Statistical and Application 统计学与应用, 2013, 2, 114-118
http://dx.doi.org/10.12677/sa.2013.24017 Published Online December 2013 (http://www.hanspub.org/journal/sa.html)

Robust Generalized Method of Moment Estimator for the Contaminated Stochastic Logistic Diffusion Model*

Xuemei Hu, Jiali Ma

School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing
Email: huxuem@163.com

Received: Oct. 17th, 2013; revised: Nov. 12th, 2013; accepted: Nov. 23rd, 2013

Abstract: This paper considers the contaminated stochastic Logistic diffusion model, and proposes robust generalized method of moment estimator. Simulations show that the proposed robust generalized method of moment estimator behaves better than the usual maximum likelihood estimator.

Keywords: The Contaminated Stochastic Logistic Diffusion Model; Robust Generalized Method of Moment Estimator

随机 Logistic 扩散污染模型的稳健广义矩估计*

胡雪梅，马家丽

重庆工商大学数学与统计学院，重庆
Email: huxuem@163.com

收稿日期：2013 年 10 月 17 日；修回日期：2013 年 11 月 12 日；录用日期：2013 年 11 月 23 日

摘　要：本文首次研究了維勒过程受污染的随机 Logistic 扩散模型，提出了稳健的广义矩估计方法。模拟表明所建立的稳健广义矩估计优于通常的极大似然估计。

关键词：随机 Logistic 扩散污染模型；稳健广义矩估计

1. 引言

考虑随机 Logistic 扩散模型

$$\begin{aligned} &\mathrm{d}N(t) = N(t)\left(\alpha - \beta N(t)\right)\mathrm{d}t + \sigma N(t)\mathrm{d}B(t), \\ &N(0) = N_0 > 0, t \in [0,T], \end{aligned} \tag{1}$$

其中：$N(t)$ 表示在时刻 t 的种群容量，$\alpha > 0$ 表示自然出生率，$\beta > 0$ 表示死亡率，α/β 表示负载能力，通常表示环境资源能够支撑的最大种群数，$\sigma > 0$ 表示噪音对 $N(t)$ 的动态影响，$B(t)$ 表示标准的布朗运动，初始条件 N_0 与 $B(t)$ 独立。我们假设 N_0 是一个正随机变量，而且存在一个 $p > 2$ 使得

$$E\left[N_0^p\right] < \infty .$$

在生物医学和生态学研究中，模型经常用来建模细胞种群的增长，例如，人类癌细胞和中国仓鼠卵巢细胞的增长等。[1,2]研究了模型(1)的正解的存在性、唯一性和全局吸引性，并对模型参数建立了极大似然估计。[3]考虑了轻微不同的模型，即

*基金项目：国家自然科学基金项目重庆市科委自然科学基金项目(No.11101452)；重庆市科委自然科学基金项目 CQ CSTC(No.2012 jjA00035)。

$$\mathrm{d}N(t)=N(t)\left[\alpha+\beta N(t)\right]\mathrm{d}t+\sigma N(t)\mathrm{d}B(t),N(0)=N_0>0,t\in[0,T],\alpha>0,\beta>0,$$

证明了不管 $\sigma>0$ 多么小，解都不会在有限时间内爆炸等。

实际上，在许多应用领域收集到的观察数据经常包含一个或多个非典型的观察值。这些非典型的观察值称为奇异值。例如，种群密度数据的增量中如果存在大的奇异值和高峰，则表明具有标准正态增量的模型(1)是误指的模型。实际上，增量服从标准正态分布只描述了大部分数据，但没有包括那些遵循不同方式的增量。因此，本文我们考虑维纳过程受到污染的随机 Logistic 扩散模型，其增量来自标准正态分布的一个邻域，即

$$B(t)-B(t-1)=\varepsilon_t\sim G_\varepsilon=(1-\varepsilon)\mathcal{N}(0,1)+\varepsilon G, \tag{2}$$

这里 $0\le\varepsilon<1$，G 是一个未知的分布。根据[4]，模型(1)~(2)可以产生大的奇异值和高的峰度。在这种情况下，如果继续使用极大似然原理对模型(1)~(2)进行统计推断，[5-8]等的研究表明，我们将会得到有偏的估计和误导的检验结果。此外，[9]的研究表明，对均值 reversion 参数采用的各种极大似然(ML)方法都具有严重的有限样本估计偏差。为了修正由于模型误指引起的偏差，[10]提出了稳健模拟基估计方法，[11]提出了稳健的广义矩估计并发展了一种 灵活的算法。但目前尚未有人对模型(1)~(2)作稳健分析。因此，本文对模型(1)~(2)发展了稳健的广义矩阵估计方法。在模拟中，我们把稳健的广义矩阵估计和极大似然估计进行了比较，发现所提出的稳健广义矩估计优于通常的极大似然估计。

本文结构如下：第 2 部分建立了随机 Logistic 模型的极大似然估计；第 3 部分建立了维纳过程受到污染的随机 Logistic 模型的稳健广义矩阵估计；第 4 部分比较了极大似然估计和稳健广义矩阵估计的有限样本表现。

2. 极大似然估计

为了方便，我们用 N_t 定义 $N(t)$，用 F_θ 表示结构模型(1)~(2)，用 μ 表示下面的辅助模型 $\tilde{F}_\mu$

$$N_t=\left(1+\alpha-\beta N_{t-1}\right)N_{t-1}+\sigma N_{t-1}\varepsilon_t \tag{3}$$

中的参数 $\mu=\left(\alpha,\beta,\sigma^2\right)$，模型误差 ε_t 是一个标准的正态随机变量。显然，模型(3)是模型(1)的一个粗糙离散化版本。现考虑辅助模型 $\tilde{F}_\mu$ 中参数 μ 的渐近极大似然估计。假设 $\left\{N_t\right\}_{t=1,\cdots,m}$ 由模型(1)~(2)产生，给定 $F_{t-1}=\left\{N_l,l\le t-1\right\}$，得到条件概率密度函数

$$f\left(N_t\middle|F_{t-1};\mu\right)=\frac{1}{\sigma N_{t-1}\sqrt{2\pi}}\times\exp\left\{-\frac{1}{2\sigma^2N_{t-1}^2}\left[N_t-\left(1+\alpha-\beta N_{t-1}\right)N_{t-1}\right]^2\right\}$$

和条件似然函数

$$f\left(\mu\middle|N_0,\cdots,N_m\right)=\prod_{t=1}^{m}\frac{1}{\sigma N_{t-1}\sqrt{2\pi}}\times\exp\left\{-\frac{1}{2\sigma^2N_{t-1}^2}\left[N_t-\left(1+\alpha-\beta N_{t-1}\right)N_{t-1}\right]^2\right\}.$$

那么，条件对数似然函数为

$$-\frac{m}{2}\log(2\pi)-\frac{m}{2}\log\sigma^2-\sum_{t=1}^{m}\log N_{t-1}-\frac{1}{2\sigma^2}\sum_{t=1}^{m}\frac{\left[N_t-\left(1+\alpha-\beta N_{t-1}\right)N_{t-1}\right]^2}{N_{t-1}^2}. \tag{4}$$

因此，β,α 和 σ^2 的渐近极大似然估计分别为

$$\tilde{\beta}=\frac{m\sum_{t=1}^{m}N_t-\sum_{t=1}^{m}N_{t-1}\sum_{t=1}^{m}\frac{N_t}{N_{t-1}}}{\left(\sum_{t=1}^{m}N_{t-1}\right)^2-m\sum_{t=1}^{m}N_{t-1}^2}, \tag{5}$$

$$\tilde{\alpha}=\frac{1}{m}\sum_{t=1}^{m}\frac{N_t}{N_{t-1}}+\tilde{\beta}\left(\frac{1}{m}\sum_{t=1}^{m}N_{t-1}\right)-1, \tag{6}$$

$$\tilde{\sigma}^2=\frac{1}{m}\sum_{t=1}^{m}\frac{\left[N_t-\left(1+\tilde{\alpha}-\tilde{\beta}N_{t-1}\right)N_{t-1}\right]^2}{N_{t-1}^2}. \tag{7}$$

众所周知，当假设的模型与潜在的模型存在偏差时，渐近的极大似然估计 $\tilde{\mu}=\left(\tilde{\alpha},\tilde{\beta},\tilde{\sigma}^2\right)$ 不稳健。因此，下面我们对模拟(1)~(2)引入稳健的广义矩估计方法。

3. 稳健的广义矩估计

3.1. 定义

[11]提出了建立稳健广义矩估计的算法，这里我们基于这种算法建立模型(1)~(2)中参数 θ 的稳健广义矩估计。众所周知，极大似然方法是采用似然导数

作为矩条件的广义矩方法。因此，根据

$$\frac{\partial \log f\left(\mu \middle| N_0,\cdots,N_m\right)}{\partial \alpha}=0,\frac{\partial \log f\left(\mu \middle| N_0,\cdots,N_m\right)}{\partial \beta}=0$$

和

$$\frac{\partial \log f\left(\mu \middle| N_0,\cdots,N_m\right)}{\partial \sigma^2}=0,$$

$\mu=\left(\alpha,\beta,\sigma^2\right)$ 的渐近极大似然估计 $\tilde{\mu}=\left(\tilde{\alpha},\tilde{\beta},\tilde{\sigma}^2\right)$ 可以写成具有以下矩条件

$$\begin{aligned}&\sum_{t=1}^{m} h\left(N_t,N_{t-1};\mu\right)\\&=\sum_{t=1}^{m}\begin{pmatrix} N_t/N_{t-1}-(1+\alpha)-\beta N_{t-1} \\ N_t-(1+\alpha)N_{t-1}+\beta N_{t-1}^2 \\ \left[N_t/N_{t-1}-\left(1+\alpha-\beta N_{t-1}\right)\right]^2-\sigma^2 \end{pmatrix}=0\end{aligned} \tag{8}$$

的广义矩估计。如果 μ_* 是真参数，则

$$E_F h\left(N_t,N_{t-1};\mu_*\right)=0. \tag{9}$$

注意到进入正交条件(8)的观察 N_{t-1} 是二次多项式。因此，对于模型参数的某些选择来说，潜在 GMM 估计的影响函数在某些污染的方向是很陡的。根据[11]，一个广义矩估计稳健的充要条件是正交条件有界。因此，稳健广义矩估计是一个用截尾的正交条件取代经典的正交条件的广义矩估计。具体来讲，对所有 δ_x，一个统计泛函 $\tilde{a}$ 的影响函数

$$IF\left(x;\tilde{a},F_{\theta_0}\right)=\lim_{\varepsilon\to 0}\frac{\tilde{a}\left((1-\varepsilon)F_{\theta_0}+\varepsilon\delta_x\right)-\tilde{a}\left(F_{\theta_0}\right)}{\varepsilon}$$

描述了一个统计量在假设模型分布 F_{θ_0} 的单点污染 δ_x 的线性渐近偏差。一个无界的影响函数暗含了一个统计量在模型的单点污染下的无界渐近偏差。因此，一个统计泛函稳健的自然要求是影响函数的有界性。而广义矩估计的影响函数是

$$\begin{aligned}&IF\left(x;\tilde{a},F_{\theta_0}\right)=\\&-S_0 E_{\theta_0}\frac{\partial h^T\left(N_t,N_{t-1};a\left(\theta_0\right)\right)}{\partial a}W_0 h\left(N_t,N_{t-1};a\left(\theta_0\right)\right),\end{aligned} \tag{10}$$

其中，

$$\begin{aligned}W_0&=V_0^{-1},V_0:\\&=E_{\theta_0}\left[h\left(N_t,N_{t-1};a\left(\theta_0\right)\right)h^T\left(N_t,N_{t-1};a\left(\theta_0\right)\right)\right]\end{aligned}$$

和

$$\begin{aligned}&S_0\left(W_0\right):=\\&E_{\theta_0}\left[\frac{\partial h^T\left(N_t,N_{t-1};a\left(\theta_0\right)\right)}{\partial a}W_0 E_{\theta_0}\frac{\partial h\left(N_t,N_{t-1};a\left(\theta_0\right)\right)}{\partial a^T}\right]^{-1}.\end{aligned}$$

由(10)我们可以看出：1) 广义矩估计的影响函数与模型 $h\left(\cdot;a\left(\theta_0\right)\right)$ 的正交函数线性相关；2) 广义矩估计的影响函数有界当且仅当潜在模型的正交函数有界。根据 Ronchetti & Trojani (2001, p46)，如果自标准范数

$$\left\|h\left(N_t,N_{t-1};a\left(\theta_0\right)\right)\right\|_{V_0^{-1}}:=\left\|V_0^{-1/2}h\left(N_t,N_{t-1};a\left(\theta_0\right)\right)\right\|$$

有界，则稳健广义矩估计满足

$$\left\|IF\left(x;\tilde{a},F_{\theta_0}\right)\right\|_{\Sigma_0^{-1}}:=\left\|\Sigma_0^{-1/2}IF\left(x;\tilde{a},F_{\theta_0}\right)\right\|\le c,$$

其中

$$\begin{aligned}&\Sigma_0=\\&\left[E_{\theta_0}\frac{\partial h^T\left(N_t,N_{t-1};a\left(\theta_0\right)\right)}{\partial a}V_0 E_{\theta_0}\frac{\partial h\left(N_t,N_{t-1};a\left(\theta_0\right)\right)}{\partial a^T}\right]^{-1}.\end{aligned}$$

因此，为了建立一个具有有界自标准影响函数的广义矩估计，我们引入 Huber 函数

$$\omega_c(y)=\begin{cases}\min\left(1,\dfrac{c}{\|y\|}\right), & y\ne 0,\\ 1, & y=0,\end{cases} \tag{11}$$

和一个新的映射

$$\begin{aligned}&h_c^{A,\tau}\left(N_t,N_{t-1};\mu\right)=\\&A\left[h\left(N_t,N_{t-1};\mu\right)-\tau\right]w_c\left(A\left[h\left(N_t,N_{t-1};\mu\right)-\tau\right]\right),\end{aligned} \tag{12}$$

其中给定的界 $c\ge\sqrt{H}$，满秩矩阵义 $A\in R^H\times R^H$ 和向量 $\tau\in R^H$

$$E_F h_c^{A,\tau}\left(N_t,N_{t-1};\mu_*\right)=0, \tag{13}$$

$$E_F h_c^{A,\tau}\left(N_t,N_{t-1};\mu_*\right)\left[h_c^{A,\tau}\left(N_t,N_{t-1};\mu_*\right)\right]'=I \tag{14}$$

决定。因此，$h_c^{A,\tau}\left(N_t,N_{t-1};\mu\right)$ 有界。稳健广义矩估计由下式

$$\tilde{\mu}_R=\arg\min_{\mu} g\left(\left\{N_t\right\};\mu\right) \tag{15}$$

定义，其中

$$g(\{N_t\};\mu)=$$

$$\left[\frac{1}{m}\sum_{t=1}^{m}h_c^{A,\tau}(N_t,N_{t-1};\mu)\right]'\left[\frac{1}{m}\sum_{t=1}^{m}h_c^{A,\tau}(N_t,N_{t-1};\mu)\right].$$

3.2. 计算稳健广义矩估计的四个步骤

第一步：

1) 取参数$\theta=(\alpha,\beta,\sigma^2)$的初值$\theta_0=(\alpha_0,\beta_0,\sigma_0^2)$：这里取极大似然估计 $\tilde{\mu}$ 作为初值。

2) 取 $\tau_0=0$ 。

3) 取 A_0 满足：

$$(A_0'A_0)^{-1}=\left[\frac{1}{m}\sum_{t=1}^{m}h(N_t,N_{t-1};\theta_0)\right]'\left[\frac{1}{m}\sum_{t=1}^{m}h(N_t,N_{t-1};\theta_0)\right].$$

第二步：

1) 取$c=2$，计算 $w_c\left(A_0\left[h(N_t,N_{t-1};\theta_0)-\tau_0\right]\right)$，其中

$$w_c\left(A_0\left[h(N_t,N_{t-1};\theta_0)-\tau_0\right]\right)=\begin{cases}\min\left(1,\dfrac{c}{\left\|A_0\left[h(N_t,N_{t-1};\theta_0)\right]-\tau_0\right\|}\right), & \text{如果}A_0\left[h(N_t,N_{t-1};\theta_0)\right]-\tau_0\neq 0,\\ 1, & \text{如果}A_0\left[h(N_t,N_{t-1};\theta_0)\right]-\tau_0=0.\end{cases}$$

2) 计算新值 τ_1：

$$\tau_1=\frac{E_{\theta_0}h(N_t,N_{t-1};\theta_0)w_c\left(A_0\left[h(N_t,N_{t-1};\theta_0)-\tau_0\right]\right)}{E_{\theta_0}w_c\left(A_0\left[h(N_t,N_{t-1};\theta_0)-\tau_0\right]\right)}.$$

3) 计算新值 A_1，取 A_1 满足

$$(A_1'A_1)^{-1}=\frac{1}{m}\sum_{t=1}^{m}\left(h(N_t,N_{t-1};\theta_0)-\tau_0\right)\left(h(N_t,N_{t-1};\theta_0)-\tau_0\right)'w_c^2\left[A_0\left(h(N_t,N_{t-1};\theta_0)-\tau_0\right)\right].$$

第三步：

由正交函数 $h_c^{A_1,\tau_1}(N_t,N_{t-1};\theta_0)$ 计算参数 θ 的第一步稳健广义矩估计估计：

$$\theta_1=\arg\min_{\mu}\left[\frac{1}{m}\sum_{t=1}^{m}h_c^{A_1,\tau_1}(N_t,N_{t-1};\theta_0)\right]'\times\left[\frac{1}{m}\sum_{t=1}^{m}h_c^{A_1,\tau_1}(N_t,N_{t-1};\theta_0)\right].$$

第四步：

如果$\|\theta_1-\theta_0\|<$容忍限，迭代停止，最优值取 θ_1；否则的话，用 A_1,τ_1 取代 A_0,τ_0，迭代第二步和第三步直到由有界正交函数序列 $h_c^{A_k,\tau_k}(N_t,N_{t-1};\theta_k)$，$k\in Z^+$ 计算得到的稳健广义矩估计满足$\|\theta_k-\theta_{k-1}\|<$容忍限 $k\in Z^+$，迭代停止，最优的稳健广义矩估计取 θ_k。

4. 模拟研究

本节我们对模型(1)~(2)中参数 $(\alpha,\beta,\sigma^2)=(0.6,0.9,0.16)$ 的极大似然估计和稳健广义矩估计进行了比较，其中

$$\varepsilon\sim G_\varepsilon=(1-\varepsilon)\mathcal{N}(0,1)+\varepsilon G,\ \varepsilon=0.05$$

和 $G\sim\mathcal{N}(0,3^2)$。先由这模型分别产生样本容量为 $m=100,300,600$ 的 5000 个数据集 $N_t,t=1,\cdots,m$；然后采用第 2 节的方法计算极大似然估计，采用第 3.2 节的方法计算稳健广义矩估计，这里我们取的容忍限为 e^{-6}。

表1展示了极大似然估计(ML)和稳健广义矩估计(RGMM)的偏差和均方误差。从表中我们可以看出，在模型误指的条件下，极大似然估计的偏差和均方误差都很大，而稳健广义矩估计的偏差和均方误差要比极大似然估计小得多，特别是偏差很小。这种现象表明稳健广义矩估计可以修正模型误指带来的极大偏差，而极大似然估计不能修正模型误差带来的偏差。

Table 1. Biases and mean squared errors of ML and RGMM estimators

表 1. 极大似然估计(ML)和稳健广义矩估计(RGMM)的偏差和均方误差

$\theta=\theta_0$	m	Bias (ML)	Bias (RGMM)	MSE (ML)	MSE (RGMM)
$\alpha=0.6$	100	0.5562	0.1178	0.4142	0.1280
$\beta=0.9$	100	0.8496	0.2452	0.8852	0.2651
$\sigma^2=0.16$	100	0.1589	0.0643	0.0160	0.0035
$\alpha=0.6$	300	0.5674	0.1015	0.4210	0.1261
$\beta=0.9$	300	0.8671	0.2310	0.8826	0.2668
$\sigma^2=0.16$	300	0.1587	0.0636	0.0152	0.0026
$\alpha=0.6$	600	0.5669	0.0915	0.4015	0.1245
$\beta=0.9$	600	0.8646	0.2012	0.8764	0.2610
$\sigma^2=0.16$	600	0.1586	0.0610	0.0132	0.0018

因此，在模型误指的情况下，稳健广义矩估计明显优于通常的极大似然估计。

参考文献 (References)

[1] Jiang, D.Q., Zhang, B.X., Wang, D.H. and Shi, N.Z. (2007) Existence, uniqueness, and global attractivity of positive solutions and MLE of the parameters to the Logistic equation with random perturbation. *Science in China Series A*: *Mathematics*, **50**, 977-986.

[2] Jiang, D.Q. and Shi, N.Z. (2005) A note on nonautonomous Logistic equation with random perturbation. *Journal of Mathematical Analysis and Application*, **303**, 164-172.

[3] Mao, X., Marion, G. and Renshaw, E. (2002) Environmental Brownian noise suppresses explosions in population dynamics. *Stochastic Processes and their Applications*, **97**, 95-110.

[4] Czellar, V., Karolyi, G.A. and Ronchetti, E. (2007) Indirect robust estimation of the short-term interest rate process. *Journal of Empirical Finance*, **14**, 546-563.

[5] Ait-Sahalia, Y. (2003) Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, **70**, 223-262.

[6] Fan, J.Q. (2005) A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, **20**, 317-337.

[7] Hurn, A., Jeisman, J. and Lindsay, K. (2007) Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Financial Econometrics*, **5**, 390-455.

[8] Nielsen, J.N., Madsen, H. and Young, P.C. (2000) Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control*, **24**, 83-94.

[9] Phillips, P.C.B. and Yu, J. (2007) Maximum likelihood and Gaussian estimation of continuous time models in finance. *Cowles Foundation Discussion Papers* 1597, Cowles Foundation for Research in Economics, Yale University.

[10] Genton, M.G. and Luna, X.D. (2000) Robust simulation-based estimation. *Statistics and Probability Letters*, **48**, 253-259.

[11] Ronchetti, E. and Trojani, F. (2001) Robust inference with GMM estimators. *Journal of Econometrics*, **101**, 37-69.