第 17 卷 第 4 期　　自 动 化 学 报　　Vol. 17, No. 4
1991 年 7 月　　ACTA AUTOMATICA SINICA　　July, 1991

残差初值的自适应估计

张裕明　吴　林
(哈尔滨工业大学)

摘　　要

本文通过对工程上常用的残差初值选取方法——均值法统计特性的深入分析，揭示了残差初值问题的重要性．在研究了均值法和最小二乘法等两种简单残差初值估计方法统计特性的基础上，得到了一种简单、实用、附加计算量小、统计特性优良的残差初值估计方法．该方法的特点是能自动根据系统参数的变化来调整残差初值算法的参数，保证残差初值估计的优良统计特性．

关键词：残差初值，残差平方和，自适应估计，鲁棒估计，统计特性．

一、引　　言

残差初值对系统参数的估计精度有不可忽视的影响，Pandit、吴贤铭[1]、Åström[2]以及安鸿志等[3]已注意和研究过此问题．Pandit、吴贤铭[1]注意到，小样本时残差初值取零会影响 ARMA (p, q) 模型参数的最小二乘估计精度，建议小样本时采用极大似然法估计参数以提高精度． Åström[2] 则把残差初值作为自由变量，与系统参数一起进行优化．安鸿志等[3]则提出采用最小二乘估计残差初值的思想，但没有给出具体方法．由于极大似然法比较复杂，而残差初值作为自由变量增加了优化问题的维数，从而造成计算量增大和对收敛性的影响，因此这些方法尚未在工程上广泛使用． 目前工程上广泛应用的残差初值选取方法仍是最古老最简单的均值法[1,4]，也就是通常取零为残差初值的方法．

对于均值法的性质，本文通过理论分析和数字仿真进行了比较详细的研究，结果表明：利用均值法求得的残差平方和，与实际残差平方和的误差的统计特性直接依赖于系统参数．当系统噪声项算子多项式的零点趋向单位圆以及多项式阶增加时，该统计特性变差．鉴此，本文提出了一种能自动根据系统参数调整残差初值估计算法参数的自适应残差初值估计算法，可以保证残差平方和统计特性的均匀性，从而提高了系统参数最小二乘估计的鲁棒性． 该算法的特点是具有较好的统计特性，并且较 Pandit、吴贤铭和 Åuström 诸方法简便，附加计算量小，故具有推广应用价值．

本文于 1988 年 12 月 8 日收到．

二、系统描述

为讨论方便，先将本文涉及的 MA，ARMA 和 CARMA. 模型以统一形式加以描述. 设各类模型的形式为

1) MA: $y_t=\theta(B)\hat{\varepsilon}_t$, $t=1,2,\cdots,N$. (2.1)

2) ARMA: $\varphi(B)y_t=\theta(B)\hat{\varepsilon}_t$, $t=p+1,\cdots,N$. (2.2)

3) CARMA: $A(B)y_t-B(B)u_t=C(B)\hat{\varepsilon}_t$, $t=p+1,\cdots,N$. (2.3)

其中 $\hat{\varepsilon}_t(0,\sigma^2)$ 为正态白噪声；y_t 为系统输出；u_t 为系统输入；$A(B)$，$B(B)$，$C(B)$，$\varphi(B)$，$\theta(B)$ 均为滞后算子多项式；$A(B)$，$C(B)$，$\varphi(B)$，$\theta(B)$ 为首 1 滞后算子多项式且设其零点均在单位圆外. 而 $A(B)$，$\varphi(B)$，$\theta(B)$ 的阶分别为 p,p,q. 显然，可得到上述各类模型的统一描述如下

$$C_t=\theta(B)\hat{\varepsilon}_t=\left(1-\sum_{j=1}^{q}\theta_jB^j\right)\hat{\varepsilon}_t,\quad t=p+1,\cdots,N. \tag{2.4}$$

其中 C_t 和 $\theta(B)\hat{\varepsilon}_t$ 分别表示 (2.1)—(2.3) 式的左边和右边部分;而对于 MA 模型，其 $p=0$. 把计算 C_t 时所需的系统参数用 Φ 表示,则系统 (2.4) 的参数为 (Φ,θ).

三、均值法的统计特性分析

系统 (2.4) 的参数 (Φ,θ) 的最小二乘估计 (Φ_{LS},θ_{LS}) 可通过极小残差平方和得到

$$\Phi_{LS},\theta_{LS}:\min_{\Phi,\theta}J(\Phi,\theta)=\min_{\Phi,\theta}\sum_{t=p+1}^{N}\hat{\varepsilon}_t^2(\Phi,\theta). \tag{3.1}$$

设 $(\Phi^{(i)},\theta^{(i)})$ 是 (3.1) 式中由优化算法给出的第 i 步迭代参数. 显然，$J(\Phi^{(i)},\theta^{(i)})$ 对 Φ_{LS}，θ_{LS} 起直接作用. 下面将通过推导 $J(\Phi^{(i)},\theta^{(i)})$ 的计算公式来证明 J 不但与 $\Phi^{(i)}$，$\theta^{(i)}$ 有关,而且与残差初值向量

$$\delta=(\delta_1,\cdots,\delta_q)^\tau=(\hat{\varepsilon}_p,\hat{\varepsilon}_{p-1},\cdots,\hat{\varepsilon}_{p-q+1})^\tau \tag{3.2}$$

有关,并研究残差初值 δ 对 J 的影响方式,从而证明残差初值估计问题的重要性. 为方便起见,下面的讨论中将 $\Phi^{(i)}$，$\theta^{(i)}$ 简写成 Φ,θ.

由 (2.4) 式可得

$$\hat{\varepsilon}=(\hat{\varepsilon}_{p+1},\cdots,\hat{\varepsilon}_N)^\tau=\mathbf{s}+A\delta, \tag{3.3}$$

其中 $(N-p)$ 维列向量 $\mathbf{s}$ 和 $(N-p)\times q$ 维矩阵 A 满足下面的递推方程:

$$\begin{cases}\mathbf{s}(i)=C_{p+i}+\sum_{j=1}^{q}\theta_j\mathbf{s}(i-j),\\ A(i)=\sum_{j=1}^{q}\theta_jA(i-j)+(\theta_i,\cdots,\theta_{i+q-1}),\end{cases} \tag{3.4}$$

且递推初值是

$$\begin{cases}\mathbf{s}(1)=C_{p+1},\ \mathbf{s}(K)=0,\ (K\leqslant 0),\\ A(1)=(\theta_1,\cdots,\theta_q),\ A(K)=0,\ (K\leqslant 0).\end{cases} \tag{3.5}$$

而 $\theta_K = 0$ $(K \geqslant q+1)$，$A(i)$ 为 A 的第 i 行。于是

$$J = \hat{\varepsilon}^\tau \varepsilon = \boldsymbol{s}^\tau \boldsymbol{s} + \delta^\tau A^\tau A \delta + 2\boldsymbol{s}^\tau A \delta. \tag{3.6}$$

可见，J 不但与系统参数 Φ，θ 有关，而且与 δ 也有关。显然，$\boldsymbol{s}^\tau \boldsymbol{s}$ 是采用均值法选取残差初值(此时残差均值为零)时对应的残差平方和 $J|_0$，其与实际残差平方和 $J|_\delta$ 的差为

$$\Delta J|_0 = J|_0 - J|_\delta = -\delta^\tau A^\tau A \delta - 2\boldsymbol{s}^\tau A \delta. \tag{3.7}$$

其统计特性由定理 3.1 给出。

定理 3.1. 设 δ 给定，$\hat{\varepsilon} \sim N(0, \sigma^2 I)$，则

$$\Delta J|_0 \sim N(\delta^\tau A^\tau A \delta,\ 4\sigma^2 \delta^\tau A^\tau A \delta).$$

证明。推导 (3.3) 式和 (3.7) 式可得

$$\Delta J|_0 = \delta^\tau A^\tau A \delta - 2\hat{\varepsilon}^\tau A \delta.$$

可见 $\Delta J|_0$ 为正态随机向量 $\hat{\varepsilon}$ 的线性函数，故 $\Delta J|_0$ 为正态分布，且

$$E(\Delta J|_0) = \delta^\tau A^\tau A \delta,$$
$$E(\Delta J|_0 - E(\Delta J|_0))^2 = E(-2\hat{\varepsilon}^\tau A \delta)^2 = 4\sigma^2 \delta^\tau A^\tau A \delta.$$

证毕。

显然，应考虑各种 δ 的平均。由于 $\delta \sim N(0, \sigma^2 I)$，故

$$E(\Delta J|_0) = \sigma^2 \mathrm{tr}(A^\tau A), \tag{3.8}$$
$$E(\Delta J|_0 - E(\Delta J|_0))^2 = 4\sigma^4 \mathrm{tr}(A^\tau A). \tag{3.9}$$

可见，估计量 $J|_0$ 的统计特性由 $\mathrm{tr}(A^\tau A)$ 决定。下面通过 $\mathrm{tr}(A^\tau A)$ 与 $\theta(B)$ 的关系来研究 $\theta(B)$ 参数对 $J|_0$ 统计特性的影响。设 $\theta(B)$ 的根为 $1/\alpha_j$ $(j = 1, 2, \cdots, q)$，则由 (3.4) 式可得下面的齐次方程：

$$\left(\prod_{j=1}^{q} (1 - \alpha_j B)\right) A(i) = 0,\ N - p \geqslant i \geqslant q + 1. \tag{3.10}$$

当 $\theta(B)$ 无重根时有 $A(i) = (\alpha_1^{i-q}, \alpha_2^{i-q}, \cdots, \alpha_q^{i-q}) G (i = q, q+1, \cdots, N-p)$，而 G 由方程 (3.10) 的初值条件决定，与 i 无关。经推导可得

$$\begin{aligned}
\mathrm{tr}(A^\tau A) &= \sum_{i=1}^{N-p} \mathrm{tr}(A^\tau(i) A(i)) \\
&= \sum_{i=1}^{q-1} A(i) A^\tau(i) + \sum_{i=q}^{N-p} \sum_{j_1=1}^{q} \sum_{j_2=1}^{q} Q_{j_1 j_2} \alpha_{j_1}^{i-q} \alpha_{j_2}^{i-q} \\
&= \sum_{i=1}^{q-1} A(i) A^\tau(i) + \sum_{j_1=1}^{q} \sum_{j_2=1}^{q} Q_{j_1 j_2} [1 - (\alpha_{j_1} \alpha_{j_2})^{N-p-q+1}] / (1 - \alpha_{j_1} \alpha_{j_2}).
\end{aligned} \tag{3.11}$$

而

$$\lim_{N \to \infty} \mathrm{tr}(A^\tau A) = \sum_{i=1}^{q-1} A(i) A^\tau(i) + \sum_{j_1=1}^{q} \sum_{j_2=1}^{q} Q_{j_1 j_2} / (1 - \alpha_{j_1} \alpha_{j_2}). \tag{3.12}$$

其中 $Q_{j_1 j_2}$ 是矩阵 GG^τ 的 j_1 行 j_2 列元素。

在 (3.11) 式和 (3.12) 式中，$A(i)(i = 1, \cdots, q)$ 和 $Q_{j_1 j_2}(j_1 = 1, \cdots, q;\ j_2 = 1, \cdots, q)$ 均由 $\theta(B)$ 决定，且可证它们均是 α_j $(j = 1, \cdots, q)$ 的零次幂系数为零的多项式。这样，当所有的 $\alpha_j \to 0 (j = 1, \cdots, q)$ 时，$A(i)$ $(i = 1, \cdots, q-1)$，$Q_{j_1 j_2}$ $(j_1 = 1,$

$\cdots,q$; $j_2=1,\cdots,q$) 均趋于零,于是有 $\mathrm{tr}(A^\tau A)\to 0$, $\lim\limits_{N\to\infty}\mathrm{tr}(A^\tau A)\to 0$. 当有 $|\alpha_j|\to 1$ 时, $1-\alpha_j^2\to 0$ (如该 α_j 为复根, 则 $1-\alpha_j\bar{\alpha}_j\to 0$), 而在一般情况下, $Q_{jj}\neq 0$, 故此时 $\lim\limits_{N\to\infty}\mathrm{tr}(A^\tau A)\to\infty$.

从上面的讨论可对 $J|_0$ 的统计特性得到如下结论: 1) $J|_0$ 直接依赖于 $\theta(B)$ 的参数; 2) 当所有 $\alpha_j\to 0$ 时, $\mathrm{tr}(A^\tau A)\to 0$, $\lim\limits_{N\to\infty}\mathrm{tr}(A^\tau A)\to 0$, $J|_0$ 的统计特性良好; 3) 当有靠近单位圆的 α_j 时,一般来说, $\mathrm{tr}(A^\tau A)$ 和 $\lim\limits_{N\to\infty}\mathrm{tr}(A^\tau A)$ 均较大, $J|_0$ 的统计特性较差.

由于 $A(i)$ 是以 q 时刻为起点作衰减(零输入响应)的, 故研究 $\mathrm{tr}(A^\tau A)$ 的收敛应以 q 为起始点. 为此,令

$$D_N=\frac{\left(\lim\limits_{N\to\infty}\mathrm{tr}(A^\tau A)-\mathrm{tr}(A^\tau A)\right)}{\lim\limits_{N\to\infty}\sum\limits_{i=q}^{N-p}\mathrm{tr}(A^\tau(i)A(i))},\tag{3.13}$$

D_N 反映了 $\mathrm{tr}(A^\tau A)$ 与其极限的相对误差. 设 $\alpha_{\max}=\max\limits_{1\leqslant j\leqslant q}|\alpha_j|=|\alpha_m|(1\leqslant m\leqslant q)$, 则当 N 足够大时可得到

$$D_N\doteq\frac{Q_{mm}\alpha_{\max}^{2(N-p-q+1)}/(1-\alpha_{\max}^2)}{Q_{mm}/(1-\alpha_{\max}^2)}=\alpha_{\max}^{2(N-p-q+1)}.\tag{3.14}$$

如 α_m 为复根,上式的 Q_{mm} 应换成与 $\alpha_m\bar{\alpha}_m$ 对应的 Q 阵元素,其结果不变. 可见, $\mathrm{tr}(A^\tau A)$ 的收敛速度主要由 $\alpha_{\max}$ 决定.

对于 $\theta(B)$ 有重根的情况, 也可以得到与无重根时类似的结果. 限于篇幅, 这里只对其中的关键问题进行说明.

不妨设 $\alpha_{q-1}=\alpha_q$,则 (3.10) 式解的形式为 $A(i)=(\alpha_1^{i-q},\cdots,\alpha_{q-1}^{i-q},(i-q)\alpha_q^{i-q})G$. 对于任意的 $\gamma:1/|\alpha_q|>\gamma>1$, 可使用罗比塔法则得到 $\lim\limits_{i\to\infty}[(i-q)/\gamma^{(i-q)}]=0$, 故当 i 足够大时恒有

$$|\alpha_{j_1}^{i-q}(i-q)\alpha_q^{i-q}|\leqslant|\alpha_{j_1}^{i-q}(\gamma\alpha_q)^{i-q}|,\tag{3.15}$$

$$|(i-q)^2\alpha_q^{2(i-q)}|\leqslant|(\gamma\alpha_q)^{2(i-q)}|.\tag{3.16}$$

因此, $\lim\limits_{N\to\infty}\sum\limits_{i=q}^{N-p}Q_{j_1q}\alpha_{j_1}^{(i-q)}\alpha_q^{(i-q)}(i-q)$ 和 $\lim\limits_{N\to\infty}\sum\limits_{i=q}^{N-p}Q_{qq}(i-q)^2\alpha_q^{2(i-q)}$ 均存在, 且 (3.15) 式和 (3.16) 式左边各项跟无重根时一样随 $|\alpha_{j1}|$, $|\alpha_q|$ 的增减而增减. 基于以上两点, 可得到与无重根时相同的三条结论.

对于 D_N, 在有重根时得不到 (3.14) 式的简练结果. 但由于当 i 充分大时,使(3.15)式、(3.16)式成立的 γ 可以任意靠近 1,故把 $\gamma\alpha_q$ 视为新的第 q 个零点不会影响 $\alpha_{\max}$. 这样, N 充分大时, D_N 的分子仍主要由 $\alpha_{\max}$ 对应项决定,而 D_N 的分母是定值,故这时仍可得到衰减速度随 $\alpha_{\max}$ 增加而降低,随 $\alpha_{\max}$ 减小而增大的结果.

为了进一步了解 $\Delta J|_0$ 统计特性与 $\theta(B)$ 的关系,表 1 给出了一些数字模拟结果,其中 N 是样本数, $\sigma^2=1$, 而 $E(\Delta J|_0)$ 和 $E(\Delta J|_0-E(\Delta J|_0))^2$ 是基于 100 次实现的结果平均得到的.

表 1　$\hat{\delta}$ 与 $\Delta J|\hat{\delta}_{LS}$ 统计特性的数字模拟

统计量 / 特性 / N / α_j	$\Delta J\|_0$ 均值 10	$\Delta J\|_0$ 均值 100	$\Delta J\|_0$ 根方差 10	$\Delta J\|_0$ 根方差 100	$\Delta J\|_{\hat{\delta}_{LS}}$ 均值 10	$\Delta J\|_{\hat{\delta}_{LS}}$ 均值 100	$\Delta J\|_{\hat{\delta}_{LS}}$ 根方差 10	$\Delta J\|_{\hat{\delta}_{LS}}$ 根方差 100	$\Delta J\|_{\hat{\delta}_A}$ 均值 10	$\Delta J\|_{\hat{\delta}_A}$ 均值 100	$\Delta J\|_{\hat{\delta}_A}$ 根方差 10	$\Delta J\|_{\hat{\delta}_A}$ 根方差 100	$\frac{E\|\hat{\delta}_{LS}-\delta\|}{q}$ 10	$\frac{E\|\hat{\delta}_{LS}-\delta\|}{q}$ 100	$\frac{E\|\hat{\delta}_A-\delta\|}{q}$ 10	$\frac{E\|\hat{\delta}_A-\delta\|}{q}$ 100
−0.3	0.08	0.08	.5	.5	−.86	−.86	1.7	1.7	0	0	.46	.46	2.3	2.3	.71	.71
0.7	0.65	0.65	2.5	2.5	−1.0	−1.0	1.7	1.7	−.87	−.88	1.5	1.5	.85	.86	.66	.66
−0.95	4.5	8.0	9.9	16	−1.2	−1.5	1.8	2.2	−1.2	−1.2	1.9	2.2	.39	.34	.36	.27
−0.2, −0.5	0.57	0.57	1.6	1.6	−1.8	−1.8	2.6	2.6	0.36	0.36	1.4	1.4	28	28	.71	.71
−0.3, −0.9	7.2	8.5	14	17	−2.0	−2.1	2.8	3.2	−.68	−.74	1.8	2.1	6.2	6.1	.52	.51
0.8, 0.9	73	112	134	209	−2.0	−2.5	2.5	3.4	−1.6	−2.2	2.1	3.0	.98	.83	.52	.51
−0.8, −0.8	37	42	67	79	−2.1	−2.5	2.9	3.5	−1.5	−1.8	2.4	2.9	1.2	1.3	.52	.57
±j0.9	3.1	3.6	5.2	5.7	−1.4	−1.4	2.2	2.0	−1.4	−1.4	2.1	1.9	.48	.47	.43	.42
−.3,.8,.9	34	51	62	94	−2.8	−3.3	3.4	4.2	−.92	−1.4	1.8	2.5	15	14	.58	.57
−.5,−.6,−.7	40	41	72	73	−2.8	−3.1	3.6	4.0	−.86	−.91	2.1	2.2	17	17	.56	.56

从表 1 还可看到，即使无 α_j 靠近单位圆，但随 q 的增加，$\Delta J|_0$ 的统计特性仍明显变差．产生这种现象的原因是由于 q 的增加导致了 $|Q_{i_1i_2}|$ 的增加和 $\mathrm{tr}(A^\tau A)$ 求和项数 $q\times q$ 的增加．限于篇幅，在此不详细论述．

四、残差初值的自适应估计及统计特性

$\Delta J|_0$ 统计特性与 q 和 α_j 的关系影响了 Φ_{LS}，θ_{LS} 的精度和鲁棒性．为了改善 Φ_{LS}，θ_{LS} 的精度和鲁棒性，本节提出了残差初值的自适应估计 $\hat{\delta}_A$．考虑到 $\hat{\delta}_A$ 是在零均值和 $\hat{\delta}_{LS}$ 的基础上发展起来的，故拟先对 $\hat{\delta}_{LS}$ 的算法和性质进行研究，然后提出 $\hat{\delta}_A$．

1. 残差初值的最小二乘估计 $\hat{\delta}_{LS}$

δ 的最小二乘估计 $\hat{\delta}_{LS}$ 由下面准则确定：

$$\hat{\delta}_{LS}:\ \min_{\delta\in R^q}\|\boldsymbol{s}+A\delta\|_2^2, \tag{4.1}$$

可得

$$\hat{\delta}_{LS}=-(A^\tau A)^{-1}A^\tau\boldsymbol{s}. \tag{4.2}$$

令

$$\Delta J|_{\hat{\delta}_{LS}}=J|_{\hat{\delta}_{LS}}-J|_{\delta}, \tag{4.3}$$

则 $\hat{\delta}_{LS}$ 和 $\Delta J|_{\hat{\delta}_{LS}}$ 的统计特性由定理 4.1 和 4.2 给出．

定理 4.1. 设 $\hat{\varepsilon}\sim N(0,\sigma^2I)$，则

$$\hat{\delta}_{LS}\sim N(\delta,\sigma^2(A^\tau A)^{-1}).$$

定理 4.2. 设 $\hat{\varepsilon}\sim N(0,\sigma^2I)$，则

$$E(\Delta J|_{\hat{\delta}_{LS}}) = -q\sigma^2,$$
$$E(\Delta J|_{\hat{\delta}_{LS}} - E(\Delta J|_{\hat{\delta}_{LS}}))^2 \leqslant 2\sigma^4 q.$$

证明．为书写简便，令 $\Lambda = A(A^\tau A)^{-1}A^\tau$．由 (3.6) 式、(4.2) 式和 (3.3) 式可得

$$\Delta J|_{\hat{\delta}_{LS}} = -\hat{\varepsilon}^\tau \Lambda \hat{\varepsilon},$$

故

$$E(\Delta J|_{\hat{\delta}_{LS}}) = -\sigma^2 \mathrm{tr}\Lambda.$$

由 Λ 的幂等性可知 Λ 的特征值为幂等，即 $\lambda_j^2 = \lambda_j$，于是有 $\lambda_j = 1$ $(j = 1, 2, \cdots, \mathrm{rank}\Lambda)$．通过将 A 中 q 个线性无关的行组成的块作为 A 的前 q 行对 A 进行分块，可以证明 $\mathrm{rank}\Lambda = q$．这样，由韦达定理即可得 $\mathrm{tr}\Lambda = q$，即有

$$E(\Delta J|_{\hat{\delta}_{LS}}) = -q\sigma^2.$$

而

$$(\Delta J|_{\hat{\delta}_{LS}} - E(\Delta J|_{\hat{\delta}_{LS}}))^2 = (\hat{\varepsilon}^\tau \Lambda \hat{\varepsilon})^2 + \sigma^4 q^2 - 2\sigma^2 q(\hat{\varepsilon}^\tau \Lambda \hat{\varepsilon}).$$

由 $\hat{\varepsilon} \sim N(0, \sigma^2 I)$ 可知，$\hat{\varepsilon}_{k_1}\hat{\varepsilon}_{k_2}$ 的概率

$$P(\hat{\varepsilon}_{k_1}\hat{\varepsilon}_{k_2}) = P(\hat{\varepsilon}_{k_1})P(\hat{\varepsilon}_{k_2}),\ (k_1 \neq k_2),$$

也即 $\hat{\varepsilon}_{k_1}$ 和 $\hat{\varepsilon}_{k_2}$ $(k_1 \neq k_2)$ 是独立的．由此独立性和 $\hat{\varepsilon}_k$ 的四阶矩 $E(\hat{\varepsilon}_k)^4 = 3\sigma^4$，并把 $(\hat{\varepsilon}^\tau \Lambda \hat{\varepsilon})^2$ 按 $\hat{\varepsilon}_k^4$，$\hat{\varepsilon}_{k_1}^3\hat{\varepsilon}_{k_2}$，$\hat{\varepsilon}_{k_1}^2\hat{\varepsilon}_{k_2}^2$，$\hat{\varepsilon}_{k_1}^2\hat{\varepsilon}_{k_2}\hat{\varepsilon}_{k_3}$ 及 $\hat{\varepsilon}_{k_1}\hat{\varepsilon}_{k_2}\hat{\varepsilon}_{k_3}\hat{\varepsilon}_{k_4}$ 分解 $(k_1 \neq k_2 \neq k_3 \neq k_4)$，可证明

$$E(\hat{\varepsilon}^\tau \Lambda \hat{\varepsilon})^2 = 2\sigma^4 \sum_{i=1}^{N-p} (\Lambda_{ii})^2 + \sigma^4 q^2.$$

于是

$$E(\Delta J|_{\hat{\delta}_{LS}} - E(\Delta J|_{\hat{\delta}_{LS}}))^2 = 2\sigma^4 \sum_{i=1}^{N-p} (\Lambda_{ii})^2.$$

而由 $\Lambda\Lambda = \Lambda$ 可证 $0 \leqslant \Lambda_{ii} \leqslant 1$，故有

$$E(\Delta J|_{\hat{\delta}_{LS}} - E(\Delta J|_{\hat{\delta}_{LS}}))^2 \leqslant 2\sigma^4 \mathrm{tr}\Lambda = 2\sigma^4 q.$$ 证毕．

由定理 4.2 知，$J|_{\hat{\delta}_{LS}}$ 具有良好的统计特性，且其统计特性与 $\theta(B)$ 的具体参数无关．但从定理 4.1 知，当所有 $|\alpha_j| \to 0 (j = 1, \cdots, q)$ 时，$\hat{\delta}_{LS}$ 的协方差阵 $\sigma^2(A^\tau A)^{-1}$ 将会由于 $A^\tau A$ 趋于零而趋于无穷．表 1 给了一些过程的 $\hat{\delta}_{LS}$ 及 $\Delta J|_{\hat{\delta}_{LS}}$ 统计特性的数字模拟结果，其中 N，σ^2 和计算统计结果所用的实现次数与 $\delta = 0$ 时的模拟计算相同．

2. 残差初值的自适应估计 $\hat{\delta}_A$

通过对 $\hat{\delta}_{LS}$，$\Delta J|_{\hat{\delta}_{LS}}$ 和 $\Delta J|_0$ 统计特性的研究知：1）当所有 $\alpha_j \to 0$ 时，$\hat{\delta}_{LS}$ 统计特性较差，$\Delta J|_0$ 统计特性良好，优于 $\Delta J|_{\hat{\delta}_{LS}}$；2）$|\alpha_j| \to 1$ 时，$\hat{\delta}_{LS}$ 的统计特性良好，$\Delta J|_0$ 统计特性很差，远劣于 $\Delta J|_{\hat{\delta}_{LS}}$．这样，对 $\hat{\delta} = 0$ 和 $\hat{\delta}_{LS}$，当所有 $\alpha_j \to 0$ 时就应选前者，而有 $|\alpha_j| \to 1$ 时应选 $\hat{\delta}_{LS}$．本文的残差初值自适应估计 $\hat{\delta}_A$ 即是一种根据 α_j 而自动调整估计算法参数，保证估计质量的残差初值鲁棒估计量：

$$\hat{\delta}_A:\ \min_{\delta \in R^q}(\|s + A\delta\|_2^2 + \delta^\tau P \delta^\tau). \tag{4.4}$$

其中 $P \geqslant 0$ 为 $q \times q$ 加权阵，可得

$$\hat{\delta}_A = -(A^\tau A + P)^{-1}A^\tau s. \tag{4.5}$$

令

$$\Delta J|_{\hat{\delta}_A} = J|_{\hat{\delta}_A} - J|_{\delta}, \tag{4.6}$$

则 $\hat{\delta}_A$ 和 $\Delta J|_{\hat{\delta}_A}$ 的统计特性由定理 4.3 和定理 4.4 给出.

定理 4.3. 设 $\hat{\varepsilon} \sim N(0, \sigma^2 I)$，则

$$\hat{\delta}_A \sim N((A^\tau A + P)^{-1}A^\tau A\delta,\ \sigma^2(A^\tau A + P)^{-1}A^\tau A(A^\tau A + P)^{-1}).$$

定理 4.4. 设 $\hat{\varepsilon} \sim N(0, \sigma^2 I)$，则

$$E(\Delta J|_{\hat{\delta}_A}) = \delta^\tau B\delta + \sigma^2 \mathrm{tr} C,$$

$$E(\Delta J|_{\hat{\delta}_A} - E(\Delta J|_{\hat{\delta}_A}))^2 = 2\sigma^4 \sum_{i=1}^{N-p} C_{ii}^2 + \sigma^2\delta^\tau DD^\tau\delta.$$

其中 $B = A^\tau A(A^\tau A + P)^{-1}A^\tau A(A^\tau A + P)^{-1}A^\tau A$
$\quad - 2A^\tau A(A^\tau A + P)^{-1}A^\tau A + A^\tau A.$

$C = A(A^\tau A + P)^{-1}A^\tau A(A^\tau A + P)^{-1}A^\tau - 2A(A^\tau A + P)^{-1}A^\tau,$

$D = -2A^\tau A(A^\tau A + P)^{-1}A^\tau A(A^\tau A + P)^{-1}A^\tau$
$\quad + 4A^\tau A(A^\tau A + P)^{-1}A^\tau - 2A^\tau.$

证明. 将 (4.5) 式代入 (4.6) 式，可得

$$\Delta J|_{\hat{\delta}_A} = \delta^\tau B\delta + \hat{\varepsilon}^\tau C\hat{\varepsilon} + \delta^\tau D\hat{\varepsilon}.$$

其中 B，C，D同上所述. 于是 $E(\Delta J|_{\hat{\delta}_A}) = \delta^\tau B\delta + \sigma^2\mathrm{tr}C$. 故

$$E(\Delta J|_{\hat{\delta}_A} - E(\Delta J|_{\hat{\delta}_A}))^2 = (\hat{\varepsilon}^\tau C\hat{\varepsilon})^2 + \delta^\tau D\hat{\varepsilon}\hat{\varepsilon}^\tau D^\tau\delta$$
$$+ \sigma^4\mathrm{tr}^2 C + 2\hat{\varepsilon}^\tau C\hat{\varepsilon}\delta^\tau D\hat{\varepsilon} - 2\hat{\varepsilon}^\tau C\hat{\varepsilon}\sigma^2\mathrm{tr}C - 2\delta^\tau D\hat{\varepsilon}\sigma^2\mathrm{tr}C.$$

由定理 4.2 的证明和 $E(\hat{\varepsilon}_k)^3 = 0$ 可分别证明

$$E(\hat{\varepsilon}^\tau C\hat{\varepsilon})^2 = 2\sigma^4\sum_{i=1}^{N-p} C_{ii}^2 + \sigma^4\mathrm{tr}C,$$

$$E(\hat{\varepsilon}^\tau C\hat{\varepsilon}\delta^\tau D\hat{\varepsilon}) = 0.$$

显然有 $E(\delta^\tau D\hat{\varepsilon}\hat{\varepsilon}^\tau D^\tau\delta) = \delta^\tau D\sigma^2 I D^\tau\delta = \sigma^2\delta^\tau DD^\tau\delta$, $E(\sigma^4\mathrm{tr}^2 C) = \sigma^4\mathrm{tr}^2 C$, $E(\hat{\varepsilon}^\tau C\hat{\varepsilon}\sigma^2\mathrm{tr}C) = \sigma^4\mathrm{tr}^2 C$，$E(\delta^\tau D\hat{\varepsilon}\sigma^2\mathrm{tr}C) = 0$. 于是

$$E(\Delta J|_{\hat{\delta}_A} - E(\Delta J|_{\hat{\delta}_A}))^2 = 2\sigma^4\sum_{i=1}^{N-p} C_{ii}^2 + \sigma^2\delta^\tau DD^\tau\delta.$$

证毕.

$\hat{\delta}_A$ 和 $J|_{\hat{\delta}_A}$ 的统计性质均与 P 有关. 由于 $\hat{\delta}_A$ 和 $J|_{\hat{\delta}_A}$ 的统计特性与加权阵 P 的关系比较复杂，得到确定 P 的一般方法比较困难. 但 P 的选取已足以达到引入 $\hat{\delta}_A$ 的目的，即

$$P = (K - 1)A^\tau A. \tag{4.7}$$

将 (4.7) 式代入 (4.5) 式，有

$$\hat{\delta}_A = -(A^\tau A)^{-1}[A^\tau\hat{\varepsilon} - A^\tau A\delta]/K,$$

则可得

$$E(\hat{\delta}_A) = \delta/K, \tag{4.8}$$

$$E(\hat{\delta}_A - E\hat{\delta}_A)(\hat{\delta}_A - E\hat{\delta}_A)^\tau = \sigma^2(A^\tau A)^{-1}/K^2, \tag{4.9}$$

$$\begin{cases} B = (1/K^2 - 2/K + 1)A^\tau A, \\ C = (1/K^2 - 2/K)A(A^\tau A)^{-1}A^\tau, \\ D = (-2/K^2 + 4/K - 2)A^\tau. \end{cases} \tag{4.10}$$

其中

$$K = 1\Big/\prod_{j=1}^{q} |\alpha_j| = 1/|\theta_q| > 1. \tag{4.11}$$

当 $|\alpha_j| \to 0$ 时，$K \to \infty$，可证

$$\lim_{K\to\infty} E(\Delta J|_{\hat\delta_A}) = \delta^\tau A^\tau A\delta, \tag{4.12}$$

$$\lim_{K\to\infty} E(\Delta J|_{\hat\delta_A} - E(\Delta J|_{\hat\delta_A}))^2 \leqslant 4\sigma^2\delta^\tau A^\tau A\delta. \tag{4.13}$$

当 $|\alpha_j| \to 1$ 时，$K \to 1$，可证

$$\lim_{K\to 1} E(\Delta J|_{\hat\delta_A}) = -\sigma^2 q, \tag{4.14}$$

$$\lim_{K\to 1} E(\Delta J|_{\hat\delta_A} - E(\Delta J|_{\hat\delta_A}))^2 \leqslant 2\sigma^4 q. \tag{4.15}$$

可见，$\Delta J|_{\hat\delta_A}$ 的统计特性能自动跟随 $\theta(B)$ 的变化而得到期望的结果，同时，$\hat\delta_A$ 克服了 $\alpha_j \to 0$ 时 $\hat\delta_{LS}$ 协方差阵变大的缺点，达到了引入 $\hat\delta_A$ 的目的。

如果取

$$\begin{cases} P = (K-1)I, \\ K = 1\Big/\prod_{j=1}^{q} |\alpha_j| = 1/|\theta_q|, \end{cases} \tag{4.16}$$

则结果较 $P = (K-1)A^\tau A$ 更好(在此不作详细讨论)。表 1 给了 $\hat\delta_A$，$\Delta J|_{\hat\delta_A}$ 统计特性的数字模拟结果。N，σ^2 以及实现次数同前，P 按 (4.16) 式选取。

3. 应用

Godolphin[5] 采用极大似然法对 $y_t = \varepsilon_t - 0.95\varepsilon_{t-1}$ $(\varepsilon_t \sim N(0,1))$ 的 12 组 $N=49$ 的样本进行参数估计，得到的 $\hat\beta$ 之均值为 0.77。而采用非线性最小二乘法，以 $\hat\delta_A$ 作为残差初值估计，结果为 0.97。

五、计 算 量

从以上讨论知 $\hat\delta_A$ 的计算过程是：1) 由递推式 (3.4) 计算向量 s 和矩阵 A；2) 由 (4.5) 计算 $\hat\delta_A$。该过程乘法次数约为

$$KK = (N-p)(1.5q^2 + 2.5q). \tag{5.1}$$

可见，计算 $\hat\delta_A$ 所需附加计算量与 $(N-p)$ 成正比。为减小计算量，可以选用 $N_1(\leqslant N)$ 来求 $\hat\delta_A$，并保证要求的精度得到保证。下面给出确定 N_1 的简便方法。

由于精度指标一般以相对形式给出，而 $J|_{\hat\delta_A}$ 的精度优于 $J|_0$，故采用下面准则求 N_1 是可行的(见 (3.14) 式)：

$$N_1 = \min N: \{\alpha_{\max}^{2(N-p-q)} \leqslant \varepsilon\}, \tag{5.2}$$

表 2

$\alpha_{\max}$	0.40	0.50	0.60	0.70	0.80	0.90	0.95
N_1-p-q	4	5	7	10	16	33	68

其中 $\varepsilon > 0$. 令 $\varepsilon = 1/1000$，则表 2 给出了 $\alpha_{\max}$ 为各种值时对应的 $N_1 - p - q$. 可见，一般情况下 N_1 较小，求 $\hat{\delta}_A$ 不会产生大的附加计算量.

结　论

1）使用均值法选取残差初值，对应的残差平方和的误差直接依赖于 $\theta(B)$ 的参数，而当 q 增加及 $|\alpha_j| \to 1$ 时，该误差增大.

2）$\hat{\delta}_A$ 通过对 $\theta(B)$ 参数的适应保证了 $\hat{\delta}_A$ 及对应残差平方和在各种 $\theta(B)$ 参数下的良好特性.

3）$\hat{\delta}_A$ 计算步骤简单，易于程序实现.

4）附加计算量小且可随 $\alpha_{\max}$ 自动调整.

5）不增加优化问题的维数，也不影响其收敛性.

由于 $\hat{\delta}$ 具有统计特性优良、计算量小、简单且不影响优化过程等特点，故 $\hat{\delta}_A$ 是一种取代残差均值法的较为理想的实用残差初值算法.

参　考　文　献

[1] Pandit, S. M. and Wu, S. M., Time Series and System Analysis with Applications, John Wiley and Sons (1983), Chapter Four.

[2] Åuström, K. J., Maximum Likelihood and Prediction Error Methods, *Automatica*, **16**(1980), 551—574.

[3] 安鸿志、陈兆国、杜金观、潘一民，时间序列分析与应用，科学出版社，(1983)，217—221.

[4] Priestley, M. B., Spectral Analysis and Time Series, Academic Press (1981), 360.

[5] Godolphin, E. J., On the Maximum Likelihood Estimation of the Parameters of a Gaussian Moving Average Process, Biomatrika, 69(1982), 443—451.

AN ADAPTIVE ESTIMATION OF INITIAL RESIDULS

Zhang Yuming　　Wu Lin

(Harbin Institute of Technology)

ABSTRACT

The statistics of the average method, a widely used one for estimating initial residuls, is analyzed in detail in this paper. By the statistics, the importance of estimating initial residuls is shown. Based on the comparison between the average and the least square method (the two kinds of simplest methods for estimating initial residuls), a simple applicable method with good statistics and little additional compuration burden for estimating initial residuls is proposed. This method can regulate its parameters to keep the statistics good by itself according to the system parameters.

Key words: Initial residuls; residul square sum; adaptive estimation; robust estimation; statistics.