排序响应的异质交互模型研究

王群勇　徐　伟

（南开大学经济学院数量经济研究所）

研究目标：扩展传统的交互模型，并估计家庭中幸福感相互影响在不同群体间的差异。**研究方法**：将异质特征加入传统的排序变量的社会交互模型，定量地描述外溢效应的异质性。**研究发现**：蒙特卡洛模拟结果表明，复合似然估计对包括异质内生系数在内的所有系数的估计具有渐进一致和渐进正态的良好性质，大样本下信息准则能够一致地选择出正确模型。应用该模型对 CFPS2010 的家庭数据的实证分析表明，幸福感在家庭内部存在相互影响，女性更容易受到其他家庭成员幸福感的影响，从而健康状况和教育水平对女性的幸福感影响倾向于更大。**研究创新**：将交互模型和影响的异质性引入到幸福感研究中。**研究价值**：模型为研究各领域外溢的异质性提供了分析工具。

关键词　社会交互模型　异质性　排序响应

中图分类号　F064.1　　**文献标识码**　A

一、问题提出

本研究聚焦于排序变量中个体的空间相关问题。一方面，在经济学、管理学、社会学、政治学等研究中，很多变量属于排序变量（如社会经济等级、个人幸福感、满意度、贷款风险等级等）；在微观调查数据中，这类排序变量尤其多见。对于这类变量一般用 0、1、2 等数值表示不同的类别，数值本身的大小并无意义，但数值的排序是有意义的。线性模型拟合这类变量存在预测值不合理的情况，因此对于这类变量的典型计量模型是排序选择模型。另一方面，个体的相依关系无处不在，但传统交互模型中存在一个较强的假设：内生影响是同质的，其他个体平均以相同程度（即内生系数）影响目标个体。事实是其他个体与目标之间的作用很可能是非对称的，内生影响也可能存在异质性：比如在家庭生活中，亲子和夫妻之间的相互作用不同，父母对子女、丈夫对妻子的情感上的影响都要比各自另一个方向上的影响要更大（Larson 和 Almeida，1999）。

目前在交互模型这一领域，研究者们对连续性变量的研究相当丰富（Lee，2004；LeSage，1997；Wang 和 Lee，2013），但对离散选择变量、特别是排序多选择变量的相依模型估计性质的研究还相当缺乏。Wang 等（2013）研究了空间 Probit 模型的估计和性质并给出了完整的数学上的证明。Bhat（2014）则详尽地介绍了空间回归范畴内排序响应模型的设定和估计。由于主要研究领域是交通行为和需求分析，他和其他研究者先后利用这一类模型研究了交通行为（Stop-making）模式（Bhat 和 Zhao，2002）、非机动车事故（Narayanamoorthy 等，2013）以及交通方式的选择问题（Bhat，2000；Paleti 等，2013）并识别出了其中重要的空间相依性质，强调考虑空间相依对交通政策制定的重要影响。此外，Bhat 等

（2017）还研究了在非 Logit/Probit 框架下，即扰动项非正态的空间排序响应模型。在异质性研究方面，Lacombe（2010）应是最早引入多个权重矩阵和内生系数的研究者。随后 Aquaro 等（2015）完备地讨论了在连续情形下异质内生参数模型的设定和估计，但该模型内包含多个可能存在的内生系数，对于大样本来说这一设定并不经济，需要进行一些简化。

本文基于 Bhat（2014）等和 Aquaro（2015）等研究，将内生异质性引入排序交互模型，用两个内生交互项体现个体之间相互影响的异质性。本文的主要贡献在于：（1）讨论了带有内生异质性的排序交互模型的估计方法，利用蒙特卡洛模拟讨论了参数似然估计量的统计分布性质；（2）讨论了排序响应变量的模型选择问题；（3）利用这种模型研究了幸福在家庭内的溢出效应的性别异质性。

二、模型设定与估计

根据传统离散变量模型的设定，以潜变量和矩阵形式的模型设定如下：

$$\begin{aligned} y^* &= \lambda_1 W_1 y^* + \lambda_2 W_2 y^* + X\beta_0 + V \\ y_i &= j \text{ 如果 } \psi_{j-1} < y_i^* \leqslant \psi_j (j = 1,2,3,\cdots,J) \end{aligned} \tag{1}$$

这里我们假设阈值 ψ 对于所有个体间都是相同的，因此它只与类别有关，与个体无关。约定 $\psi_0 = -\infty$，$\psi_J = +\infty$，这样剩下 $J-1$ 个值把数轴分成了 J 段。这 $J-1$ 个待估参数满足：

$$\psi_1 < \psi_2 < \cdots < \psi_{J-2} < \psi_{J-1} \tag{2}$$

与连续情形有所区别的是，为了保证参数的可识别性，这里 X 不含常数项。一个最简单的情形为 $V \sim N(0, I_n)$。$W_2 y^*$ 叫作异质内生项，W_2 用以体现个体相互影响的溢出效应的非对称性。通常设定 $W_2 = BW_1$，其中 B 是由表示群体属性的虚拟变量构成的对角矩阵。例如，A 类（个体 1）和 B 类（个体 2、3）接收其他个体的影响能力有差异，模型设定如下：

$$\begin{bmatrix} y_1^* \\ y_2^* \\ y_3^* \end{bmatrix} = \lambda_1 \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} y_1^* \\ y_2^* \\ y_3^* \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 & & \\ & 0 & \\ & & 0 \end{bmatrix} \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} y_1^* \\ y_2^* \\ y_3^* \end{bmatrix} + X\beta \tag{3}$$

个体 1 以 $\lambda_1 + \lambda_2$ 的平均能力接收来自其他个体的影响，而个体 2、3 都以 λ_1 的能力接收来自其他个体的影响。考虑识别问题，$W_1 \neq cW_2$（其中 c 是不为 0 的常实数），这等同于要求在两类不同群体或不同层次内都必须至少有一个观测值。

把式（1）右侧的内生项移到左边，可得：

$$\begin{aligned} Sy^* &= X\beta_0 + V \\ S &= I - \lambda_1 W_1 - \lambda_2 W_2 \end{aligned} \tag{4}$$

因为 V 是多元正态分布的，所以 $y^* \sim N(S^{-1}X\beta_0, S^{-1}S'^{-1})$，对应的密度函数为：

$$l(\theta) = (2\pi)^{-\frac{q}{2}} \mid S^{-1}S'^{-1} \mid^{\frac{-1}{2}} \exp\left\{-\frac{1}{2}(y^* - S^{-1}X\beta_0)'(S^{-1}S'^{-1})^{-1}(y^* - S^{-1}X\beta_0)\right\} \tag{5}$$

似然函数为：

$$L(\theta)=\int_{D_{y^*}} l(\theta)dy^*$$
$$D_{y^*}=\{y^*:\psi_{y_i-1}<y_i^*\leqslant\psi_{y_i}\} \tag{6}$$

但是，直接计算上述积分是不经济的，尤其是在 n 非常大的情形下；传统的 MLE 并不适用。为了解决这一问题，一系列数值模拟技术应运而生，最为常见的一种是复合似然函数（Composite Likelihood，CL）技术。构建配对复合似然（pairwise Composite Likelihood，pCL）函数（Bhat，2014）：

$$L_{pCML}(\theta)=\prod_{i=1}^{N-1}\prod_{s=i+1}^{N}L_{is}$$

其中：

$$L_{is}=P\{y_i,y_s\}=[\Phi_2(\varphi_i,\varphi_s,v_{is})-\Phi_2(\varphi_i,\mu_s,v_{is})-\Phi_2(\mu_i,\varphi_s,v_{is})+\Phi_2(\mu_i,\mu_s,v_{is})]$$
$$\varphi_i=\frac{\psi_{y_i}-[S^{-1}X\beta_0]_i}{\sqrt{(S'S)_{ii}^{-1}}}\qquad \mu_i=\frac{\psi_{y_i-1}-[S^{-1}X\beta_0]_i}{\sqrt{(S'S)_{ii}^{-1}}}\qquad v_{is}=\frac{(S'S)_{is}^{-1}}{\sqrt{(S'S)_{ii}^{-1}}\sqrt{(S'S)_{ss}^{-1}}} \tag{7}$$

因为 y^* 是联合多元正态分布的，所以任意二维变量的联合分布都仍然是正态分布的，此时标准化变量服从标准二元正态分布，相关系数 v_{is} 的计算如式（6）。Φ_2（a，b，ρ）表示相关系数为 ρ 的二元标准正态分布在由点（a，b）构成的方形区域上的累积值，则 L_{is} 表示的是相关系数为 v_{is} 的二元标准正态分布在由（φ_i，φ_s）、（φ_i，μ_s）、（μ_i，φ_s）、（μ_i，μ_s）四点围成的区域上的累积值。

三、微观交互的特征

本文研究的微观交互影响一个突出的特征就是“组/小团体”的存在。组是社交关系的一个概念，其特点是组内的个体存在关联，而与组外的任一个体都不存在关联。从这个概念来看，与传统的空间计量模型相比微观交互模型有其特殊之处：在微观样本中通常存在大小不一的多个组，而传统空间模型只有一个。因此在计算上，微观交互模型有如下两个优势：

（1）交互模型中的联结矩阵是多个分块矩阵在对角线上罗列而成，这样 S 矩阵的逆可以分块求得，并且这些块通常都很小（一般小于 10×10），比直接求逆更快。如在计算方差协方差矩阵以及平均边际影响时不得不计算 S^{-1} 和 $(S'S)^{-1}$，而 $n\times n$ 维矩阵直接求逆会极大降低运算速度（>10s），分块后运算时间可忽略不计。

（2）通常设定下，这些组可以被视为是某一系统的大量重复观测，或者说组之间是独立的。因此理论上 CL 中的对一共有 n（$n-1$）/2 个，但 Bhat（2014）指出没有必要计算这么多的对；与目标距离最近的一些观测值提供了更多信息，因此在成对的复合似然函数计算中，只需计算有限的对。考虑联结矩阵的伪对角性质，在交互模型的估计中只需考虑同组构成的对。由于组规模小，这一值通常远远小于全部对的个数。

记 $\mathbb{C}_i$ 为 W_1 第 i 行中元素不为零的列之标号集合，则有：

$$L_{CML}(\theta)=\prod_{i=1}^{N-1}\prod_{\substack{s=i+1\\ n'\in\mathbb{C}_i}}^{N}L_{is} \tag{8}$$

待估参数的方差协方差矩阵为：

$$\mathrm{var}(\theta)=\frac{\hat{H}^{-1}\hat{J}\hat{H}'^{-1}}{\tilde{W}} \tag{9}$$

其中 $\widetilde{W}$ 为构成的对的个数：

$$\hat{H}=-\frac{1}{\widetilde{W}}\sum_{i}^{N-1}\sum_{\substack{s-i\mid 1\\ s\in\mathbb{C}_i}}^{N}\frac{\partial^2\log L_{is}}{\partial\theta\partial\theta'}\Bigg|_{\theta=\hat{\theta}_{(ML}} \tag{10}$$

由于天然存在大量重复观测，在计算参数方差协方差矩阵中的 $\hat{J}$ 时，不必进行网格取样，利用各组就能近似计算。考虑 R 个组 $N=\sum_{r=1}^{R}m_r$ 个个体的情形，网格点 $\widetilde{N}$ 天然存在：R 个组每组内任意某个体就作为网格点，组内所有其他个体都作为该网格点的“最近的”个体，这样形成一个近似于重复独立观测的含有 R 个子样本，用这个子样本去近似计算 $\hat{J}$，即有：

$$\hat{J}=\frac{1}{R}\sum_{\tilde{s}=1}^{R}\frac{1}{C_{m_{\tilde{s}}}^{2}}\left[\sum_{l,l'\in\{1,2,\cdots,m_{\tilde{s}}\}}\left(\frac{\partial\log L_{ll'}}{\partial\theta}\right)_{\tilde{s}}\right]\left[\sum_{l,l'\in\{1,2,\cdots,m_{\tilde{s}}\}}\left(\frac{\partial\log L_{ll'}}{\partial\theta'}\right)_{\tilde{s}}\right]\Bigg|_{\theta=\hat{\theta}_{(ML}} \tag{11}$$

其中 $\sum_{l,l'\in\{1,2,\cdots,m_{\tilde{s}}\}}\left(\frac{\partial\log L_{ll'}}{\partial\theta}\right)_{\tilde{s}}=\sum_{l=1}^{\widetilde{m}_{\tilde{s}}-1}\sum_{l'=l+1}^{m_{\tilde{s}}}\left(\frac{\partial\log L_{ll'}}{\partial\theta}\right)$ 是组内成对个体一阶导数的和。$C_{m_{\tilde{s}}}^{2}$ 是组合数，代表在规模为 $m_{\tilde{s}}$ 的组中成对观测值的数量。

针对模型估计需要说明三点。第一，在复合似然函数中使用 3 个、4 个甚至更多观测值做联合边缘分布在理论上都是可行的，然而使用成对观测值，对应的是二维正态分布，可图形表示的同时计算又简便些。在微观交互框架下，当存在仅含两个人的组时，选择三维或更多做联合分布会发现这些组内没有足够的观测值，二维则确保成对数据都属于同一个组，是较为折中的选择。第二，在空间范畴内，确定多少个对参与计算的通常做法是，选择与目标点最近的 k 个个体去成对计算，然后选择一个最小化方差协方差矩阵的迹的 k 值作为最终的结果。一方面这样做不经济（需要多次计算，在样本量很大时是费时的）；另一方面，“距离最近”的定义是模糊的，社交关系中的“距离”是层次的、分类的而不是空间意义上的连续距离变量。例如对属于某三口之家内的目标个体来说，其他两位家庭成员可看作是与他“距离最近”的，但当 k 超过家庭成员数时，如何选择另一个观测值和目标构成对就并不确定了。本文的做法是，仅将家庭成员作为与目标距离“最近的个体”，其他不属于目标家庭的个体均不与目标构成对参与计算。因为家庭规模可能不同，每个组构成的对的数量也不一样。这样做并不保证一定最优，但却是一种合乎逻辑、计算简便的方法。第三，pCLE 只是其中一种估计法，因为组的存在分块复合似然（Eidsvik 等，2014）也是一种选择。

在 Probit 类模型中，系数大小，甚至系数本身的符号都没有什么含义。要解释非线性模型的系数，通常借用平均边际影响的概念。对某个体 i 全部因变量 x_k 同时变动一个单位时对它潜变量的总影响为 $\beta_k\sum_u[S^{-1}]_{iu}$，与 S^{-1} 的行和有关。但在非线性模型中，我们通常考虑外生变量对取不同类别概率的变动。在交互模型中理论上观测值 i 取 j 类的概率为：

$$\begin{aligned}P\{y_i=j\}&=\int_{\psi_{j-1}}^{\psi_j}\frac{1}{\sqrt{2\pi}(S'S)_{ii}^{-1}}\exp\left\{-\frac{1}{2}\frac{(v-[S^{-1}X\beta_0]_i)^2}{(S'S)_{ii}^{-1}}\right\}dv\\&=\Phi\left(\frac{\psi_j-[S^{-1}X\beta_0]_i}{\sqrt{(S'S)_{ii}^{-1}}}\right)-\Phi\left(\frac{\psi_{j-1}-[S^{-1}X\beta_0]_i}{\sqrt{(S'S)_{ii}^{-1}}}\right)\end{aligned} \tag{12}$$

表明不仅第 i 观测值本身的自变量、其他个体自变量通过 S^{-1} 的作用也将影响观测值 i 取 j 类的概率。以 $PE_{i,j}(x_k)$ 表示全部个体第 k 自变量对个体 i 的取类别 j 的总影响（Partial Effect to an Observation i），则有：

$$\begin{aligned} PE_{i,j}(x_k) &= \sum_{s=1}^{N} \frac{\partial P\{y_i = j\}}{\partial x_k^{(s)}} \\ &= \left[\phi\left(\frac{\psi_{j-1} - [S^{-1}X\beta_0]_i}{\sqrt{(S'S)_{ii}^{-1}}}\right) - \phi\left(\frac{\psi_j - [S^{-1}X\beta_0]_i}{\sqrt{(S'S)_{ii}^{-1}}}\right)\right]\frac{a'S^{-1}i}{\sqrt{(S'S)_{ii}^{-1}}}\beta_k \end{aligned} \tag{13}$$

其中，向量 $a'=$（0，0，…，1，0，…，0）为第 i 列元素为 1、其他元素全为 0 的行向量，$a'S_n^{-1}i$ 表示求 S_n^{-1} 第 i 行元素的和。则自变量 x_k 对应响应取 j 类 APE 为：

$$APE_j(x_k) = \frac{1}{N}\sum_{i=1}^{N} PE_{i,j}(x_k) \tag{14}$$

通过对 $PE_{i,j}$（x_k）的分解，可求得直接效应和间接效应；对 $PE_{i,j}$（x_k）在不同类别的个体间求平均，可分别求得各自的 ΛPE_j（x_k）。式（13）和式（14）表明 ΛPE 是非线性的，其大小除与 S^{-1}的行和有关外还与其他变量的值有关，但可以确定的是，行和大的个体其 APE 有大的倾向。

按照式（13）和式（14）并不能得到 ΛPE 的标准差，因此一个更一般的做法是 Ferdous 和 Bhat（2013）中提到的模拟过程：记 CLE 为 $\hat{\theta}=$（$\hat{\lambda}_1$，$\hat{\lambda}_2$，$\hat{\psi}_1$，$\hat{\psi}_2$，$\hat{\psi}_3$，$\hat{\psi}_4$，$\hat{\beta}$）是某个估计值，据此可计算出潜变量服从的多元正态分布的均值向量和方差协方差矩阵：

$$\mathrm{S} = \mathrm{I_n} - \lambda_1 \mathrm{W} - \lambda_2 \mathrm{BW} \tag{15}$$

$$\hat{\mathrm{B}} = S^{-1}X\hat{\beta} \tag{16}$$

$$\hat{\Xi} = (S'\mathrm{S})^{-1} \tag{17}$$

从 N（$\hat{B}$，$\hat{\Xi}$）中抽取含 N 个元素的潜变量 y_i^*，根据 $\hat{\psi}_1-\hat{\psi}_4$ 这些 y_i^* 能够立刻被标上 1～5 的号。重复 T 次后，对每一个个体 y_i 属于 1～5 类的概率（比例）就可得。对关心的外生变量，施加一个单位的变动，再次计算新的、属于 1～5 类的概率，两者之差可近似地看作是平均总边际效应。以此方式能得到 5 个（通常不同）的数值，就是相应类别的 x 的平均总边际效应。对于离散变量的“伪平均边际效应”的计算要复杂些：为了求个体从 a 类（基准类）变动到 b 类的 APE，先要把所有个体设定在 a 类（基准类）上，然后再设定在 b 类上，比较前后概率的差异。此外一个值得注意的问题是，在应用中自变量通常是多维的。本文的做法是对每个个体计算可能的边际影响，然后再在人际间做平均（而不是在其他变量的均值处计算）。

APE 的标准差可由 Bootstrap 得到。前述结论已知参数 θ 服从的是均值为 $\hat{\theta}_{CL}$、方差协方差矩阵为 $H^{-1}JH^{-1}/\widetilde{W}$ 的多维正态分布，因此每次都从这样一个正态分布中抽取一个 θ 的样本，按照上述方法计算 APE，重复若干次即可得到经验的标准差。

四、蒙特卡洛模拟

本节利用蒙特卡洛模拟考察异质交互模型的参数估计量的有限样本特征。数据生成过程（DGP）设定为：

$$\begin{aligned} y^* &= \lambda_1 W y^* + \lambda_2 BW y^* + X\beta_0 + V \\ V &\sim N(0, I_n) \end{aligned} \tag{18}$$

作为对比的同质 DGP 为：

$$y^* = \lambda_1 W y^* + X\beta_0 + V$$
$$V \sim N(0, I_n) \tag{19}$$

主要参数设定如下：权重矩阵 $W = I_R \otimes \frac{i'_m i_m - I_m}{m-1}$，其中 R 是组的个数，m 是每组包含的成员个数。B 是代表异质个体的矩阵，由 0 和 1 在对角线上 1∶1 随机混合（混合比例 0.5 固定）。$W_2 = BW$，相当于取出 W 中一半的行，此时 W_2 仍然满足所有对于权重矩阵的要求（行和为 1，对角线为 0）。外生变量 $X \sim N(1, I_n)$。参数真值设定为：$\lambda_1 = 0.4$，$\lambda_2 = 0.2$，$\beta_0 = 1$，端点 $\psi_1 = -0.4448$，$\psi_2 = 0.6828$，$\psi_3 = 2.4065$，$\psi_4 = 5.0953$[①]。

潜变量生成后，按照如下规则生成观测值：

$$y_i = \begin{cases} 1, & y_i^* \leqslant \psi_1 \\ 2, & \psi_1 < y_i^* \leqslant \psi_2 \\ 3, & \psi_2 < y_i^* \leqslant \psi_3 \\ 4, & \psi_3 < y_i^* \leqslant \psi_4 \\ 5, & y_i^* > \psi_1 \end{cases} \tag{20}$$

1. 对参数的还原

根据由式（18）生成的 y、x 以及联结矩阵 W、B，按照 CLE 方法估计参数（λ_1，λ_2）、端点（ψ_1，ψ_2，ψ_3，ψ_4）以及 β，重复 200 次，结果如表 1 所示。

表 1　　模拟结果

R	30			60			120		
m	3	5	10	3	5	10	3	5	10
λ_1	0.3977 (0.0933)	0.3964 (0.0928)	0.3895 (0.0949)	0.4012 (0.0701)	0.3891 (0.0663)	0.3920 (0.0611)	0.3989 (0.0491)	0.3923 (0.0449)	0.3957 (0.0423)
λ_2	0.2111 (0.1188)	0.2037 (0.1069)	0.1979 (0.0784)	0.1928 (0.0782)	0.2074 (0.0719)	0.2030 (0.0502)	0.1943 (0.0517)	0.2034 (0.0487)	0.1995 (0.0372)
ψ_1	−0.4243 (0.4853)	−0.4229 (0.3968)	−0.4523 (0.3355)	−0.4675 (0.2923)	−0.4405 (0.2791)	−0.4642 (0.2191)	−0.4736 (0.2294)	−0.4619 (0.1817)	−0.4434 (0.1698)
ψ_2	0.7799 (0.4917)	0.7648 (0.4374)	0.7067 (0.3607)	0.6998 (0.3238)	0.6990 (0.2789)	0.6806 (0.2279)	0.6826 (0.2323)	0.6617 (0.1808)	0.6790 (0.1712)
ψ_3	2.6222 (0.6388)	2.5360 (0.5380)	2.4588 (0.4050)	2.4635 (0.3895)	2.4691 (0.3300)	2.4250 (0.2808)	2.4212 (0.2718)	2.3849 (0.2087)	2.4143 (0.2102)
ψ_4	5.6785 (1.1645)	5.3995 (0.8256)	5.2182 (0.6002)	5.2591 (0.6090)	5.2422 (0.5150)	5.1447 (0.3818)	5.1565 (0.4579)	5.0717 (0.3195)	5.1190 (0.3058)
β	1.0669 (0.2032)	1.0418 (0.1570)	1.0235 (0.0919)	1.0231 (0.1159)	1.0324 (0.0875)	1.0110 (0.0650)	1.0134 (0.0836)	0.9998 (0.0604)	1.0069 (0.0564)

注：数值为抽样均值，小括号内为 RMSE。

① 这些数字是随机选择的，它们接近潜变量 y^* 分布的 10、30、65、90 百分位数，以此方式确定端点能够几乎保证每一个分类下都有相当数量的观测值。

表 1 表明在模型正确设定下，CLE 都比较接近真实值，均方根误差（RMSE）也都随样本量的增大减小。但注意到在样本量偏小时，参数标准差比较大，实际非零的参数 λ_2 也只有边际显著，因此为了能够正确地识别出非零的异质内生系数，样本量不能太小；从模拟结果来看，样本量至少要达到 100。表 2 给出不同样本量下一个抽样样本按照式（9）～式（11）计算出的各个参数标准差，可以看出网格法计算的标准差与抽样标准差比较接近，而单独按照 Hessian 矩阵计算的标准差则与抽样标准差有一定差异；网格法计算的标准差更为稳健。

表 2　　样本方差计算

样本	参数	抽样标准差	H^{-1}/W	$H^{-1}JH^{-1}/W$
$m=3$ $R=30$	λ_1	0.0932	0.0629	0.0782
	λ_2	0.1186	0.0764	0.1149
	ψ_1	0.4864	0.2961	0.4148
	ψ_2	0.4776	0.3995	0.4953
	ψ_3	0.5946	0.5032	0.6004
	ψ_1	1.0098	0.7662	0.8978
	β	0.1937	0.1666	0.1929
$m=10$ $R=120$	λ_1	0.0421	0.0134	0.0447
	λ_2	0.0374	0.0118	0.0373
	ψ_1	0.1709	0.0407	0.1574
	ψ_2	0.1724	0.0405	0.1629
	ψ_3	0.2108	0.0500	0.1916
	ψ_1	0.3052	0.0740	0.2664
	β	0.0557	0.0150	0.0471

2. 渐近性质

通常条件下的 CLE 有着良好性质：不仅一致，而且是渐近正态的。但在空间回归框架下的多元排序 Probit 模型的估计性质，相关理论研究并不充分。本部分将通过模拟来展示 CLE 的有限样本性质，通过增加样本量，还可一窥其大样本性质。

图 1 是表 1 中模拟结果的图示，直方图代表重复 200 次的参数抽样分布，实线表示抽样分布，虚线表示具有相同均值标准差的正态分布理论密度。可以看出有限样本下所有参数都呈现出近似正态的分布，参数分布的直方图和密度估计都与同均值方差的正态分布比较接近。从图 2 可以看出，参数还呈现一致的特征：给定 m 随 R 增加，参数标准差都随样本量增加而降低，分布趋于集中。

3. 参数检验

LR 检验的基本形式为：

$$LR = 2[\log L(\hat{\theta};y) - \log L(\tilde{\theta};y)] \tag{21}$$

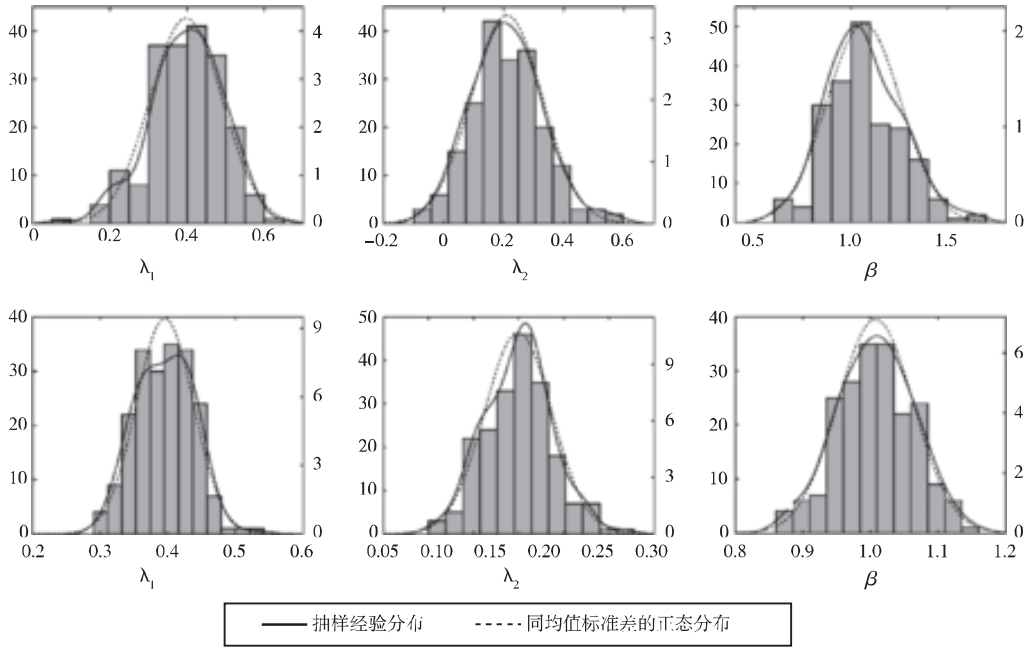

图 1　参数的渐近正态性质

注：上层为小样本情形（$m=3$，$R=30$），下层为大样本情形（$m=10$，$R=120$）。

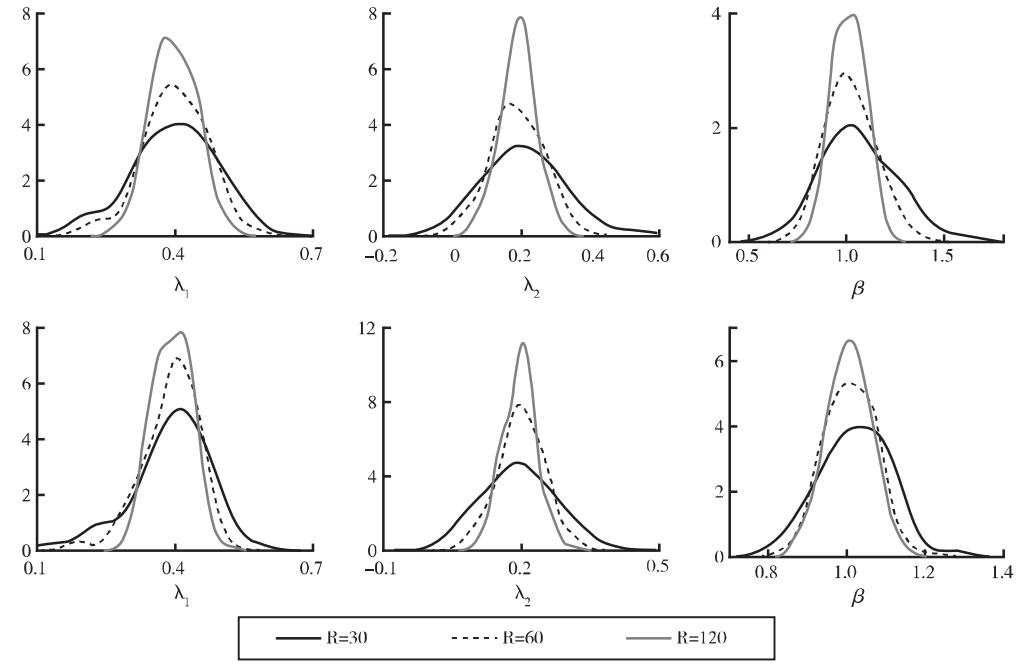

图 2　参数的一致特征

注：上层为小样本情形（$m=3$，$R=30$、60、120），下层为大样本情形（$m=10$，$R=30$、60、120）。

当 $\log L(\hat{\theta};\ y)$ 表示无约束 pCL 对数值，$\log L(\tilde{\theta};\ y)$ 表示有约束值时，上式就是 CL 框架下的 LR 检验，但此时该统计量不服从卡方分布（因为 $J\neq H$）。在原假设 H_0：$\tau=\tau_0$

下，根据 Pace 等（2011）的结论 Bhat（2011）给出了一个调整形式：

$$\mathrm{aLR}=\frac{[g_\tau(\theta)]'[H_\tau(\theta)]^{-1}\{H(\theta)J^{-1}(\theta)H(\theta)\}_\tau[H_\tau(\theta)]^{-1}g_\tau(\theta)}{[g_\tau(\theta)]'[H_\tau(\theta)]^{-1}g_\tau(\theta)}\mathrm{LR}\sim\chi^2(d) \tag{22}$$

d 表示约束的个数（τ 的维度），$g_\tau(\theta)$ 表示梯度向量中对应的 $d\times1$ 列，$H_\tau(\theta)$ 表示 $H(\theta)$ 中对应的 $d\times d$ 块。式（22）是对式（21）的标准化，当 $J=H$ 时式（22）就退化成了式（21）的形式。

4. 模型选择

显然，如果真实模型中存在异质的交互影响但设定的模型忽略这种异质性，模型会有设定错误问题。因此，接下来的一个问题是如何判断是否存在异质性特征。这部分将分别使用不同样本量的样本去研究模型选择问题。模拟将按照如下步骤进行：按照异质 DGP 式（17）生成 y，W，B 和 x，确定已知的真值 λ_1，λ_2 和 β_0，然后考虑如下几个估计模型：

模型 1：

$$y^*=\lambda\mathrm{W}y^*+\lambda'\mathrm{BW}y^*+x\beta \tag{23}$$

模型 2：

$$y^*=\lambda\mathrm{W}y^*+x_n\beta \tag{24}$$

模型 3：

$$y^*=x\beta \tag{25}$$

在误差项独立正态分布的假定下，模型 1 是完全正确设定，其他两个模型都在一定程度上与真实情形相悖：模型 2 缺少异质内生项，模型 3 则完全忽略内生性。作为对比，实际 DGP 为式（18）时，模型 2 是完全正确的设定，模型 1 有冗余异质项，模型 3 则完全忽略内生性。分别计算两种 DGP 下每一个模型的信息准则 AIC 和 BIC，并计算在模拟 200 次样本中能正确选择出模型的概率。

信息准则的基本形式为：

$$\mathrm{AIC}=-2\log L(\hat\theta;y)+2\mathrm{trace}[J(\hat\theta)H^{-1}(\hat\theta)] \tag{26}$$

$$\mathrm{BIC}=-2\log L(\hat\theta;y)+\mathrm{trace}[J(\hat\theta)H^{-1}(\hat\theta)]\cdot\log N \tag{27}$$

其中 $\log L(\hat\theta;y)$ 表示相应对数似然函数在估计量 $\hat\theta$ 处的值，$J(\hat\theta)$ 表示梯度向量的方差协方差矩阵，$H(\hat\theta)$ 表示 Hessian 矩阵。依据估计方法的不同，三者可以有不同的形式，如：$\log L(\hat\theta;y)$ 是 pCL 对数值，$J(\hat\theta)$ 和 $H(\hat\theta)$ 的计算如式（9）和式（10）。这样定义的 AIC 是 Varin 和 Vidoni（2004）基于 CL 框架下提出的形式，BIC 是 Katsikatsou 和 Moustaki（2016）提出的形式。关于 BIC 的其他算法，可以参考 Gao 和 Song（2012），与式（27）指标相比该指标有额外的对模型稀疏性的惩罚因子，式（27）是忽略这一惩罚因子的简化形式。

结果表明真实情形为异质 DGP 时 AIC 优些，而真实情形为同质 DGP 时 BIC 优些（见表 3）；小样本下以信息准则判断并不准确，但随着样本量增加，两者能够一致地选择出正确模型。因此在大样本下怀疑内生异质性存在时，可使用信息准则选择相对最好的模型。

表 3 信息准则与模型选择

选出正确模型的比例（%）	异质 DGP			同质 DGP		
$m=3$	$R=30$	$R=60$	$R=120$	$R=30$	$R=60$	$R=120$
AIC	71.5	93.0	100.0	77.0	82.0	100
BIC	45.0	78.0	100.0	91.0	97.5	100

五、应用：幸福感外溢的异质性

许多文献都发现了幸福感存在相互影响的特征（Fowler 和 Christakis，2008；刘斌等，2012；Tumen 和 Zeydanli，2014；王群勇和徐伟，2019），但很少有文献去研究这种相互影响的非对称性。一些文献研究了这种内生影响在不同性别人群中的差异，如刘斌等（2012）发现各种估计方法下，女性群体的内生系数都更显著地高，结论是幸福感更容易传递给女性。一些研究夫妻间幸福感影响的文献也发现女方更容易受男方的影响（Larson 和 Almeida，1999；Carr 等，2014）。上述文献大都不是在交互模型框架下进行的，本部分将基于上述提出的异质交互模型对这一问题进行验证性研究。

文中使用的数据来自中国家庭追踪调查（China Family Panel Studies，CFPS）2010 年的截面数据。CFPS 提供了被调查者的家庭连结信息，使我们可以考察家庭成员之间的相互影响。我们将样本限定在家庭成员不少于两人、均为成人（16 岁及以上），且都有完整有效回答的家庭，最终我们得到 4876 户共 11859 个个体的观测值。在这些家庭中，家庭成员数最小为 2 人，最大为 9 人；采访家庭中主要为两人或三人家庭。本文选择健康状况、教育水平、城乡、性别、年龄、年龄的平方和对数的个人年收入（对数形式）等作为自变量，表 4 列出了这些变量的描述性统计结果。

表 4 主要变量的描述性统计

离散变量	均值	标准差	取值与占比
主观幸福感	3.8588	1.0106	1：非常不幸福（2.74%） 2：比较不幸福（5.73%） 3：说不上幸福不幸福（25.55%） 4：比较幸福（34.97%） 5：非常幸福（31.01%）
性别	0.5065	0.5000	0：女（49.47%） 1：男（50.53%）
教育水平	1.9115	0.7825	1：文盲/半文盲（30.42%） 2：小学（20.36%） 3：初中（27.52%） 4：高中（14.10%） 5：大专/本科/硕士（7.60%）
连续变量	均值	标准差	取值范围
健康状况①	5.0098	1.3251	[1，7]
年龄	49.9048	16.1152	[16，109]
对数收入	6.6638	3.8595	[0，13.4588]

① 这是采访员对回答者健康状况的评估，取值为 1～7 的整数。1 表示不健康，7 表示很健康，本文中视为连续变量，但将其看作离散变量也并不改变之后的结论。

模型设定如式（18）和式（20）。其中 W 是表示家庭关系的联结矩阵（类似空间计量中的权重矩阵），矩阵满足全部对权重矩阵的要求（行和为 1，对角线元素为 0）。B 是性别这一虚拟变量列向量构成的对角阵，BW 相当于提出 W 对应属于男性群体的行。对男性群体，其内生系数为 $\lambda_1+\lambda_2$，对女性群体，内生系数只有 λ_1。X 表示表 5 中列出的健康状况、教育水平、年龄、年龄的平方/100 以及对数收入（不包含常数项）。使用成对复合似然函数估计模型，并与其他设定模型相比，结果如表 5。

表 5　　估计结果

模型类别		模型 1		模型 2		模型 3	
内生系数							
$\hat{\lambda}_1$		0.3126	(0.0168)	0.2334	(0.0086)	—	
异质内生参数 $\hat{\lambda}_2$		−0.1582	(0.0266)	—		—	
健康状况		0.1456	(0.0074)	0.1309	(0.0076)	0.1451	(0.0076)
教育水平	小学	0.1372	(0.0280)	0.1126	(0.0277)	0.1153	(0.0273)
	初中	0.1715	(0.0269)	0.1438	(0.0267)	0.1594	(0.0263)
	高中	0.1681	(0.0326)	0.1430	(0.0324)	0.1643	(0.0320)
	大专/本科/硕士	0.0999	(0.0437)	0.0691	(0.0437)	0.0988	(0.0434)
年龄		−0.0226	(0.0024)	−0.0296	(0.0027)	−0.0280	(0.0026)
年龄的平方/100		0.0284	(0.0026)	0.0342	(0.0028)	0.0332	(0.0027)
对数收入		−0.0002	(0.0026)	−0.0039	(0.0026)	−0.0037	(0.0025)
$-\log L$		13700.96		13710.86		13945.52	
$tr(H^{-1}J)$		21.09		19.86		18.38	
AIC		27444.12		27461.43		27927.80	
BIC		27599.81		27608.01		28063.46	

注：小括号中数值为标准差。

不同模型的估计结果表明，模型设定对外生系数 β 的估计影响有限，但不同内生系数设定对模型指标有一定影响。从信息准则看，模型 3 与模型 1、模型 2 差异较大，是最先被舍弃的，这表明建模时应当至少包含一个内生影响项。模型 1 和模型 2 的信息准则差异较小，但模型 1 略优。此外，在原假设 H_0：$\lambda_2=0$ 下计算调整的似然比统计量为 13.4444（p=0.0003），拒绝原假设，也支持异质内生项非零的结论。模型选择的结果支持内生系数的存在，并且是包含异质内生系数的模型 1。在模型 1 设定下，幸福感在不同性别间的外溢确实存在一定差异：对男性群体来说，内生系数为 0.1544，对女性群体这一系数则为 0.3126。二者的差为 λ' 的值，标准差表明它在 1% 的显著性水平下统计显著，说明这种差异不可忽略。外生参数的估计都符合直觉，其他条件不变时，健康状况每下降一个单位或教育水平每上升一个等级，个体响应幸福感都有提升的倾向；绝对收入并未对个体幸福感有显著影响，同时年龄（对潜变量）的影响呈现明显 U 形特征，拐点在 47 岁，这都与前人研究结论一致（金江，2010；Cuñado 和 Gracia，2012；Blanchflower 和 Oswald，2008）。

因为是非线性模型，所以表 5 中外生变量的系数大小没有具体含义，为了对模型得到的结果进行解释，需计算外生变量对响应的平均边际影响（APE）。表 6 和表 7 分别给出的是根据模型 1 的回归结果计算的健康和教育对男性群体和女性群体的总平均边际影响（计算过程见前文）。当全部个体健康状况上升一个等级时，从 APE 的角度来看，男性群体取高响应的概率提升 6 个百分点，女性群体取高响应的概率提升 7 个百分点，二者相差约 1 个百分点并且差异是显著的（抽样标准差为 0.1394）。对不同教育水平进行类似的分析，可知全部个体的教育水平从基准组（文盲/半文盲）提升至相应水平时，女性群体高水平幸福感响应的概率提升都更大些。总体上看，女性群体更容易受其他家庭成员幸福感的影响，因此当全部个体的健康或教育水平提升时，女性幸福感提升的倾向更大、受益更多，这一结论是同质模型所不能捕捉到的。

表 6　　健康和教育对男性群体的 APE

		响应（单位：%）				
		1	2	3	4	5
健康状况		−0.7605 (0.0549)	−1.2488 (0.0740)	−3.6019 (0.1826)	−0.5706 (0.0922)	6.1818 (0.3367)
教育水平	小学	−0.9591 (0.2051)	−1.4137 (0.2915)	−3.4469 (0.7356)	0.1558 (0.1016)	5.6839 (1.1852)
	初中	−1.1320 (0.2054)	−1.6905 (0.2791)	−4.2358 (0.7269)	0.0805 (0.1120)	6.9779 (1.1692)
	高中	−1.1239 (0.2022)	−1.6798 (0.2835)	−4.2040 (0.7258)	0.0802 (0.1258)	6.9274 (1.1832)
	大专 本科/硕士	−0.7032 (0.3290)	−1.0216 (0.4777)	−2.4701 (1.2413)	0.1662 (0.1011)	4.0287 (2.0260)

注：括号中为 100 次抽样标准差。

表 7　　健康和教育对女性群体的 APE

		响应（单位：%）				
		1	2	3	4	5
健康状况		−0.9248 (0.0635)	−1.3715 (0.0718)	−3.8198 (0.1818)	−0.8104 (0.1017)	6.9265 (0.3394)
教育水平	小学	−1.1291 (0.2412)	−1.5323 (0.3137)	−3.7105 (0.7925)	−0.0952 (0.1095)	6.4671 (1.3631)
	初中	−1.3306 (0.2472)	−1.8313 (0.3042)	−4.5333 (0.7836)	−0.2397 (0.1027)	7.9348 (1.3499)
	高中	−1.3215 (0.2427)	−1.8191 (0.3089)	−4.4989 (0.7883)	−0.2389 (0.1436)	7.8784 (1.3842)
	大专/本科/硕士	−0.8281 (0.3810)	−1.1080 (0.5140)	−2.6426 (1.3335)	−0.0011 (0.1461)	4.5799 (2.3100)

注：同表 6。

六、现存问题与讨论

本文考察了排序响应变量中带有异质性的交互模型，利用蒙特卡洛模拟方法考察了配对复合似然估计量的有限样本特征。模拟结果表明配对符合似然估计能够还原出包括一致内生参数在内的所有参数，它们的分布是渐进一致和渐进正态的。大样本下信息准则能够选择出正确的模型。在应用部分，异质交互模型的回归结果表明，基于家庭的幸福感外溢存在异质性，女性更容易受其他家庭成员感受的影响，健康状况和教育水平对女性的幸福感影响要比男性有更大的倾向，这些结论与使用其他方法得到的结论一致。

本文提出的异质性模型可以考察溢出效应的异质特征，但仍有一定的局限。其一，该异质性模型依赖于理论和后验（统计检验），未能提出一个良好的、事前统计量（类似于 Moran's I）去检验不同群体间的内生系数是否有显著差异。其二，模拟结果表明 CLE 的一致性和渐进正态性，但数学上严格的证明还需要进一步研究。Wang 等（2013）的证明启示我们，如果相依程度随距离的增加而迅速减小，那么 CL 估计量就是一致的，因此有理由猜想满足这一要求的异质模型的 CL 估计量依然满足渐进一致和渐进正态。另一个启发性工作是 Jin（2010）完成的，他给出了多维正态分布 CLE 的良好性质——这可视为潜变量已知情形，而本文中的模型可视为潜变量未知情形。其三，由于最优化是对多个参数同时进行，与连续因变量情形相比程序运行的时间大大增加①。因此，如何优化程序、提升估计效率也将是十分重要的。

在本文研究的基础上，未来可以在如下几个方面进行扩展。其一，实践中模型的误差项很可能是有偏的、异质的，研究误差项非标准设定下模型的估计和性质是未来的一个方向。其二，该模型可向面板扩展，扩展后能将短面板的微观数据全部囊括进来，有效解决数据量不足的问题。向时间维度扩展要特别注意时空两个维度上的关联性、关联的异质性以及面板数据中常见的固定效应和随机效应。其三，CL 被指定为等权重，但一个更有效的形式为不等权重，此时涉及权重的选择问题，可作为未来研究的方向之一。其四，本文将影响的范围限定在家庭内部，当然个体的相互影响完全可以扩展至其他范围。

参 考 文 献

[1] Larson R. W., Almeida D. M., 1999, *Emotional Transmission in the Daily Lives of Families: A New Paradigm for Studying Family Process* [J], Journal of Marriage and Family, 61 (1): 5～20.

[2] Lee L-F., 2004, *Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models* [J], Econometrica, (6): 1899.

[3] Lesage J. P., 1997, *Bayesian Estimation of Spatial Autoregressive Models* [J], International Regional Science Review, 20 (1～2): 113～129.

[4] Wang W., Lee L F., 2013, *Estimation of Spatial Autoregressive Models with Randomly Missing Data in the Dependent Variable* [J], The Econometrics Journal, 16 (1): 73～102.

[5] Wang H., Iglesias E. M., Wooldridge J. M., 2013, *Partial Maximum Likelihood Estimation of Spatial Probit Models* [J], Journal of Econometrics, 172 (1): 77～89.

① 连续情形下，对数似然函数可经变形成为只含有内生参数的“浓缩”对数似然函数（concentrated log-likelihood），最优化过程等价于只对两个参数进行有约束最优。

[6] Bhat C. R., 2014, *The Composite Marginal Likelihood (CML) Inference Approach with Applications to Discrete and Mixed Dependent Variable Models* [J], Foundations and Trends in Econometrics, 7 (1): 1～117.

[7] Bhat C. R., Zhao H., 2002, *The Spatial Analysis of Activity Stop Generation* [J], Transportation Research Part B: Methodological, 36 (6): 557～575.

[8] Narayanamoorthy S., Paleti R., Bhat C. R., 2013, *On Accommodating Spatial Dependence in Bicycle and Pedestrian Injury Counts by Severity Level* [J], Transportation Research Part B: Methodological, 55: 245～264.

[9] Bhat C. R., 2000, *A multi-level Cross-classified Model for Discrete Response Variables* [J], Transportation Research Part B: Methodological, 34 (7): 567～582.

[10] Paleti R., Bhat C. R., Pendyala R. M., et al., 2013, *Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects* [J], Transportation Research Record, 2343 (1): 86～94.

[11] Bhat C. R., Astroza S., Hamdi A. S., 2017, *A Spatial Generalized Ordered-response Model with Skew Normal Kernel Error Terms with An Application to Bicycling Frequency* [J], Transportation Research Part B: Methodological, 95: 126～148.

[12] Lacombe D. J., 2010, *Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques* [J], Geographical Analysis, 36 (2): 105～118.

[13] Aquaro M., Bailey N., Pesaran M. H., 2015, *Quasi Maximum Likelihood Estimation of Spatial Models with Heterogeneous Coefficients* [R], USC～INET, 64.

[14] Eidsvik J., Shaby B. A., Reich B. J., et al., 2014, *Estimation and Prediction in Spatial Models with Block Composite Likelihoods* [J], Journal of Computational and Graphical Statistics, 23 (2): 295～315.

[15] Tumen S., Zeydanli T., 2015, *Is Happiness Contagious? Separating Spillover Externalities from the Group-Level Social Context* [J], Journal of Happiness Studies, 16 (3): 719～744.

[16] 刘斌、李磊、莫骄：《幸福感是否会传染》[J]，《世界经济》2012 年第 6 期。

[17] 王群勇、徐伟：《你快乐所以我快乐——家庭内幸福感溢出效应研究》[J]，《中国经济问题》2019 年第 4 期。

[18] Fowler J. H., Christakis N. A., 2008, *Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis over 20 Years in the Framingham Heart Study* [R], BMJ, 337.

[19] Carr D., Freedman Vicki A., Cornman Jennifer C., et al., 2014, *Happy Marriage, Happy Life? Marital Quality and Subjective Well-being in Later Life* [J], Journal of Marriage and Family, 76 (5): 930～948.

[20] 金江：《主观幸福的经济学初探》[D]，武汉大学博士学位论文，2010。

[21] Cuñado J., De Gracia F. P., 2012, *Does Education Affect Happiness? Evidence for Spain* [J], Social Indicators Research, 108 (1): 185～196.

[22] Blanchflower D. G., Oswald A. J., 2008, *Is Well-being U-shaped Over the Life Cycle?* [J], Social Science & Medicine, 66 (8): 1733～1749.

[23] Jin Z., 2010, *Aspects of Composite Likelihood Inference* [D], Graduate Department of Statistics, University of Toronto.

[24] Cristiano V., Paolo V., 2005, *A Note on Composite Likelihood Inference and Model Selection* [J], Biometrika, 92 (3): 519～528.

[25] Katsikatsou M., Moustaki I., 2016, *Pairwise Likelihood Ratio Tests and Model Selection Criteria for Structural Equation Models with Ordinal Variables* [J], Psychometrika, 81 (4): 1046～1068.

[26] Gao X., Song P. X. K., 2010, *Composite Likelihood Bayesian Information Criteria for Model Selection in High-Dimensional Data* [J], Journal of the American Statistical Association, 105 (492): 1531～1540.

[27] Pace L., Salvan A., Sartori N., 2011, *Adjusting Composite Likelihood Ratio Statistics* [J],

Statistica Sinica, 21 (1): 129～148.

[28] Bhat C. R., 2011, *The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-based Unordered Response Choice Models* [J], Transportation Research Part B: Methodological, 45 (7): 923～939.

[29] Ferdous N., Bhat C. R., 2013, *A Spatial Panel Ordered-response Model with Application to the Analysis of Urban Land-use Development Intensity Patterns* [J], Journal of Geographical Systems, 15 (1): 1～29.

A Study on Hetero Social Interaction Model for Ordered Response

Wang Qunyong Xu Wei

(Institute of Statistical and Econometrics, School of Economics, Nankai University)

Research Objectives: To expand conventional social interaction model, and estimate the difference of happiness spillover between various groups. **Research Methods**: To add hetero terms into conventional models with ordered responses, and quantitatively describe the heterogeneity of interaction. **Research Findings**: Pairwise composite likelihood estimate (pCL) are of good properties. According to MCMC results, all parameters, including the hetero one, are asymptotically consistent and normally distributed. Information criteria can help pick up correct model if the sample size is large enough. Using the model to study the gender differences of happiness spillover, the results show that females are more sensitive than their male counterparts. As a result, health condition and level of education are more influential on females. **Research Innovations**: To combine social interaction model and heterogeneous influences in happiness study. **Research Value**: This model provides an econometric tool to describe hetero spillover effect in various fields.

Key Words: Social Interaction Model; Heterogeneity; Ordered Response

JEL Classification: C31; I31

(责任编辑：王喜峰)