误差分量动态经济模型未知参数的几种处理方法

童光荣

由确定性作用转移到随机作用的形式所产生的误差分量模型的估计方法与常规形式是不完全一样的，特别是对于建立在模型的自回归部分相互作用下的序列相关性之上的随机作用的动态模型更是如此。这里我们不妨考虑干扰项的二阶结构的特殊情形给出几种具体的技巧性处理方法。

1. 模型

我们假定的误差分量动态模型的形式为：

$$Y_{it} = \sum_{i=1}^{p} \delta_j Y_{it-j} + \sum_{k=1}^{k} X_{kit}\beta_k + u_{it} \qquad (1.1)$$

$$u_i = \mu_i + v_{it},\ i = 1, 2, \cdots, N,\ t = 1, 2, \cdots, T$$

记成矩阵形式，就有：

$$Y = Y_{[P]} + X\beta + U \qquad (1.2)$$

$$EU = 0, VU = \Omega = \sigma_v^2 W_N + (\sigma_v^2 + T\sigma_\mu^2) B_N$$

$$= \sigma_v^2 [W_N + \frac{1}{Q^2} B_N]$$

其中 $Q^2 = \sigma_v^2 / (\sigma_v^2 + T\sigma_\mu^2)$；$Y_{[P]}$ 是关于滞后内生变量的 $NT \times P$ 维的观测值矩阵；X 是关于外生变量的 $NT \times K$ 维的观测值矩阵。

2. 处理方法

2.1. λ— 类估计法

根据 Maddala 定义，我们把所有属于一般类的使用误差分量模型的估计方法称为"λ— 类"估计法。

λ— 类估计量是按转换模型：

$$(W_N + \sqrt{\lambda} B_N) Y = (W_N + \sqrt{\lambda} B_N) Y_{[P]} \delta + (W_N + \sqrt{\lambda} B_N) \times \beta + (W_N + \sqrt{\lambda} B_N) u \qquad (1.3)$$

的最小二乘估计来计算。

对于每一个 $\lambda \in [0, +\infty]$，我们可以得到 $\hat{\delta}(\lambda), \hat{\beta}(\lambda)$。这类估计量包括误差分量模型所有的古典估计量。譬如当 $\lambda = 0$ 时的 WITHIN 估计量，当 $\lambda = 1$ 时的最小二乘估计，当 $\lambda = Q^2$ 时的工具变量估计，当 $\lambda = \infty$ 时的中间估计等等。

如果 $\delta = 0$，所有的这些估计量是一致的，如果 $\delta \neq 0$，几乎所有 λ— 类估计量是渐近无偏的。我们先不考虑这些无偏性的结构，而首先注意最初的动态误差分量模型，即

$P = 1, \delta_{[1]} = \delta \neq 0, \beta = 0$ 时，其模型为：

$$Y = \delta Y_{it-1} + \mu_i + v_i \qquad (1.4)$$

跟以前表述的方式一样，可以得到递归方程如下：

$$Y_{it} = \delta_{i0} + \frac{1-\delta^t}{1-\delta}\mu_j + \sum_{j=0}^{t-1} \delta v_{it-j} \qquad (1.5)$$

为了近似计算的方便，我们将假定 $Y_{i0}, i = 1, 2, \cdots, N$ 是独立同分布的特征变量，其二阶矩为 EY_{i0}^2，它与 μ_i 的相关系数是：$EY_{i0}\mu_j$。

那么，毫不疑问 $EY_{i0}{}^2$ 和 $EY_{i0}\mu_{i0}$ 的任何 λ— 估计量是渐近无偏的。根据强大数定量，我们可以计算 $\delta_\infty(\lambda) = \lim_N \delta(\lambda)$。但是由于 $\delta_\infty(\lambda)$ 的复杂性，这里一般不能绘出它的具体表达式，然而，根据初始值不同的假定，以一个简单的例子可以说明无偏性的问题。例如，我们取一形式如同 (1.4) 的纯自回归模型，就有：当 $T = 10, P = \sigma_\mu^2 / (\sigma_N{}^2 + \sigma_v^2) = 0.5, \delta = 0.9$ 时，如果

$$Y_{i0} = u_i + v_{i0},$$

$$P \lim_{N\to\infty} \hat{\delta}(0) - \delta = -0.080$$

如果 $Y_{i0} = \dfrac{u_i}{1-\delta} + \dfrac{v_{i0}}{\sqrt{1-\delta^2}}$ 时，那么

$$P \lim_{N\to\infty} \hat{\delta}(0) - \delta = -0.243$$

这是都是对 WITHIN 估计而言的。对于最小二乘估计，Y_{i0} 在以上相同的假定之下分别有：

$$P \lim_{N\to\infty} \hat{\delta}(1) - \delta = 0.176$$

$$P \lim_{N\to\infty} \hat{\delta}(1) - \delta = 0.095$$

这些例子清楚地说明，初始值不同的假定对这些估计无偏性的影响是非常重要的。

在这里，我们可以得出下面的结果：对任意的 $EY_{i0}\mu_i$，$\delta(\lambda)$ 是 λ 的递增函数：

$$\lim_N \hat{\delta}(0) < \lim_N \hat{\delta}(Q^2) < \lim_N \hat{\delta}(1) < \lim_N \hat{\delta}(\infty)$$

由此清楚地知道存在一个 $\lambda^* \in [0, Q^2]$，使得 $\lim_N \hat{\delta}(\lambda^*) = \delta$。SEVESTER 和 TROGNON 曾得出 λ^* 的值如下：

$$\lambda^* = k(1-\rho) / [\frac{1-\delta^T}{a-\delta} \frac{EY_{i0}\mu_i}{\sigma^2} + K(1-\rho+t\rho)] \qquad (1.6)$$

其中，$K = (T - 1 - T\delta + \delta^T) / T(1-\delta)^2$

$$\rho = \sigma_\mu{}^2 / (\sigma_\mu{}^2 + \sigma_v{}^2),$$

$$\sigma^2 = \sigma_\mu{}^2 + \sigma_v{}^2$$

一般的情况下，$\lambda^* \neq Q^2$，这就确定了一个这样模型的工具

变量估计的半 — 不一致性。

2.2. 工具变量和推广的瞬间估计方法

为了简便起见,我们考虑 $P = k = 1$ 时的情形,初始值定义为:

$$Y_{i0} = C + k_1\mu_i + k_2 v_{i0} = c + u_{i0}, u_{i0} = k_1\mu_i + k_2 v_{i0}$$

$$Y_{i1} = \delta Y_{i0} + b x_{i1} + \mu_i + v_{i1}$$

$$\vdots$$

$$Y_{iT} = \delta Y_{it-1} + b x_{iT} + \mu_i + v_{i1}$$

这个模型出现以 $T + 1$ 个内生变量 $(Y_{i0}, \cdots, Y_{iT})$ 和 $T + 1$ 个外生变量(常数项和 $x_{i1} \cdots x_{iT}$)的三角形模型。这个 $T + 1$ 个方程可以简记为矩阵的形式: $A Y_i - B X_i = \eta$ (2.2)

其中: $Y_i = (Y_{i0}, Y_{i1}, \cdots, Y_{iT})'$,

$X_i' = (1, X_{i1}, \cdots, X_{iT})$

$\eta' = (u_{i0}, \mu_i + v_{i1}, \cdots, \mu_i + v_{iT})$

A 和 B 是 $(T + 1) \times (T + 1)$ 阶的矩阵。

例如,当 $T = 2$ 时, A 和 B 是:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -\delta & 1 & 0 \\ 0 & -\delta & 1 \end{bmatrix}; \quad B = \begin{bmatrix} c & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{bmatrix}$$

按(2.1)和(2.2)的要求,我们可以重新记为: $\Delta D_i = \eta$,这里 $\Delta = (A, -B), D_j' = (Y_i', X_i')$,其中 $E\eta = 0, V\eta = \Omega$,

$$\Omega = \begin{bmatrix} w & \tau' \\ \tau & v \end{bmatrix}; v = \sigma_\mu^2 J_T + \sigma_v^2 I_T$$

$$w = k_1^2\sigma_\mu^2 + k_2^2\sigma_v^2, \tau' = k_1\sigma_\mu^2(1, \cdots, 1)$$

这样对特殊的模型都可以表示成任意的自回归误差分量模型。动态误差分量模型(1.1)的具体形式可能简述得更加严密:

$$\Delta D' = U' \qquad (2.3)$$

其中 $D = (\tilde{Y}, \tilde{X})$ (2.4)

这里 $\tilde{Y}$ 是 $N \times (T + p)$ 维内生变量的观测值矩阵, $\tilde{X}$ 是 $N \times K(T + p)$ 维的外生变量矩阵, U 是 $N \times (T + p)$ 维的干扰变量矩阵, Δ 是 $T \times (T + p + K(T + P))$ 维的系数矩阵。

那么方程(2.3)和(2.4)表示的是误差分量动态模型为三角的联立方程组。这个模型的估计可以通过取以目标函数的极小值作为技术工具变量而得到,即有目标函数:

$$T_r \quad \Delta R \Delta' \qquad (2.5)$$

这里 $$R = \begin{bmatrix} Y'H(H'H)^{-1}H'Y & Y'H \\ H'Y & H'H \end{bmatrix}$$

$$= D'H H(H'H)^{-1}H'D$$

方差 — 协方差矩阵 Ω 的特殊结构蕴含技术工具变量产生一致性的估计。

如果与初始观测值有关的方程可以从方程组中舍弃,那么合理地选择 H,估计的步骤可以根据以下的(2.6)式实施而得到,譬如在下面的模型中:

$$Y_{it} = \delta Y_{it-1} + \beta X_{it} + \mu_i + v_i \qquad (2.6)$$

以 $X_{it}, X_{it-1}, X_{it-2} \cdots$ 作为工具变量估计的工具变量。

2.3. 极大似然法

当干扰向量是正态的,对估计的问题自然首先考虑的是极大似然的准则。但是条件似然函数在使用的时候总假定初始观测值是固定的,这样极大似然估计对参数组合应用范围很广来说等价于最小二乘估计,而且它们是不一致的。

当似然函数考虑到初始观测值的密度函数是无条件的,将是什么样的情况呢?我们给出一个无条件的极大似然估计量作为例子来讨论。

假设我们考虑的模型为:

$$Y_{it} = \delta Y_{it-1} + \beta X_{it} + \gamma Z_{it} + \mu_{it} + v_{it} \qquad (3.1)$$

初始值假设为:

$$Y_{i0} = \varphi Z_i + u_{i0}$$

不同误差都假定为多维正态的,特殊效用 μ_i 适合于分解为下面的概率回归式,即:

$$\mu_i = \psi u_{i0} + \epsilon_i \qquad (3.3)$$

这里 ϵ_i 是与 u_{i0} 独立的。

在这模型里 $(u_{i0}, \epsilon_i, v_{i1}, \cdots, v_{iT})$ 是服从正态分布,

即 $N(0, diag(\sigma_u^2, \sigma_\epsilon^2, \sigma_v^2 e_T'))$,而 log — 似然函数是:

$$L_{NT}(\delta, \beta, \gamma, \varphi, \psi, \sigma_u^2, \sigma_\epsilon^2, \sigma_v^2)$$

$$= -\frac{NT}{2}\ln 2\pi - \frac{N}{2}\ln det\sigma_u^2 - \frac{1}{2}\sum_j \tau_i' \Omega^{-1}\tau_i - \frac{1}{2\sigma_u^2}\sum_i u_{i0}^2$$

其中:

$$\tau_i' = (Y_{i1} - \delta Y_{i0} - \beta X_{i1} - \gamma Z_i - \psi u_{i0}, \cdots, Y_{iT} - \delta Y_{iT-1} - \beta X_{iT} - \gamma Z_i - \psi u_{i0}),$$

$$u_{i0} = Y_{i0} - YZ_i\Omega = \sigma_v^2(I_T - \frac{J_T}{T}) + (\sigma_v^2 + T\sigma_\epsilon^2)\frac{J_T}{T}$$

$$= \sigma_v^2 W_N + (\sigma_v^2 + T\sigma_\epsilon^2)B_N$$

于是我们可以得到下面的一些似然方程:

$$\frac{\partial L}{\partial\delta} = \frac{1}{\sigma_v^2}\sum_i Y_{i(-1)}'W\tau_i + \frac{1}{(\sigma_v^2 + T\sigma_\epsilon^2)}\sum_i Y_{i(-1)}'B\tau_i = 0 \qquad \text{(i)}$$

$$\frac{\partial L}{\partial\beta} = \frac{1}{\sigma_v^2}\sum_i X_i'W\tau_i + \frac{1}{(\sigma_v^2 + T\sigma_\epsilon^2)}\sum_i X_i'B\tau_i = 0 \qquad \text{(ii)}$$

$$\frac{\partial L}{\partial\gamma} = \frac{1}{(\sigma_v^2 + T\sigma_\epsilon^2)}\sum_i Z_i e_T'\tau_i = 0 \qquad \text{(iii)}$$

$$\frac{\partial L}{\partial\varphi} = \frac{1}{(\sigma_v^2 + T\sigma_\epsilon^2)}\sum_i u_{i0} e_T'\tau_i = 0 \qquad \text{(iv)}$$

$$\frac{\partial L}{\partial\varphi} = \frac{1}{(\sigma_v^2 + T\sigma_\epsilon^2)}\sum_i \psi Z_i e_T'\tau_i + \frac{1}{\sigma_v^2}\sum_i Z_i u_{i0} \qquad \text{(v)}$$

$$\frac{\partial L}{\partial\sigma_v^2} = -\frac{N(T-1)}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\sum_i \tau_i'W\tau_i = 0 \qquad \text{(vi)}$$

$$\frac{\partial L}{\partial(\sigma_v^2 + T\sigma_\epsilon^2)} = -\frac{N}{2(\sigma_v^2 + T\sigma_\epsilon^2)} + \frac{1}{2(\sigma_v^2 + T\sigma_\epsilon^2)^2}\sum_i \tau_i'W\tau_i = 0 \qquad \text{(vii)}$$

$$\frac{\partial L}{\partial\sigma_u^2} = -\frac{N}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}\sum_i u_{i0}^2 = 0 \qquad \text{(viii)}$$

其中: $e_T' = (1, \cdots, 1); X_i' = (X_{i1}, \cdots, X_{iT}); Y_{i(-1)}' = (Y_{i0}, \cdots, Y_{iT-1})$

上述方程组中的(iii),(v),(viii)方程式蕴含了 σ_u^2 和 φ 的极大似然估计量在式(3.3)中是最小二乘估计。如果 $\hat{\sigma}_{i0}$ 是方程(3.3)的残差,那么其他的极大似然估计量是方程组中的(i),(ii),(iv),(vi)和(vii)式的解,其中 σ_{i0} 在 τ_i 里由 $\hat{\sigma}_{i0}$ 来代替。

在这种情形下,初始观测值定义为:

$$Y_{i0} = \varphi Z_i + \alpha X_{i0} + u_{i0} \qquad (3.4)$$

而且(3.1)是不变的,变量 X_{i0} 不包含在自回归方程中。但是可以推得一个技术估计量等价于考虑由一个扩充的自回归模型所得出的极大似然估计量,这扩充的模型就是:

$$Y_{it} = \delta Y_{it-1} + \beta X_{it} + \gamma Z_i + \rho X_{i0} + \psi u_{i0} + v_{it}$$

在这个扩充的模型中 $\delta, \beta, \gamma, \psi, \rho, \sigma_\epsilon^2, \sigma_v^2$,的极大似然估计量是

渐近等价于真实模型的极大似然估计。

假如$\hat{\alpha}$是(3.4)式中α的最小二乘估计，$\hat{\rho}^*$和$\hat{\Psi}^*$是扩充模型里ρ和Ψ的极大似然估计，那么，$\hat{\alpha}^* = \hat{\alpha} + \hat{\rho}^* / \hat{\Psi}^*$是渐近的有效估计。

2.4 Π— 矩阵法

用于误差分量动态模型未知参数处理的Π— 矩阵法最早是CHAMBERLAN在80年代中期利用嵌入信息的模型估计问题而提出的，它近似于研究个体观测值的同分布和独立性的联系。并且以计算二阶矩为基础，以及把推广的均值法与渐近的最小均方法作为工具。

下面我们考虑的误差分量动态模型，其表达式中仅限于一个滞后变量在依变量里且只有一个外生变量，即：

$$Y_{it} = \delta Y_{it-1} + \beta X_{it} + \mu_i + v_{it} \qquad (3.5)$$

它可以表示为如下形式：

$$Y_{it} = \delta Y_{i0} + \beta \sum_{j=0}^{t-1} \delta X_{it-j} + \mu_i \frac{1-\delta}{1-\delta} + \sum_{j=0}^{t-1} \delta v_{it-j} \qquad (3.6)$$

Y_{it}的线性性是以$Y_{i0}, Y_{i1}, \cdots, X_{iT}$的一种线性组合为条件的。如果个体作用$\mu_i$是与序列$(X_{i1}, \cdots, X_{iT})$不相关，那么就有：

$$E^*(Y_{it}/Y_{i0}, X_{i1}, \cdots, X_{iT}) = \delta Y_{i0} + \beta \sum_{j=0}^{t-1} \delta X_{it-j} + \frac{1-\delta}{1-\delta} \frac{E\mu_i Y_{i0}}{VY_{i0}} (Y_{i0} - E(Y_{i0})) \qquad (3.7)$$

具体地说观测值的实际结构就是原义的线性性有一般的形式：

$$E^*(Y_{it}/Y_{i0}, X_{i1}, \cdots, X_{iT}) = \Pi_{ot} Y_{i0} + \sum_{j=1}^{T} \Pi_{j,t} X_{ij} + \Pi_{T+1,t} \qquad (3.8)$$

简约式参数Π取决于下面形式里的参数结构形式：

i) $\Pi_{0,t} = \delta + \frac{1-\delta}{1-\delta} \frac{E\mu_i Y_{i0}}{VY_{i0}}$,

ii) $\Pi_{j,t} = \beta \delta^{-j}, \quad j = 1, 2, \cdots, t$,

iii) $\Pi_{j,t} = 0, \quad j = t+1, \cdots, T$,

iv) $\Pi_{T+1,t} = -\frac{1-\delta}{1-\delta} \frac{E\mu_i Y_{i0}}{VY_{i0}} EY_{i0}$。

Π的非简约的一致估计就是上述关系式(i)至(iv)以外的任何形式，它们可以由Y_{it}对$Y_{i0}, X_{i1}, \cdots, X_{iT}$使用最小二乘法而得到。同样地公式(3.8)是一确定性模型，最小二乘估计是Π的非简约最好的线性无偏估计。

渐近最小均方估计法可以显示出结构参数$\delta, \beta, \gamma_0 = \frac{E\mu_i Y_{i0}}{VY_{i0}}, S_0 = \gamma_0 EY_{i0}$的一致估计。

现在我们来考虑以下线性结构的渐近最小均方法。首先假定线性结构方程如下：

1) $\Pi_{0,1} = \delta + r_0$,

2) $\Pi_{0t} = \delta \Pi_{0,1} + r_0, \quad t = 2, \cdots, T$

3) $\Pi_{t,t} = \beta, \quad t = 1, \cdots, T$

4) $\Pi_{j,t} = \delta \Pi_{j+1,t}, \quad t = 2, \cdots, T; \ j = 1, \cdots, t-1$,

5) $\Pi_{j,t} = 0, \quad t = 1, \cdots, T; j = T+1, \cdots, T$,

6) $\Pi_{T+1,1} = -S_0$,

7) $\Pi_{T+1,t} = \delta \Pi_{T+1,t-1} - S_0, \quad t = 2, \cdots, T$

综合以上的关系式在结构参数$(\delta, \beta, \gamma_0, S_0)$里产生一个线性模型。例如：若$T = 3$，我们就有：

$$[\Pi_{01}\Pi_{02}\Pi_{03}\Pi_{11}\Pi_{12}\Pi_{13}\Pi_{21}\Pi_{22}\Pi_{23}\Pi_{31}\Pi_{32}\Pi_{33}\Pi_{41}\Pi_{42}\Pi_{43}]'$$

$$= \begin{bmatrix} 1 & \Pi_{02} & \Pi_{03} & 0 & \Pi_{22} & \Pi_{23} & 0 & 0 & \Pi_{33} & 0 & 0 & 0 & 0 & \Pi_{41} & \Pi_{42} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}' \times (\delta \quad \beta \quad \gamma_0 \quad S_0)'$$

显然上式可以简记为：

$$\Pi = Z(\Pi)\Gamma \qquad (3.9)$$

其中：Π是元素为Π_{ij}的$(T+2) \times 1$维向量，$Z(\Pi)$是依赖Π的$(T+2) \times 4$维矩阵，Γ是结构参数向量。

如果$\hat{\Pi}$是除任何限制以外Π的一致最小二乘估计，我们可以写出有限矩的线性模型：

$$\hat{\Pi} = Z(\hat{\Pi})\Gamma + \epsilon \qquad (3.10)$$

这里当$N \to +\infty$时，$\epsilon \to 0$。

显然，Γ的一致估计是：

$$\hat{\Gamma} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'\hat{\Pi}$$

其中$\hat{Z} = Z\hat{\Pi}$。

这个估计量$\hat{\Gamma}$是渐近无效的，因为它没考虑ϵ的渐近方差。$\hat{\Pi}$的渐近方差S可以通过模型(3.8)式得到，$\epsilon = \hat{\Pi} - Z(\hat{\Pi})\Gamma$的渐近方差是一矩阵$\sum(S, \Gamma)$。$S$由估计量$\hat{S}$代替，(它为模型(3.8)上的最小二乘估计)，Γ由估计量$\hat{\Gamma}$代之，因此$\sum(S, \Gamma)$的一致估计量就是$\hat{E} = E(\hat{S}, \hat{\Gamma})$，Γ的最后估计量是：

$$\hat{\Gamma} = (\hat{Z}'\hat{Z}^{-1}\hat{Z})^{-1}\hat{Z}'\hat{Z}^{-1}\hat{\Pi}$$

Π矩阵法一个重要的优势就是它可以应用于当动态模型有P阶滞后时的自回归部分里，另外，干扰v_{it}的二阶假设不受限制。此外，一些异方差S的估计量可以由使用S的一个估计量的任意限制性的假定引进，具体地就是以粗略的异方差来估计最小二乘估计的方差。

这种近似的方法同样可以非常容易用于一些区域检验。譬如模型(3.1)的所有假设是资料生成过程所需要的，

$\sqrt{N}(\hat{\Pi} - \hat{Z}\hat{\Gamma})$依分布收敛于$N(0, \sum)$。

所以$\xi = N(\hat{\Pi} - \hat{Z}\hat{\Gamma})\hat{\Sigma}^{-1}(\hat{\Pi} - \hat{Z}\hat{\Gamma}) \to \chi^2_T((T+2)+4)$。

如果(3.5)式所引入的假设是不满足的，则ξ趋于$+\infty$。

这个置信水平为α的区域检验的临界域是：

$C = \{\xi > \chi^2_{1-\alpha}(T(T+2)-4)\}$。

参考文献:

1. 童光荣：《关于嵌入信息动态模型的两个问题》，载《经济评论》，1999(5)。

2. Chamberlain G. (1984): "Panel data" in Z. Griliches and M. D. Intriligatoreds. Handbook of Econometrics North-Hollond.

3. Gourieroux C., Monfort A and A-Trognon (1984): "Asymptotic Least squares Application to qualitative models", Document de travail INSEE-ENSAE N8108.

4. Sevestre P and A. Trognon (1990): "Consistent estimation methods for dynamic error components models", Mimeo.

5. Sevestre P. and A. Trognon (1985): "A note on autoregressive error components models", Journal of Econometrics vol29, pp. 231～245.

6. Maddala G. S. (1971): "The use of variance components models in pooling cross-seetion and time series data". Econometrica. vol. 39. pp. 341～358.

(作者单位：武汉大学经济学系　武汉　430072)

(责任编辑：金　萍)