武汉大学金融系2022秋季学期

金融学/金融工程专业博士生方法论专题课

第三讲：固定效应与面板回归

授课人：刘 岩

2022年10月10日

本讲内容

1. 面板数据
2. 面板回归模型与固定效应
3. 聚类标准误
4. 动态面板回归模型的GMM估计

面板数据

面板数据的结构

- 包含多个个体，并且同一个体有一系列不同时间观测点的数据
  - 截面和时间两个维度：个体（截面）维度$i = 1,2,\ldots,N$和时间维度$t = 1,2,\ldots,T$

ID	YEAR	INC	EDU	AGE	GENDER
1	2017	800	3	23	1
1	2018	1000	4	24	1
1	2019	1200	5	25	1
2	2017	1200	5	30	0
2	2018	1250	6	31	0
2	2019	1300	7	32	0

- 数据样本单位：个体-时间，即一个样本/一行数据以个体-时间为单位

面板数据的结构

- 并不是所有包含个体和时间两个维度的数据都是面板数据
- 混合截面(pooled cross-section)数据：没有跟踪记录同一个个体，观测点属于不同个体，可以看作是横截面数据的简单合并

YEAR	INC	EDU	AGE	GENDER
2017	800	3	23	1
2018	1000	4	24	1
2019	1200	5	25	1
2017	1200	5	30	0
2018	1250	6	31	0
2019	1300	7	32	0

- 样本单位不具有个体-时间二维结构，例如债券发行数据，单位是债券

面板数据的分类

- 短面板和长面板
  - 短面板是指个体维度N较大，时间维度T较小
  - 长面板是指数据的N较小，T较大
- 平衡面板与非平衡面板
  - 平衡面板：每个个体都有相同时间T的观测点
  - 非平衡面板：有部分个体没有相同时间T的观测点
  - 若非平衡面数据由随机原因造成的，那么处理方法和平衡面板一样，但是如果数据缺失由非随机原因造成的，则必须考虑缺失的原因：如样本选择偏差

面板数据的信息来源

- 两个维度的信息
  - 不同个体间的差异和同一个个体在不同时间上的差异
- 分解：总变动total variation = 个体间变动between variation + 个体内变动within variation
- $\bar{\bar{X}}$是X_{it}在数据里总的平均值，$\bar{X}_i$是X_{it}在个体i中的平均值

$$s_0^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \bar{\bar{X}}\right)^2$$

$$s_B^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{X}_i - \bar{\bar{X}}\right)^2 \quad s_W^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)^2$$

$$s_0^2 = s_B^2 + s_W^2$$

面板回归模型与固定效应

面板数据因果关系分析的直观理解

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + \theta TALENT_i + \varphi LUCK_{it}$$

- 收入INC_{it}、受教育程度EDU_{it}：可观测随时间变化的变量
- 性别$GENDER_i$：可观测的不随时间变化的变量
- 个人天赋$TALENT_i$：不可观测且不随时间变化的变量
- 个人运气$LUCK_{it}$：不可观测且随时间变化的

面板数据因果关系分析的直观理解

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + e_{it}$$

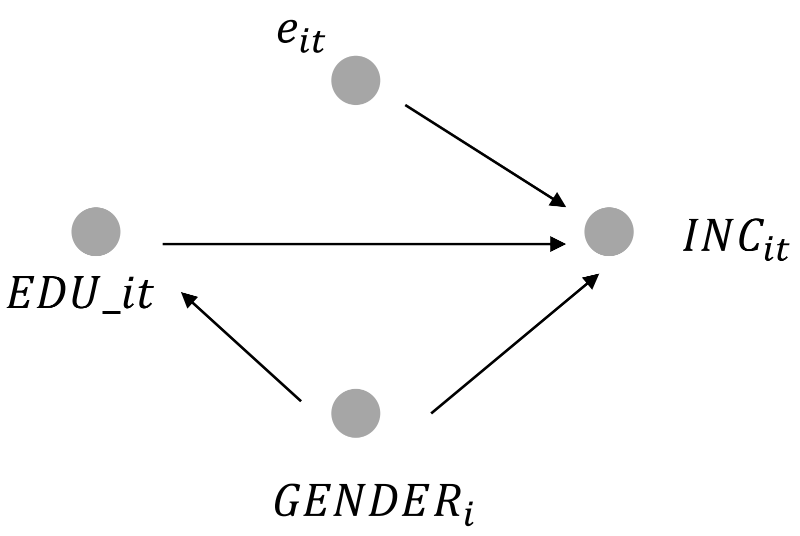
- 上式把所有不可观测的因素，包括$TALENT$和$LUCK$都归于干扰项e，那么要得到β的正确估计，需要EDU和e不相关，即EDU与天赋和运气都不相关

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + \alpha_i + u_{it}$$

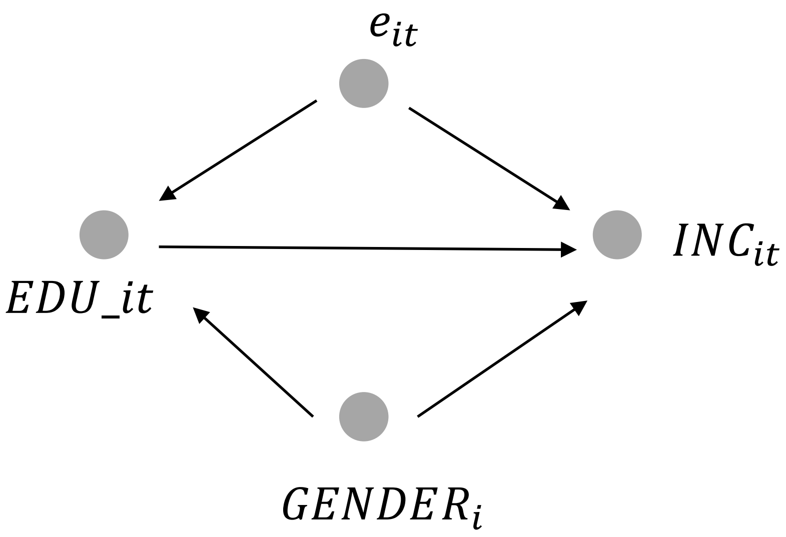
- 将干扰项e_{it}分解为α_i和u_{it}，α_i(个体效应)是个体不可观测且不随时间变化的因素$\theta TALENT_i$，u_{it}是个体不可观测且随时间变化的因素$\varphi LUCK_{it}$
- 此时α_i控制了天赋因素，要正确估计β只需要满足EDU与$LUCK$不相关

面板数据因果关系分析的直观理解

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + e_{it}$$



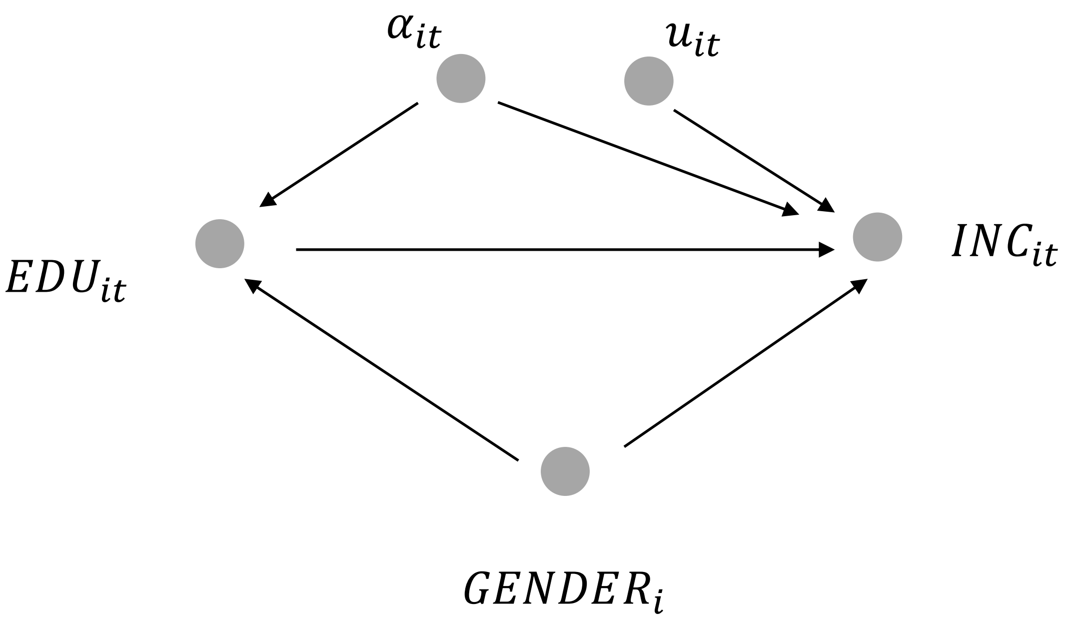
不存在混淆路径



存在混淆路径

面板数据因果关系分析的直观理解

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + \alpha_i + u_{it}$$



面板数据的变量路径图：通过控制不可观测且不随时间变化的变量**截断混淆路径**

面板回归模型基本假设

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + \alpha_i + u_{it}$$

- 假设：不可观测且随时间变化的变量u_{it}与可观测变量不相关
  - $\mathbb{E}(u_{it}|EDU_{it}, GENDER_{it}) = 0$
  - 在本例中，即为$\mathbb{E}(LUCK_{it}|EDU_{it}, GENDER_{it}) = 0$
- 面板回归三类模型的关键差别在于：对个体不可观测且不随时间变化的变量α_i的假设
  - 混合截面模型：个体效应不存在，$\alpha_i \equiv 0$
  - 随机效应模型：个体效应存在，但与解释变量不相关
  - 固定效应模型：个体效应存在，且与解释变量相关

混合截面模型

- 假设 α_i不存在，即不存在会影响收入的不可观测且不随时间变化的因素，本例中$\theta TALENT_i$假设为零

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + u_{it}$$

- 假设此时也满足$\mathbb{E}(u_{it}|EDU_{it}, GENDER_{it}) = 0$
- 这只是简单的横截面数据在时间上的叠加
- 若α_i此时存在并且与可观测变量相关，就会导致缺失变量问题

随机效应模型

- 假设α_i存在，但α_i与可观测变量不相关：$\mathbb{E}(\alpha_{it}|EDU_{it}, GENDER_{it}) = 0$，此时将$\alpha_i$放进干扰项不会造成估计误差

$$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + e_{it}$$
$$e_{it} = \alpha_i + u_{it}$$

- 由于同一个个体的干扰项e_{it}在不同时间包含了相同的α_i，即干扰项在同一个个体内是相关的，其相关系数为：

$$\mathrm{corr}(e_{it}, e_{it-s}) = \mathrm{corr}(\alpha_i + u_{it}, \alpha_i + u_{it-s}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2}$$

随机效应模型

- 当干扰项已知时，采用GLS估计，现将模型转换为同方差：

$$INC_{it}^{*} = \alpha^{*} + \beta EDU_{it}^{*} + \gamma GENDER_{i}^{*} + e_{it}^{*}$$
$$INC_{it}^{*} = INC_{it} - \theta \overline{INC_{i}}$$
$$\theta = 1 - \frac{\sigma_{\alpha}}{\sqrt{\sigma_{\alpha}^{2} + T\sigma_{u}^{2}}}$$
$$e_{it}^{*} = e_{it} - \theta \overline{e_{i}}$$

- 对转换之后的模型使用OLS，估计出来的$\hat{\beta}^{RE}$，在α_i与可观测值无关的情况下是无偏、一致且有效的估计量

固定效应模型

- 假设α_i存在，且α_i与可观测变量相关：$\mathbb{E}(\alpha_i|EDU_{it}, GENDER_{it}) \neq 0$，需要把$\alpha_i$看作解释变量处理
- 若没有把α_i看作解释变量，而是作为干扰项的一部分处理，即$INC_{it} = \alpha + \beta EDU_{it} + \gamma GENDER_i + e_{it}$
- 对上述模型求条件期望值：

$$
\begin{aligned}
&\mathbb{E}(INC_{it}|EDU_{it}, GENDER_i) \\
&= \alpha + \beta EDU_{it} + \gamma GENDER_i + \mathbb{E}(e_i|EDU_{it}, GENDER_i) \\
&= \alpha + \beta EDU_{it} + \gamma GENDER_i + \mathbb{E}(\alpha_i + u_{it}|EDU_{it}, GENDER_i) \\
&= \alpha + \beta EDU_{it} + \gamma GENDER_i + \theta \mathbb{E}(TALENT_i|EDU_{it}, GENDER_i)
\end{aligned}
$$

固定效应模型

- 假设$TALENT_i$与可观测变量EDU_{it}、$GENDER_i$的相关关系表示如下：
$$\mathbb{E}(TALENT_i|EDU_{it}, GENDER_i) = \phi_o + \phi_1 EDU_{it} + \phi_2 GENDER_i$$
- 将相关关系代入原式得
$$\mathbb{E}(INC_{it}\ |EDU_{it}, GENDER_i) \\ = (\alpha + \theta\phi_0) + (\beta + \theta\phi_1)EDU_{it} + (\gamma + \theta\phi_2)GENDER_i$$
- 可以看出，若将α_i看作干扰项，估计出来的系数是$\beta + \theta\phi_1$，而不是β，会有缺失变量$TALENT_i$对INC_{it}与EDU_{it}相关的部分
- 存在缺失变量误差

固定效应模型

$$INC_{it} = \alpha + \underbrace{\beta EDU_{\mathrm{it}} + \gamma GENDER_i}_{\text{解释变量}} + \underbrace{\alpha_i + u_{it}}_{\text{干扰项}}$$

- 固定效应模型：认为每个个体对应一个与其他解释变量相关的固定个体效应α_i
- 随机效应模型：认为个体效应是从某个分布中随机抽取，与其他解释变量无关
  - 假设过强，现实数据中几乎不会出现

固定效应模型的估计

- 面板回归模型的一般形式

$$Y_{it} = \boldsymbol{X}_{it}'\boldsymbol{\beta} + e_{it} = \boldsymbol{X}_{it}'\boldsymbol{\beta} + \alpha_i + \delta_t + u_{it}$$

- α_i：个体固定效应，δ_t：时间固定效应，u_{it}：与$\boldsymbol{X}_{it}$无关的残差项
  - 并不一定假设u_{it}在个体和时间两个维度的相关性为0，更常见的情况是u_{it}在t维度有相关性，并在截面聚类维度有相关性，见第三节讨论
- 对于大N小T型数据，δ_t并不带来任何估计的挑战，可以直接引入固定数量的时间虚拟变量进行估计
  - 原因在于OLS估计$\widehat{\boldsymbol{\beta}}^{OLS}$的渐近性质由$N$趋近于无穷保证，而$T$固定
    - 注意，引入δ_t后，$\boldsymbol{X}_{it}$中所有宏观变量，都与δ_t完全共线性，无法估计其系数
  - 若是小N大T型数据，或大N大T型数据，则引入时间虚拟变量会导致解释变量同时趋近于无穷，无法获得好的估计结果

固定效应模型的差分估计

- 第一种直观的估计方法：回归方程两端取差分（δ_t以虚拟变量形式并入$\boldsymbol{X}_{it}$中）

$$\Delta Y_{it} = \Delta \boldsymbol{X}'_{it}\boldsymbol{\beta} + \Delta \alpha_i + \Delta u_{it} = \Delta \boldsymbol{X}'_{it}\boldsymbol{\beta} + 0 + \Delta u_{it}$$

  - 这种方法不是特别常用：差分运算会去除个体变量的趋势性低频信息

- 第二种最常用估计方法：回归两端减去个体层面均值

$$Y_{it} - \bar{Y}_i = (\boldsymbol{X}'_{it} - \bar{X}'_i)\boldsymbol{\beta} + \underbrace{(\alpha_i - \bar{\alpha}_i)}_{\equiv 0} + (u_{it} - \bar{u}_i) = (\boldsymbol{X}'_{it} - \bar{X}'_i)\boldsymbol{\beta} + v_{it}$$

  - 此时残差项序列v_{it}自然带有组内时间相关性，计算标准误时需要进行聚类调整
  - STATA命令：xtreg, fe

高维固定效应

- 面板回归中，经常会使用除个体和时间之外的固定效应，但也经常存在误用，即固定效应嵌套情况
  - 截面维度：例如企业面板数据it，已经取过企业固定效应α_i，则再取行业或者地区固定效应$\alpha_{\text{ind}}, \alpha_{\text{reg}}$是没有意义的，因为与$\alpha_i$完全共线性
    - 此处假设企业没有更换行业与地区
  - 时间维度：已经取时间固定效应δ_t，再取特殊时段（如金融危机）或季节效应（1-4季度虚拟变量），没有意义
    - 后一种情况是指数据本身的时间单位短于或等于季度
- 面板回归使用高维固定效应的前提：数据单位本身是高维的，如数据单位是企业-行业-地区-时间$ijrt$，即企业在不同行业、地区、时间的行为/结果变量$Y_{ijrt}, \boldsymbol{X}_{ijrt}$，此时可以加高维固定效应
  - 甚至是高维交互固定效应，如α_{jt}, α_{rt}等；STATA命令：reghdfe，查阅说明文档

聚类标准误

聚类相关的定义

- 聚类标准误差(cluster standard error)：回归模型残差项在同一个聚类内的干扰项是相关的，但不同聚类间的干扰项是不相关的，回归系数所对应的标准误
- 例子：同一个企业不同年份的观测点是一个聚类，同一个企业不同年份观测点的干扰项很可能是相关的，因为影响同一家企业经营的干扰因素有连续性，但不同企业观测点的干扰项可能是不相关的

聚类相关的定义

- 可以将聚类g的T个观测点用向量表示为：

$$\boldsymbol{Y}_g = \boldsymbol{X}_g\boldsymbol{\beta} + \boldsymbol{e}_g$$

$$\boldsymbol{Y}_g = \left(Y_{g1}, Y_{g2}, \dots, Y_{gT}\right)', \boldsymbol{X}_g = \left(\boldsymbol{X}'_{g1}, \boldsymbol{X}'_{g2}, \dots, \boldsymbol{X}'_{gT}\right)', \boldsymbol{e} = \left(e_{g1}, e_{g2}, \dots, e_{gT}\right)'$$

- 聚类g内的干扰项结构为：

$$\mathbb{E}\left(\boldsymbol{e}_g\boldsymbol{e}'_g\right) = \boldsymbol{\Omega}_g = \begin{bmatrix} \sigma_{g1}^2 & \sigma_{g21}^2 & \sigma_{g31}^2 & \cdots & \sigma_{gT1}^2 \\ \sigma_{g12}^2 & \sigma_{g2}^2 & \sigma_{g32}^2 & \cdots & \sigma_{gT2}^2 \\ \sigma_{g13}^2 & \sigma_{g23}^2 & \sigma_{g33}^2 & \cdots & \sigma_{gT3}^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{g1T}^2 & \sigma_{g2T}^2 & \sigma_{g3T}^2 & \cdots & \sigma_{gT}^2 \end{bmatrix}$$

聚类相关的定义

- 如果有G个聚类，可以将G个聚类线性相关矩阵$\boldsymbol{Y}_g=\boldsymbol{X}_g\boldsymbol{\beta}+\boldsymbol{e}_g, g=1,2,\ldots,G$进一步叠加，表示为：

$$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{e}$$

$$\boldsymbol{Y}=(\boldsymbol{Y}_1',\boldsymbol{Y}_2',\ldots,\boldsymbol{Y}_G')',\boldsymbol{X}=(\boldsymbol{X}_1',\boldsymbol{X}_2',\ldots,\boldsymbol{X}_G')',\boldsymbol{e}=(\boldsymbol{e}_1',\boldsymbol{e}_2',\ldots,\boldsymbol{e}_G')'$$

- 其干扰项方差矩阵结构为：

$$\mathbb{E}(\boldsymbol{e}\boldsymbol{e}')=\boldsymbol{\Omega}_{\text{cluster}}=\begin{bmatrix}\boldsymbol{\Omega}_1 & 0 & 0 & \ldots & 0\\ 0 & \boldsymbol{\Omega}_2 & 0 & \ldots & 0\\ 0 & 0 & \boldsymbol{\Omega}_3 & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & 0 & \ldots & \boldsymbol{\Omega}_G\end{bmatrix}$$

- 对角线外为0反映聚类间干扰项不相关；对角线是聚类g的方差矩阵$\boldsymbol{\Omega}_g$

理解聚类相关

- 例子：假设要估计全市中学生期末考试的平均成绩，通过在不同学校进行抽样，抽取100名学生的成绩，得到100个观测点$(Y_1, Y_2, \dots, Y_{100})$
- 对上述的观测点求均值，得到样本均值$\frac{1}{N}\sum_{i=1}^{N} Y_i$
- 样本均值是总体均值的无偏估计：

$$\mathbb{E}(\bar{Y}) = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(Y_i) = \mathbb{E}(Y_i)$$

- 若100个学生是独立抽样，则样本均值的方差为：

$$\mathrm{var}(\bar{Y}) = \mathrm{var}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{var}(Y_i) = \frac{\sigma^2}{N}$$

理解聚类相关

- 上述方差计算的时候假设了100个观测值是独立的
- 但是学生的成绩分布可能不是独立的——两种可能的情况
- 第一种情况：若得到的观测值实际上是同一个人的成绩，那么它们的成绩完全相关，样本均值的方差为

$$\mathrm{var}(\bar{Y}) = \mathrm{var}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right) = \frac{1}{N^2} N^2\sigma^2 = \sigma^2$$

  - 此时的方差是独立分布时的分布的N倍

理解聚类相关

- 第二种情况：同一所学校所有学生的成绩可能相关，但是不同学校之间的学生的成绩不相关
- 那么在A学生和B学生的成绩相关的情况下，B学生的成绩提供了多少新的信息呢？
- 我们可以考虑估计只有常数项的回归方程：

$$Y_i = \alpha + e_i, \qquad i = 1,2,\dots,N, \qquad \mathbb{E}(e_i) = 0$$

- 此时$\mathbb{E}(Y_i) = \alpha$，此时可以理解为α为全市学生成绩的平均值，e_i为个体学习成绩的差异

理解聚类相关

- 此时用OLS估计系数$\hat{\alpha}$,可以得到$\hat{\alpha} = \frac{1}{N}\sum_i^{100} Y_i$
- 若此时样本是独立抽样，则干扰项满足同方差$\mathbb{E}\left(e_i^2\right) = \sigma^2, \mathbb{E}\left(e_i e_j\right) = 0$，而估计系数的方差为$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$，此时$\boldsymbol{X} = [\mathbf{1} \quad \mathbf{1} \quad \ldots \quad \mathbf{1}]'$，因此$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\mathbf{1}'\mathbf{1})^{-1} = \frac{\sigma^2}{N}$
- 因此只有常数项的回归方程在同方差的情况下，得到的结果和通常使用的样本均值估计方法所得的结果一致

理解聚类相关

- 如果同一个学校的学生成绩是相关的，可以把同一个学校g的学生$t=1,\dots,T$当成一个聚类，则可以把回归方程写成：

$$Y_{gt}=\alpha+e_{gt},\qquad g=1,2,\dots,G,\qquad t=1,2,\dots,T$$

- 此时干扰项可以分解为两部分，$e_{gt}=c_g+v_{gt}$，其中c_g为聚类因素造成的学习成绩的差异，v_{gt}为学生个人因素造成的学习成绩差异
  - 同一个学校的学生的成绩会由于c_g而产生相关关系

忽略聚类相关造成参数估计准确度被高估的程度

- 考虑一个单变量的回归方程：

$$Y_{gt} = \beta_0 + \beta_1 X_{gt} + e_{gt}, \qquad g = 1,2,\dots,G, \qquad t = 1,2,\dots,T$$

$$e_{gt} = c_g + v_{gt}$$

$$\mathbb{E}(c_g) = 0, \qquad \mathrm{var}(c_g) = \sigma_c^2, \qquad \mathbb{E}(c_g c_h) = 0, g \neq h$$

$$\mathbb{E}(v_{gt}) = 0, \qquad \mathrm{var}(v_{gt}) = \sigma_v^2, \qquad \mathbb{E}(v_{gt} v_{gs}) = 0, t \neq s$$

$$\mathbb{E}(v_{gt} v_{hs}) = 0, g \neq h$$

- 其中$\mathbb{E}(c_g c_h) = 0, g \neq h$与$\mathbb{E}(v_{gt} v_{hs}) = 0, g \neq h$表示聚类间无关

忽略聚类相关造成参数估计准确度被高估的程度

- 聚类内两个观测点t,s的干扰项的相关系数为：

$$\rho_e=\frac{\mathrm{cov}(e_{gt},e_{gs})}{\sigma_{e_{gt}}\sigma_{e_{gs}}}=\frac{\mathrm{var}(c_g)}{\mathrm{var}(c_g)+\mathrm{var}(v_{gt})}=\frac{\sigma_c^2}{\sigma_c^2+\sigma_v^2}$$

- 此时，聚类g干扰项的方差结构为：

$$\mathrm{var}(\boldsymbol{e}_g)=\mathbb{E}(\boldsymbol{e}_g\boldsymbol{e}_g')=\boldsymbol{\Omega}_g=(\sigma_c^2+\sigma_v^2)\begin{bmatrix}1 & \rho_e & \rho_e & \cdots & \rho_e\\ \rho_e & 1 & \rho_e & \dots & \rho_e\\ \rho_e & \rho_e & 1 & \dots & \rho_e\\ \dots & \dots & \dots & \dots & \dots\\ \rho_e & \rho_e & \rho_e & \cdots & 1\end{bmatrix}$$

简单Moulton因子

- 假设每个聚类的解释变量都是一样的，即$X_{gt}=X_g$，且每个聚类的规模相同
- 考虑一个组内解释变量完全相同的例子：研究班级人数对学生学习成绩的影响

$$Score_{gt}=\alpha+\beta Size_g+e_{gt},\qquad t=1,2,\dots,T$$

- 简单Moulton因子

$$\frac{\mathrm{var}(\hat{\beta}^{OLS}_{\text{cluster}})}{\mathrm{var}(\hat{\beta}^{OLS}_{\text{homo}})}=1+(T-1)\rho_e$$

- 其中T为聚类规模，ρ_e为聚类内干扰项的相关系数

推导简单Moulton因子

$$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_G)', \boldsymbol{X}_g = \underbrace{(X_g, X_g, \dots, X_g)'}_{T个X}$$

$$\mathrm{var}\left(\hat{\beta}^{OLS}_{\mathrm{cluster}}\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix} \boldsymbol{\Omega}_1 & 0 & 0 & \dots & 0 \\ 0 & \boldsymbol{\Omega}_2 & 0 & \dots & 0 \\ 0 & 0 & \boldsymbol{\Omega}_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \boldsymbol{\Omega}_G \end{bmatrix}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$\boldsymbol{\Omega}_g = (\sigma_c^2 + \sigma_v^2)\begin{bmatrix} 1 & \rho_e & \rho_e & \cdots & \rho_e \\ \rho_e & 1 & \rho_e & \dots & \rho_e \\ \rho_e & \rho_e & 1 & \dots & \rho_e \\ \dots & \dots & \dots & \dots & \dots \\ \rho_e & \rho_e & \rho_e & \cdots & 1 \end{bmatrix}$$

推导简单Moulton因子

$$\boldsymbol{X}'\begin{bmatrix}\boldsymbol{\Omega}_1 & 0 & 0 & \dots & 0\\ 0 & \boldsymbol{\Omega}_2 & 0 & \dots & 0\\ 0 & 0 & \boldsymbol{\Omega}_3 & \dots & 0\\ \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \dots & \boldsymbol{\Omega}_G\end{bmatrix}\boldsymbol{X}=\sum_{g=1}^{G}\boldsymbol{X}_g'\boldsymbol{\Omega}_g\boldsymbol{X}_g$$

$$=\sum_{g=1}^{G}(X_g,X_g,\dots,X_g)(\sigma_c^2+\sigma_v^2)\begin{bmatrix}1 & \rho_e & \rho_e & \cdots & \rho_e\\ \rho_e & 1 & \rho_e & \dots & \rho_e\\ \rho_e & \rho_e & 1 & \dots & \rho_e\\ \dots & \dots & \dots & \dots & \dots\\ \rho_e & \rho_e & \rho_e & \cdots & 1\end{bmatrix}(X_g,X_g,\dots,X_g)'$$

$$=(\sigma_c^2+\sigma_v^2)\sum_{g=1}^{G}\left(T{X_g}^2+(T-1)\rho_e{X_g}^2\right)$$

推导简单Moulton因子

$$\text{var}(\hat{\beta}^{OLS}_{\text{homo}}) = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\begin{bmatrix} \boldsymbol{\Omega} & 0 & 0 & \ldots & 0 \\ 0 & \boldsymbol{\Omega} & 0 & \ldots & 0 \\ 0 & 0 & \boldsymbol{\Omega} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \boldsymbol{\Omega} \end{bmatrix}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}$$

$$\boldsymbol{\Omega} = (\sigma_c^2 + \sigma_v^2)\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

推导简单Moulton因子

$$\boldsymbol{X}'\begin{bmatrix}\boldsymbol{\Omega} & 0 & 0 & \dots & 0\\ 0 & \boldsymbol{\Omega} & 0 & \dots & 0\\ 0 & 0 & \boldsymbol{\Omega} & \dots & 0\\ \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \dots & \boldsymbol{\Omega}\end{bmatrix}\boldsymbol{X} = \sum_{g=1}^{G}\boldsymbol{X}_g'\boldsymbol{\Omega}\boldsymbol{X}_g$$

$$= \sum_{g=1}^{G}(X_g, X_g, \dots, X_g)(\sigma_c^2 + \sigma_v^2)\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \dots & 0\\ 0 & 0 & 1 & \dots & 0\\ \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \cdots & 1\end{bmatrix}(X_g, X_g, \dots, X_g)'$$

$$= (\sigma_c^2 + \sigma_v^2)\sum_{g=1}^{G}TX_g^2$$

推导简单Moulton因子

$$\frac{\text{var}(\hat{\beta}^{OLS}_{\text{cluster}})}{\text{var}(\hat{\beta}^{OLS}_{\text{homo}})} = \frac{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix} \boldsymbol{\Omega}_1 & 0 & 0 & \dots & 0 \\ 0 & \boldsymbol{\Omega}_2 & 0 & \dots & 0 \\ 0 & 0 & \boldsymbol{\Omega}_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \boldsymbol{\Omega}_G \end{bmatrix}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}}{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix} \boldsymbol{\Omega} & 0 & 0 & \dots & 0 \\ 0 & \boldsymbol{\Omega} & 0 & \dots & 0 \\ 0 & 0 & \boldsymbol{\Omega} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \boldsymbol{\Omega} \end{bmatrix}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}}$$

$$= \frac{(\sigma_c^2 + \sigma_v^2)\sum_{g=1}^{G}(T{X_g}^2 + (T-1)\rho_e {X_g}^2)}{(\sigma_c^2 + \sigma_v^2)\sum_{g=1}^{G} T{X_g}^2} = 1 + (T-1)\rho_e$$

简单Moulton因子

- 简单Moulton因子：$\frac{\text{var}(\hat{\beta}_{\text{cluster}}^{OLS})}{\text{var}(\hat{\beta}_{\text{homo}}^{OLS})} = 1 + (T - 1)\rho_e$，其中$T$为聚类规模，$\rho_e$为聚类内干扰项的相关系数
- 假设$\rho_e = 1$，即组内的所有干扰项是完全相关的，此时Moulton因子为T
- 当聚类内观测数T增加时，Moulton因子增加
- 很小的聚类内相关系数也能导致一个很大的Moulton因子
  - 例如，若1000个观测点分属于10个聚类，组内相关系数为0.1，此时简单Moulton因子为$1 + (100 - 1) \times 0.1 = 10.9$，即忽略聚类相关会导致方差低估

广义Moulton因子

- 广义Moulton因子允许聚类内解释变量不完全相同，但是彼此相关，并且允许不同聚类的规模不同
- 广义Moulton因子：$\frac{\text{var}(\hat{\beta}_{\text{cluster}}^{OLS})}{\text{var}(\hat{\beta}_{\text{homo}}^{OLS})} = \left[\frac{V(T_g)}{\bar{T}} + \bar{T} - 1\right]\rho_x\rho_e$
- 其中，$\rho_x = \frac{\sum_g \sum_t \sum_{s \neq t}(X_{gt} - \bar{X})(X_{gs} - \bar{X})}{V(X_{gt}) \sum_g T_g(T_g - 1)}$是组内各个解释变量之间的相关系数，$T_g$为群组$g$的规模，$V(T_g)$为群组规模的方差，$\bar{T}$为各个群组的平均规模

广义Moulton因子的新增的性质

- 当各个聚类规模方差$V(T_g)$较大或者解释变量X_{gt}聚类内的相关系数ρ_x很大时，干扰项的聚类相关方差会对标准误差造成更大的影响
- 解释变量的聚类内相关性ρ_x与干扰项的聚类内相关性ρ_e一样，都会影响估计系数方差，当解释变量聚类内相关系数ρ_x为0时，结果和同方差的估计值一样，此时干扰项的聚类方差不影响标准误差

处理方法

- 其一，采用GLS的方法，根据干扰项的聚类方差的结构，对模型进行调整之后转换
  - 但是很难预先知道干扰项的聚类方差的结构，因此实际很少采用该方法
- 其二，使用OLS进行估计，并且估计出聚类方差下OLS估计值的标准误差，这是比较常用的方法
- STATA面板回归中，xtreg, vce(robust)自动取截面单位聚类稳健标准误
  - 同时考虑异方差及残差项序列相关
  - 等价于vce(cluster)，即vce(cluster panel-id)缺省值
  - 另外可以使用vce(cluster groupvar)指定更大的聚类范围，如在行业或地区取聚类标准误；对于大N大T型数据，还可以在截面和时间两个维度取双向聚类，如跨国面板数据同时在国家和时间取twoway cluster，见Cameron & Miller (2015, J Human Resouce)

处理方法

- OLS估计系数的方差为：

$$\operatorname{var}\left(\hat{\beta}^{OLS}\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbb{E}[\boldsymbol{e}\boldsymbol{e}']\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 把聚类方差矩阵代入，可以得到：

$$\operatorname{var}\left(\hat{\beta}^{OLS}_{\text{cluster}}\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix} \boldsymbol{\Omega}_1 & 0 & 0 & \dots & 0 \\ 0 & \boldsymbol{\Omega}_2 & 0 & \dots & 0 \\ 0 & 0 & \boldsymbol{\Omega}_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \boldsymbol{\Omega}_G \end{bmatrix}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 此时我们用估计值$\widehat{\boldsymbol{\Omega}}_g$来代替$\boldsymbol{\Omega}_g$

处理方法

- 用残差值$\hat{e}_{gt}^2$代替σ_{gt}^2，用$\hat{e}_{gt}\hat{e}_{gs}$来代替σ_{gtgs} 来计算聚类稳健协方差矩阵估计值
- 因此

$$\widehat{\boldsymbol{\Omega}}_g = \hat{\boldsymbol{e}}_g\hat{\boldsymbol{e}}_g' = \begin{bmatrix} \hat{e}_{g1}^2 & \hat{e}_{g1}\hat{e}_{g2} & \hat{e}_{g1}\hat{e}_{g3} & \cdots & \hat{e}_{g1}\hat{e}_{gT} \\ \hat{e}_{g2}\hat{e}_{g1} & \hat{e}_{g2}^2 & \hat{e}_{g2}\hat{e}_{g3} & \cdots & \hat{e}_{g2}\hat{e}_{gT} \\ \hat{e}_{g3}\hat{e}_{g1} & \hat{e}_{g3}\hat{e}_{g2} & \hat{e}_{g3}^2 & \cdots & \hat{e}_{g3}\hat{e}_{gT} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{e}_{gT}\hat{e}_{g1} & \hat{e}_{gT}\hat{e}_{g2} & \hat{e}_{gT}\hat{e}_{g3} & \cdots & \hat{e}_{gT}^2 \end{bmatrix}$$

处理方法

$$\text{var}\left(\hat{\beta}^{OLS}_{\text{cluster}}\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix} \widehat{\boldsymbol{\Omega}}_1 & 0 & 0 & \dots & 0 \\ 0 & \widehat{\boldsymbol{\Omega}}_2 & 0 & \dots & 0 \\ 0 & 0 & \widehat{\boldsymbol{\Omega}}_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \widehat{\boldsymbol{\Omega}}_G \end{bmatrix}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}[\boldsymbol{X}_1', \boldsymbol{X}_2', \dots, \boldsymbol{X}_G']\begin{bmatrix} \widehat{\boldsymbol{\Omega}}_1 & 0 & 0 & \dots & 0 \\ 0 & \widehat{\boldsymbol{\Omega}}_2 & 0 & \dots & 0 \\ 0 & 0 & \widehat{\boldsymbol{\Omega}}_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \widehat{\boldsymbol{\Omega}}_G \end{bmatrix}\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \dots \\ \boldsymbol{X}_G \end{bmatrix}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{g=1}^{G}\boldsymbol{X}_g'\widehat{\boldsymbol{\Omega}}_g\boldsymbol{X}_g(\boldsymbol{X}'\boldsymbol{X})^{-1} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{g=1}^{G}\boldsymbol{X}_g'\hat{\boldsymbol{e}}_g\hat{\boldsymbol{e}}_g'\boldsymbol{X}_g(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

聚类相关方差Stata实例

student	school	score	student	school	score
1	M	71	16	W	86
2	M	72	17	W	87
3	M	73	18	W	88
4	T	74	19	R	89
5	T	75	20	R	90
6	T	76	21	R	91
7	Q	77	22	U	92
8	Q	78	23	U	93
9	Q	79	24	U	94
10	L	80	25	S	95
11	L	81	26	S	96
12	L	82	27	S	97
13	G	83	28	A	98
14	G	84	29	A	99
15	G	85	30	A	100

聚类相关方差Stata实例

- 计算聚类相关系数
- loneway score school

Intraclass correlation	Asy. S.E.	[95% Conf. Interval]	
0.98798	0.00677	0.97471	1.00125

Estimated SD of school effect	9.064583
Estimated SD within school	1
Est. reliability of a school mean(evaluated at n=3.00)	0.99596

聚类相关方差Stata实例

- 由前面可知，总体均值的估计等价于估计以下的回归方程：

$$Y_{gt} = \alpha + c_g + v_{gt}$$

- 由上述运行结果可知：组内相关系数为$\rho_e = 0.98798$
- 此时Moulton因子=$1 + (3 - 1) \times 0.98798 = 2.97596$
- 这就意味着假设同方差而忽略聚类相关，估计系数的方差会变成真实值的1/2.97596

聚类相关方差Stata实例

- sum score

Variable	Obs	Mean	Std. Dev.	Min	Max
score	30	85.5	8.803408	71	100

- Std. Err.$(\bar{Y}) = \sqrt{Var(\hat{\alpha})} = \frac{\hat{\sigma}}{\sqrt{N}} = \frac{8.803408}{\sqrt{30}} = 1.607275$

- mean score

	Mean	Std. Err.	[95% Conf. Interval]	
score	85.5	1.607275	82.21275	88.78725

聚类相关方差Stata实例

- reg score

score	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
_cons	85.5	1.607275	53.20	0.000	82.21275	88.78725

- 这三个命令默认的都是观测点的干扰项不相关，但是其实它有较大的组内相关，因此正确的标准误差应该为：
- $\mathrm{Std.Err.}\left(\hat{\beta}^{OLS}_{\mathrm{cluster}}\right) = \mathrm{Std.Err.}\left(\hat{\beta}^{OLS}_{\mathrm{homo}}\right) \times \sqrt{Moulton\ factor} = 1.607275 \times \sqrt{2.97596} = 2.7727$

聚类相关方差Stata实例

- reg score, cluster(school)

score	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
_cons	85.5	2.872281	29.77	0.000	79.00245	91.99755

- 由此可以看出，如果干扰项存在正的聚类相关，而我们忽略了这个相关性，就会低估标准误差，高估了参数的准确性

聚类方差运用常见问题

- 由于聚类方差的复杂性，到目前为止都没有选择聚类的统一的标准
  - 因为当样本数量一定时，规模更大的的聚类考虑更加广泛的相关项，偏差较小，但是同时更少的聚类使得方差更大，估计更加不准确。
- 一般认为在没有由于聚类数量过少引发问题的情况下，尽量使用更大的聚类，提高方差估计的准确性
  - 聚类稳健标准误的大样本渐近性质要求聚类数G趋于无穷，因此如果G过小，则不满足标准误估计的渐近要求

聚类方差运用常见问题

- 如果聚类内为正相关，则聚类方差会比同方差大，使用聚类方差后，回归系数的方差会变大
- 如果聚类内为负相关，则聚类方差会比同方差小
- 如果同时还存在异方差，此时聚类方差与同方差的差异，不仅取决于聚类内的相关性，还取决于异方差

聚类方差运用常见问题

- 即使在模型中加入了聚类固定效应，也不一定就控制了聚类相关项，不一定不需要使用聚类方差
- 例如考虑模型$Y_{it}=\alpha+\beta X_{it}+f_i+e_{it}$
- 如果企业在样本期间经历了一些意外事件，这些事件会有一些持续性，就会导致企业内的自相关，但是它们并不属于企业的固定效应

动态面板回归模型的GMM估计

动态面板模型

- 动态面板模型：
  - 动态：模型中包含了因变量的滞后项
  - 有个体的固定效应
  - 可以有一些自变量是内生的
  - 除了固定效应之外的误差项e，可以异方差，可以序列相关
  - 可以有前定的但不是完全外生的变量；前定变量：w_{it}与e_{it}不相关，但可以与e_{it-1}及更高阶滞后相关
  - 大N小T，即个体数量要足够多，但时间不用太长；如果时间足够长的话，动态面板误差不会太大，用固定效应即可

动态面板模型

- 动态面板回归(dynamic panel regression)模型的最基本形式：
$$Y_{it} = \rho Y_{it-1} + \boldsymbol{X}'_{it}\boldsymbol{\beta} + \alpha_i + u_{it}$$
- u_{it}表示残差项且与所有回归变量无当期相关性
- 回归方程右侧可加入Y_{it} 的更多滞后项
- 由于个体效应α_i会影响Y_{it}，从而在$t+1$期的回归方程中导致解释变量Y_{it}与α_i相关，因此α_i一定是个体固定效应
- 传统的固定效应估计不适用于动态面板模型，需要使用GMM (generalized method of moments)估计方法，因此该类模型由常简称为面板GMM（方法）

动态面板的内生性问题

- 当面板回归模型具有动态特征时，静态面板的简单变换方法无法消除固定效应的影响，因而 OLS 估计不可行
- 作一阶差分：回归方程两边取差分后得

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \Delta \boldsymbol{X}'_{it} \boldsymbol{\beta} + \Delta u_{it}$$

- 但$\mathrm{cov}(\Delta Y_{it-1}, \Delta u_{it}) \neq 0$，原因在于$u_{it-1}$对$Y_{it-1}$有影响
- 通过组内均值的形式消除固定效应α_i，同样会引入$Y_{it-1} - \bar{Y}_i$与$u_{it} - \bar{u}_i$间的相关性问题，因为$\bar{u}_i = \frac{1}{T}\sum_s u_{is}$自然与$Y_{it-1}$相关，而$\bar{Y}_i = \frac{1}{T}\sum_s Y_{is}$自然与$u_{it}$相关

基本解决思路：滞后项作为工具变量

- 差分法在动态面板下无法克服固定效应带来的问题：根源在于被解释变量的动态性 $\Rightarrow t$期回归变量Y_{it-1}与残差差分项中u_{it-1}的相关性
- 解决思路：找一个工具变量，使其与差分方程中回归变量相关，但与残差无关
- 最简单的工具变量：Y_{it-2}，它和ΔY_{it-1}相关，但和Δu_{it}无关
- 基本假设1：u_{it}没有序列相关性

差分GMM

- Arrelano & Bond (1991, RES) 提出把所有的滞后项全部引入为工具变量，使用 GMM 方法进行估计
  - 由于这一方法的基础是对差分方程进行估计，故称为差分GMM；同时文献中也大量使用 AB 方法这一名称
- 如果 Y_{it}和解释变量的一阶自相关很接近 1，则可能存在弱工具变量问题
  - $\Delta Y_{it} = Y_{it} - Y_{it-1}$几乎完全由回归方程中的余项决定，与 Y_{it-1}的相关系数会很小
  - 如果工具变量和内生变量的相关性较弱，那么2SLS的偏差会很严重

系统GMM

- 针对上述问题高自相关性带来的弱工具变量问题，Arrelano & Bover (1995) 提出了另外一种选取工具变量的方法：水平回归与差分滞后

$$Y_{it} = \rho Y_{it-1} + \boldsymbol{X}'_{it}\boldsymbol{\beta} + \alpha_i + u_{it}$$

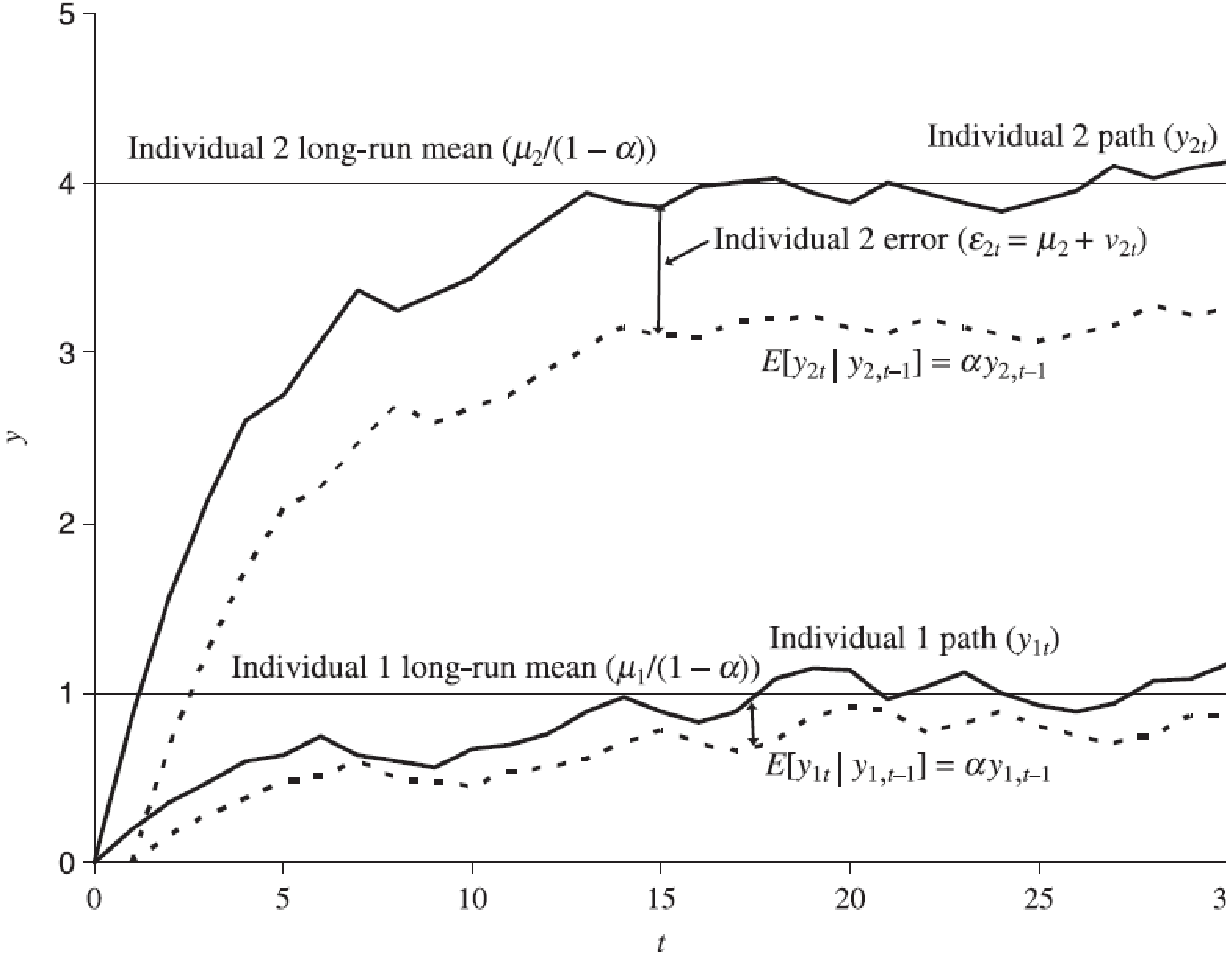
- 如果被解释变量差分滞后ΔY_{it-1}与固定效应α_i没有相关性，那么ΔY_{it-1}也可以作为水平回归的工具变量

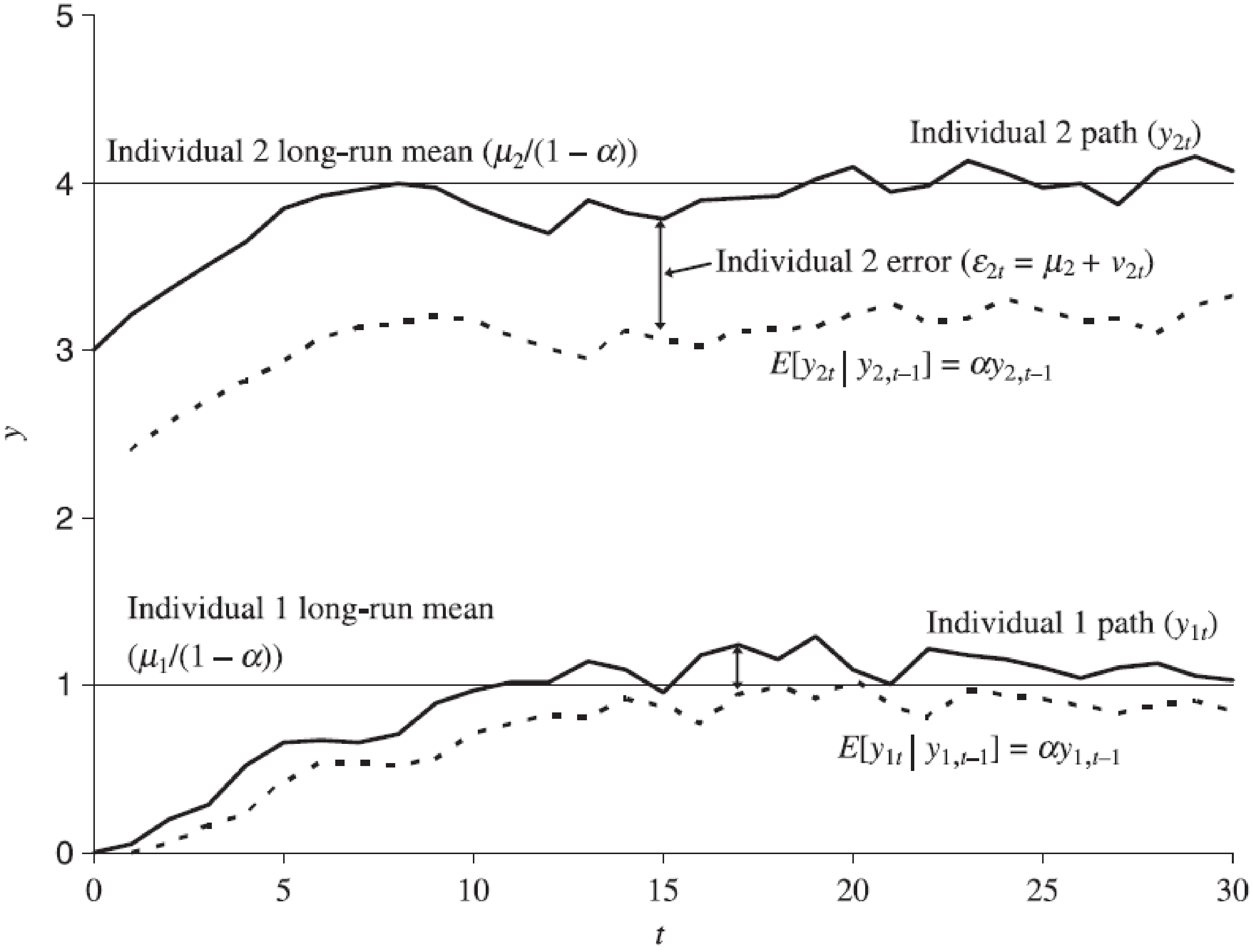
系统GMM

- 基本假设2：$\text{cov}(\Delta Y_{it-1}, \alpha_i) = 0$
- 在此假设下，差分回归与滞后工具变量可以同水平回归与差分滞后工具变量相结合，形成一个更完善的 GMM 估计系统
- 这一方法被 Blundell & Bond (1998) 称为系统 GMM
  - Blundell & Bond 给出的模拟结果显示，当样本自相关很高时，系统 GMM 的有限样本偏差要比差分 GMM 好
- Blundell & Bond 同时指出了基本假设2的实质：Y_{it}关于其长期均值的暂时偏离与u_i无关

$$\mathbb{E}\left[\left(Y_{it} - \frac{\alpha_i}{1-\rho}\right)\alpha_i\right] = 0$$

Roodman (2009, Ox. Bu. Eco & Stat)：不满足与满足假设2示例





注意与前页模型的对应：图中$\alpha = \rho$，$\mu = \alpha_i$；Roodman (2009)是面板GMM技术参考宝典

系统GMM

- 定义样本矩阵：*表示样本变换（如差分），L表示水平值

$$\boldsymbol{Y}_i = \begin{bmatrix} \boldsymbol{Y}_i^* \\ \boldsymbol{Y}_i^L \end{bmatrix}, \qquad \boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{X}_i^* \\ \boldsymbol{X}_i^L \end{bmatrix}$$

- $\boldsymbol{Z}_i$ 表示所有工具变量构成的矩阵
- 给定变换矩阵 $\boldsymbol{H}_i$，定义

$$\boldsymbol{Q}_{XZ} = \sum_i \boldsymbol{X}_i'\boldsymbol{Z}_i, \boldsymbol{Q}_{ZY} = \sum_i \boldsymbol{Z}_i'\boldsymbol{Y}_i, \boldsymbol{W} = \boldsymbol{Q}_{XZ}\boldsymbol{A}\boldsymbol{Q}_{XZ}', A = \left(\sum_i \boldsymbol{Z}_i'\boldsymbol{H}_i\boldsymbol{Z}_i\right)^{-1}$$

- 回归系数$\boldsymbol{\theta} = [\rho, \boldsymbol{\beta}']'$的GMM估计为

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{W}^{-1}\boldsymbol{Q}_{XZ}\boldsymbol{A}\boldsymbol{Q}_{ZY}$$

系统GMM估计

- 动态面板 GMM 估计分为一步法与两步法。
- 在一步法中，变换矩阵$\boldsymbol{H}_{1i}$取特定常数值（如单位阵），对应的加权矩阵$\boldsymbol{A}_1$直接计算可得
- 在两步法中，变换矩阵$\boldsymbol{H}_{2i}$由一步法估计值$\widehat{\boldsymbol{\theta}}_1$计算残差构造而得：

$$\hat{u}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\theta}}_1, \qquad \boldsymbol{H}_{2i} = \widehat{\boldsymbol{u}}_i'\widehat{\boldsymbol{u}}_i$$

- 再由$\boldsymbol{H}_{2i}$计算新的加权矩阵$\boldsymbol{A}_2$，及相应的$\boldsymbol{W}_2$，最终得到两步法估计值

$$\widehat{\boldsymbol{\theta}}_2 = \boldsymbol{W}_2^{-1}\boldsymbol{Q}_{XZ}\boldsymbol{A}_2\boldsymbol{Q}_{ZY}$$

系统GMM估计：一步法

- 在同方差假设下，一步法得到的参数估计（向量）$\widehat{\boldsymbol{\theta}}_1$具有下列协方差矩阵：$\hat{\sigma}_1^2 \boldsymbol{W}_1^{-1}$，其中$\hat{\sigma}_1^2$表示残差（变换后）的同方差估计值
- 一步法下，可以使用 White 的异方差稳健标准误（robust standard error）方法估计$\widehat{\boldsymbol{\theta}}_1$的协方差矩阵：

$$\boldsymbol{W}_1^{-1}\boldsymbol{Q}_{XZ}\boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{A}_1\boldsymbol{Q}_{XZ}'\boldsymbol{W}_1^{-1}$$

- 其中$\boldsymbol{A}_2$是两步法中使用的加权矩阵，由一步法残差构造

系统GMM估计：两步法

- 两步法中，同方差假设下$\widehat{\boldsymbol{\theta}}_2$具有协方差矩阵$\boldsymbol{W}_2^{-1}$。
- 但由于一步法残差构造的两步法加权矩阵有很大的有限样本偏误，利用$\boldsymbol{W}_2^{-1}$所做统计推断均存在很大偏差
  - 150左右的样本系数即可保证绝大部分系数1%显著
- Windermeijer (2005) 提出了经过偏误纠正的加权矩阵估计值$A_{2,\text{robust}}$，对应$\widehat{\boldsymbol{\theta}}_2$的协方差矩阵称为偏误纠正（或WC-robust）协方差阵
- WC-robust 协方差阵得到的统计推断结果具有较好的有限样本性质；综合水平与一步法稳健标准误结果类似或略高
  - STATA命令xtabond/xtdpdsys, twostep vce(robust)即提供WC-robust协方差矩阵及相应标准误

模型设定检验

- 对动态面板 GMM 的模型设定检验通常包括两类
- 首先是对残差项 ε_{it} 自相关性的检验，基本假设 1 要求ε_{it}无自相关，故$\Delta\varepsilon_{it}$不会具有 2 阶及以上自相关
- 其次是对工具变量数目过多带来的过渡识别进行检验，通常可以进行一个 Sargan 检验：原假设为不存在过渡识别，因此当以较小的 p-值拒绝原假设时，需要考虑缩减工具变量的数量
  - 但其他方面的模型设定问题也可能引起 Sargan 检验拒绝原假设，如被解释变量滞后期设定不足等
  - Roodman (2009)提出的经验规则是保证工具变量综述约等于面板的截面个体数
    - 通过限制工具变量（水平值、差分值）滞后期来减少工具变量个数

解释变量设定

- 前面的讨论中我们假定解释变量 $\boldsymbol{X}_{it}$ 完全外生：与 u_{it} 及其滞后项均无相关性
- 上述假设可以放松：允许存在前定变量和内生变量。
- 前定变量：$\boldsymbol{W}_{it}$ 与 u_{it} 不相关，但可以与 u_{it-1} 及更高阶滞后相关。
- 内生变量：$\boldsymbol{W}_{it}$ 与 u_{it} 相关及更高阶滞后相关
- 在 GMM 估计设定中，可以指定哪些变量是前定变量或内生变量，软件会相应的引入这些变量的对应水平滞后或者差分滞后作为工具变量，进入 GMM 估计程序

示例：金融发展对经济增长的影响

- Beck & Levine (2004, JBF)及Arcand et al. (2015, J Econ Growth)
- 跨国面板回归

$$Y_{it} = \rho Y_{it-1} + \beta F_{it} + \boldsymbol{X}'_{it}\boldsymbol{\phi} + \alpha_i + u_{it}$$

- 其中Y_{it}为对数人均GDP，F_{it}为金融发展指标
- 根据Solow模型及衍生的增长收敛理论，人均产出增速随着人均产出绝对值的提高而降低，即$Y_{it} - Y_{it-1} = \eta Y_{it-1} + \cdots$，故跨国增长回归自然需要考虑$Y_{it}$的滞后项，即动态面板回归模型