

第 17 章 非线性面板

17.1 面板二值选择模型

对于面板数据, 如果被解释变量为虚拟变量, 则为“面板二值选择模型”(binary choice model for panel data)。

二值选择行为可通过“潜变量”(latent variable)来概括其净收益。净收益大于 0, 则选择做; 否则, 不做。

假设净收益为

$$y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + \varepsilon_{it} \quad (i = 1, \dots, n; \quad t = 1, \dots, T)$$

y_{it}^* 不可观测, u_i 为个体效应, 解释变量 \mathbf{x}_{it} 不含常数项。选择规则:

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0 \\ 0 & \text{if } y_{it}^* \leq 0 \end{cases}$$

给定 $\mathbf{x}_{it}, \boldsymbol{\beta}, u_i$, 则有

$$\begin{aligned}
P(y_{it} = 1 \mid \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) &= P(y_{it}^* > 0 \mid \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\
&= P(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i + \varepsilon_{it} > 0 \mid \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\
&= P(\varepsilon_{it} > -u_i - \mathbf{x}_{it}'\boldsymbol{\beta} \mid \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\
&= P(\varepsilon_{it} < u_i + \mathbf{x}_{it}'\boldsymbol{\beta} \mid \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) \\
&= F(u_i + \mathbf{x}_{it}'\boldsymbol{\beta})
\end{aligned}$$

$F(\cdot)$ 为 ε_{it} 的 cdf, 并假设 ε_{it} 的密度函数关于原点对称。

如果 $\varepsilon \sim N(0, 1)$, 则为 Probit 模型,

$$P(y_{it} = 1 \mid \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \Phi(u_i + \mathbf{x}_{it}'\boldsymbol{\beta})$$

如果 ε 服从逻辑分布，则为 Logit 模型，

$$P(y_{it} = 1 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \Lambda(u_i + \mathbf{x}_{it}'\boldsymbol{\beta}) = \frac{e^{u_i + \mathbf{x}_{it}'\boldsymbol{\beta}}}{1 + e^{u_i + \mathbf{x}_{it}'\boldsymbol{\beta}}}$$

面板二值模型的估计方法包括混合回归、随机效应估计与固定效应估计。

如果 $u_1 = u_2 = \cdots = u_n$ ，则为混合回归(pooled probit 或 pooled logit)，可作为截面数据处理，应使用以面板为聚类的聚类稳健标准误(cluster-robust standard errors)。

17.2 面板二值选择模型的随机效应估计

更一般地，允许个体效应的存在，不同的个体拥有不同的 u_i 。

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

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)。

但非线性面板不便使用 FGLS，故使用 MLE。

假设 $u_i \sim N(0, \sigma_u^2)$ ，记其密度函数为 $g(u_i)$ 。

以 Logit 模型为例。

给定 u_i ，则个体 i 的条件分布为(推导过程参考第 11 章)，

$$f(y_{i1}, y_{i2}, \dots, y_{iT} \mid \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \prod_{t=1}^T [\Lambda(u_i + \mathbf{x}_{it}' \boldsymbol{\beta})]^{y_i} [1 - \Lambda(u_i + \mathbf{x}_{it}' \boldsymbol{\beta})]^{1-y_i}$$

但 u_i 不可观测。记 $(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$ 的联合密度为 $f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$ ，并进行如下分解，

$$f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) = f(y_{i1}, y_{i2}, \dots, y_{iT} \mid u_i) \cdot g(u_i)$$

在 $(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$ 的联合密度中，将 u_i 积分积掉，可得 $(y_{i1}, y_{i2}, \dots, y_{iT})$ 的边缘密度，

$$\begin{aligned} f(y_{i1}, y_{i2}, \dots, y_{iT}) &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) du_i \\ &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT} | u_i) \cdot g(u_i) du_i \\ &= \int_{-\infty}^{+\infty} \left\{ \prod_{t=1}^T [\Lambda(u_i + \mathbf{x}'_{it} \boldsymbol{\beta})]^{y_i} [1 - \Lambda(u_i + \mathbf{x}'_{it} \boldsymbol{\beta})]^{1-y_i} \right\} \cdot g(u_i) du_i \end{aligned}$$

上式无解析解，Butler and Moffitt (1982) 提出使用 “Gauss-Hermite quadrature” 方法进行数值积分。

Stata 的默认方法为在 12 个点上进行 “adaptive Gauss-Hermite quadrature” 计算。

此积分的精度依赖于数值计算的点数，故 Stata 提供命令 `quadchk` 来检验积分的精度，即使用其他计算点数，比较结果的稳定性。

最大化此似然函数即得到对 β 的“随机效应 Logit 估计量”(Random Effect Logit)。

如果将逻辑分布 $\Lambda(\cdot)$ 改为正态分布 $\Phi(\cdot)$ ，则为 “随机效应 Probit 估计量” (Random Effect Probit)。

在估计时已将 u_i 积分掉，故得不到对个体效应 u_i 的估计，也无法预测 y_i 的发生概率或计算解释变量的边际效应。

解决方法之一是假设 $u_i = 0$ 。

由于 u_i 的存在，同一个体不同期的扰动项之间仍存在自相关，

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

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，

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

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。如果 $\rho = 0$ ，则 $\sigma_u^2 = 0$ ，不存在个体随机效应，应选择混合回归。

如果拒绝“ $H_0: \rho = 0$ ”，则认为应采用随机效应模型；反之，则支持混合回归。

17.3 面板二值选择模型的固定效应估计

在面板二值模型中，如果 u_i 与 \mathbf{x}_{it} 相关，则为固定效应模型，

$$P(y_{it} = 1 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = F(u_i + \mathbf{x}_{it}'\boldsymbol{\beta})$$

其中， $F(\cdot)$ 为 $\Phi(\cdot)$ 或 $\Lambda(\cdot)$ 。随机效应模型或混合回归将不一致。

对于线性面板，可通过组内变换($y_{it} - \bar{y}_i$)或一阶差分($y_{it} - y_{i, t-1}$)消去固定效应 u_i 。

对于非线性面板，这些变换不起作用。

即使在模型中增加个体虚拟变量(LSDV 法)，仍得不到一致估计(除非 $T \rightarrow \infty$)。因为当 $n \rightarrow \infty$ 时，待估计的个体效应 $\{u_i\}_{i=1}^n$ 的个数也随之增加，这些 $\{u_i\}_{i=1}^n$ 称为“伴生参数”(incidental parameters)。

每个 u_i 只能由个体 i 的 T 个观测值来估计，而 T 并不增加。对于短面板，当 $n \rightarrow \infty$ 而 T 为有限值，对 $\{u_i\}_{i=1}^n$ 的估计不一致。

对于 u_i 的不一致估计还会“污染”对 β 的估计，导致对 β 的估计也不一致，称为“伴生参数问题” (incidental parameters problem)。

在线性面板中，可通过组内变换或差分变换解决伴生参数问题。对于固定效应的面板 Probit，无法解决此伴生参数问题。

对于固定效应的面板 Logit，可通过寻找 u_i 的“充分统计量” (sufficient statistic)，然后在给定此充分统计量的条件下进行“条件最大似然估计” (conditional MLE)。

考虑总体参数 θ 与统计量 W 。如果统计量 W 包含了样本中所有可用来估计 θ 的信息，则称 W 为参数 θ 的充分统计量。

换言之，给定 W 之后，任何根据样本计算的其他统计量都不可能提供关于 θ 的额外信息。

样本数据在给定充分统计量 W 后的条件分布将不再依赖于 θ ；否则，控制了 W 之后的条件分布仍将包含 θ 的信息。

Chamberlain (1980)提出使用 $n_i \equiv \sum_{t=1}^T y_{it}$ 作为 u_i 的充分统计量，并计算在给定 n_i 情况下的条件似然函数(此条件似然函数不再依赖于 u_i)，然后进行条件 MLE 估计。

由于逻辑函数的指数形式，故 Logit 模型存在 u_i 的充分统计量。

以两期模型为例，即 $T = 2$ 。对于个体 i ，只有三种可能，即 $n_i = y_{i1} + y_{i2} = 0, 1$ ，或 2 。

$$(1) \quad n_i = y_{i1} + y_{i2} = 0。$$

必然有 $y_{i1} = y_{i2} = 0$ ，即 $P(y_{i1} = 0, y_{i2} = 0 | n_i = 0) = 1$ ，其对数似然函数为 $\ln 1 = 0$ ，对整个样本的对数似然函数没有贡献。

由于此条件似然函数为常数，故此观测值不包含可用于估计 β 的信息，等于损失了此观测值。

$$(2) \quad n_i = y_{i1} + y_{i2} = 2。$$

必然有 $y_{i1} = y_{i2} = 1$ ，即 $P(y_{i1} = 1, y_{i2} = 1 | n_i = 2) = 1$ 。这些观测值不包含任何有助于估计 $\boldsymbol{\beta}$ 的信息，也可忽略。

(3) $n_i = y_{i1} + y_{i2} = 1$ 。或者 $(y_{i1} = 0, y_{i2} = 1)$ ，或者 $(y_{i1} = 1, y_{i2} = 0)$ 。分别计算其条件概率如下。

$$P(y_{i1} = 0, y_{i2} = 1 | n_i = 1) = \frac{P(y_{i1} = 0, y_{i2} = 1)}{P(y_{i1} = 0, y_{i2} = 1) + P(y_{i1} = 1, y_{i2} = 0)}$$

$$P(y_{i1} = 1, y_{i2} = 0 | n_i = 1) = \frac{P(y_{i1} = 1, y_{i2} = 0)}{P(y_{i1} = 0, y_{i2} = 1) + P(y_{i1} = 1, y_{i2} = 0)}$$

假设在给定 u_i 与 \mathbf{x}_{i1} 的条件下， y_{i1} 与 y_{i2} 相互独立，则

$$P(y_{i1} = 0, y_{i2} = 1) = \frac{1}{1 + e^{u_i + \mathbf{x}'_{i1}\boldsymbol{\beta}}} \cdot \frac{e^{u_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}{1 + e^{u_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}$$

$$P(y_{i1} = 1, y_{i2} = 0) = \frac{e^{u_i + \mathbf{x}'_{i1}\boldsymbol{\beta}}}{1 + e^{u_i + \mathbf{x}'_{i1}\boldsymbol{\beta}}} \cdot \frac{1}{1 + e^{u_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}$$

代入方程可得，

$$\begin{aligned} P(y_{i1} = 0, y_{i2} = 1 | n_i = 1) &= \frac{e^{u_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}{e^{u_i + \mathbf{x}'_{i1}\boldsymbol{\beta}} + e^{u_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}} \\ &= \frac{e^{\mathbf{x}'_{i2}\boldsymbol{\beta}}}{e^{\mathbf{x}'_{i1}\boldsymbol{\beta}} + e^{\mathbf{x}'_{i2}\boldsymbol{\beta}}} = \frac{e^{(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \boldsymbol{\beta}}}{1 + e^{(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \boldsymbol{\beta}}} = \Lambda[(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \boldsymbol{\beta}] \end{aligned}$$

上式分子与分母的 e^{u_i} 同时消去。同理，

$$P(y_{i1} = 1, y_{i2} = 0 | n_i = 1) = \Lambda[-(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \boldsymbol{\beta}] = 1 - \Lambda[(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \boldsymbol{\beta}]$$

定义虚拟变量 $d_i = 1$ ，如果 $(y_{i1} = 0, y_{i2} = 1)$ ； $d_i = 0$ ，如果 $(y_{i1} = 1, y_{i2} = 0)$ 。

个体 i 的条件对数似然函数为：

$$\ln L_i(\boldsymbol{\beta}) = \{d_i \ln \Lambda[(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \boldsymbol{\beta}] + (1 - d_i) \ln \{1 - \Lambda[(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \boldsymbol{\beta}]\}\} \cdot \mathbf{1}(n_i = 1)$$

小结:

- (1) 给定 n_i 的条件似然函数不再依赖于 u_i (已在推导中消去);
- (2) 此条件似然函数仍为Logit模型, 只是解释变量变为 $(\mathbf{x}_{i2} - \mathbf{x}_{i1})$;
- (3) 不随时间变化的变量将无法识别其系数, 因为 $(x_{i2} - x_{i1}) = 0$ 。
固定效应的似然函数不含积分, 无须进行数值积分。

对于 $T > 2$, 可计算给定 $n_i = 1, n_i = 2, \dots$, 或 $n_i = T - 1$ 的条件似然函数。

固定效应模型的缺点是, 将损失所有 $n_i = 0$ 或 $n_i = T$ 的观测值。

而且，无法估计个体效应 u_i （已被消去），故无法预测 y_i 的发生概率或解释变量的边际效应。

解决方法之一是假设 $u_i = 0$ 。

如何在固定效应模型与混合回归之间选择？可使用豪斯曼检验。

原假设为不存在个体效应，即“ $H_0 : u_i = u$ ”。

如果原假设成立，则固定效应模型与混合回归都一致。

反之，如果原假设不成立，则固定效应一致，而混合回归不一致。

如果二者的系数估计值相差较大(以二次型来衡量), 则倾向于拒绝原假设, 接受存在个体效应的替代假设。

对于固定效应与随机效应之间的选择, 也可进行豪斯曼检验。

17.4 面板二值选择模型的 **Stata** 实例

17.5 面板泊松回归

考虑被解释变量为计数变量的面板数据。

例：若干企业在一段时间内每年获得专利的个数

例：全国各省在几年内每年发生矿难的数目

例：世界各国近几十年每年发生示威游行的次数

例：一些病人在一段时间内的每期发病次数

对于个体 i ，时期 t ，记被解释变量为 Y_{it} ，假设 $Y_{it} = y_{it}$ 的概率由参数为 λ_{it} 的泊松分布所决定：

$$P(Y_{it} = y_{it} \mid \mathbf{x}_{it}) = \frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!} \quad (y_{it} = 0, 1, 2, \dots)$$

$\lambda_{it} > 0$ 为泊松到达率。为保证 λ_{it} 非负，假设

$$\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i) = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}) \cdot \exp(u_i) \equiv v_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$$

\mathbf{x}_{it} 不含常数项，而 $v_i \equiv \exp(u_i)$ 为乘积形式的个体效应 (multiplicative individual effects)。

如果 $v_1 = v_2 = \dots = v_n$ ，不存在个体效应，为混合泊松回归(pooled

Poisson), 可作截面数据处理, 须使用聚类稳健标准误。

一般地, 允许个体效应的存在, 即不同个体拥有不同的 v_i 。

如果 v_i 与所有解释变量 \mathbf{x}_{it} 均不相关, 则为“随机效应模型”(Random Effects Model, 简记 RE)。

如果 v_i 与某个解释变量相关, 则为“固定效应模型”(Fixed Effects Model, 简记 FE)。

首先考虑随机效应模型, 进行 MLE 估计。

记 v_i 的密度函数为 $g(v_i)$ 。

假设样本为 iid，在给定 v_i 情况下，个体 i 的条件分布为，

$$f(y_{i1}, y_{i2}, \dots, y_{iT} | v_i) = \prod_{t=1}^T \left(\frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!} \right) = \prod_{t=1}^T \left(\frac{e^{-v_i \exp(\mathbf{x}'_{it} \boldsymbol{\beta})} [v_i \exp(\mathbf{x}'_{it} \boldsymbol{\beta})]^{y_{it}}}{y_{it}!} \right)$$

但 v_i 不可观测。记 $(y_{i1}, y_{i2}, \dots, y_{iT}, v_i)$ 的联合密度为 $f(y_{i1}, y_{i2}, \dots, y_{iT}, v_i)$ ，并进行如下分解，

$$f(y_{i1}, y_{i2}, \dots, y_{iT}, v_i) = f(y_{i1}, y_{i2}, \dots, y_{iT} | v_i) \cdot g(v_i)$$

将 v_i 积分积掉，可得到 $(y_{i1}, y_{i2}, \dots, y_{iT})$ 的边缘密度，

$$\begin{aligned} f(y_{i1}, y_{i2}, \dots, y_{iT}) &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT}, v_i) dv_i \\ &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT} | v_i) \cdot g(v_i) dv_i \end{aligned}$$

由于 $v_i = \exp(u_i) > 0$ ，常选择 v_i 服从 Gamma 分布(指数分布与卡方分布为 Gamma 分布的特例)。

假设 $v_i \sim \text{Gamma}(1/\alpha, \alpha)$ ， $\alpha > 0$ ，则 $E(v_i) = 1$ ， $\text{Var}(v_i) = \alpha$ 。

将 $\text{Gamma}(1/\alpha, \alpha)$ 的概率密度代入，可得负二项分布的概率密度，然后进行 MLE 估计。

参数 α 为个体异质性 v_i 的方差，故可对 “ $H_0: \alpha = 0$ ” (Stata 称

为“ $\alpha = 0$ ”)进行 LR 检验来判断是否存在个体异质性，即究竟应使用随机效应泊松回归还是混合泊松回归。

另一方法是，假设 $u_i \sim N(0, \sigma_u^2)$ ，但无解析解，一般用“Gauss-Hermite quadrature”方法进行数值积分；也需用 Stata 命令 `quadchk` 检验数值积分的精度。

可检验“ $H_0: \sigma_u = 0$ ” (Stata 称为“`sigma_u = 0`”)判断是否存在个体异质性。

固定效应的面板泊松模型：

可使用 $n_i \equiv \sum_{t=1}^T y_{it}$ 作为 v_i 的充分统计量，计算在给定 n_i 情况下的条件似然函数，然后进行条件 MLE 估计。

如果 $n_i = 0$ ，则 $y_{i1} = \dots = y_{iT} = 0$ ，个体 i 的观测值对于条件似然函数的贡献为 0，此观测值将被去掉，损失样本容量。

另一缺点是，无法估计不随时间而变的变量系数。

在固定效应与随机效应泊松回归之间选择时，可用豪斯曼检验。

17.6 面板负二项回归

泊松回归假设均等分散，即方差等于期望。如果存在过度分散，即方差大于期望，可考虑负二项回归。

如果不存在个体异质性，可进行混合负二项回归。

对于随机效应的面板负二项模型，假设个体异质性服从 $\text{Beta}(r, s)$ 分布（取值范围为 $[0, 1]$ ），则可将个体异质性积分掉，然后进行 MLE 估计。

对于固定效应的面板负二项模型，可考虑在给定 $n_i \equiv \sum_{t=1}^T y_{it}$ (个体异质性的充分统计量) 情况下的条件似然函数，然后进行条件

MLE 估计。

固定效应负二项回归的奇特之处在于，它可以估计不随时间而变的变量系数。

在固定效应与随机效应负二项回归之间选择，可用豪斯曼检验。

如何在泊松回归与负二项回归之间进行选择，这涉及到在稳健性与有效性之间的权衡(前者稳健，而后者更有效)。

17.7 面板计数模型的 **Stata** 实例

17.8 面板 Tobit

考虑归并数据(censored data)的面板模型。假设

$$y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + \varepsilon_{it}$$

y_i^* 不可观测，扰动项 $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ ， u_i 为个体效应。

假定在 0 处存在左归并(left-censored)。假设可以观测到

$$y_{it} = \begin{cases} y_{it}^*, & \text{若 } y_{it}^* > 0 \\ 0, & \text{若 } y_{it}^* \leq 0 \end{cases}$$

如果 $u_1 = u_2 = \cdots = u_n$ ，则可直接进行混合 Tobit 回归，但须使用聚类稳健标准误。

更一般地，允许个体效应的存在。

如果 u_i 与解释变量 \mathbf{x}_{it} 不相关，则为随机效应模型(RE)；反之，则为固定效应模型(FE)。

对于固定效应的 Tobit 模型，由于找不到个体异质性 u_i 的充分统计量，故无法进行条件 MLE 估计。如果直接在混合 Tobit 回归中加入个体虚拟变量(LSDV 法)，也不一致。

考虑随机效应的 Tobit 模型。在给定个体效应 u_i 的情况下，个体 i 的条件分布为：

$$\begin{aligned}
& f(y_{i1}, y_{i2}, \dots, y_{iT} \mid u_i) \\
&= \prod_{t=1}^T \left[1 - \Phi((\mathbf{x}'_{it} \boldsymbol{\beta} + u_i) / \sigma_\varepsilon) \right]^{I(y_{it}=0)} \left[\frac{1}{\sigma_\varepsilon} \phi((y_{it} - \mathbf{x}'_{it} \boldsymbol{\beta} - u_i) / \sigma_\varepsilon) \right]^{I(y_{it}>0)}
\end{aligned}$$

推导方法与第 14 章第 5 节类似。

但 u_i 不可观测。假设 $u_i \sim N(0, \sigma_u^2)$ ，记其概率密度函数为 $g(u_i)$ 。

记 $(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$ 的联合密度为 $f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i)$ ，进行分解：

$$f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) = f(y_{i1}, y_{i2}, \dots, y_{iT} \mid u_i) \cdot g(u_i)$$

将 u_i 积分掉，可得 $(y_{i1}, y_{i2}, \dots, y_{iT})$ 的边缘密度，

$$\begin{aligned} f(y_{i1}, y_{i2}, \dots, y_{iT}) &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT}, u_i) du_i \\ &= \int_{-\infty}^{+\infty} f(y_{i1}, y_{i2}, \dots, y_{iT} | u_i) \cdot g(u_i) du_i \end{aligned}$$

上式积分无解析解，一般使用“Gauss-Hermite quadrature”方法进行数值积分；也需用 Stata 命令 quadchk 检验此积分的精度。

可通过检验“ $H_0: \sigma_u = 0$ ” (Stata 称为“sigma_u = 0”)判断是否存在个体异质性。

定义同一个体不同期扰动项的自相关系数为，

$$\rho \equiv \text{Corr}(u_i + \varepsilon_{it}, u_i + \varepsilon_{is}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} \quad (t \neq s)$$

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。

如果 $\rho = 0$ ，则 $\sigma_u^2 = 0$ ，不存在个体随机效应，应选择混合回归。

17.9 面板随机前沿模型（选读）