时变系数空间自回归面板数据模型的极大似然估计*

邓 明

内容提要: 本文对扰动项存在跨时期的异方差、但不存在序列相关的时变系数空间自回归模型提出了极大似然的估计方法，并证明了该估计量的一致性，同时，证明了该估计量渐进服从正态分布，由此说明该估计量具有优良的大样本性质。同时，我们还对本文所提出估计量的小样本性质进行了数值模拟。本文研究表明，估计量虽然在 N 较小时偏差较大，但是随着 N 的不断增加，估计量偏差减小，体现了比较优良的渐进性质。同时，估计量的偏差会随着时期数的增加而变大，这说明本文所提出的估计方法适用于个体数较多、时期数较少的短面板数据。

关键词: 时变系数; 空间自回归模型; 极大似然估计

DOI: 10.19343/j.cnki.11-1302/c.2016.09.012

中图分类号: F222.3　　**文献标识码**: A　　**文章编号**: 1002-4565(2016)09-0096-08

The Maximum Likelihood Estimation of Time-varying Coefficient Spatial Autoregression Panel Data Model

Deng Ming

Abstract: This paper researches the time-varying coefficient spatial autoregression panel data model, whose error has heteroscedasticity over time but without time serial correlation. This paper proposes a maximum likelihood estimation (MLE) for this model and proves the consistency and asymptotic normality of this MLE method, which testifies the favorite large sample properties of the MLE method. Meanwhile, we use Monte Carlo method to simulate the small properties of the MLE method, which shows that the estimator has large bias when N is small while the bias becomes smaller and smaller over the growth of N. Furthermore, we also find that the bias will increase over the growth of T, which shows that the MLE method is more suitable for the short panel data with small T and large N.

Key words: Time-varying Coefficient; Spatial Autoregression Model; Maximum Likelihood Estimation

一、引言

“地理学第一定律”(Tobler´s First Law of Geography)认为对空间个体观测到的数据均存在空间相关性，而且这种空间相关性随距离而衰减，即距离越远，空间相关性越小，反之亦然(Tobler，1970[1])。个体观测数据空间相关性的存在，使得计量经济理论中关于样本相互独立的古典假定不再成立，因此，传统的计量方法在估计基于空间数据的计量模型时会存在估计偏误。空间计量模型通过引入空间权重矩阵(Spatial Weight Matrix)，将研究对象的空间相关性、异质性引入了模型中，较好地解决了这一问题。

与一般的计量理论研究一样，空间计量经济模型也是首先从截面模型入手，然后逐渐扩展到空间面板数据模型。与一般的截面数据模型一样，空间截面数据模型同样存在自由度损失、无法处理个体异质性等问题(Hsiao，2003[2]; Baltagi，2001[3])。而空间面板数据模型则比较好地解决了空间截面数据

* 本文获国家自然科学基金青年项目“人口老龄化下的技术进步方向与要素收入份额”(71503220)、教育部人文社会科学研究一般项目“空间似无关回归模型、参数估计、设定检验及其应用”(13YJC910003)、福建省自然科学基金项目“基于样本数据内生的空间权重矩阵: 理论与应用”(2014J01270)资助。

模型所存在上述问题,并且解决了普通面板数据模型只考虑个体异质性而忽略个体空间相关性的问题。与一般的面板数据研究类似,当前对空间面板数据模型的理论和应用研究主要集中在固定效应模型和随机效应模型,这两类模型处理空间异质性和空间相关性的方式是: 使用空间权重矩阵来控制空间相关性,使用个体截距项来控制空间(个体) 异质性。在这两类常见的空间面板数据模型中,待估计参数(包括解释变量系数和空间自回归系数) 均不会随观测个体或时期的变化而变化,也就是说,系数是不变的。但是,个体截距项仅仅能反映出不由解释变量所解释因素的个体差异,但不能反映出解释变量对被解释变量作用的个体差异,此时,也就是说,很多时候,空间异质性并不能完全由个体截距项的差异来解释,此时,可以允许解释变量系数和空间自回归系数随个体或时期变动而变动,构建变系数空间面板数据模型(Elhorst, 2003) [4]。

变系数空间面板数据模型有两种类型: 一种是系数随观测个体的变动而变动,但同一观测个体在不同时期的系数是相同的,Elhorst(2003) [4] 根据系数变动特征又将这类变系数空间面板数据模型分为固定系数模型和随机系数模型。在 Elhorst 归纳的固定系数和随机系数的空间面板数据模型中,虽然系数随观测个体变动而变动,但系数与观测个体或是个体的空间位置之间并无某种函数关系。另一种系数随空间个体变动而变动的空间面板数据模型是 Brunsdon 等(1996) [5] 提出的"地理加权回归"(Geographically Weighted Regressor, GWR) 模型,这种模型利用相邻个体的信息估计个体参数,因此模型系数是观测个体地理位置的函数。

另一种变系数空间面板数据模型的系数随时期的变动而变动,但同一时期上所有空间个体的系数不随个体变动而变动。Anselin(1988) [6] 最早基于 SUR 模型分析了此类变系数空间计量模型,所以也将其定义为 spatial SUR 模型。在其模型中,同一观测个体的系数不仅在不同时期上存在差异,而且其观测值还存在序列相关(但不存在个体相关性),因此需要对观测个体和扰动项的空间关系作进一步设定。类似于一般的截面数据空间回归模型,Anselin 提出的 spatial SUR 模型有两种设定方式,一种设定方式是空间自回归形式:

$$Y_t = \rho_t WY_t + X_t\beta_t + \varepsilon_t; E(\varepsilon_t\varepsilon'_s) = \sigma_{ts}I_N \quad (1)$$

另一种是空间误差形式:

$$Y_t = X_t\beta_t + \varepsilon_t; \varepsilon_t = \lambda_t W\varepsilon_t + u_t; E(u_t u'_s) = \omega_{ts}I_N \quad (2)$$

对于式(1) 的空间自回归形式的 spatial SUR 模型,Anselin 建议采用极大似然或是工具变量的方法进行估计; 对于式(2) 的空间误差形式的 spatial SUR 模型,Anselin 认为同样可以使用极大似然的方法进行估计。在其基础上,邓明和钱争鸣(2013) [7] 利用贝叶斯估计方法对空间自回归形式的 Spatial SUR 模型进行了估计。进一步,邓明(2013) [8] 基于 Kapoor 等(2007) [9] 所构建的误差合成空间面板数据模型建立了一种更为一般化的时变系数的空间面板数据模型,其模型利用扰动项中的空间个体成分将不同时期的方程联系起来,同时其解释变量系数和空间自回归系数均随时期变动而变动,但不随空间个体的变动而变动,因而同样是一种时变系数空间计量模型。邓明(2013) [8] 对该模型所采用的估计方法是基于 FGLS 和 GM 的多阶段估计,在保证了估计量良好渐进性质的同时还大大降低了运算量。邓明和钱争鸣(2014) [9] 将空间自回归形式和空间误差形式的 Spatial SUR 模型结合起来,构建了一种广义的 Spatial SUR 模型,同样采用一种多阶段估计策略对模型系数进行了估计。

综上来看,基于 SUR 模型是目前研究者分析时变系数空间面板数据模型的主要手段,所使用的估计策略也多是多阶段的估计策略。但是如果不同观测个体或者时期上的扰动项不存在相关性,那么就无从构建 SUR 系统,从而上述这些多阶段估计也无从展开。本文的研究即基于这样的背景而展开,本文的研究对象是具有时变系数的空间自回归面板数据模型,同时允许扰动项方差也随观察个体的变动而变动,但与 SUR 模型不同的是,扰动项并不存在序列相关性。对于本文所构建的模型,我们采用极大似然估计的方法对模型参数进行了估计,分析了所得到估计量的大样本性质,并利用 Monte Carlo 模拟讨论了估计量的小样本性质。

二、模型设定

本文研究的时变系数模型建立在传统的空间自回归模型基础之上,假设得到空间个体 $i = 1, 2, \cdots, N$ 在时刻 $t = 1, 2, \cdots, T$ 的观测,本文的时变系数空间自回归模型形式如下:

$$y_{it} = \lambda_t \sum_{j=1}^{N} w_{ij} y_{jt} + \beta'_t x_{it} + \varepsilon_{it}; i = 1, 2, \cdots, N; t = 1, 2, \cdots, T \quad (3)$$

其中，λ_t 为观测时期 t 的空间自回归系数，$x_{it} = (x_{it,1}, x_{it,2}, \cdots, x_{it,k})'$ 为 $k \times 1$ 阶的解释变量向量，$\beta_t = (\beta_{t,1}, \beta_{t,2}, \cdots, \beta_{t,k})'$ 为时期 t 的 $k \times 1$ 的解释变量系数向量，w_{ij} 为空间权重矩阵 W 中的元素 (i, j)。令 $y_i = (y_{i1}, y_{12}, \cdots, y_{iT})'$，$w_i = (w_{i1}, w_{i2}, \cdots, w_{iN})'$。将每个观测值按个体和时期堆积起来，得到如下的矩阵形式：

$$y_i = \lambda W_i y_i + B x_i + \varepsilon_i, i = 1, 2, \cdots, N \quad (4)$$

其中，$\lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_T)$ 为 $T \times T$ 的空间自回归系数矩阵，$W_i = (w'_i, w'_i, \cdots, w'_i)'$ 是将 T 个 w'_i 堆积而成的空间权重矩阵，$\beta = diag(\beta'_1, \beta'_2, \cdots, \beta'_T)$，$x_i = (x'_{i1}, x'_{i2}, \cdots, x'_{iT})'$，$\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$。令 $S_i(\lambda) = I_T - \lambda W_i$；假定 λ_0 为 λ 的真实估计值，则 $S_i(\lambda_0) = I_T - \lambda_0 W_i$。对于上述模型，为了便于模型的估计，我们作如下假设：

假设 1：T 是一个固定的正整数，N 同样为正整数，$N >> T$。

假设 2（对间权重矩阵 W 的设定）：2(a) W 为已知的、外生的矩阵，且不随时期变化；2(b) W 的主对角线元素为 0；2(c) W 有可计算的实特征根；2(d) $(I_N - \rho_t W)$ 为非奇异矩阵对于所有的 $|\lambda_t| < 1$ 均是成立的；2(e) W 和 $(I_N - \lambda_t W)$ 的行列加总均是一致有界的。

假设 1 是为了解决待估计参数过多导致的估计困难，因为本文模型中的待估计参数个数为 $T(k+2)$ 个，而且正如后文的模拟结果所显示的那样，如果时期数较多，估计量的偏误会较大。假设(a) 则是一个简易化的处理，我们自然可以允许空间权重矩阵随时期发生变动，但这样并不会对估计量性质带来实质影响；而且，基于地理距离的空间权重通常不会随时间发生变化，而基于经济关系或其他社会、文化关系的空间权重矩阵在本文所设定的短面板中也很难有太大的变动。假设(b) 中 W 的主对角线元素为 0 表示观测个体与自身是不相邻的。假设(c) 中实特征根的存在性和可计算性是为了保证后文中提出的似然函数的可计算性。假设(d) 非奇异性的设定是为了保证模型设定是完备的，以确保方程的稳定性。假设(e) 最早是由 Kelejian 和 Prucha (1999)[11] 提出的，用以将空间相关性的程度限制在一个可处理的范围内，以便采用相关的极限理论来分析估计量的渐进性质。

假设 3（关于误差项 ε 的假设）：(a) 对于所有的 $1 \leqslant t \leqslant T, 1 \leqslant i \leqslant N, N > 1, E(\varepsilon_{it}) = 0, Var(\varepsilon_{it}) = \sigma_t^2$，也就是说，扰动项存在不同时期上的异方差，但不同个体上的方差是相同的，同时，扰动项在不同时期和个体上均是相互独立的；(b) $0 < \sigma_i^2 < a < \infty$，且 ε_{it} 有任意高于四阶的矩存在，即 $E|\varepsilon_{it}|^{4+\zeta} < \infty, \zeta > 0$；(3c) 扰动项分布服从正态分布。

假设 3 是计量经济学中的一般假设，存在任意高于 4 阶的矩是为了保证能够运用中心极限定理以及方差估计的存在（Kelejian 和 Prucha，2001[12]）。

假设 4：对于任意 T、i 以及 $\lambda \in \Theta$，$S'_i(\lambda) S_i(\lambda)$ 的最小特征值大于 0，即

$$\lambda_{\min} S'_i(\lambda) S_i(\lambda) > 0 \quad (5)$$

假设 5：$T(k+2) \times 1$ 维系数矩阵 $\theta = (\lambda_1, \lambda_2, \cdots, \lambda_T, \beta'_1, \beta'_2, \cdots, \beta'_T, \sigma_1^2, \sigma_2^2, \cdots, \sigma_T^2)' \in \Theta$ 是 $T(k+2)$ 维欧氏空间 $R^{T(k+2)}$ 的一个子集。Θ 为一个有界闭集，并且包含了参数向量 θ 的真值 θ_0，θ_0 在 Θ 的内部且一致有界。

假设 6（关于解释变量的假设）：(a) 对于任意给定的 $t = 1, 2, \cdots, T$，x_{it} 是严格外生的，即 $E(x_{it}\varepsilon_{i\tau}) = 0$；(b) 对于任意 t 和 τ，有 $N^{-1} \sum_{i=1}^{N} x_{it}\varepsilon_{i\tau} \rightarrow_p 0$；(c) 对于任意 t 和 τ，协方差矩阵 $E(x_{it}x'_{i\tau}) = \sum_{t\tau}$ 是有界的，而且不会随着观测个体的变动而变动，同时，$\sum_{tt}$ 是非奇异矩阵；(d) $N^{-1} X'_t X_\tau \rightarrow_p \sum_{t\tau}$，其中 $X_t = (x_{1t}, x_{2t}, \cdots, x_{Nt})'$；(e) $\sup_t [\lambda_{\max}(N^{-1} X'_t X_\tau)] < \upsilon$，$\inf_t [\lambda_{\min}(N^{-1} X'_t X_\tau)] > 0$，其中，$\lambda_{\max}(N^{-1} X'_t X_\tau)$ 表示 $(N^{-1} X'_t X_\tau)$ 的最大特征根。

假设 6 中解释变量非随机且有界的设定是一个不失一般性的简便处理，同时考虑到了解释变量之间可能存在的个体相关性，这也是对 Anselin (1988)[3] spatial SUR 模型的一个扩展。

三、参数的极大似然估计量及其性质

（一）参数估计与识别

在假设 2 成立时，可以将式(4) 写成如下的形式：

$$y_i = (I - \lambda W_i)^{-1}(B x_i + \varepsilon_i), i = 1, 2, \cdots, N \quad (6)$$

根据假设 3(e)，我们可以得到 $y_1, y_2, \cdots, y_N$ 的

联合密度函数为:

$$\prod_{t=1}^{T}\frac{|S(\lambda)|}{\left|\sum_{\sigma_t^2}I_T\right|^{1/2}(2\pi)^{T/2}}\cdot\exp\left(-\frac{(y_t-\lambda_t\tilde{y}_t-x_t\beta_t)'(y_t-\lambda_t\tilde{y}_t-x_t\beta_t)}{2\sigma_t^2}\right)\quad(7)$$

其中,$S(\lambda)=(I_T-N^{-1}\sum_{i=1}^{N}(\lambda W_i))$,$\sum_{\sigma_t^2}=diad(\sigma_1^2,\sigma_2^2,\cdots,\sigma_T^2)$,$\beta_t=(\beta_{1t},\beta_{2t},\cdots,\beta_{Nt})'$。$x_t=(x_{1t},x_{2t},\cdots,x_{Nt})'$为时期$t$的$N\times k$阶的解释变量矩阵,$y_t=(y_{1t},y_{2t},\cdots,y_{Nt})'$为时期$t$的$N\times 1$阶的被解释变量矩阵,$\tilde{y}_t=(\tilde{y}_{1t},\tilde{y}_{2t},\cdots,\tilde{y}_{Nt})'$,$\tilde{y}_{it}=\sum_{j=1}^{N}w_{ij}y_{jt}=w'_iy_t$。根据式(6)我们可以得到如下的对数似然函数:

$$\ln(\theta)=-\frac{NT}{2}\ln(2\pi)-\frac{N}{2}\sum_{t=1}^{T}\ln\sigma_t^2+\ln|S(\lambda)|-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\lambda_t\tilde{y}_t-x_t\beta_t)'(y_t-\lambda_t\tilde{y}_t-x_t\beta_t)}{\sigma_t^2}\quad(8)$$

因此,参数 $\theta=(\lambda_1,\lambda_2,\cdots,\lambda_T,\beta'_1,\beta'_2,\cdots,\beta'_T,\sigma_1^2,\sigma_2^2,\cdots,\sigma_T^2)'$ 的极大似然估计为:

$$\hat{\theta}=\underset{\theta\in\Theta}{\operatorname{argmax}}\ln(\theta)\quad(9)$$

最大化对数似然函数(8)的一阶条件为:

$$\frac{\partial\ln(\theta)}{\partial\lambda_t}=-\sum_{i=1}^{N}tr\left[(I_T-\lambda W_i)^{-1}E_{tt}W_i\right]+\frac{\tilde{y}'_t(y_t-\lambda_t\tilde{y}_t-x_t\beta_t)}{\sigma_t^2},t=1,2,\cdots,T\quad(10)$$

$$\frac{\partial\ln(\theta)}{\partial\beta_t}=\frac{x'_t(y_t-\lambda_t\tilde{y}_t-x_t\beta_t)}{\sigma_t^2},t=1,2,\cdots,T\quad(11)$$

$$\frac{\partial\ln(\theta)}{\partial\sigma_t^2}=-\frac{T}{2\sigma_t^2}+\frac{1}{2\sigma_t^4}(y_t-\lambda_t\tilde{y}_t-x_t\beta_t)'(y_t-\lambda_t\tilde{y}_t-x_t\beta_t),t=1,2,\cdots,T\quad(12)$$

其中,E_{tt} 为 $T\times T$ 的矩阵,其第(t,t)个元素为1,其他元素为 0。对式(9)的对数似然函数求二阶导,即可得到如下的海塞因矩阵:

$$H(\theta_0)=E_0\left[-\frac{1}{N}\cdot\frac{\partial^2\ln(\theta_0)}{\partial\theta\partial\theta'}\right]=\begin{pmatrix}H_{11}&H_{12}&H_{13}\\ \cdot&H_{22}&H_{23}\\ \cdot&\cdot&H_{33}\end{pmatrix}$$

$$=\begin{pmatrix}E_0\left[-\frac{1}{N}\cdot\frac{\partial^2\ln(\theta_0)}{\partial\lambda\partial\lambda'}\right]&E_0\left[-\frac{1}{N}\cdot\frac{\partial^2\ln(\theta_0)}{\partial\lambda\partial\beta'}\right]&E_0\left[-\frac{1}{N}\cdot\frac{\partial^2\ln(\theta_0)}{\partial\lambda\partial(\sigma^{2})'}\right]\\ \cdot&E_0\left[-\frac{1}{N}\cdot\frac{\partial^2\ln(\theta_0)}{\partial\beta\partial\beta'}\right]&E_0\left[-\frac{1}{N}\cdot\frac{\partial^2\ln(\theta_0)}{\partial\beta\partial(\sigma^{2})'}\right]\\ \cdot&\cdot&E_0\left[-\frac{1}{N}\cdot\frac{\partial^2\ln(\theta_0)}{\partial(\sigma^2)\partial(\sigma^{2})'}\right]\end{pmatrix}\quad(13)$$

其中,$\beta=(\beta'_1,\beta'_2,\cdots,\beta'_T)'$,$\sigma^2=(\sigma_1^2,\sigma_2^2,\cdots,\sigma_T^2)'$。$H_{11}$ 为一个 $T\times T$ 矩阵,其形式为:

$$H_{11}=(\Psi_0\odot\Psi'_0)+diag\left[\frac{1}{\sigma_t^2}\cdot\frac{1}{N}\sum_{i=1}^{N}E_0(\tilde{y}_{it}^2)\right]\quad(14)$$

其中,

$$\frac{1}{N}\sum_{i=1}^{N}E_0(\tilde{y}_{it}^2)=\frac{1}{N}\sum_{i=1}^{N}w'_i(I_T-\lambda_0W_i)^{-1}\left[BE(x_ix'_i)B'+\sum_{\sigma_t^2}\right](I_T-\lambda_0W_i)^{-1}w_i=\frac{1}{N}\sum_{i=1}^{N}\varphi'_{0t}\left[BE(x_ix'_i)B'+\sum_{\sigma_t^2}\right](I_T-\lambda_0W_i)^{-1}\varphi_{0t}\quad(15)$$

其中,$\Psi_0=N^{-1}\sum_{i=1}^{N}(W_i(I_T-\lambda_0W_i)^{-1})$,$\varphi'_{0t}$ 为 Ψ_0 的第 t 行向量,$\odot$为矩阵的 Hadamard 积。H_{12} 为一个 $T\times kT$ 矩阵,$H_{12}=diag_t(\sigma_t^{-2}E_0(N^{-1}\tilde{y}'x_t))$,其中,

$$E_0(N^{-1}\tilde{y}'x_t)=E_0(N^{-1}\sum_{i=1}^{N}\tilde{y}_{it}x'_{it})=w'_i(I_T-\lambda_0W_i)\beta E(x_ix'_{it})=\varphi'_{0t}(\sum_{\sigma_{1t}^2}\beta_1,\sum_{\sigma_{2t}^2}\beta_2,\cdots,\sum_{\sigma_{Tt}^2}\beta_T)'\quad(16)$$

$H_{13}=diag_t(N^{-1}\sum_{i=1}^{N}\sigma_t^{-2}w'_i(I_T-\lambda_0W))e_t=diag_t(\sigma_t^{-2}\varphi_{0,tt})$ 为一个 $T\times T$ 矩阵。$H_{22}=diag_t(\sigma_t^{-2}\sum_{\sigma_{tt}^2})$ 为一个 $Tk\times Tk$ 矩阵。$H_{23}=0$,$H_{33}=diag_t(1/2\sigma_t^4)$。对于一个给定的 T,当 $N\to\infty$ 时,在假设 1、2、4、5 和 6 下,当 λ_t 被识别时,β'_t 和 $\sigma_t^{2\prime}$ 是可被识别的。因此,模型参数能否被识别的前提是 λ_t 能够被识别。为了推导出 λ_t 的识别条件,我们将海塞因矩阵 $H(\theta_0)$ 改写成如下的形式:

$$H(\theta_0)=\begin{pmatrix}H_{11}&H_{12}^*\\ \cdot&H_{22}^*\end{pmatrix}\quad(17)$$

其中,$H_{12}^*=(H_{12},H_{13})$ 为一个 $H_{12}^*=(H_{12},H_{13})$ 为一个 $T\times(Tk+T)$ 矩阵,$H_{22}^*=(H_{22},H_{33})$ 为一个 $(Tk+T)\times(Tk+T)$ 矩阵。$\sqrt{N}\hat{\lambda}_N$ 的渐进协方差矩阵为可以表示为$(H_{11.2}^*)^{-1}$,其中 $H_{11.2}^*=H_{11}-H_{12}^*(H_{22}^*)^{-1}H_{21}^*=H_{11}-H_{12}H_{22}^{-1}H_{21}-H_{13}H_{33}^{-1}H_{31}$。根据式(14)和(15)可得,

$$H_{11}=(\Psi_0\odot\Psi'_0)+diag[\sigma_t^{-2}\varphi'_{0t}(B\sum_{xx}B'+\sum_{\sigma_t^2})\varphi_{0t}] \quad (18)$$

其中，矩阵 $B\sum_{xx}B'$ 的第 (m,n) 个元素为 $\beta'_m\sum_{mn}\beta_n$，$\varphi_{0t}\sum_{\sigma_t^2}\varphi_{0t}=\sum_{m=1}^{T}\sigma_m^2\varphi_{0,tm}^2$，因此式(18)可以进一步改写为如下形式：

$$H_{11}=(\Psi_0\odot\Psi'_0)+diag[\sum_{m=1}^{T}(\sigma_m^{-2}/\sigma_t^{-2})\varphi_{0,tm}^2]+diag[\sigma_t^{-2}\sum_{m=1}^{T}\sum_{n=1}^{T}\varphi_{0,tm}\varphi_{0,tn}\beta'_m\Sigma_{mn}\beta_m] \quad (19)$$

同理可得，

$$H_{12}H_{22}^{-1}H_{21}=diag\begin{bmatrix}\sigma_t^{-2}(N^{-1}\sum_{i=1}^{N}(w'_i(I_T-\lambda_0W_i)^{-1}BE(x_ix'_{it})))\sum_{tt}^{-1}\\(N^{-1}\sum_{i=1}^{N}(E(x_ix'_{it})B'(I_T-\lambda_0W_i)^{-1}w_i))\end{bmatrix}$$

$$=diag[\sigma_t^{-2}\sum_{m=1}^{T}\sum_{n=1}^{T}\varphi_{0,mt}\varphi_{0,nt}\beta'_m\Sigma_{mt}\Sigma_{tt}^{-1}\Sigma_{nt}\beta'_n] \quad (20)$$

$$H_{13}H_{33}^{-1}H_{31}=2diag[\varphi_{0,tt}^2] \quad (21)$$

将式(19)~(21)代入到 $H_{11.2}^*$ 的表达式中可得，

$$H_{11.2}^*=(\Psi_0\odot\Psi'_0)+diag[-\varphi_{0,tt}^2+\sum_{n=1,n\neq t}^{T}(\sigma_n^2/\sigma_t^2)\varphi_{0,nt}^2]+diag[\sigma_t^{-2}\sum_{m=1}^{T}\sum_{n=1}^{T}\varphi_{0,mt}\varphi_{0,nt}\beta'_m\Sigma_{mt}\Sigma_{tt}^{-1}\Sigma_{nt}\beta'_n] \quad (22)$$

此时，λ_t 能够被识别(亦即所有参数可被识别)的条件即为：$Rank(H_{11.2}^*)=T$。因此，当 T 较大时，识别 λ_t 需要满足：

$$\lambda_{\min}(H_{11.2}^*)>\kappa>0,\lambda_{\max}(H_{11.2}^*)<\zeta<\infty,\forall T \quad (23)$$

其中，$\lambda_{\min}(\Lambda)$ 和 $\lambda_{\max}(\Lambda)$ 分别表示矩阵 Λ 的最小和最大特征根。当 T 不是无穷大时，上述条件中的第一个是一个普通的秩条件。因此我们重点分析第二个条件是否得到满足。考虑到矩阵 $H_{11.2}^*$ 为一个对称的正定矩阵，因此有：

$$\lambda_{\max}(H_{11.2}^*)=|\lambda_{\max}(H_{11.2}^*)|\leqslant\|H_{11.2}^*\|_\infty \quad (24)$$

将式(22)代入上式可得：

$$\|H_{11.2}^*\|\leqslant\|\Psi_0\odot\Psi'_0\|_\infty+\sup_t\|-\varphi_{0,tt}^2+\sum_{n=1,n\neq t}^{T}(\sigma_n^2/\sigma_t^2)\varphi_{0,nt}^2\|+\sup_t\|\sigma_t^{-2}\sum_{m=1}^{T}\sum_{n=1}^{T}\varphi_{0,mt}\varphi_{0,nt}\beta'_m\Sigma_{mt}\Sigma_{tt}^{-1}\Sigma_{nt}\beta'_n\| \quad (25)$$

其中，$\|\Psi_0\|_1=\max_{1\leqslant i\leqslant N}(\sum_{i=1}^{N}|\varphi_{0,ij}|)$，$\|\Psi_0\|_\infty=\max_{1\leqslant j\leqslant N}(\sum_{i=1}^{N}|\varphi_{0,ij}|)$。在下面的分析之前，我们先引入下面的引理：

引理 1：在假设 2 成立时，令 $\Psi_0=N^{-1}\sum_{i=1}^{N}(W_i(I_T-\lambda_0W_i)^{-1})$，则 $\|\Psi_0\|_1=\max_{1\leqslant i\leqslant N}(\sum_{i=1}^{N}|\varphi_{0,ij}|)$ 和 $\|\Psi_0\|_\infty=\max_{1\leqslant j\leqslant N}(\sum_{i=1}^{N}|\varphi_{0,ij}|)$ 是有界的。

证明：根据假设(2)，我们有 $\Psi_0=W+W\Psi_0W+W\Psi_0W\Psi_0W+\cdots$。从而有：

$$\|\Psi_0\|_1\leqslant\|W\|_1+\|W\|_1^2\|\Psi_0\|_1+\|W\|_1^3\|\Psi_0\|_1^2+\cdots$$

由于 $\|\Psi_0\|_1^\tau=(\sup_t|\lambda_t|)^\tau$，同时，在假设 2 下，我们有，$\sup_t|\lambda_t|\|W\|_1<1$，因此可以得到，$\|\Psi_0\|_1\leqslant\|W\|_1(1-\sup_t|\lambda_t|\|W\|_1)^{-1}$。同理可得：$\|\Psi_0\|_\infty\leqslant\|W\|_\infty(1-\sup_t|\lambda_t|\|W\|_\infty)^{-1}$。根据假设 2 可知，$\|W\|_1$ 和 $\|W\|_\infty$ 是有界的，而 $1-\sup_t|\lambda_t|\|W\|_1$ 和 $1-\sup_t|\lambda_t|\|W\|_\infty$ 均大于零。由此可得，$\|\Psi_0\|_1=\max_{1\leqslant i\leqslant N}(\sum_{i=1}^{N}|\varphi_{0,ij}|)$ 和 $\|\Psi_0\|_\infty=\max_{1\leqslant j\leqslant N}(\sum_{i=1}^{N}|\varphi_{0,ij}|)$ 是有界的。

当 $\|\Psi_0\|_1$ 和 $\|\Psi_0\|_\infty$ 均是有界时，则 $\|\Psi_0\odot\Psi'_0\|_\infty$ 同样也是有界的。而对于式(25)右边的第二项，有：

$$\|-\varphi_{0,tt}^2+\sum_{n=1,n\neq t}^{T}(\sigma_n^2/\sigma_t^2)\varphi_{0,nt}^2\|_\infty\leqslant\frac{\sup_t(\sigma_t^2)}{\inf_t(\sigma_t^2)}\cdot\sum_{n=1}^{T}\varphi_{0,nt}^2<\zeta<\infty \quad (26)$$

上式表明式(25)右边的第二项同样是有界的，最后我们来看式(25)右边的最后一项，可以得到：

$$\|\sigma_t^{-2}\sum_{m=1}^{T}\sum_{n=1}^{T}\varphi_{0,mt}\varphi_{0,nt}\beta'_m\Sigma_{mt}\Sigma_{tt}^{-1}\Sigma_{nt}\beta'_n\|_\infty$$
$$\leqslant\sigma_t^{-2}\sum_{m=1}^{T}\sum_{n=1}^{T}|\varphi_{0,mt}||\varphi_{0,nt}|\|\beta'_m\Sigma_{mt}\Sigma_{tt}^{-1}\Sigma_{nt}\beta'\|_\infty$$
$$\leqslant\sigma_t^{-2}\sum_{m=1}^{T}\sum_{n=1}^{T}|\varphi_{0,mt}||\varphi_{0,nt}|\cdot\sup_{m,n}\|\beta'(\Sigma_{mn}-\Sigma_{mt}\Sigma_{tt}^{-1}\Sigma_{nt})\beta'\|_\infty$$
$$\leqslant\sigma_t^{-2}\sup_m\|\beta_m\|_1\cdot\sup_n\|\beta_n\|_\infty\cdot\sup_{m,n}\|(\Sigma_{mn}-\Sigma_{mt}\Sigma_{tt}^{-1}\Sigma_{nt})\|_\infty\cdot(\sum_{n=1}^{T}|\varphi_{0,mt}|) \quad (27)$$

根据前文的假设 3、5、6 可得，$\inf\sigma_t^2>0$；同时，$\sup_m\|\beta_m\|_1$、$\sup_n\|\beta_n\|_\infty$、Σ_{mn} 和 $\sum_{tt}^{-1}$ 均存在且均是有界的；此外，在前文的设定条件下，

$sip_t \sum_{n=1}^{T} |\varphi_{0,nt}| = \|\Psi_0\|_\infty$ 也是有界的。因此,式(25) 右边的最后一项是有界的,式(23) 是成立的,从而我们可以得到如下的命题 1。

命题 1: 在假设 1、2、4、5 和 6 成立时,且 $\lambda_{\min}(H_{11.2}^*) > \kappa > 0, \forall T$ 时,式(4) 给出的时变系数空间面板数据模型的极大似然估计是可识别的。

(二) 估计量的性质

我们首先来讨论由式(9) 得到的极大似然估计量的一致性。对于真实参数 θ_0,我们有: $\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmax}} E_0 [N^{-1}\ln(\theta)]$。根据命题1 可知,上述 θ_0 是唯一存在的。因此可以得到:

$$\lim_{N\to\infty} E_0 [N^{-1}\ln(\theta_0)] \geqslant \lim_{N\to\infty} E_0 [N^{-1}\ln(\theta)] \quad (28)$$

在假设 4 成立时,有,

$$p\lim_{N\to\infty} [N^{-1}\ln(\hat{\theta})] = \lim_{N\to\infty} E_0 [N^{-1}\ln(\theta)] \quad (29)$$

联立式(28) 和(29) 可得,

$$p\lim_{N\to\infty} [N^{-1}\ln(\hat{\theta})] = \lim_{N\to\infty} E_0 [N^{-1}\ln(\theta)] \quad (30)$$

由于 θ_0 是最大化 $E_0 [N^{-1}\ln(\theta)]$ 所得到的唯一值,因此 $p\lim_{N\to\infty} [N^{-1}\ln(\hat{\theta})] = \theta_0$。从而可以得到如下的命题 2:

命题 2: 在假设 1 ~6 成立时,且 $\lambda_{\min}(H_{11.2}^*) > \kappa > 0, \forall T$ 时,式(3) 给出的时变系数空间面板数据模型的极大似然估计量 $\hat{\theta}$ 是一个一致估计量,即 $p\lim_{N\to\infty} [N^{-1}\ln(\hat{\theta})] = \theta_0$。

接下来我们讨论估计量的渐进性质。我们将式(8) 的极大似然函数的一阶条件在均值处展开可得:

$$\frac{1}{\sqrt{N}} \cdot \frac{\partial \ln(\hat{\theta})}{\partial \theta} = \frac{1}{\sqrt{N}} \frac{\partial \ln(\theta_0)}{\partial \theta} - \sqrt{N} H(\bar{\theta})(\hat{\theta} - \theta_0) = 0 \quad (31)$$

其中,$H(\bar{\theta}) = -N^{-1}(\partial^2 \ln(\bar{\theta}) / \partial\theta\partial\theta')$,根据前面得到的 $p\lim_{N\to\infty} [N^{-1}\ln(\bar{\theta})] = \theta_0$,我们可以得到:

$$H(\bar{\theta}) \to_p E_0 \left(-\frac{1}{N} \cdot \frac{\partial^2 \ln(\theta_0)}{\partial\theta\partial\theta'} \right) = H(\theta_0) \quad (32)$$

根据前文分析,$H(\theta_0)$ 是一个正定矩阵。因此,根据式(31) 可得,$\sqrt{N}(\hat{\theta} - \theta_0)$ 极限分布是由 $N^{-1/2}(\partial \ln(\theta_0) / \partial\theta)$ 的极限分布决定的,$\partial \ln(\theta_0) / \partial\theta$ 同样可以由前文的式(10) ~(12) 给出。根据式(11) 可得,对解释变量系数的一阶导仅仅是在被解释变量前面乘上外生的空间权重矩阵,因此,解释变量系数 β 的极大似然估计量的渐进性质与一般的面板数据的极大似然估计的渐进性质没有区别,均是渐进无偏的和渐进服从正态分布①。因此,此处我们重点讨论时变的空间自回归系数 λ_t 的渐进性质。根据式(10),我们可以得到下式:

$$\frac{1}{\sqrt{N}} \frac{\partial \ln(\theta_0)}{\partial \lambda_t} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{it} \quad (33)$$

其中,$\xi_{it} = e'_t \Psi_0 \zeta_i \zeta_{it} - \varphi_{0,tt}$,$e_t$ 为第 t 个元素为 1、其他元素为 0 的 T 维列向量,$\zeta_i = \varepsilon_i / \sigma_0 = (\zeta_{i1}, \zeta_{i2}, \cdots, \zeta_{iT})$。根据前文对误差项的假设可知,$E(\zeta_i) = 0$,$E(\zeta_i \zeta_{it}) = e_i$,$Var(\zeta_i) = I_T$。同样可得,$E(\xi_{it}) = 0$,并且有: $E(\xi_{it}^2) = e'_t \Psi_0 E(\zeta_i \zeta'_i \zeta_{it}^2) \Psi'_0 e_t - \varphi_{0,tt}^2$ 由此,我们可以得到:

$$Var(\xi_{it}) = \varphi_{0,tt}^2 [E(\zeta_{it}^4) - 2] + \sum_{s=1}^{T} \varphi_{0,ts}^2 \quad (34)$$

$$E\left[\frac{1}{\sqrt{N}} \frac{\partial \ln(\theta_0)}{\partial \lambda_t}\right] = E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{it}\right] = 0 \quad (35)$$

根据前文的假设 3 可知,不同观测个体的扰动项是相互独立的,因此对于不同观测个体而言,ξ_{it} 同样是相互独立的。此外,根据命题(3b) 并结合中心极限定理可得,$N^{-1/2}(\partial \ln(\theta_0) / \partial \lambda_t)$ 渐进服从均值为 0 的正态分布,$\sqrt{N}(\hat{\lambda}_t - \lambda_{t0})$ 同样渐进服从均值为 0 的正态分布; 此外,前文已经分析得到,$\sqrt{N}\hat{\lambda}_N$ 的渐进协方差矩阵为可以表示为 $(H_{11.2}^*)^{-1}$,因此,$\sqrt{N}(\hat{\lambda}_t - \lambda_{t0})$ 的渐进协方差矩阵同样为 $(H_{11.2}^*)^{-1}$。由此我们可以得到如下的命题 3:

命题 3: 在假设 1 ~6 成立时,且 $\lambda_{\min}(H_{11.2}^*) > \kappa > 0, \forall T$ 时,式(3) 给出的时变系数空间面板数据模型的空间自回归系数的极大似然估计量 $\hat{\lambda}_t$ 是一个一致估计量,即当 $N \to \infty$ 时,有:

$$\sqrt{N}(\hat{\lambda}_t - \lambda_{t0}) \to_d N [0, (H_{11.2}^*)^{-1}] \quad (36)$$

上述命题在 $T \to \infty$ 时也是成立的,只要满足 $\sum_s^T \varphi_{0,ts}^2$ 是有界的这一条件即可,而这一条件在矩阵 Ψ_0 是行列加总是一致有界时是成立的,而后者在假设 2(e) 下是成立的。此外,即使扰动项不满足正态分布,根据 Kelejian 和 Prucha(2001) [12] 的结论,只要 ε_{it} 有任意高于四阶的矩存在,即 $E|\varepsilon_{it}|^{4+\zeta} <$

① 如果在解释变量 X 中包含了元素为 1 的列向量,那么本文的模型可以看成是固定效应模型,关于固定效应空间面板数据模型的极大似然估计可以参见 Lee 和 Yu(2010) [13] 的研究结论。

∞，$\zeta > 0$，$\hat{\lambda}_t$ 同样会渐进服从正态分布。

四、Monte Carlo 模拟

(一) Monte Carlo 模拟设计

前文已经证明，本文对时变系数空间自回归面板数据模型所提出的极大似然估计量具有较好的大样本性质，下面我们对该估计量进行 Monte Carlo 数值模拟，讨论其小样本性质。模拟过程设计如下:

(1) 假设模型中只有两个解释变量，一个为常数项，一个为外生的解释变量 x，数据生成过程如下所示:

$$y_{it} = \alpha_t + \lambda_t \sum_{j=1}^{N} w_{ij} y_{jt} + \beta_t x_{it} + \varepsilon_{it}; i = 1, 2, \cdots, N; t = 1, 2, \cdots, T \quad (37)$$

(2) 两个解释变量相对应的真实参数分别记为 α_t 和 β_t。两个变量均服从独立的标准正态分布 $N(0, 1)$。

(3) 前文的大样本性质是基于 $N \to \infty$ 得到的，因此在模拟中，设置相对较小的 T，考虑 $T = 2$ 和 $T = 5$ 两种情况。样本容量 N 分别取 10、25、50 和 100 四种情况。

(4) 当 $T = 2$ 时，待估计参数共有 6 个①，分别为 $(\lambda_1, \lambda_2, \alpha_1, \alpha_2, \beta_1, \beta_2)$，下标表示不同时期，其参数初始值设为 $(0.4, 0.6, 0.5, 1, 1.5, 2)$；当 $T = 5$ 时，待估计参数共有 15 个，分别为 $(\lambda_1, \cdots, \lambda_5, \alpha_1, \cdots, \alpha_5, \beta_1, \cdots, \beta_5)$，设定其初始值分别为 $(0.3, 0.4, 0.5, 0.6, 0.7; 1.6, 1.8, 2, 2.2, 2.4; 0.6, 0.8, 1, 1.2, 1.4)$。

(5) 空间权重矩阵 W 为 Rook 相邻型空间权重矩阵，且随机生成②。

(6) 假定扰动项有两种生成过程，一种是服从正态分布，即 $\varepsilon_{it}/\sigma_{0t} \sim i.i.d.\ N(0,1)$。另一种是非正态分布，假定其生成形式为: $\varepsilon_{it}/\sigma_{0t} \sim i.i.d.\ [\chi^2(2) - 2]/2$，并设定在两种扰动项生成过程下，$\sigma_{0t}^2 = 1$。

(二) Monte Carlo 模拟结果

根据上面的 Monte Carlo 模拟设计，使用 Matlab 软件进行了 1000 次独立的数值模拟，并记录下来系数估计值的模拟结果。基于所得到的各个系数的 1000 个模拟结果，计算每个系数模拟值的均值和样本均方误平方根(RMSE)，其中样本均方误平方根定义如下:

$$RMSE = \left[(\hat{\theta}_{0.5} - \theta_0)^2 + \left(\frac{\hat{\theta}_{0.75} - \hat{\theta}_{0.25}}{1.35} \right) \right]^2 \quad (38)$$

其中，θ_0 是待估计参数的真实值，$\hat{\theta}_{0.25}$、$\hat{\theta}_{0.5}$ 和 $\hat{\theta}_{0.75}$ 分别是系数估计 1000 次值模拟结果的上四分位数、中位数和后四分位数。

根据模拟结果可以得到如下结论: 第一，当样本量较少时($N = 10$)，根据前文估计方法得到的参数估计量模拟值与参数真值之间偏差较大，RMSE 的度量结果表明，最高的是 $T = 5$ 时非正态分布下 β_{40} 的 RMSE，达到了 49.9%，这个偏差是比较大的，这说明尽管本文提出的估计量具有很好的渐进性质，但在小样本下，依然具有较大的估计偏误。但是随着观测个体数量的增加，估计量的偏误不断减小，当 $N = 50$ 且时期数较少($T = 2$)时，绝大部分系数估计量的量的 RMSE 在 10% 以下; 而当 $N = 100$ 时，所有变量在各种情形下的 RMSE 均在 10% 以内，这反映了本文所提出的极大似然估计具有较好的渐进性质。第二，对于同样的 N，随着 T 的增加，估计量的 RMSE 有所增加。例如当 $N = 50$ 且 $T = 2$ 时，绝大部分变量的 RMSE 在 10% 以下，而当 $N = 50$ 且 $T = 5$ 时，RMSE 则不能控制在 10% 以内。这说明本文提出的时变系数空间面板数据模型的估计方法虽然能估计空间面板数据模型的时变系数，但随着时期数的增加，待估计参数增加，估计量的偏差有所增加。第三，对于同样的 N 和 T，正态分布下的 RMSE 要小于非正态分布下的 RMSE，这说明虽然扰动项的正态分布不是估计量渐进服从正态分布的必要条件，但是，在小样本下，扰动项的正态分布能够减少估计量的偏差。

五、结束语

空间计量模型尤其是空间面板数据模型是当前计量经济模型的一个重要分支，当前对空间面板数据模型的研究侧重于常系数模型的研究，而对变系数模型的研究相对较少。本文构建了一种系数(包括空间自回归系数和解释变量系数)随时期变动、但不随观测个体变动的空间自回归面板数据模型，种模型的应用价值在于可以考察变量

---

① 此处没有包含扰动项方差，如果加上扰动项方差，待估计参数共 $4T$ 个，下文将介绍扰动项的设定。

② 其具体生成过程参见郭鹏辉(2009)[14]、邓明(2013)[5]。

之间的关系以及个体之间的空间关系的动态特征。已有的研究主要侧重于利用似无关回归模型(SUR)来研究时变系数的空间面板数据模型(邓明和钱争鸣,2013[7];邓明,2013[8];邓明和钱争鸣,2014[10]),而估计策略也多是多阶段估计。但是,构建 SUR 模型需要扰动项之间存在序列相关性,当扰动项仅存在时期上的异方差但不存在序列相关性时,则可以简化分析,直接对模型进行极大似然估计。

本文对扰动项不存在序列相关、仅仅存在时期异方差的时变系数空间自回归模型提出了极大似然的估计方法,同时,也证明了本文所提出的极大似然估计量是一个渐进服从正态分布的一致估计量,由此说明该估计量所具备的优良的大样本性质。本文还对估计量的小样本性质进行了数值模拟,模拟结果表明,估计量虽然在 N 较小时偏差较大,但是随着 N 的不断增加,估计量偏差减小,体现了比较优良的渐进性质。此外,模拟结果还表明,估计量的偏差会随着时期数的增加而增加,这也说明本文所提出的极大似然估计方法适用于观测个体较多但时期数较少的短面板数据,这也是我们在假设 1 中要求 $N>>T$ 的原因之一。

参考文献

[1] Tobler W R. A Computer Movie Simulating Urban Growth in the Detroit Region [J]. Economic Geography, 1970, 46(2): 234-240.

[2] Hsiao C. Analysis of panel data (Second Edition) [M]. Cambridge University Press, Cambridge, UK, 2003.

[3] Baltagi B H. Econometric Analysis of Panel Data (Second Edition) [M]. John Wiley & Sons, Chichester, United Kingdom, 2001.

[4] Elhorst J P. Specification and Estimation of Spatial Panel Data Models [J]. International Regional Science Review, 2003, 26(3): 244-268.

[5] Brunsdon C, Fotheringham A S, Charlton M E. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity [J]. Geographical Analysis, 1996, 28(4): 281-298.

[6] Anselin L. Spatial Econometrics: Methods and Models [M]. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1988.

[7] 邓明,钱争鸣. 我国省际知识生产及其空间溢出的动态时变特征 [J]. 数理统计与管理, 2013(4): 111-123.

[8] 邓明. 时变系数的空间误差合成模型——基于 FGLS 和 GM 的多阶段迭代估计 [J]. 数量经济技术经济研究, 2013(4): 111-123.

[9] Kapoor M, Kelejian H H, Prucha I R. Panel Data Models with Spatially Correlated Error Components [J]. Journal of Econometrics, 2007, 140(1): 97-130.

[10] 邓明,钱争鸣. 混合形式的变系数空间面板数据模型:一个多阶段估计 [J]. 数理统计与管理, 2014(3): 571-585.

[11] Kelejian H H, Prucha I R. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model [J]. International Economic Review, 1999, 40(2): 509-533.

[12] Kelejian H H, Prucha I R. On the Asymptotic Distribution of the Moran I Test Statistic with Applications [J]. Journal of Econometrics, 2001, 104(2): 219-257.

[13] Lee L F, Yu J. Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects [J]. Journal of Econometrics, 2010, 154(2): 165-185.

[14] 郭鹏辉. 一类动态空间固定效应模型研究 [D]. 厦门大学博士学位论文, 2009.

作者简介

邓明,男,2010 年毕业于厦门大学统计系,获经济学博士学位,现为厦门大学经济学院副教授、硕士生导师。研究方向为公共经济、空间计量经济。

(责任编辑: 曹 麦)