

泛函差分

嘉树

2026 年 1 月

说明：本文是对 [Bonhomme \(2012\)](#) 的介绍，讲述了怎么在非线性面板模型中构造不含有固定效应的矩条件，类比于线性面板中的「组内变换」。本质上这个构造过程是希尔伯特投影定理的一个应用。阅读本文需要掌握面板模型以及一些泛函分析的基本知识。本文并不是对原文的忠实还原，如有疑问请参考原文。

1 背景介绍：短面板下的次要参数问题

在非线性模型中，「次要参数问题」(incidental parameter problem, IPP) 是一个由来已久的问题，它指的是，当次要参数的数量会随着样本量增加而增加时，使用极大似然估计「共同参数」得到的估计量一般不是一致的，产生的偏误被称为「次要参数偏误」。在非线性面板模型中，人们常常会加入固定效应，而这些固定效应就是次要参数，当个体数 N 比较大而时期数 T 比较小 (短面板)，如用极大似然去估计共同参数，很有可能是不一致的。这个问题必须被认真对待。

人们希望能绕开对次要参数的估计，这在某些特殊的模型中是可能的，譬如静态 logit 模型，通过条件于次要参数的充分统计量，可以得到不含有次要参数的条件似然函数，从而得到共同参数的一致估计量。然而，对于一般的非线性模型，我们很难找到次要参数的充分统计量。统计学家们的处理方式是，将似然函数分解为两个部分，一部分只包含共同参数，另一部分只包含次要参数，于是对两部分参数的估计是独立的。但这种分解很多时候也难以得到^①。

另一种路径是，对次要参数 (固定效应) 的分布做出一些假设，然后就可以对条件于固定效应和解释变量的似然函数做积分 (关于固定效应的积分)，得到仅仅条件于解释变量的似然函数，也就避开了对次要参数的估计^②。这种方法因为其复杂性并未被人们广泛采用。

我们希望能找到一种完全剔除了次要参数的估计方法，就像在线性面板模型中使用的「组内变换」(within transformation) 一样。这种方法要足够普遍，能适用于任何非线性模型，且无需对固定效应的分布做出任何假设。

2 模型设定

设可观测的数据是内生变量 y_{it} 和严格外生变量的向量 x_{it} , $i = 1, \dots, N$, $t = 1, \dots, T$ 。令 α_i 表示个体 i 特定的不可观测的固定效应的向量。记 $y_i = (y_{i1}, \dots, y_{iT})'$, $x_i = (x'_{i1}, \dots, x'_{iT})$ 。假设 y_i, x_i, α_i 这些变量在不同的个体之间是联合独立且同分布的 (i.i.d.)。我们考虑「大 N 小 T 」的情况，即 $N \rightarrow \infty$ 而 T 固定。

① 当精确分离不可得时，[Cox and Reid \(1987\)](#) 提出了近似条件似然方法。不过，这样得到的估计量还是有渐进误差的，阶数为 $O(1/T^2)$ ，对比之下，直接用极大似然估计得到的估计量收敛率为 $O(1/T)$ ；尽管偏误更小，对于固定的 T ，这个改进可能并不会让人满意。计量中，[Dhaene and Jochmans \(2015\)](#) 开发的「面板拆分版刀切法」(split panel jackknife) 也是广受好评的处理 IPP 的方法，但严格来讲它也只能降低偏误的阶数，不能得到固定 T 下的一致估计量。

② 所谓的「相关随机效应」(correlated random effects) 就是这种思路，将固定效应建模为解释变量的一个参数形式，它介于固定效应假设和随机效应假设之间。如果我们不愿意对固定效应做出太强的假设，那么还有一些半参数的方法可以考虑，譬如使用筛法去逼近固定效应的分布。

假设我们已知 y_i 条件于 x_i, α_i 的分布, 记作 $f_{y|x,\alpha}(\cdot|x_i, \alpha_i; \theta_0)$ 。 $f_{y|x,\alpha}$ 的函数形式已知, 但参数 $\theta_0 \in \Theta$ 未知。记 α_i 条件于 x_i 的分布为 $f_{\alpha|x}(\cdot|x_i)$ 。令 \mathcal{A} 表示所有可能的 α_i 的集合。那么, 排除了 α_i 的影响后, y_i 条件于 x_i 的分布可以写成❸

$$f_{y|x}(\cdot|x; \theta_0, f_{\alpha|x}) = \int_{\mathcal{A}} f_{y|x,\alpha}(\cdot|x, \alpha; \theta_0) f_{\alpha|x}(\alpha|x) d\alpha \quad (1)$$

我们的目标是, 对 (1) 给出的一般模型, 得到排除了 α_i 而含有共同参数 θ_0 的矩条件, 然后能使用 GMM 得到共同参数的一致估计量。其一致性来源于我们免于估计次要参数而带来的偏误且有正确的矩条件。

3 矩条件的构造

为描述方法, 我们引入一个记号。考虑一个线性泛函 $L_{\theta,x}(\cdot)$, 对于给定的 θ 和 x , 它将关于 α 的 (任意) 函数 $g(\alpha)$ 映射为关于 y 的函数 $[L_{\theta,x}g](y)$:

$$[L_{\theta,x}g](y) = \int_{\mathcal{A}} f_{y|x,\alpha}(y|x, \alpha; \theta) g(\alpha) d\alpha$$

于是 (1) 可以写成

$$f_{y|x}(\cdot|x; \theta_0, f_{\alpha|x}) = L_{\theta_0,x} f_{\alpha|x}(\cdot|x) \quad (2)$$

线性算子 $L_{\theta_0,x}$ 将固定效应的分布映射为响应变量的分布。

假设我们找到了一个函数 $\varphi(\cdot, x, \theta)$, 它满足对任意 $g(\alpha)$ 都有

$$\int_{\mathcal{Y}} \varphi(y, x, \theta) [L_{\theta,x}g](y) dy = 0 \quad (3)$$

积分区域 \mathcal{Y} 是响应变量 y 的取值范围。这个式子表明, φ 和 $L_{\theta,x}$ 的整个值域是正交的 (在希尔伯特空间的内积意义下)。因为 (3) 对任意 $g(\alpha)$ 都成立, 所以当取 $g(\alpha) = f_{\alpha|x}(\cdot|x)$ 时, 就有

$$\begin{aligned} \mathbb{E}[\varphi(y_i, x_i, \theta_0)|x_i = x] &= \int_{\mathcal{Y}} \varphi(y, x, \theta_0) f_{y|x}(y|x; \theta_0, f_{\alpha|x}) dy \\ &= \int_{\mathcal{Y}} \varphi(y, x, \theta_0) [L_{\theta_0,x} f_{\alpha|x}](y) dy = 0 \end{aligned} \quad (4)$$

于是我们就得到了一个不含有固定效应的 (条件) 矩约束, 我们可以认为 φ 的作用就类似于线性模型中「差分掉」固定效应的影响。现在的问题是, 我们如何找到这样的 φ ?

3.1 离散情况

作为示范, 我们先考虑离散的情况。假设 y_i 和 α_i 都服从离散分布, 且支撑集是有限集。设 N_y 和 N_α 分别是 y_i 和 α_i 的支撑集中元素的个数。那么 (2) 就变成了一个线性方程组, 其中 $f_{y|x}$, $f_{\alpha|x}$ 和 $L_{\theta_0,x}$ 都会变成条件概率:

$$f_{y|x} = \begin{bmatrix} \Pr(y_i = \underline{y}_1|x_i = x) \\ \vdots \\ \Pr(y_i = \underline{y}_{N_y}|x_i = x) \end{bmatrix}_{N_y \times 1} \quad f_{\alpha|x} = \begin{bmatrix} \Pr(\alpha_i = \underline{\alpha}_1|x_i = x) \\ \vdots \\ \Pr(\alpha_i = \underline{\alpha}_{N_\alpha}|x_i = x) \end{bmatrix}_{N_\alpha \times 1}$$

❸ 因为我们并未对 α_i 的分布 $f_{\alpha|x}$ 做出任何假设, 这个模型是半参数化的。特别地, α_i 可以和 x_i 有任意相关性, 它们自己也可以有任意分布, 这和经典的线性面板模型所青睐的固定效应假设一致。

$$L_{\theta_0, x} = \begin{bmatrix} \Pr(y_i = \underline{y}_1 | x_i = x, \alpha_i = \underline{\alpha}_1; \theta) & \cdots & \Pr(y_i = \underline{y}_1 | x_i = x, \alpha_i = \underline{\alpha}_{N_\alpha}; \theta) \\ \vdots & \ddots & \vdots \\ \Pr(y_i = \underline{y}_{N_y} | x_i = x, \alpha_i = \underline{\alpha}_1; \theta) & \cdots & \Pr(y_i = \underline{y}_{N_y} | x_i = x, \alpha_i = \underline{\alpha}_{N_\alpha}; \theta) \end{bmatrix}_{N_y \times N_\alpha}$$

线性算子 $L_{\theta_0, x}$ 此时是一个矩阵，它的值域是它的列向量所张成的 N_y 维向量空间。为了构造和这个向量空间正交的向量 $\varphi \in \mathbb{R}^{N_y}$ ，使用「组内」投影矩阵

$$W_{\theta, x} = I_{N_y} - L_{\theta, x} L_{\theta, x}^\dagger$$

其中 I_{N_y} 是 N_y 维单位矩阵， $L_{\theta, x}^\dagger$ 是 $L_{\theta, x}$ 的伪逆。

$L_{\theta, x}$ 的列向量张成的向量空间是 $L_{\theta, x}$ 的值域在 \mathbb{R}^{N_y} 中的正交补空间^④，即对任意 $h \in \mathbb{R}^{N_y}$ ，都有 $W_{\theta, x} h \in \mathbb{R}^{N_y}$ 正交于 $L_{\theta, x}$ 的值域。 $[W_{\theta, x} h](y)$ 可视作定义在支撑集 $\{\underline{y}_1, \dots, \underline{y}_{N_y}\}$ 上的函数，当 $y = \underline{y}_j$ 时， $[W_{\theta, x} h](y)$ 取 $W_{\theta, x} h$ 的第 j 个分量。于是 (4) 就变成了

$$\mathbb{E}([W_{\theta, x} h](y_i) | x_i = x) = (W_{\theta, x} h)'(L_{\theta, x} f_{\alpha|x}) = 0$$

这个式子对任意 $h \in \mathbb{R}^{N_y}$ 都成立，我们可以选择 \mathbb{R}^{N_y} 的一组标准正交基来构造出 N_y 个矩条件^⑤，然后用 GMM 估计 θ_0 。

本质上，这个方法就是找到一个线性算子 $W_{\theta, x}$ ，当对 $f_{y|x} = L_{\theta_0, x} f_{\alpha|x}$ 两边左乘 $W_{\theta, x}$ 时，便得到

$$W_{\theta, x} f_{y|x} = W_{\theta, x} L_{\theta_0, x} f_{\alpha|x} = 0$$

从而消除掉任何关于 α_i 的影响。因为我们不知道任何关于函数 $f_{\alpha|x}$ 的信息，所以 $W_{\theta, x}$ 必须完全正交于 $L_{\theta, x}$ 的值域。

3.2 一般情况

当 y_i 或 α_i 是连续分布或支撑集是无限集时，上一小节的方法不再适用。此时， $L_{\theta, x}$ 是一个积分算子，它的值域是定义在 \mathcal{Y} 上的所有平方可积函数所张成的希尔伯特空间 $L^2(\mathcal{Y})$ 。

我们引入一些记号。给定两个权重函数 $\pi_\alpha > 0$ 和 $\pi_y > 0$ ，分别定义在 \mathcal{A} 和 \mathcal{Y} 上，定义两个 \mathcal{A} 和 \mathcal{Y} 上的平方可积函数所构成的空间：

$$\begin{aligned} \mathcal{G}_\alpha &= \left\{ g: \mathcal{A} \rightarrow \mathbb{R}, \int_{\mathcal{A}} g(\alpha)^2 \pi_\alpha(\alpha) d\alpha < \infty \right\} \\ \mathcal{G}_y &= \left\{ g: \mathcal{Y} \rightarrow \mathbb{R}, \int_{\mathcal{Y}} g(y)^2 \pi_y(y) dy < \infty \right\} \end{aligned}$$

它们都是「希尔伯特空间」(Hilbert space)，即完备的内积空间，内积定义为

$$\begin{aligned} \langle g_1, g_2 \rangle_\alpha &= \int_{\mathcal{A}} g_1(\alpha) g_2(\alpha) \pi_\alpha(\alpha) d\alpha \\ \langle g_1, g_2 \rangle_y &= \int_{\mathcal{Y}} g_1(y) g_2(y) \pi_y(y) dy \end{aligned}$$

由此可以定义范数 $\|g\|_\alpha = \sqrt{\langle g, g \rangle_\alpha}$ 和 $\|g\|_y = \sqrt{\langle g, g \rangle_y}$ 。

于是线性积分算子 $L_{\theta, x}: \mathcal{G}_\alpha \rightarrow \mathcal{G}_y$ 的作用是将 \mathcal{G}_α 中的函数映射为 \mathcal{G}_y 中的函数。记 $L_{\theta, x}$ 的值域为 $\mathcal{R}(L_{\theta, x}) \subseteq \mathcal{G}_y$ ：

$$\mathcal{R}(L_{\theta, x}) = \{L_{\theta, x} g: g \in \mathcal{G}_\alpha\}$$

④ 要保证这个正交补空间非空，需要 $L_{\theta, x}$ 的值域是 \mathbb{R}^{N_y} 的严格子空间。这是这个方法的一个必要识别条件，否则无法给出有意义的矩条件。

⑤ 这 N_y 个矩条件或许有些是冗余的，但只要有效矩条件的数量不小于 θ 的维度，这个模型就是可识别的。

令 $\bar{\mathcal{R}}(L_{\theta,x})$ 是 $\mathcal{R}(L_{\theta,x})$ 在 \mathcal{G}_y 中的闭包。由此可以定义投影算子 $W_{\theta,x}$ ：对任意 $h \in \mathcal{G}_y$ 有

$$W_{\theta,x}h = h - \text{Proj}_{\pi_y}[h|\bar{\mathcal{R}}(L_{\theta,x})]$$

其中 $\text{Proj}_{\pi_y}[h|\bar{\mathcal{R}}(L_{\theta,x})]$ 是 h 在 $\bar{\mathcal{R}}(L_{\theta,x})$ 上的正交投影：

$$\text{Proj}_{\pi_y}[h|\bar{\mathcal{R}}(L_{\theta,x})] = \underset{g \in \bar{\mathcal{R}}(L_{\theta,x})}{\operatorname{argmin}} \|h - g\|_y$$

希尔伯特投影定理将保证正交投影的存在性和唯一性。

定理 1. 设 $f_{y|x} \in \mathcal{G}_y$ 。以下两个结论成立且等价：

$$\begin{aligned} W_{\theta_0,x}f_{y|x} &= 0 \\ \mathbb{E}(\pi_y(y_i)[W_{\theta_0,x}h](y_i)|x=x_i) &= 0 \quad \forall h \in \mathcal{G}_y \end{aligned} \quad (5)$$

这个定理从理论上给出了关于 θ_0 的条件矩约束。我们可以据 (5) 构造出无条件矩，然后使用 GMM 对 θ_0 进行估计。

至此，我们完成了对「泛函差分」基本思想的介绍，它的内核本质上就是找正交投影算子，而这通过希尔伯特投影定理得到保证。不过其实有一件很重要的点我们略过了，就是定义了 \mathcal{G}_α 和 \mathcal{G}_y 的权重函数 π_α 和 π_y （以及 h ）该如何选择的问题。尽管从理论上，任意的 π_α 和 π_y 都能构造出投影算子 $W_{\theta,x}$ ，但不同的选择会对矩条件的估计效率产生影响。要达到最好的估计效率是一件很困难的事，原文第 4 节做了一些讨论，这里不展开介绍。

3.3 投影算子的离散近似

如果 $f_{y|x}$ 和 $f_{\alpha|x}$ 是离散分布（且支撑集是有限集），那么我们可以直接使用第 3.1 小节中的方法来构造投影矩阵 $W_{\theta,x}$ 。但当 $f_{y|x}$ 和 $f_{\alpha|x}$ 是连续分布时，投影算子一般没有显式表达式，此时，我们可以进行离散化近似处理。这里进行简单介绍。

把「无限维投影」近似成「有限维投影」，近似来自两个方面：其一，希尔伯特空间中内积是连续积分；其二， $\bar{\mathcal{R}}(L_{\theta,x})$ 是无限维子空间。离散化近似同时把这两个「无限」替换为有限维对象。

用抽样近似 \mathcal{G}_y 内积（把积分变求和）

在 $\mathcal{G}_y = L^2(\pi_y)$ 中，

$$\langle g_1, g_2 \rangle_y = \int_{\mathcal{Y}} g_1(y)g_2(y) \pi_y(y) dy \quad \|u\|_y^2 = \int_{\mathcal{Y}} u(y)^2 \pi_y(y) dy$$

抽取 N_y 个样本点 $y_s \sim \pi_y$ ($s = 1, \dots, N_y$)^⑥，取近似

$$\|h - g\|_y^2 = \int_{\mathcal{Y}} (h(y) - g(y))^2 \pi_y(y) dy \approx \frac{1}{N_y} \sum_{s=1}^{N_y} (h(y_s) - g(y_s))^2$$

即「 \mathcal{G}_y 意义下的投影」可用「在采样点 $\{y_s\}$ 上的平方损失最小化」来近似。

⑥ 这里假设 π_y 是一个密度函数，但也可以推广到一般函数。

用有限维空间近似值域 $\bar{\mathcal{R}}(L_{\theta,x})$

第二步是用有限个可计算的函数去近似 $\bar{\mathcal{R}}(L_{\theta,x})$ 。使用由模型 $f_{y|x,\alpha}$ 诱导的「自然基」(natural bases):

$$\mu_s(y) = \int_{\mathcal{A}} \frac{1}{\pi_{\alpha}(\alpha)} f_{y|x,\alpha}(y|x, \alpha; \theta) f_{y|x,\alpha}(y_s|x, \alpha; \theta) d\alpha \quad s = 1, \dots, N_y \quad (6)$$

我们用 μ_1, \dots, μ_{N_y} 张成的有限维空间 $\text{span}\{\mu_1, \dots, \mu_{N_y}\}$ 去近似 $\bar{\mathcal{R}}(L_{\theta,x})$ 。若直接积分困难，可以用数值积分方法（如重要性采样）近似 (6)^⑦。

当离散采样点数越来越大时，近似误差会趋于 0，由此得到的 GMM 估计量的一致性和渐进正态性不会受到影响。

4 总结

本文介绍了「泛函差分」方法的基本思想。它的内核是通过正交投影算子完全去掉固定效应的影响，得到不含有固定效应的矩条件。这个方法相当普适，因为它适用于任何非线性模型，且无需对次要参数的分布做出任何假设。

原文还对该方法的识别条件进行了详细讨论^⑧，并通过数值模拟验证了方法的有效性，感兴趣的读者可以阅读原文。

⑦ 这实际上又对 α 做了一次离散化。如此，线性算子 $L_{\theta,x}$ 的离散近似就成了一个矩阵，回到了第 3.1 小节的离散情况。

⑧ 主要论点是让线性算子 $L_{\theta,x}$ 的值域是 \mathcal{G}_y 的严格子空间，从而保证正交补空间非空，即通过投影算子能得到有意义的矩条件。

参考文献

Bonhomme, S. (2012): “Functional Differencing,” *Econometrica*, 80(4), 1337–1385.

Cox, D. R. and N. Reid (1987): “Parameter Orthogonality and Approximate Conditional Inference,” *Journal of the Royal Statistical Society (Series B)*, 49(1), 1–18.

Dhaene, G. and K. Jochmans (2015): “Split-Panel Jackknife Estimation of Fixed-Effect Models,” *Review of Economic Studies*, 82(3), 991–1030.