

离散选择模型下需求函数的估计

本文是IO中需求估计的一派方法的起点，文中除了介绍BLP的计量结构和算法，也强调了改写模型的意义以及其与之前的经典模型（McFadden的 Multinomial Logit以及Nested Logit）的对比，强调BLP引入的随机系数的模型是如何优于MNL以及Nested Logit。其可以被认为是Berry 1994的一种特例，但由于给出了数值积分以及压缩映射的算法，使具体的算法可以得到应用，因此更为著名。BLP一个另外的贡献是也刻画了供给端，并指出联合估计需求和供给往往得到更高的有效性。

需求端的估计

想说清楚BLP的优势，必须解释为什么之前的Logit有问题。

Multinomial Logit

在经典的Multinomial Logit模型中，借助于第一类极值分布，市场份额可以被表示为（在以下所有的数学表达中我们都不单独强调价格，而将其当作x的一部分）

$$s_{ij} \equiv P \left(j \in \arg \max_{k \in C} u_{ik} \right) = \frac{\exp(x'_j \beta)}{\sum_{k \in C} \exp(x'_k \beta)} \quad (1)$$

- 该表达假设了没有外部选择，即数据中的所有消费者都选择了J个产品中的有且只有一个。
- 价格可能有内生性：BLP的模型中（最早的应该是Berry 94）引入了 ξ 代表研究者无法观测但消费者可以观测到的因素，如果这一部分存在，价格就存在内生性问题，在BLP给的实证例子中，价格效应的系数甚至有可能是正的。一个更通俗的解释在于，存在一些不在X中的变量系统性的影响着人们的偏好，例如口碑和售后服务等无法被X捕捉的因素。
- 如果X捕捉的维度过多，这会影响估计的有效性。为了提升有效性，往往可以将不关心的几个维度合成为一个维度，这样只需要估计这一个系数，会使得估计的准确性提高。BLP中引入 δ 的本意也是如此。
- IIA问题，即这样的需求下计算的交叉价格弹性只与商品的价格和市场份额有关，与其他的商品无关，BLP给的一个例子是，如果劳斯莱斯和奇瑞QQ的市场份额相同，那么他们对宝马的交叉价格弹性相同，但显然宝马和劳斯莱斯都属于豪华车，他们之间的替代效应应该更多一些。

Multinomial 的改进I: Nested Logit

显然，如何引入替代效应和如何解决价格内生性是最本质的问题。

为了解决IIA的问题（即尝试引入某种替代效应），之后一系列Nested Logit的文献认为可以对误差项进行假设，即出现了类似下面的表述，其中 g 是一个人为划分的种类，这样的误差项显然使得一个消费者 i 对同一种类下的不同品牌的 ϵ 相关，从而构造出了同一种类下的替代效应。

$$\varepsilon_{ij} = (1 - \lambda_g)\nu_{ig} + \lambda_g e_{ij} \quad (2)$$

估计的算法也会相应的复杂，需要MLE以及Loop Search。但Nested Logit的问题在于分类引入了人为的干预，且当Nest的结构比较复杂时，Nest的顺序会显著影响结果（以汽车为例，先按豪华-平价分，再按运动-舒适分，还是反之）。

Multinomial 的改进II：Unobserved Quality

而为了解决价格内生的问题，Berry 1994最早提出可以加入不可被观测的质量 ξ 来解决，也是在这篇论文中，Berry指出 ξ 要解决这两个问题的核心假设：

- 需要有外部选择及其份额（多少人没有买）
- 需要many markets（市场数量足够多，则可以以每一个市场为观测，做OLS）

Berry使用 δ 来代替产品的效用，这样我们不再关心如何估计一个 β 的向量，减少了估计的参数。Berry讨论了三种情况，简单的logit或者nested logit的情况，或者更一般的函数假设。

在简单的logit假设下（即随机项iid的服从第一类极值分布）仍然类似上面的市场份额，此时因为有外部选择，可以写为

$$s_{jt} = \frac{\exp(\delta_{jt})}{1 + \sum_k \exp(\delta_{kt})} \quad (3)$$

从而使用所谓的Berry Reversion得到 $\delta_{jt} = \ln s_{jt} - \ln s_{0t}$ 。又由定义，

$$\delta_{jt} = x_{jt}\beta + \xi_{jt}. \quad (4)$$

所以可以对 JT 维度（每个产品每个市场下）进行OLS或使用panel。它的nested logit版本在只有一层nest的情况下（即只需要额外多估计 λ ）的算法可以认为是每固定一个 λ ，仍然可以显示解出市场份额，然后用simple logit的办法做一次估计，然后得到误差，最后搜索 λ 。最后如果 ϵ 的分布是任意形式，那么求解市场份额（3）就会缺乏显示解，可能需要数值积分求解。求解的目的是得到 δ 的估计，然后使用GMM估计 β （使用类似 $\mathbb{E}[(\delta - x'_t\beta)|z_t] = 0$ 的矩条件）。

值得注意的是Berry 1994 中没有给出显示解的情况（3），以及直接使用Nested Logit的情况（2）可以理解为将Nest的思路与不可观测的质量的结合，只不过（2）用的是ad-hoc的分类模型（因此继承了Nested Logit的所有缺点，例如依赖Nest顺序），而（3）直接假设随即项满足任何分布（因此自然允许不同产品之间相关，也自然的包含了（2），但是因为过于一般，相当于对 J 个产品的随机项的相关系数矩阵进行估计，参数个数太多）。

BLP 1995

随机系数需求是一系列尝试沿用Nested Logit的思路解决IIA的模型，本质是想要用更容易理解的方式引入随机项之间的相关性（而不是用Berry 1994的情况（3）中一个待估计的ad hoc的相关系数矩阵）。BLP本质上是Berry 1994的一种特例，虽然最初的形式上看着并不像，这也正是我认为他包装很好的地方。其假设效用函数是柯布道格拉斯型，这使得 $x_{jt}\beta_i$ 中包含收入和价格的一项可能是 $\alpha_i \ln(y_i - p_{jt})$ 而不是线性项，不过这一点并不核心，Nevo使用拟线性的效用函数得到线性的 $y_i - p_{jt}$ 一项，这一项并不影响计算的本质，且适用于不同的研究产品，如线性型代表没有财富效应，可能适用于价格比较便宜的产品（如Nevo研究的即食麦片），而对数型可能更适用于比较贵的产品（如BLP研究的汽车）

$$\begin{aligned} u_{ijt} &= x_{jt}\beta_i + \xi_{jt} + \epsilon_{ijt} \\ u_{i0t} &= \epsilon_{i0t} \\ \beta_i &= \beta + \sigma_k v_k \\ (\epsilon_{i0t}, \dots, \epsilon_{iJt}) &\sim \text{i.i.d. EV} \\ v_k &\sim \text{i.i.d. N}(0, 1) \end{aligned} \quad (5)$$

经过转化可以写为

$$\begin{aligned} u_{ijt} &= x_{jt}\beta + \xi_{jt} + \mu_{ijt} \\ u_{i0t} &= \mu_{i0t} \\ (\mu_{i0t}, \dots, \mu_{iJt}) &\sim \text{Some Distribution} \end{aligned} \quad (6)$$

其中将 μ 看作新的随机项，那么这一项不是iid分布的，可以认为是Berry 94的特殊分布的情况，我们可以计算其相关系数，

$$\begin{aligned} \text{corr}(\mu_{ij}, \mu_{il}) &= \frac{\sum_k \sigma_k^2 x_{jk} x_{lk}}{\sqrt{\left(\sum_k \sigma_k^2 x_{jk}^2 + \sigma_1^2\right) \cdot \left(\sum_k \sigma_k^2 x_{lk}^2 + \sigma_1^2\right)}} \\ \text{corr}(\mu_{ij}, \mu_{i0}) &= \frac{\sigma_e^2}{\sum_k \sigma_k^2 x_{jk}^2 + \sigma_1^2}, \end{aligned} \quad (7)$$

这样的相关系数的特点是当产品j和产品k的产品特征接近时，corr比较大，这就捕捉了替代效应。到这里实际上可以揭示BLP模型对Berry1994的最大的改进：他给了一个更加容易理解的故事，而且允许使用更少的参数来刻画不同产品的随机项之间的相关性。具体来说即，BLP假设每个消费者对产品的各个指标的taste不同（注意是对指标k而不是产品，例如他如果看重车的马力，那么他对所有品牌的汽车都看重马力，所以天然的他在肌肉跑车之间的替代效应就会更强），而taste满足一个随机分布，每一个性能指标k都有一个对均值的偏离 v_{ik} ，iid，其方差是 σ_k^2 。事实上这也使得要估计的参数个数大大下降，我们只需要估计taste向量 β 所在的分布的均值，以及 σ_k 即可（共 K 个），而不再需要估计 $\frac{J(J-1)}{2}$ 个参数刻画的相关矩阵。所以（5）到（6）的变化可以认为是BLP模型假设的核心。在这样的假设下，套用Berry 1994的算法，对 s_{jt} 的表达式进行数值积分，有

$$s_{jt} = \sum_{i=1}^{ns} \frac{\bar{s}}{\bar{f}(v_i, \theta)} f_j(v_i, \theta) = C_1 \sum_{i \in \mathbb{I}_{\theta, \theta}} \frac{\exp(\delta_{jt} + \mu_{jt})}{1 + \sum (\delta_{jt} + \mu_{jt})}. \quad (8)$$

其中分划 v 在原文中的推荐方法是抽样，然后按密度概率接受或拒绝，使用如下的迭代

$$T(s, \theta, P) [\delta_j] = \delta_j + \ln(s_j) - \ln[s_j(p, x, \delta_j, P; \theta)] \quad (9)$$

来更新 δ 的求解。证明了 T 是一个压缩映射（模小于1），这保证了使用11式进行更新的序列一定是收敛的。因此对需求函数的计算可以被以下单纯型算法刻画：

1. 猜测 σ
2. 得到此 σ 下的估计值 δ
 - 猜测 δ_0
 - 重复更新 $\delta_i = T^i(\delta_0)$ 直到收敛
3. 使用 δ 做 GMM 估计 β

$$\min(\delta - x\beta)'z$$

BLP自己知道两层搜索是一个非常复杂的计算行为，因此给了一些实践的tips，比如尽量让初值比较准确，做法之一是先自己悄悄做一个Logit，将估计的结果作为 δ_0 。此外BLP estimator的无偏也需要很严格的条件保证，在误差分析部分会再次提到。作为BLP的一个改进，这部分算法被Mathematical Programming with Equilibrium Constraints（MPEC算法）取代，该算法的核心思路是将上述问题变成约束最小化的问题，即

$$\begin{aligned} \min_{\delta, \bar{\beta}, \sigma_k} \quad & g(\delta; \bar{\beta})' W g(\delta; \bar{\beta}) \\ \text{s.t.} \quad & s_j = \int \frac{\exp(\delta_j + \sum_k x_{jk} \nu_{ik} \sigma_k)}{1 + \sum_q \exp(\delta_q + \sum_k x_{qk} \nu_{ik} \sigma_k)} dF_\nu(\nu_i) \quad \text{for all } j. \end{aligned} \quad (10)$$

该算法不使用 δ 的唯一性，而是直接对 δ 进行搜索，可能会带来搜索空间过大的问题。但是往往MPEC被认为是比BLP更有意义的实践算法。

结合供给端的估计

供给端的刻画是标准的Nash Bertrand，即厂商调整其生产的所有产品的价格使得利润最大化，得到的矩阵表达是 $p = mc + \Delta^{-1}(p)s(p)$

其中 mc 代表边际成本，由看得到的变量 w 和看不到的 ω 刻画： $\log(mc_j) = w_j\gamma + \omega_j$

因此有

$$\omega_j = \log(p_j - e_j' \Delta(p)^{-1} s(p)) - w_j\gamma \quad (11)$$

如果假设IV足够好，即其同时正交于demand shock ϵ 和supply shock ω ，那么可以联合估计，也即在GMM中多加一维：

$$g(\delta; \bar{\beta}, \gamma) = \frac{1}{J} \begin{pmatrix} \sum_j (\delta_j - \mathbf{x}_j' \bar{\beta}) h(z_j) \\ \sum_j (\log(p_j - e_j' \Delta(p)^{-1} s(p)) - w_j\gamma) h(z_j) \end{pmatrix} \quad (12)$$

误差分析

在Berry和Pakes的另一篇论文中给出了 δ 的统计量的渐进性质，一致性要求抽样的次数 ns 和消费者样本数 n 都必须足够大。随着产品数量 J 也增多，越来越多的产品的被选择概率趋于0，因此 J 的大小事实上决定了 n 和 ns 的增长速度的下界。这部分似乎容易被忽略，但也是其短板的主要来源。BLP使用BP的方法将协方差矩阵分解为三部分

- 产生产品特性变量时的数据生产过程 (x, s, w 以及 ξ 的测量误差，本身被GMM解决)
- 消费者抽样时的数据生产过程 (即market share本身作为选择概率的积分，可能有测量误差，没有解决)
- 数值模拟 (因为使用了数值积分，肯定有抽样的问题)。可以使用蒙特卡洛模拟 (例如重复抽10000次计算其方差) 来估算这部分的误差 (可以估计，没有解决)

第二部分误差的内涵是两层搜索的精度必须足够好才可以保证BLP 估计量的无偏性，例如传统的要求对 δ 的估计要求精度通常在 $1e-13$ 或更精确，对 σ 的估计精度则可以稍微放松。对数据的要求也比较高，要求消费者数量，抽样次数和市场个数对产品个数的增长速度足够大。事实上BLP自己的论文后面带的实证部分几乎没有人看 (也有很多人说根本就是错的)，原因之一应该是市场个数对产品数量太少的问题 (20年一共2217条观测，而产品有900多种)。另外其每个产品的市场份额对外部选择都太小 (其每一年的 s_0 都在0.8以上)，这天然的使得准确估计 δ 本身就比较困难，更别提使用估算的 δ 去估计 β 。

Reference

Nevo, Aviv. "A practitioner's guide to estimation of random-coefficients logit models of demand." *Journal of economics & management strategy* 9.4 (2000): 513-548.

Xing Li. 'Lecture Notes Section 3: Logit Models'

Xing Li. 'Lecture Notes Section 4: Market-level Demand Model'

Berry, Steven, James Levinsohn, and Ariel Pakes. "Automobile Prices in Market Equilibrium." *Econometrica* 63, no. 4 (1995): 841-90.

Berry, Steven. "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242-262.

Frank Pinter. 'Demand Estimation Notes'

M. Shum. 'Lecture Notes: Discrete-Choice Demand Models'

Qi Wu. 'Lecture Notes: Frontier of Empirical IO'

<https://zhuanlan.zhihu.com/p/111889827>

另外顺带发现一个看起来很不错的Empirical IO的课件 [ECON5630 Topics in Empirical Industrial Organization \(kohei-kawaguchi.github.io\)](https://github.com/kohei-kawaguchi/ECON5630_Topics_in_Empirical_Industrial_Organization).

以及一个老哥replicate BLP的notebook [BLP Replication - Ivan Li \(ivan-li.com\)](https://ivan-li.com/).

