

文章编号:1003—207(2008)02—0132—08

缺失数据下 ARMA(1,1)模型的估计方法

田 萍, 张屹山

(吉林大学商学院吉林大学数量经济研究中心, 辽宁 吉林 130012)

摘要:近几十年以来, 国际上在对“风险的处理和效益的优化”这两个现代金融学的中心议题的分析和处理过程中, 金融时间序列的计量学模型及其相应的分析越来越起到非常重要的作用。对于线性时间序列模型如 AR(p), MA(q), ARMA(p, q)等, 已经为我们所熟知。具体到模型的参数估计在数据没有缺失时, 也有很多经典的办法, 如最小二乘法、极大似然法等。但是当数据在中间有缺失时, 上述方法将无能为力。本文将详细讨论在数据有缺失时的 ARMA(1,1)模型, 即 $Z_t = \alpha Z_{t-1} + \epsilon_t - \beta \epsilon_{t-1}$ 的参数的估计方法。

关键词:缺失数据; ARMA(1,1)模型; 似然函数; EM 算法

中图分类号:0212 文献标识码:A

1 引言

在金融研究及实际的投资决策中, 实证研究一直占据非常重要的地位, 而在绝大多数实证研究中, 时间序列数据等相关的数据是进行实证研究的前提条件。但是, 处理数据时的一个非常棘手的问题就是数据的缺失。毫无疑问, 对这类问题处理不当, 就会影响研究结果的准确性。人们对数据缺失问题较早的处理方法主要有以下几种:

当数据量很大, 并且其它的部分很完整时, 在保证剩余数据仍为连续的情况下, 抛开缺失的数据进行实证研究; 当数据量不够大时, 可以利用其它的信息把缺失的数据补上, 比如利用求平均数法, 线性插值法等等。

上述这些方法都是较为直观易懂的, 但它们又有各自明显的缺点, 如数据受到信息完整性等某些限制是不能够抛开的, 而前述两种直观的补数据的办法显然又没有充分利用到已知的一些信息, 使得相应的估计结果不够准确。到目前为止, 国内外在对数据有缺失的情况下的各种模型的参数估计及其相关问题都有所讨论, 如国际上 James 等(2004)^[6]综述了在有数据缺失时的一些常用方法并进行了比较; Andrea(1994)^[2]考察了数据缺失时参数估计结

果的偏差问题; Sik-Yum Lee(2003)^[8]对非线性结构方程应用 EM 算法给出了有缺失数据时的参数估计方法; 国内林静等(2006)^[11]讨论了基于 MCMC 稳态模拟的贝叶斯经验费率厘定信用模型问题; 顾龙全等(2006)^[3]考察了指数分布场合恒加试验缺失数据的 Bayes 统计分析方法; 庞新生(2005)^[4]讨论了多重插补处理缺失数据方法的理论基础; 闫长华和张忠占(2005)^[5]给出了缺失数据下线性 EV 模型的参数估计的方法。

对于目前国际上比较流行的线性时间序列模型, 在有数据缺失情况时, 国内外也有很多相关的研究成果: Machado 等(1998)^[11]简述了有数据缺失时自回归模型的参数估计问题; Ljung, G, M, (1982)^[2]和 Pernzer, J. and Shea, B. (1997)^[13]分别给出了自回归滑动平均模型的高斯估计方法和应用卡尔曼滤波的似然估计方法; 国内田萍(2003)^[2]等对具有缺失数据的 AR(p)模型的估计方法进行了探讨。值得一提的是马春生(Chunsheng Ma, 2002)年发表在《时间序列分析杂志》上的文章《缺失数据下 ARMA(1,1)模型的极大似然估计》, 处理的就是具有缺失数据时 ARMA(1,1)的参数估计方法。相对于本文, 该文应用的是极大似然的思想利用矩阵的性质和运算获得参数的估计值满足一个方程组。但该方程组的缺点是形式复杂, 不能直接获得显示解, 需要通过数值算法获得解的近似值。而本文主要应用 EM 算法, 对数据从中间连续缺失一个或两个(根据对中国股票市场的了解, 缺失的数据多数为

收稿日期:2007-06-19; 修订日期:2008-03-24

基金项目:社科项目资助(07BJY168)

作者简介:田萍(1976—), 女(汉族), 辽宁省人, 吉林大学商学院应用经济研究所讲师, 博士, 研究方向:数量经济学。

一个到两个)的情况推导出零均值的 ARMA(1,1)模型的估计方法。我们给出的是参数估计值的显示解的算法, 相对而言使估计结果更清晰易得。最后我们就一个数据缺失的情况进行实证分析。

2 一个数据缺失时的参数的估计方法

本节主要研究了只有一个缺失数据时的 ARMA(1,1)模型参数的估计方法。设时间序列 $z_{-1}, z_0, z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n$ 是来源于如下的零均值的 ARMA(1,1)过程的样本, 其中 z_i 为缺失的数据。该 ARMA(1,1)模型形式如下

$$Z_t = \alpha Z_{t-1} + \epsilon_t - \beta \epsilon_{t-1} \quad (1)$$

其中 α 与 β 为实值待估参数, 设 $\{\epsilon_t\}$ 是均值为 0 方差为 1 的高斯过程, 即 ϵ_t 是服从标准正态分布的随机变量, 并且其与 $\{\epsilon_s, s < t\}$ 及 $\{z_s, s < t\}$ 独立。

2.1 基于观察值的似然函数的计算

应用模型(1)及其相应的性质有 $z_t - \alpha z_{t-1}, t \geq 0$ 服从均值为 0 方差为 $(1+\beta^2)$ 的正态分布。在应用已知观测值 $z_0, z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n$ 条件下的似然函数记为 $L(\alpha, \beta)$, 则我们有如下的命题。

命题 1 在模型(1)下, 观测值的似然函数 $L(\alpha, \beta)$ 满足

$$\begin{aligned} L(\alpha, \beta) &= \int_{-\infty}^{\infty} L(\alpha, \beta) dz_i \\ &= \prod_{j=0}^{i-1} f(z_j - \alpha z_{j-1}) \int_{-\infty}^{\infty} \prod_{j=i}^{i+1} f(z_j - \alpha z_{j-1}) dz_i \\ &\quad \prod_{j=i+2}^n f(z_j - \alpha z_{j-1}) \\ &= (1/\sqrt{2\pi(1+\beta^2)})^n \frac{1}{a} \exp(-\frac{c-b^2/a^2}{2(1+\beta^2)}) \end{aligned} \quad (2)$$

$$D_{i-1,0} D_{n;i+2}$$

其中 $a^2 = 1 + \alpha^2$, $b = \alpha(z_{i+1} + z_{i-1})$, $c = z_{i+1}^2 + \alpha^2 z_{i-1}^2$, $D_{i-1,0} = \prod_{j=0}^{i-1} \exp\{-\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2\}$,

$$D_{n;i+2} = \prod_{j=i+2}^n \exp\{-\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2\},$$

$$L(\alpha, \beta) = \prod_{j=0}^n f(z_j - \alpha z_{j-1}).$$

这里 $f(\cdot) = \frac{1}{\sqrt{2\pi(1+\beta^2)}} \exp\{-\frac{\cdot^2}{2(1+\beta^2)}\}$ 为一正态分布的密度函数, $L(\alpha, \beta)$ 为数据 $z_0, z_1, \dots, z_{i-1}, z_i, z_{i+1}, \dots, z_n$ 对应的似然函数。

证明: 为了证明上述等式成立先验证

$$\begin{aligned} \int_{-\infty}^{\infty} \prod_{j=i}^{i+1} f(z_j - \alpha z_{j-1}) dz_i &= \int_{-\infty}^{\infty} \\ &\quad (\frac{1}{\sqrt{2\pi(1+\beta^2)}})^2 \exp\left(-\frac{(z_i - \alpha z_{i-1})^2}{2(1+\beta^2)}\right) \exp\left(-\right. \\ &\quad \left.\frac{(z_{i+1} - \alpha z_i)^2}{2(1+\beta^2)}\right) dz_i = \frac{1}{a} \frac{1}{\sqrt{2\pi(1+\beta^2)}} \exp\left(-\right. \\ &\quad \left.\frac{c-b^2/a^2}{2(1+\beta^2)}\right) \end{aligned}$$

把上式带入观测值的似然函数中, 再利用正态分布密度函数定义把左右两边打开即有命题 1 成立。

2.2 相应的 EM 算法(E 步)

为了应用 EM 算法, 还需先求出 z_i 关于 $\alpha^{(k)}$, $\beta^{(k)}$ 和 $z_i^- = \{z_1, z_2, \dots, z_n\}/\{z_i\}$ 的条件分布密度, 为

$$\begin{aligned} f(z_i^- | \alpha^{(k)}, \beta^{(k)}, z_i^-) &= L(\alpha^{(k)}, \beta^{(k)}) / L(\alpha^{(k)}, \beta^{(k)}) \\ &= \frac{\prod_{j=0}^n f(z_j - \alpha^{(k)} z_{j-1})}{1 / \sqrt{2\pi(1+\beta^{(k)2})} \prod_{j=0}^{i-1} f(z_j - \alpha^{(k)} z_{j-1})} \\ &\quad \frac{1}{\prod_{j=i+2}^n f(z_j - \alpha^{(k)} z_{j-1})} \\ &= \frac{1}{\frac{1}{a^{(k)}} \exp\left(-\frac{1}{2(1+\beta^{(k)2})}(c^{(k)2} - b^{(k)2}/a^{(k)2})\right)} \\ &= \prod_{j=i}^{i+1} f(z_j - \alpha^{(k)} z_{j-1}) \circ \exp\left\{\frac{1}{2(1+\beta^{(k)2})}[c^{(k)} - \right. \\ &\quad \left.(b^{(k)}/a^{(k)})^2]\right\} \circ a^{(k)} \sqrt{2\pi(1+\beta^{(k)2})} \end{aligned} \quad (3)$$

其中 $a^{(k)}, b^{(k)}$ 和 $c^{(k)}$ 为运用该 EM 算法由第 k 步得出的参数 $(\alpha^{(k)}, \beta^{(k)})$ 后算出的值, 其函数形式同 a, b 和 c 。而 $(\alpha^{(k)}, \beta^{(k)})$ 就是运用 EM 算法得出的参数 (α, β) 的第 k 步的近似值。

EM 算法中所需的对数似然函数的形式如下

$$\ln(L(\alpha, \beta)) = \ln\left[\prod_{j=1}^n f(z_j - \alpha z_{j-1})\right] = -(n+1)\ln(\sqrt{2\pi(1+\beta^2)}) + \sum_{j=0}^n \left[-\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2\right].$$

综合上述讨论, 我们可以对 EM 算法中的 E 步进行计算, 得到如下命题。

命题 2 在 EM 算法中,

$$\begin{aligned} E(\ln(L(\alpha, \beta)) | z_i^-, \alpha^{(k)}, \beta^{(k)}) &= -(n+1)\ln(\sqrt{2\pi(1+\beta^{(k)2})}) + \ln D_{i-1,0} + \end{aligned}$$

$$\ln D_{n,i+2} = \frac{1}{2} (z_{i+1}^2 + \alpha^2 z_{i-1}^2) / (1 + \beta^2) - (1 + \alpha^2) b^{(k)2} / [2a^{(k)4} (1 + \beta^2)] + b^{(k)} (z_{i+1} + z_{i-1}) \alpha / [a^{(k)2} (1 + \beta^2)] - \frac{1}{2} (1 + \alpha^2) (1 + \beta^{(k)2}) / [a^{(k)2} (1 + \beta^2)] \quad (4)$$

证明:由前面的结果,我们对 EM 算法的 E 步进行计算,得 $E(\ln(L(\alpha, \beta)) | z_i^-, \alpha^{(k)}, \beta^{(k)})$

$$\begin{aligned} &= \int_{-\infty}^{\infty} [- (n+1) \ln(\sqrt{2\pi(1+\beta^2)}) + \sum_{j=0}^n - \frac{1}{2} (z_j - \alpha z_{j-1})^2 / (1 + \beta^2)] \cdot a^{(k)} \\ &\cdot (1/\sqrt{2\pi(1+\beta^{(k)2})})^{-1} \exp\left\{\frac{1}{2(1+\beta^{(k)2})}[c^{(k)} - (b^{(k)}/a^{(k)})^2]\right\} \cdot \prod_{j=i}^{i+1} f(z_j - \alpha z_{j-1}) dz_i \end{aligned}$$

$$= I_1 + I_2 + I_3 + I_4$$

其中 I_1, I_2, I_3 和 I_4 分别如下

$$\begin{aligned} I_1 &= - \int_{-\infty}^{\infty} (n+1) \ln(\sqrt{2\pi(1+\beta^2)}) f(z_i^- | z_i^-, \alpha^{(k)}, \beta^{(k)}) dz_i, \\ I_2 &= \sum_{j=0}^{i-1} \int_{-\infty}^{\infty} \left[-\frac{1}{2(1+\beta^2)} (z_j - \alpha z_{j-1})^2 \right] f(z_i^- | z_i^-, \alpha^{(k)}, \beta^{(k)}) dz_i, \\ I_3 &= \sum_{j=i+2}^n \int_{-\infty}^{\infty} \left[-\frac{1}{2(1+\beta^2)} (z_j - \alpha z_{j-1})^2 \right] f(z_i^- | z_i^-, \alpha^{(k)}, \beta^{(k)}) dz_i, \\ I_4 &= \sum_{j=i}^{i+1} \int_{-\infty}^{\infty} \left[-\frac{1}{2(1+\beta^2)} (z_j - \alpha z_{j-1})^2 \right] f(z_i^- | z_i^-, \alpha^{(k)}, \beta^{(k)}) dz_i, \end{aligned}$$

其中的 $f(z_i^- | z_i^-, \alpha^{(k)}, \beta^{(k)})$ 见(3)式。

下面我们将分别计算 I_1, I_2, I_3 和 I_4 。由前面的计算可知

$$\int_{-\infty}^{\infty} \prod_{j=i}^{i+1} f(z_j - \alpha z_{j-1}) dz_j = 1/a^{(k)} \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}[c^{(k)} - (b^{(k)}/a^{(k)})^2]\right\} (1/\sqrt{2\pi(1+\beta^{(k)2})}) \quad (5)$$

或者直接应用密度函数积分为 1 的性质,可计算得

$$\begin{aligned} I_1 &= -(n+1) \ln(\sqrt{2\pi(1+\beta^2)}), \\ I_2 &= \sum_{j=0}^{i-1} \left[-\frac{1}{2(1+\beta^2)} (z_j - \alpha z_{j-1})^2 \right] = \ln D_{i-1,0}, \\ I_3 &= \sum_{j=i+2}^n \left[-\frac{1}{2(1+\beta^2)} (z_j - \alpha z_{j-1})^2 \right] = \ln D_{n,i+2}. \end{aligned}$$

以下求 I_4 , 整理得

$$I_4 = \sum_{j=i}^{i+1} \int_{-\infty}^{\infty} \left[-\frac{1}{2(1+\beta^2)} (z_j - \alpha z_{j-1})^2 \right] a^{(k)} \exp\left\{\frac{1}{2(1+\beta^{(k)2})}[c^{(k)} - (b^{(k)}/a^{(k)})^2]\right\}.$$

$$\begin{aligned} &\sqrt{2\pi(1+\beta^{(k)2})} \prod_{j=i}^{i+1} f(z_j - \alpha z_{j-1}) dz_i \\ &= I_4^1 + I_4^2, \end{aligned}$$

其中,

$$\begin{aligned} I_4^1 &= -\frac{1}{2(1+\beta^2)} \int_{-\infty}^{\infty} z_i^2 a^{(k)} \sqrt{2\pi(1+\beta^{(k)2})} \exp\left\{\frac{1}{2(1+\beta^{(k)2})}(c^{(k)} - (b^{(k)}/a^{(k)})^2)\right\} \cdot \prod_{j=i}^{i+1} \\ &f(z_j - \alpha z_{j-1}) dz_i + \frac{1}{2(1+\beta^2)} \int_{-\infty}^{\infty} 2z_i z_{i-1} a^{(k)} \sqrt{2\pi(1+\beta^{(k)2})} \exp\left\{\frac{1}{2(1+\beta^{(k)2})}(c^{(k)} - (b^{(k)}/a^{(k)})^2)\right\} \cdot \prod_{j=i}^{i+1} \\ &f(z_j - \alpha z_{j-1}) dz_i - \frac{1}{2(1+\beta^2)} \int_{-\infty}^{\infty} \alpha^2 z_{i-1}^2 a^{(k)} \sqrt{2\pi(1+\beta^{(k)2})} \\ &\exp\left\{\frac{1}{2(1+\beta^{(k)2})}(c^{(k)} - (b^{(k)}/a^{(k)})^2)\right\} \cdot \prod_{j=i}^{i+1} f(z_j - \alpha z_{j-1}) dz_i = l_1^1 + l_1^2 + l_1^3. \end{aligned}$$

为了行文清楚,我们继续做了如上的定义,继续应用(5)式以下分别求 l_1^1, l_1^2 和 l_1^3 有

$$l_1^1 = -\frac{1}{2a^{(k)2}(1+\beta^2)} \int_{-\infty}^{\infty} (a^{(k)} z_i)^2 \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}[a^{(k)} z_i - (b^{(k)}/a^{(k)})]^2\right\} (1/\sqrt{2\pi(1+\beta^{(k)2})}) da^{(k)} z_i.$$

令 $x = a^{(k)} z_i - b^{(k)}/a^{(k)}$, 及利用正态分布的性质,求得上式为

$$l_1^1 = -\frac{1+\beta^{(k)2}}{2a^{(k)2}(1+\beta^2)} - \frac{b^{(k)2}}{2a^{(k)4}(1+\beta^2)}.$$

类似的,对于 l_1^2 及 l_1^3 作同样的变换及应用正态分布的性质,有

$$l_1^2 = \frac{\alpha b^{(k)} z_{i-1}}{a^{(k)2}(1+\beta^2)}, \quad l_1^3 = -\frac{\alpha^2 z_{i-1}^2}{2(1+\beta^2)}.$$

类似的,对于 I_4^2 有

$$\begin{aligned} I_4^2 &= -\frac{1}{2(1+\beta^2)} \int_{-\infty}^{\infty} \alpha^2 z_i^2 a^{(k)} \sqrt{2\pi(1+\beta^{(k)2})} \exp\left\{\frac{1}{2(1+\beta^{(k)2})}(c^{(k)} - (b^{(k)}/a^{(k)})^2)\right\} \cdot \prod_{j=i}^{i+1} f(z_j - \alpha z_{j-1}) dz_i + \frac{1}{2(1+\beta^2)} \\ &\int_{-\infty}^{\infty} 2\alpha z_i z_{i-1} a^{(k)} \sqrt{2\pi(1+\beta^{(k)2})} \exp\left\{\frac{1}{2(1+\beta^{(k)2})}(c^{(k)} - (b^{(k)}/a^{(k)})^2)\right\} \cdot \prod_{j=i}^{i+1} \\ &f(z_j - \alpha z_{j-1}) dz_i = l_2^1 + l_2^2 + l_2^3. \end{aligned}$$

$$\begin{aligned} & \left(b^{(k)} / a^{(k)} \right)^2 \} \circ \prod_{j=i}^{i+1} f(z_j - \alpha^{(k)} z_{j-1}) dz_i \\ & - \frac{1}{2(1+\beta^2)} \int_{-\infty}^{\infty} z_{i+1}^2 a^{(k)} \sqrt{2\pi(1+\beta^{(k)2})} \exp \{ \\ & \frac{1}{2(1+\beta^{(k)2})} (c^{(k)} - (b^{(k)} / a^{(k)})^2) \} \circ \prod_{j=i}^{i+1} \\ & f(z_j - \alpha^{(k)} z_{j-1}) dz_i = l_2^1 + l_2^2 + l_3^3. \end{aligned}$$

同样为了条理清楚, 我们也做了同 I_4^1 一样的定义, 下面通过变量替换及正态分布的性质等详细的计算, 得

$$\begin{aligned} l_2^1 &= -\frac{\alpha^2(1+\beta^{(k)2})}{2a^{(k)2}(1+\beta^2)} - \frac{\alpha^2 b^{(k)2}}{2a^{(k)4}(1+\beta^2)}; \quad l_2^2 = \\ & \frac{b^{(k)} z_{i+1} \alpha}{a^{(k)2}(1+\beta^2)}; \quad l_2^3 = -\frac{z_{i+1}^2}{2(1+\beta^2)} \end{aligned}$$

综上, 我们得到

$$\begin{aligned} I_4 &= -\frac{1+\beta^{(k)2}}{2a^{(k)2}(1+\beta^2)} - \frac{b^{(k)2}}{2a^{(k)4}(1+\beta^2)} + \\ & \frac{b^{(k)} z_{i-1} \alpha}{a^{(k)2}(1+\beta^2)} - \frac{z_{i-1}^2 \alpha^2}{2(1+\beta^2)} - \frac{(1+\beta^{(k)2}) \alpha^2}{2a^{(k)2}(1+\beta^2)} - \\ & \frac{b^{(k)2} \alpha^2}{2a^{(k)4}(1+\beta^2)} + \frac{b^{(k)} z_{i+1} \alpha}{a^{(k)2}(1+\beta^2)} - \frac{z_{i+1}^2}{2(1+\beta^2)}. \end{aligned}$$

结合 I_1, I_2, I_3 故有该命题成立。

为了估计未知参数, 进一步整理式(4), 将其表示为未知参数 (α, β) 的函数, 记为 $g(\alpha, \beta)$ 。则

$$g(\alpha, \beta) = -(n+1) \ln(\sqrt{2\pi(1+\beta^2)}) - \frac{1}{2(1+\beta^2)} A \alpha^2 + \frac{1}{2(1+\beta^2)} B \alpha - \frac{1}{2(1+\beta^2)} C, \quad (6)$$

其中

$$A = z_{-1}^2 + z_0^2 + \dots + z_{i-2}^2 + z_{i-1}^2 + z_{i+1}^2 + \dots + z_{n-1}^2 + \frac{1+\beta^{(k)2}}{a^{(k)2}} + \frac{b^{(k)2}}{a^{(k)4}},$$

$$B = 2(z_{-1}z_0 + z_0z_1 + \dots + z_{i-2}z_{i-1} + z_{i+1}z_{i+2} + \dots + z_{n-1}z_n) + 2 \frac{b^{(k)} z_{i-1}}{a^{(k)2}} + 2 \frac{b^{(k)} z_{i+1}}{a^{(k)2}},$$

$$C = z_0^2 + z_1^2 + \dots + z_{i-1}^2 + z_{i+1}^2 + z_{i+2}^2 + \dots + z_n^2 + \frac{1+\beta^{(k)2}}{a^{(k)4}} + \frac{b^{(k)2}}{a^{(k)4}}.$$

2.3 EM 算法的 M 步(求解 α, β)

为了求得未知参数的最佳估计值, 应用极大似然思想, 对(6)式分别关于未知参数 α 和 β 求一阶偏导数并令其等于 0, 求解二元方程组可得新的参数的估计值 $\alpha^{(k+1)}$ 和 $\beta^{(k+1)}$ 。将新的估计值再带入公式(6), 继续用上面的方法求得解 $\alpha^{(k+2)}$ 和 $\beta^{(k+2)}$, 如此反复下去, 直到达到事先给定的精度为止。

为了解题的方便, 我们不妨设 $\sigma^2 = 1+\beta^2$, 则公式(6)化为

$$\begin{aligned} g(\alpha, \sigma^2) &= -(n+1) \ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2} A \alpha^2 + \\ & \frac{1}{2\sigma^2} B \alpha - \frac{1}{2\sigma^2} C \end{aligned} \quad (6)'$$

以下只需解方程组

$$\frac{\partial g(\alpha, \sigma^2)}{\partial \alpha} = 0$$

$$\frac{\partial g(\alpha, \sigma^2)}{\partial \sigma^2} = 0$$

整理得

$$\begin{aligned} \alpha &= \frac{1}{2} B / A \\ \sigma^2 &= \frac{1}{(n+1)} (-\frac{1}{4} B^2 / A + C) \end{aligned} \quad (7)$$

用微分法可以验证(7)式中的 α 和 σ^2 确实使(6)式达到最大。这样我们总结出估计未知参数 α 和 β 的具体步骤如下:

1)给出初值 $(\alpha^{(0)}, \beta^{(0)})$, 以 $(\alpha^{(0)}, \sigma^{(0)2})$ 参加计算;

2)利用样本及给出的初值计算公式(6)中 A, B, C 各值;

3)计算(7)式, 求出 $(\alpha^{(1)}, \sigma^{(1)2})$;

4)以 $(\alpha^{(1)}, \sigma^{(1)2})$ 为初值从 2)开始循环计算, 直到达到事先给定的精度, 再利用公式 $\sigma^2 = 1+\beta^2$ 重新计算出 β , 这就得到了参数 (α, β) 的估计值。

3 连续两个数据缺失时的参数的估计方法

本节主要讨论当有两个连续的数据 (z_{i-1}, z_i) 缺失时, 用 EM 算法求未知参数 (α, β) 的计算公式。如前假设可获得的观察值序列为 $z_{-1}, z_0, z_1, \dots, z_{i-2}, z_{i+1}, \dots, z_n$ 。

3.1 基于观察值的似然函数的计算

类似于第二节, 我们首先获得关于观测值的似然函数 $L(\alpha, \beta)$ 的定理如下:

命题 3 在观测数据 (z_{i-1}, z_i) 缺失的情况下, 实际观测样本的似然函数为

$$\begin{aligned} L(\alpha, \beta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L(\alpha, \beta) dz_{i-1} dz_i \\ &= \int_{-\infty}^{\infty} \frac{1}{a} \exp \left\{ -\frac{1}{2} (A^2 z_i + 2Bz_i + C) \right\} \cdot (1/\sqrt{2\pi(1+\beta^2)})^2 dz_i \cdot D_{i-2,0} D_{n,i+2} \\ &= (1/\sqrt{2\pi(1+\beta^2)})^{n-1} \frac{1}{aA} \exp \left\{ -\frac{1}{2(1+\beta^2)} (C - \frac{B^2}{A^2}) \right\} D_{i-2,0} D_{n,i+2} \end{aligned} \quad (8)$$

其中

$$A = 1 + \alpha^2 - \alpha^2/a^2, B = \alpha z_{i+1} + \frac{\alpha^2}{a^2} z_{i-2}, C = \alpha^2 (1 - \frac{1}{a^2}) z_{i-2}^2 + z_{i+1}^2,$$

$$D_{i-2,0} = \prod_{j=0}^{i-2} \exp\left\{-\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2\right\},$$

$$D_{n,i+2} = \prod_{j=i+2}^n \exp\left\{-\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2\right\}$$

证明: 由于

$$\prod_{j=i+1}^{i+1} f(z_j - \alpha z_{j-1}) = \left(\frac{1}{\sqrt{2\pi(1+\beta^2)}}\right)^3 \exp\left\{-\frac{1}{2(1+\beta^2)}\left[\left(az_{i-1} - \frac{\alpha(z_i + z_{i-2})}{a}\right)^2 + ((1+\alpha^2)z_i^2 - \frac{\alpha^2}{a^2}(z_i + z_{i-2})^2 - 2\alpha z_{i+1}z_i + \alpha^2 z_{i-2}^2 + z_{i+1}^2)\right]\right\}$$

则利用正态分布的性质及命题中给的变量替换, 可最后得

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=i+1}^{i+1} f(z_j - \alpha z_{j-1}) dz_{i-1} dz_i = \frac{1}{\sqrt{2\pi(1+\beta^2)}} \cdot \frac{1}{aA} \exp\left\{-\frac{1}{2(1+\beta^2)}(C - \frac{B^2}{A^2})\right\} \quad (9)$$

这样我们证明了命题 3 的正确性。

3.2 相应于两个连续缺失数据的 EM 算法(E 步)

仍然设 $\alpha^{(k)}$ 和 $\beta^{(k)}$ 为经过第 k 步运算后得到的参数的值, 由于 a, A, B 和 C 都是 (α, β) 的函数, 则经过第 k 步运算后得到的这些参数也分别记为 $a^{(k)}, A^{(k)}, B^{(k)}$ 和 $C^{(k)}$ 。记 $(z_{i-1}, z_i)^- = \{z_{-1}, z_0, \dots, z_n\} / \{z_{i-1}, z_i\}$ 。则由前面运算可以得到 (z_{i-1}, z_i) 关于 $(z_{i-1}, z_i)^-$ 和参数 $\alpha^{(k)}$ 和 $\beta^{(k)}$ 的条件概率密度为

$$f(z_{i-1}, z_i^- | (z_{i-1}, z_i)^-, \alpha^{(k)}, \beta^{(k)}) = L(\alpha^{(k)}, \beta^{(k)}) / L(\alpha^{(k)}, \beta^{(k)})$$

$$= \left(\frac{1}{\sqrt{2\pi(1+\beta^{(k)2})}}\right)^{-1} \cdot \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(C^{(k)} - \frac{B^{(k)2}}{A^{(k)2}})\right\} \cdot a^{(k)} A^{(k)} \prod_{j=i+1}^{i+1} f(z_j - \alpha^{(k)} z_{j-1}).$$

于是有

命题 4 在 E 步计算中

$$E(\ln(L(\alpha, \beta)) | (z_{i-1}, z_i)^-, \alpha^{(k)}, \beta^{(k)}) = -(n+1) \ln(\sqrt{2\pi(1+\beta^{(k)2})}) + \ln D_{i-1,0} + \ln D_{n,i+2} + D \quad (10)$$

其中, D 有如下形式

$$D = \sum_{i=1}^6 U_i. \text{ 各 } U_i, i = 1, \dots, 6 \text{ 如下}$$

$$U_1 = -\frac{(1+\alpha^2)(1+\beta^{(k)2})}{2(1+\beta^2)a^{(k)2}} - \frac{\alpha^{(k)2}(1+\alpha^2)}{2(1+\beta^2)a^{(k)4}}$$

$$z_{i-2}^2 - \frac{\alpha^{(k)2}(1+\alpha^2)}{2(1+\beta^2)a^{(k)4}} \cdot \left[\frac{(1+\beta^{(k)2})}{A^{(k)2}} + \frac{B^{(k)2}}{A^{(k)4}} + \frac{2B^{(k)}}{A^{(k)2}z_{i-2}} \right],$$

$$U_2 = -\frac{(1+\alpha^2)(1+\beta^{(k)2})}{2(1+\beta^2)A^{(k)2}} - \frac{(1+\alpha^2)B^{(k)2}}{2(1+\beta^2)A^{(k)4}},$$

$$U_3 = \frac{\alpha^{(k)}\alpha}{(1+\beta^2)a^{(k)2}} \left[\frac{(1+\beta^{(k)2})}{A^{(k)2}} + \frac{B^{(k)2}}{A^{(k)4}} + \frac{B^{(k)}}{A^{(k)2}z_{i-2}} \right],$$

$$U_4 = \frac{\alpha^{(k)}\alpha}{(1+\beta^2)a^{(k)2}} z_{i-2}^2 + \frac{\alpha^{(k)}\alpha B^{(k)}}{(1+\beta^2)a^{(k)2} A^{(k)2}} z_{i-2},$$

$$U_5 = \frac{\alpha z_{i+1} B^{(k)}}{(1+\beta^2)A^{(k)2}}, U_6 = -\frac{1}{2(1+\beta^2)}(z_{i+1}^2 + \alpha^2 z_{i-2}^2).$$

证明: 由公式(9)及密度函数的性质, 计算得

$$E(\ln(L(\alpha, \beta)) | (z_{i-1}, z_i)^-, \alpha^{(k)}, \beta^{(k)})$$

$$= -(n+1) \ln \sqrt{2\pi(1+\beta^2)} + \sum_{j=0}^{i-2} \left[-\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2 \right] + \sum_{j=i+2}^n \left[-\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2 \right] + \sum_{j=i+1}^{i+1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} -\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2$$

$$a^{(k)} A^{(k)} \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(C^{(k)} - \frac{B^{(k)2}}{A^{(k)2}})\right\}.$$

$$\sqrt{2\pi(1+\beta^{(k)2})} \prod_{j=i+1}^{i+1} f(z_j - \alpha^{(k)} z_{j-1}) dz_{i-1} dz_i$$

上式中前三项恰好是命题 4 的前三项。我们以下求最后一个加号后的式子。

$$\sum_{j=i+1}^{i+1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} -\frac{1}{2(1+\beta^2)}(z_j - \alpha z_{j-1})^2$$

$$a^{(k)} A^{(k)} \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(C^{(k)} - \frac{B^{(k)2}}{A^{(k)2}})\right\}.$$

$$\sqrt{2\pi(1+\beta^{(k)2})} \prod_{j=i+1}^{i+1} f(z_j - \alpha^{(k)} z_{j-1}) dz_{i-1} dz_i$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} -\frac{1}{2(1+\beta^2)}[(1+\alpha^2)z_{i-1}^2 + (1+\alpha^2)z_i^2 - 2\alpha z_{i-1}z_i - 2\alpha z_{i-2}z_{i-1} - 2\alpha z_{i+1}z_i + (z_{i+1}^2 + \alpha^2 z_{i-2}^2)] \cdot a^{(k)} A^{(k)} \exp$$

$$\frac{1}{2(1+\beta^{(k)2})}(C^{(k)} - \frac{B^{(k)2}}{A^{(k)2}}).$$

$$\sqrt{2\pi(1+\beta^{(k)2})} \prod_{j=i+1}^{i+1} f(z_j - \alpha^{(k)} z_{j-1}) dz_{i-1} dz_i = U_1 + U_2 + U_3 + U_4 + U_5 + U_6$$

对应于最后等号中各式子我们分别做上述各项

的定义。这其中,第二、五、六项相对较简单,我们可以直接应用正态分布的性质进行与第二节相类似的运算,得到

$$\begin{aligned} U_2 &= -\frac{(1+\alpha^2)(1+\beta^{(k)2})}{2(1+\beta^2)A^{(k)2}} - \frac{(1+\alpha^2)B^{(k)2}}{2(1+\beta^2)A^{(k)4}}, \\ U_1 &= -\frac{1}{2(1+\beta^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1+\alpha^2)z_{i-1}^2 \circ A^{(k)} \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(a^{(k)}z_{i-1} - \frac{\alpha^{(k)}(z_i+z_{i-2})}{a^{(k)}})^2\right\} dz_{i-1} \\ &\quad \circ A^{(k)} \left(\frac{1}{\sqrt{2\pi(1+\beta^{(k)2})}}\right)^2 \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(A^{(k)}z_i - \frac{B^{(k)}}{A^{(k)}})^2\right\} dz_i \\ &= -\frac{1+\alpha^2}{2(1+\beta^2)a^{(k)2}} \int_{-\infty}^{\infty} [(1+\beta^{(k)2}) + \frac{\alpha^{(k)2}(z_i+z_{i-2})^2}{a^{(k)2}}] \circ A^{(k)} \frac{1}{\sqrt{2\pi(1+\beta^{(k)2})}} \circ \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(A^{(k)}z_i - \frac{B^{(k)}}{A^{(k)}})^2\right\} dz_i \\ &= -\frac{(1+\alpha^2)(1+\beta^{(k)2})}{2(1+\beta^2)a^{(k)2}} - \frac{\alpha^{(k)2}(1+\alpha^2)}{2(1+\beta^2)a^{(k)4}z_{i-2}^2} \\ &- \frac{\alpha^{(k)2}(1+\alpha^2)}{2(1+\beta^2)a^{(k)4}} \int_{-\infty}^{\infty} (z_i^2 + 2z_{i-2}z_i) \frac{1}{\sqrt{2\pi(1+\beta^{(k)2})}} \\ &\circ A^{(k)} \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(A^{(k)}z_i - \frac{B^{(k)}}{A^{(k)}})^2\right\} dz_i \end{aligned}$$

此时 U_1 式变成了一重积分,按照类似于第二节中的方法计算可获得最后结果为

$$\begin{aligned} U_1 &= -\frac{(1+\alpha^2)(1+\beta^{(k)2})}{2(1+\beta^2)a^{(k)2}} - \frac{\alpha^{(k)2}(1+\alpha^2)}{2(1+\beta^2)a^{(k)4}z_{i-2}} \\ &- \frac{\alpha^{(k)2}(1+\alpha^2)}{2(1+\beta^2)a^{(k)4}} \circ [\frac{(1+\beta^{(k)2})}{A^{(k)2}} + \frac{B^{(k)2}}{A^{(k)4}} + \frac{2B^{(k)}}{A^{(k)2}z_{i-2}}] \end{aligned}$$

依照相同办法,再分别计算 U_3, U_4 有

$$\begin{aligned} U_3 &= \frac{\alpha^{(k)}\alpha}{(1+\beta^2)a^{(k)2}} \left[\frac{(1+\beta^{(k)2})}{A^{(k)2}} + \frac{B^{(k)2}}{A^{(k)4}} + \frac{B^{(k)}}{A^{(k)2}z_{i-2}} \right], \\ U_4 &= \frac{\alpha^{(k)}\alpha}{(1+\beta^2)a^{(k)2}z_{i-2}^2} + \frac{\alpha^{(k)}\alpha B^{(k)}}{(1+\beta^2)a^{(k)2}A^{(k)2}z_{i-2}} \end{aligned}$$

这样我们得出了命题 4 的结论。为了方便进行参数的估计把(10)式进一步整理、化简为未知参数的函数得

$$\begin{aligned} g(\alpha, \beta) &= E(\ln(L(\alpha, \beta)) | (z_{i-1}, z_i)^-, \alpha^{(k)}, \beta^{(k)}) \\ &= -(n+1)\ln(\sqrt{2\pi(1+\beta^2)}) - \frac{1}{2(1+\beta^2)}U\alpha^2 \\ &+ \frac{1}{2(1+\beta^2)}V\alpha - \frac{1}{2(1+\beta^2)}W \end{aligned} \quad (11)$$

其中,

$$\begin{aligned} U &= z_{-1}^2 + z_0^2 + \dots + z_{i-2}^2 + z_{i+1}^2 + \dots + z_{n-1}^2 + \frac{B^{(k)2}}{A^{(k)4}} + \frac{1+\beta^{(k)2}}{A^{(k)2}} + \frac{\alpha^{(k)2}}{a^{(k)4}}(z_{i-2}^2 + \frac{2B^{(k)}}{A^{(k)2}z_{i-2}} + \frac{B^{(k)2}}{A^{(k)4}} + \dots) \end{aligned}$$

$$U_5 = \frac{\alpha z_{i+1} B^{(k)}}{(1+\beta^2)A^{(k)2}}, \quad U_6 = -\frac{1}{2(1+\beta^2)}(z_{i+1}^2 + \alpha^2 z_{i-2}^2)$$

以下我们详细计算剩余的几项

$$\begin{aligned} &\frac{1}{2(1+\beta^{(k)2})} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1+\alpha^2)z_{i-1}^2 \circ A^{(k)} \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(a^{(k)}z_{i-1} - \frac{\alpha^{(k)}(z_i+z_{i-2})}{a^{(k)}})^2\right\} dz_{i-1} \\ &\circ A^{(k)} \left(\frac{1}{\sqrt{2\pi(1+\beta^{(k)2})}}\right)^2 \exp\left\{-\frac{1}{2(1+\beta^{(k)2})}(A^{(k)}z_i - \frac{B^{(k)}}{A^{(k)}})^2\right\} dz_i \\ &= \frac{1+\beta^{(k)2}}{A^{(k)2}} - \frac{1+\beta^{(k)2}}{a^{(k)2}} \\ V &= 2(z_{-1}z_0 + z_0z_1 + \dots + z_{i-3}z_{i-2} + z_{i+1}z_{i+2} \\ &+ \dots + z_{n-1}z_n) + 2 \frac{\alpha^{(k)}}{a^{(k)2}} z_{i-2}^2 + 2(\frac{B^{(k)}}{A^{(k)2}} z_{i+1} + 2 \\ &\frac{\alpha^{(k)}B^{(k)}}{a^{(k)2}A^{(k)2}} z_{i-2} + \frac{\alpha^{(k)}}{a^{(k)2}A^{(k)2}}(1+\beta^{(k)2} + \frac{B^{(k)2}}{A^{(k)2}})) \\ W &= z_0^2 + \dots + z_{i-2}^2 + z_{i+1}^2 + \dots + z_n^2 + (1 \\ &+ \beta^{(k)2})(\frac{1}{a^{(k)2}} + \frac{1}{A^{(k)2}}) + \frac{\alpha^{(k)2}}{a^{(k)4}}(\frac{1+\beta^{(k)2}}{A^{(k)2}} + \frac{B^{(k)2}}{A^{(k)4}}) + 2 \\ &\frac{B^{(k)}}{A^{(k)2}} z_{i-2} + z_{i-2}^2) + \frac{B^{(k)2}}{A^{(k)4}} \end{aligned}$$

3.3 EM 算法的 M 步(求解 α, β)

(1)事实上,只要满足 $U > 0$,就仍然可以用和第二节相同的解法,即对 $g(\alpha, \beta)$ 分别关于 α 和 β 求一阶偏导并令其等于 0,求解即可,整理有

$$\begin{aligned} \alpha &= \frac{1}{2}V/U \\ \sigma^2 &= \frac{1}{(n+1)}(\frac{1}{4}V^2/U - W) \end{aligned} \quad (12)$$

这样我们总结出估计未知参数 α 和 β 的具体步骤如下:

- 1) 给出初值 $(\alpha^{(0)}, \beta^{(0)})$,以 $(\alpha^{(0)}, \sigma^{(0)2})$ 参加计算;
- 2) 利用样本及给出的初值计算公式(11)中各项值;
- 3) 计算(12)式,求出 $(\alpha^{(1)}, \sigma^{(1)2})$;
- 4) 以 $(\alpha^{(1)}, \sigma^{(1)2})$ 为初值从 2)开始循环计算,直到达到事先给定的精度,再利用公式 $\sigma^2 = 1 + \beta^2$ 从新计算出 β ,这就得到了参数 (α, β) 的估计值。

(2)当 $U > 0$ 不满足时,还可以用其它办法获得参数 (α, β) 的估计值。如数值计算法,或者利用下面的思想获得参数的估计值:先固定一个参数,只求另一个参数使其达到目标函数最大,然后再对该参数固定到所求的值上,求第一个参数使其在这种条件下使目标函数达到最大,依次获得参数的估计值。

当然,我们的研究结果也可以扩展成为处理

AR(p)，其误差项 ϵ 为相互独立的服从均值为零，方差是 σ^2 (未知) 模型的参数的估计方法。当 $p = 1$ 时，方法基本与前面提到的相同，当 $p \geq 2$ 时可结合本文与田萍等 2003 年 4 月发表的《缺失数据下 AR(p) 模型的估计方法》一文。当扩展成处理 ARMA($p, 1$), $p \geq 2$ 时，大体上的计算方法与本文相同，但运算及结果都较复杂。

4 实例应用

本节我们将利用前面得到的结论，分别对上证股市的综合指数与深圳综合指数对数差序列的 ARMA(1, 1) 模型进行缺失一个数据的建模，并且与没有数据缺失时的模型进行比较。

分别设深圳成指指数和上证综合指数的时间序列数据为 $x_t, y_t, t = -1, 0, \dots, n$ 。根据实际情况对序列做如下的变换：

设

$$O_t = \ln x_t - \ln y_t, t = -1, 0, \dots, n$$

对 O_t 进行中心化，令 $z_t = O_t - \bar{O}_t, t = -1, 0, 1, \dots, n$ 。其中 $\bar{O}_t = \frac{1}{n+1} \sum_{t=-1}^n O_t$ 。我们关心 z_t 随时间的变化规律。

若 z_t 服从零均值的 ARMA(1, 1) 模型，即有

$$z_t = \alpha z_{t-1} + \epsilon_t + \beta \epsilon_{t-1}, \epsilon_t$$

其中 $\epsilon_t, t = -1, 0, \dots, n$ 为相互独立的标准服从正态分布的随机变量，且 ϵ_t 与 $\{z_s, s < t\}$ 独立。当数据有缺失时，我们可以按照本章的内容进行参数估计，这里我们仅就一个数据缺失时的情况按第二节的内容进行参数估计，并与没有数据缺失时的参数估计值进行对比。所得结果见表 1。

表 1 对上证综合指数与深证成指指数变换后的模型参数估计

本文算法参数估计值		没数据缺失时的估计值	
$\hat{\alpha}$	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\beta}$
0.979121	1.54E-5	0.976316	0.054792

我们选择了上证综合指数与深证成指的 2004 年 7 月到 2005 年年底的数据作为样本，当使用本文的方法为简单仅设第 51 个数据为缺失数据，为了便于对结果的比较使用所有的样本数据用最小二乘法对该模型的参数进行了估计。当数据没有缺失时，得到的 α 的估计值为 0.976216，并且在置信水平为 0.01 甚至更小的情况下显著，但是 β 的估计系数为 0.054792，在置信水平为 0.1 甚至再大一些时也不显著。这说明对于这两个指数的关系使用 AR 模型是正确的，但其并不符合 MA 模型。再来看按

照本文给出的方法所估计的结果 α 的估计值为 0.979121，这与前面方法的结果基本相同。而 σ^2 的估计系数为 1.54E-5，其明显小于 1，这与文中 σ^2 应大于 1 相矛盾，在我们的理论没有问题的情况下，唯一的解释就是，所考察的时间序列不符合给出的模型。这样我们也得到了与前面方法相同的结论。

由此可见，我们给出的方法不但能够处理在数据有缺失时的模型的参数估计的问题，并且在估计的过程中对那些不符合的模型也能够给出提示，进而进行模型的改进。

5 结语

本文主要应用 EM 算法并结合极大似然方法推导了在有一个或连续两个数据缺失条件下的 ARMA(1, 1) 模型的参数估计公式，并利用该公式对我国的股票市场数据进行了实证分析，且与无数据缺失时得到的结果进行了对比，得到了较理想的结果。该结果适用于很多如金融、医学、统计学等用时间序列数据模型考察的实证研究问题。当然，我们的结论还很不全面，如对于一般化的 ARMA(p, q) 模型的估计方法还有待于进一步考察。

参考文献：

- [1] 林静, 韩玉启, 朱惠明. 基于 MCMC 稳态模拟的贝叶斯经验费率厘定信用模型 [J]. 中国管理科学, 2006, 14(2): 33—38.
- [2] 田萍, 董险峰, 王德辉, 等. 缺失数据下 AR(p) 模型的估计方法 [J]. 吉林大学学报(理学版), 2003, (2): 127—133.
- [3] 顾龙全, 周晓东, 汤银才. 指数分布场合恒加试验缺失数据的 Bayes 统计分析 [J]. 高校应用数学学报 A(中文版), 2006, (2): 183—190.
- [4] 庞新生. 多重插补处理缺失数据方法的理论基础探析 [J]. 统计与决策, 2005, (4): 12—14.
- [5] 闫长华, 张忠占. 缺失数据下线性 EV 模型的参数估计 [J]. 数学的实践与认识, 2005, 12: 116—122.
- [6] James L. Peugh and Craig K. Enders, Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement, Review of Educational Research, 2004, 74, (4): 525—556.
- [7] Andrea Rotnitzky, David Wypij. A Note on the Bias of Estimators with Missing Data [J]. Biometrics, 1994, 50(4): 1163—1170.
- [8] Sik-Yum Lee, Xin-Yuan Song, John C. K. Lee. Maximum Likelihood estimation of Nonlinear Structural Equation Models with Ignorable Missing Data [J]. Journal of Educational and Behavioral Statistics, 2003, 28,

- (2). :111—134.
- [9] Ke—Hai Yuan; Peter M. Bentler. Three Likelihood—Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data, *Sociological Methodology*, 2000 30: 165—200.
- [10] Hua Yun Chen, Double—Semiparametric Method for Missing Covariates in Cox Regression Models[J] . *Journal of the American Statistical Association*, 2002, 97 (458): 565—576.
- [11] José A. F. Machado; J. Roderick McCrorie; Jeremy Penzer, Estimation of Time—Series Regressions with Autoregressive Disturbances and Missing Observations,
- Econometric Theory, 14(5): 689—691.
- [12] Ljung, G. M. The likelihood function for a stationary Gaussian autoregressive moving average process with missing observations[J] . *Biometrika*, 1982, (69): 265—8.
- [13] Penzer, J. and Shea, B. The exact likelihood of an autoregressive—moving average model with incomplete data[J] . *Biometrika*, 1997, 84 .
- [14] Chunsheng Ma. Maximum Likelihood estimation of ARMA(1, 1) with missing data[J] . *Journal of time series analysis*, 2002, 23: 49—55.

Estimation Method of ARMA(1, 1) Model with Missing Data

TIAN Ping, ZHANG Qi-shan

(Business School of Jilin University, Jilin 130012, China)

Abstract: In recent decades, when analysis and process two modern financial issues of "risk management and effectiveness of the optimization" in the international community, the financial time series model and the measurement of the growing corresponding play a very important role. The linear time-series models Have been known such as the AR (p), MA (q), ARMA (p, q). There are a lot of classical methods such as the least-squares method and the maximum likelihood method can estimate parameters of many models except missing data in the middle, when the above methods will be powerless. This article will discuss in detail the data deletion ARMA (1, 1) model, that is, the parameter estimation methods of $Z_t = \alpha Z_{t-1} + \epsilon_t - \beta \epsilon_{t-1}$.

Key words: missing data; ARMA (1, 1) model; likelihood function; EM algorithm