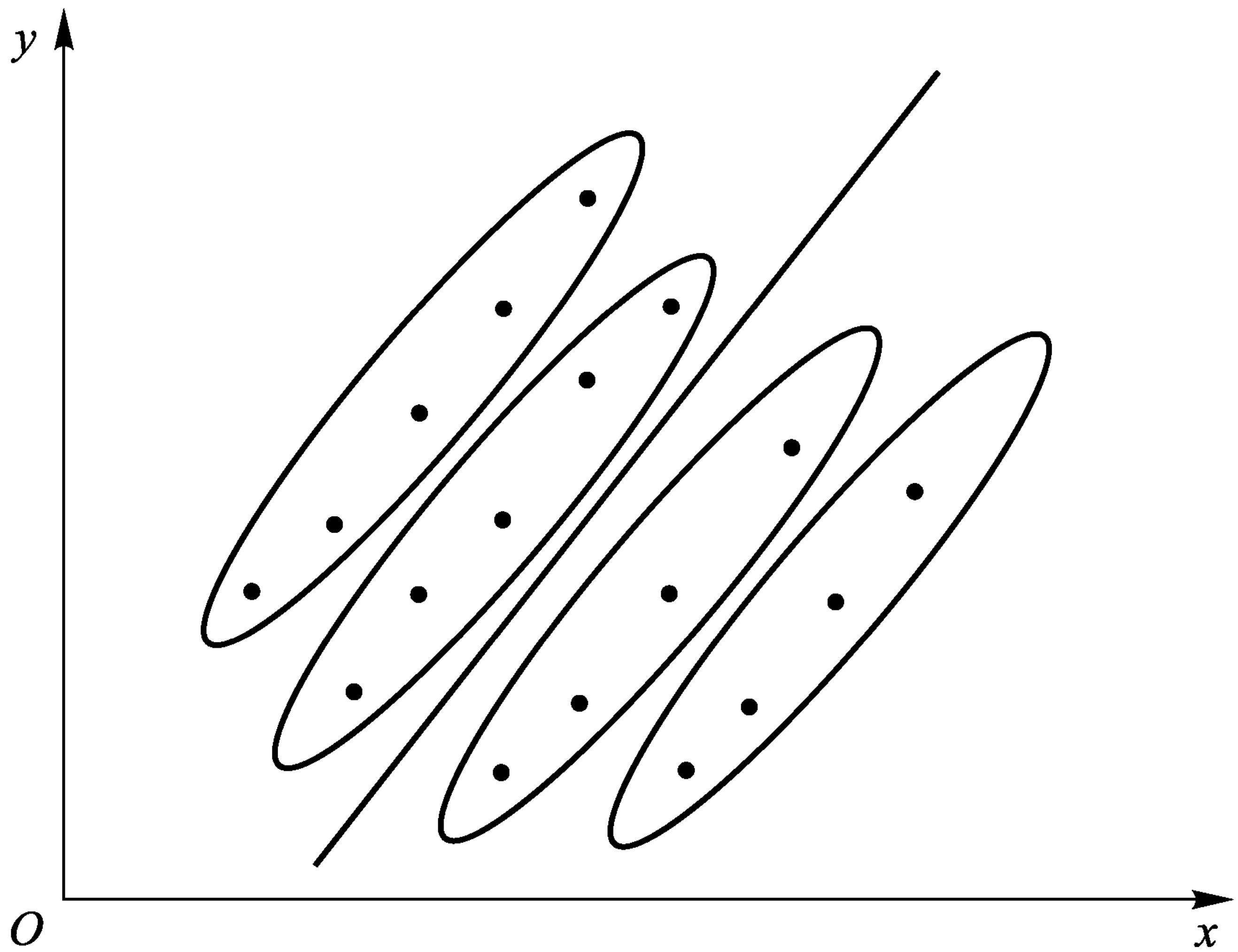
15.2 面板数据的估计策略

一个极端策略是将其看成是截面数据而进行混合回归(pooled regression)，要求样本中每位个体拥有相同的回归方程。

此策略忽略个体间不可观测或被遗漏的异质性(heterogeneity)，而该异质性可能与解释变量相关，导致估计不一致。

另一极端策略则是，为每位个体估计一个单独的回归方程。此策略忽略了个体的共性，可能没有足够大的样本容量。

实践中常采用折衷的估计策略，即假定个体的回归方程拥有相同的斜率，但可有不同的截距项，以此来捕捉异质性。


y
O
x

个体效应模型 (individual-specific effects model)

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

$\boldsymbol{z}_i$为不随时间而变(time invariant)的个体特征，比如性别；

$\boldsymbol{x}_{it}$可随个体及时间而变(time-varying)；

扰动项由$(u_i + \varepsilon_{it})$两部分构成，称为“复合扰动项”(composite error term)；不可观测的随机变量u_i是代表个体异质性的截距项。

ε_{it}为随个体与时间而改变的扰动项。假设$\{\varepsilon_{it}\}$为 iid，且与u_i不相关。

如果u_i与某个解释变量相关，则称为“固定效应模型”(Fixed Effects Model，简记 FE)。此时，OLS 不一致。

如果u_i与所有解释变量$(\boldsymbol{x}_{it}, \boldsymbol{z}_i)$均不相关，则称为“随机效应模型”(Random Effects Model，简记 RE)。

15.3 混 合 回 归

如果所有个体拥有一样的回归方程，则方程可写为

$$y_{it} = \alpha + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_{it}$$

$\boldsymbol{x}_{it}$不包括常数项。把所有数据放在一起，像对待横截面数据那样进行 OLS 回归，称为“混合回归”(pooled regression)。

应使用聚类稳健的标准误(cluster-robust standard errors)，聚类(cluster)由每位个体不同期的所有观测值所组成。

15.4 个体固定效应模型

对于固定效应模型，给定个体 i，将方程两边对时间平均：

$$\bar{y}_i = \bar{\boldsymbol{x}}_i' \boldsymbol{\beta} + \boldsymbol{z}_i' \boldsymbol{\delta} + u_i + \bar{\varepsilon}_i$$

将原方程减去平均后的方程可得：

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

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

$$\tilde{y}_{it} = \tilde{\boldsymbol{x}}_{it}'\boldsymbol{\beta} + \tilde{\varepsilon}_{it}$$

上式已将u_i消去，只要$\tilde{\varepsilon}_{it}$与$\tilde{\boldsymbol{x}}_{it}$不相关，可用 OLS 一致地估计$\boldsymbol{\beta}$，称为“固定效应估计量”(Fixed Effects Estimator)，记为$\hat{\boldsymbol{\beta}}_{\text{FE}}$。

$\hat{\boldsymbol{\beta}}_{\text{FE}}$主要使用了每个位体的组内离差信息，也称“组内估计量”(within estimator)。

即使个体特征u_i与解释变量$\boldsymbol{x}_{it}$相关，组内估计量也一致。

在作离差转换时，$z_i'\boldsymbol{\delta}$也被消掉，无法估计$\boldsymbol{\delta}$，故 FE 无法估计不随时间而变的变量之影响。

为保证$(\varepsilon_{it} - \bar{\varepsilon}_i)$与$(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)$不相关，要求第 i 个观测值满足严格外

生性，即$\mathrm{E}(\varepsilon_{it}\,|\,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT})=0$，因为$\bar{\boldsymbol{x}}_i$中包含了所有$(\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT})$的信息。扰动项须与各期解释变量均不相关(不仅仅是当期解释变量)。

在原方程中引入$(n-1)$个虚拟变量(如果没有截距项，则引入 n 个虚拟变量)来代表不同的个体，可得到同样结果。

FE 也称为“最小二乘虚拟变量模型”(Least Square Dummy Variable Model，简记 LSDV)。

正如线性回归与离差形式的回归在某种意义上是等价的。比如，

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

使用 LSDV 的好处是可以得到个体异质性u_i的估计。

LSDV 法的缺点是，如果 n 很大，须在回归方程中引入很多虚拟变量，可能超出计量软件所允许的解释变量个数。

15.5 时间固定效应

引入时间固定效应，可解决不随个体而变(individual invariant)但随时间而变(time varying)的遗漏变量问题。假设模型为

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \gamma S_t + u_i + \varepsilon_{it}$$

S_t不可观测。定义$\lambda_t \equiv \gamma S_t$，则

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \lambda_t + u_i + \varepsilon_{it}$$

将λ_t视为第 t 期独有的截距项，并将其解释为“第 t 期”对 y 的效应，故$\lambda_1, \cdots, \lambda_T$称为“时间固定效应”(time fixed effects)。

使用 LSDV 法来，对每个时期定义一个虚拟变量，把$(T-1)$个时间虚拟变量包括在回归方程中：

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \gamma_2 D2_t + \cdots + \gamma_T DT_t + u_i + \varepsilon_{it}$$

其中，时间虚拟变量$D2_t = 1$，如果$t = 2$；$D2_t = 0$，如果$t \neq 2$；以此类推。

此方程既考虑个体固定效应，又考虑时间固定效应，称为“双向固定效应”(Two-way FE)。

为节省参数，可引入时间趋势项，替代$(T-1)$个时间虚拟变量：

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \gamma t + u_i + \varepsilon_{it}$$

上式隐含较强假定，即每个时期的时间效应相等，每期均增加γ。

15.6　一阶差分法

对于固定效应模型，可对原方程两边进行一阶差分，以消去个体效应u_i(同时把$\boldsymbol{z}_i'\boldsymbol{\delta}$消掉了)，

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

对此方程使用 OLS，即得到“一阶差分估计量”(First Differencing Estimator)，记为$\hat{\boldsymbol{\beta}}_{\text{FD}}$。

只要$(\varepsilon_{it}-\varepsilon_{i,\,t-1})$与$(\boldsymbol{x}_{it}-\boldsymbol{x}_{i,\,t-1})$不相关，则$\hat{\boldsymbol{\beta}}_{\text{FD}}$一致。

此一致性条件比严格外生性假定更弱，这是$\hat{\boldsymbol{\beta}}_{\text{FD}}$的主要优点。

可以证明(参见习题)，如果$T=2$，则$\hat{\boldsymbol{\beta}}_{\text{FD}}=\hat{\boldsymbol{\beta}}_{\text{FE}}$。

对于$T>2$，如果$\{\varepsilon_{it}\}$为 iid，则$\hat{\boldsymbol{\beta}}_{\text{FE}}$比$\hat{\boldsymbol{\beta}}_{\text{FD}}$更有效率，故实践中主要使用$\hat{\boldsymbol{\beta}}_{\text{FE}}$。

对于动态面板(第 16 章)，严格外生性假定无法满足，用差分法。

15.7 随机效应模型

对于方程$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + u_i + \varepsilon_{it}$，随机效应模型假设$u_i$与解释变量$\{\boldsymbol{x}_{it}, \boldsymbol{z}_i\}$均不相关，故 OLS 一致。

但扰动项由$(u_i + \varepsilon_{it})$组成，不是球型扰动项，故 OLS 不是最有效率的，应进行 FGLS 估计。

假设不同个体之间的扰动项互不相关。由于u_i的存在，同一个体不同时期的扰动项之间仍存在自相关，

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

σ_u^2为u_i的方差，σ_ε^2为ε_{it}的方差。

当$t \neq s$时，其自相关系数为

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

自相关系数ρ不随时间距离$(t-s)$而改变。

ρ越大，则复合扰动项$(u_i + \varepsilon_{it})$中个体效应的部分(u_i)越重要。

同一个体扰动项的协方差阵为

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_u^2+\sigma_\varepsilon^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2+\sigma_\varepsilon^2 \end{pmatrix}_{T\times T}$$

整个样本的协方差阵为块对角矩阵(block diagonal matrix)，

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{\Sigma} \end{pmatrix}_{nT\times nT}$$

由于 OLS 是一致的，且其扰动项为$(u_i + \varepsilon_{it})$，故可用 OLS 的残差来估计$(\sigma_u^2 + \sigma_\varepsilon^2)$。

另一方面，FE 也一致，且其扰动项为$(\varepsilon_{it} - \bar{\varepsilon}_i)$，故可用 FE 的残差来估计$\sigma_\varepsilon^2$。

然后，用 FGLS 估计原模型，得到“随机效应估计量”(Random Effects Estimator)，记为$\hat{\boldsymbol{\beta}}_{\text{RE}}$。

具体来说，用 OLS 来估计以下“广义离差”(quasi-demeaned)模型：

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

其中，$\hat{\theta}$是$\theta \equiv 1-\dfrac{\sigma_{\varepsilon}}{(T\sigma_{u}^{2}+\sigma_{\varepsilon}^{2})^{1/2}}$的一致估计量。

可以证明，此扰动项不再有自相关。

对于随机效应模型，如果进一步假设扰动项服从正态分布，可进行 MLE 估计。

15.8　组间估计量

对于随机效应模型，还可使用“组间估计量”。

如果个体数据较不准确，可对每位个体取时间平均值，然后用平均值来回归：

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

对上式用 OLS，可得“组间估计量”(Between Estimator)，记$\hat{\boldsymbol{\beta}}_{\text{BE}}$。

由于$\{\bar{\boldsymbol{x}}_i, \boldsymbol{z}_i\}$中包含了$\{\boldsymbol{x}_{it}, \boldsymbol{z}_i\}$的信息，如果$u_i$与解释变量$\{\boldsymbol{x}_{it}, \boldsymbol{z}_i\}$相关，则$\hat{\boldsymbol{\beta}}_{\text{BE}}$不一致。故不能在固定效应模型下使用组间估计法。

15.9　拟合优度的度量

在有常数项的情况下，线性模型的R^2等于被解释变量 y 与预测值$\hat{y}$之间相关系数的平方，即$R^2 = [\text{corr}(y,\, \hat{y})]^2$。

对于面板模型，如使用混合回归，可直接用混合回归的R^2。

如使用固定效应、随机效应或组间回归，拟合优度略复杂。

给定估计量$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}})$，Stata 提供了以下三种$R^2$。

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

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

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

对于固定效应模型，建议使用组内R^2，即组内方程的R^2。

对于组间回归模型，建议使用组间R^2，即组间方程的R^2。

对于随机效应模型，这三种R^2都只是相应的相关系数平方，而非随机效应方程的R^2。

15.10　非平衡面板

非平衡面板数据并不影响计算离差形式的组内估计量(within estimator)，固定效应模型的估计可照样进行。

对于随机效应模型而言，非平衡面板数据也没有实质性影响，只要在做广义离差变换时让

$$\theta_i \equiv 1 - \frac{\sigma_\varepsilon}{(T_i \sigma_u^2 + \sigma_\varepsilon^2)^{1/2}}$$

其中，T_i为个体 i 的时间维度，就可照常进行 FGLS 估计。

非平衡面板的最大问题是，那些原来在样本中但后来丢掉的个

体，如果“丢掉”的原因是内生的(即与扰动项相关)，则会导致样本不具有代表性(不再是随机样本)，从而导致估计量不一致。

比如，低收入的人群更易从面板数据中丢掉。

15.11 究竟该用固定效应还是随机效应模型

检验原假设“$H_0: u_i$与$\boldsymbol{x}_{it}, \boldsymbol{z}_i$不相关”(即随机效应模型为正确模型)。

无论原假设成立与否，FE 都是一致的。

如果原假设不成立，则 RE 不一致。

如果H_0成立，则 FE 与 RE 估计量将共同收敛于真实的参数值，故

$(\hat{\boldsymbol{\beta}}_{\text{FE}} - \hat{\boldsymbol{\beta}}_{\text{RE}}) \xrightarrow{p} \mathbf{0}$。如果二者的差距过大，则倾向于拒绝原假设。

豪斯曼检验(Hausman, 1978)的统计量为

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

其中，K 为 $\hat{\boldsymbol{\beta}}_{\text{FE}}$ 的维度。

上述检验假设在 H_0 成立的情况下，$\hat{\boldsymbol{\beta}}_{\text{RE}}$ 最有效率。如果存在异方差，则 $\hat{\boldsymbol{\beta}}_{\text{RE}}$ 并非最有效率的估计量，故不适用异方差的情形。

解决方法之一，通过自助法计算 $\text{Var}(\hat{\boldsymbol{\beta}}_{\text{FE}} - \hat{\boldsymbol{\beta}}_{\text{RE}})$，参见第 19 章。

解决方法之二，进行以下辅助回归(Wooldridge, 2010)，

$$y_{it}-\hat{\theta}\overline{y}_i=(\boldsymbol{x}_{it}-\hat{\theta}\overline{\boldsymbol{x}}_i)'\boldsymbol{\beta}+(1-\hat{\theta})\boldsymbol{z}_i'\boldsymbol{\delta}+(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)'\boldsymbol{\gamma}+\left[(1-\hat{\theta})u_i+(\varepsilon_{it}-\hat{\theta}\overline{\varepsilon}_i)\right]$$

使用聚类稳健标准误检验原假设“$H_0:\boldsymbol{\gamma}=\mathbf{0}$”，此检验在异方差的情况下也适用。

由于总可以把原模型变换为随机效应的方程：

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

故在上面的辅助回归中，$\boldsymbol{\gamma}=\mathbf{0}$。

如果随机效应模型成立，则 OLS 一致，故$\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{\gamma}}=\boldsymbol{\gamma}=\mathbf{0}$。

如果固定效应模型成立，扰动项$\left[(1-\hat{\theta})u_i+(\varepsilon_{it}-\hat{\theta}\overline{\varepsilon}_i)\right]$与$(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)$相关(因为$u_i$与$\boldsymbol{x}_{it}$相关)，OLS 不一致，即$\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{\gamma}}=\boldsymbol{\gamma}^*\neq\boldsymbol{\gamma}=\mathbf{0}$。

拒绝“$H_0:\boldsymbol{\gamma}=\mathbf{0}$”，则意味着拒绝随机效应，接受固定效应。

对于非平衡面板，则以$\hat{\theta}_i$替代方程中的$\hat{\theta}$即可。

15.12　个体时间趋势

个体异质性还可能表现为个体的不同时间趋势。比如，在跨国面板中，各国的经济增长率可能不同。考虑以下模型：

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \gamma_i t + u_i + \varepsilon_{it}$$

$\gamma_i t$ 为个体时间趋势。

一般将γ_i视为来自某分布的随机变量(从该分布随机抽出一个观测值后，就不再随时间而变)。

此模型称为“随机趋势模型”(random trend model)。

如果y_{it}取对数形式(比如$\ln GDP_{it}$)，则γ_i可解释为在给定$(\boldsymbol{x}_{it}, \boldsymbol{z}_i)$条件下的平均增长率(即$\partial \mathrm{E}(\ln GDP_{it}) / \partial t$)，故也称“随机增长模型”(random growth model)。

首先对方程两边做差分，去掉u_i：

$$\Delta y_{it} = \Delta \boldsymbol{x}_{it}' \boldsymbol{\beta} + \gamma_i + \Delta \varepsilon_{it}$$

在形式上，此方程与标准的个体效应模型一样。

如果γ_i与解释变量$\Delta \boldsymbol{x}_{it}$不相关，可用 RE 估计此方程。

如果γ_i与解释变量$\Delta \boldsymbol{x}_{it}$相关，可用 FE 或 FD 估计此方程。