Some Computational Contribution on the Estimation Procedure of a First Order Moving Average

Dai Young Kim*

1. Introduction

Suppose we have a stationary first-order moving average model

$$y_t=\sum_{i=0}^{p}x_{ti}b_i+e_t+de_{t-1},\quad t=1,\cdots\cdots,\ T, \tag{1}$$

where $[e_t]$ is a series of independent normal variate with zero mean and finite variance σ^2, $|d|<1$, b_i's are unknown parameters of coefficients and $[x_{ti}]$ is a series of known variables. We are interested in obtaining the maximum likelihood estimators (MLE) for unknown parameters; $\mathrm{var}(y_t)=\sigma^2(1+d^2)$, correlation $(y_t, y_{t+1})=d/(1+d^2)$ and b_i's. In the simplest case when $p=0$, Durbin (1959) proposed an efficient estimator for d other than MLE because the likelihood function is a complicated form in d. He introduced an approximated likelihood equation,

$$\frac{\partial}{\partial d}[(1-d^2)^{-1}(\sum y_t^2-2d\sum y_ty_{t+1}+2d^2\sum y_ty_{t+2}-\cdots\cdots)]=0,$$

and found it as an unmanageable estimating equation. In this paper we represent the exact likelihood equation in the orthogonally transformed model, and study about the existence of solution for the likelihood equation. The uniqueness is demonstrated in the cases of more than 40 sets of generated data among which few are presented in the Appendix.

2. Notation and Useful Lemma

Matrices are denoted by capital Gothic letters; vectors (column) by lower case Gothic letters; scalars by lower case Roman letters. Dimensions of matrices, when needed, are written $A(\mathrm{m}\times\mathrm{n})$. As usual A^{-1} is the inverse of A. Sometimes in order to consider a matrix in terms of its components a_{ij}, we write $A=(a_{ij})$. For a matrix $A(\mathrm{T}\times\mathrm{T})$, we

* Senior Fellow, Korea Development Institute.

denote the i-th characteristic root of A by $h_i=h_i(A)$, with the ordering $h_1\geq\cdots\geq h_T$. $D(k_i)$ is to indicate the diagonal matrix with the given diagonal components $[k_1, \cdots, k_T]$.

The notation, $L(\boldsymbol{x})\equiv N(\boldsymbol{\mu}, G)$, indicates that the random $T\times 1$ vector $\boldsymbol{x}$ has a normal distribution, i. e., $\boldsymbol{x}$ has the density function,

$$f(\boldsymbol{x})=(2\pi)^{-\frac{T}{2}}|G|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'G^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

where G is positive definite. First we present lemmas used in this paper.

Lemma 1. Let M(T×T) be the matrix which has ones on the diagonals one element removed from the main diagonal, and has zeros elsewhere, then M can be decomposed as $M=QDQ$, where $D(h_j)$ is the T×T diagonal matrix with $h_j(M)=2\cos[j\pi/(T+1)]$ and Q(T×T)$=(q_{jk})=[2/(T+1)^{\frac{1}{2}}\sin[jk\pi/(T+1)]$.

⟨Proof⟩ Shaman (1969) showed that $h_i(M)=2\cos[j\pi/(T+1)]$, $j=1, \cdots\cdots, T$. Let $q^j=(q_{j1}, \cdots\cdots, q_{jT})'$ be the characteristic vector associated with h_j, i. e.,

$$Mq^j=h_jq^j, \quad j=1, \cdots\cdots, T, \tag{2}$$

then we obtain

$$q_{jk-1}+q_{jk+1}=h_jq_{jk}, \quad j=1, \cdots\cdots, T, \quad k=1, \cdots\cdots, T, \tag{3}$$

where $q_{j0}=q_{jT+1}=0$. The associated polynomial equation for (3) has roots

$$r_{j1}=[h_j+(h_j^2-4)^{\frac{1}{2}}]/2 \text{ and } r_{j2}=[h_j-(h_j^2-4)^{\frac{1}{2}}]/2, \tag{4}$$

so that

$$q_{jk}=(r_{j1}{}^k-r_{j2}{}^k)/(r_{j1}-r_{j2}), \quad j=1, \cdots\cdots, T, \quad k=1, \cdots\cdots, T. \tag{5}$$

Substituting $h_j=2\cos[j\pi/(T+1)]$ into (4), we obtain

$$r_{j1}=\exp[ij\pi/(T+1)] \text{ and } r_{j2}=\exp[-ij\pi/(T+1)], \tag{6}$$

where $i=\sqrt{-1}$. Because the characteristic vector is invariant under scalar multiplication, we obtain from (5) and (6) that for any real number z,

$$q_{jk}=z\frac{\exp[ijk\pi/(T+1)]-\exp[-ijk\pi/(T+1)]}{\exp[ij\pi/(T+1)]-\exp[-ij\pi/(T+1)]}=z\frac{\sin[jk\pi/(T+1)]}{\sin[j\pi/(T+1)]}.$$

Observing that

$$\sum_{k=1}^{T}\sin^2[jk\pi/(T+1)]=(T+1)/2,$$

we take $z=[2/(T+1)^{\frac{1}{2}}]\sin[j\pi/(T+1)]$, and obtain

$$q_{jk}=[2/(T+1)]^{\frac{1}{2}}\sin[jk\pi(T+1)], \quad k=1, \cdots\cdots, T, \quad j=1, \cdots\cdots, T,$$

and the proof follows immediately.

⟨Remark⟩ From the result of lemma 1 it is straightforward to present a representation of $\left[v(I+rM)\right]^{-1}_{ij}$ other than the one proposed by Shaman (1969), namely,

$$2/v(T+1)\sum_{k=1}^{T}\left(\sin\frac{ik\pi}{T+1}\sin\frac{jk\pi}{T+1}\right)\Big/\left(1+2r\cos\frac{k\pi}{T+1}\right).$$

Lemma 2. Let $\left[s_i,\ h_i\right]^{T}_{i=1}$ be a sequence of real numbers, then solving

$$\sum_{i=1}^{T} s_i/(1+xh_i)\sum_{i=1}^{T} h_i/(1+xh_i)-T\sum_{i=1}^{T} s_ih_i/(1+xh_i)^2=0, \tag{7}$$

for $x\neq 0$ and for

$$|x|<[\max_{1\le i\le T}(h_i)]^{-1}$$

is equivalent to solve

$$\sum_i s_i/(1+xh_i)\sum_i 1/(1+xh_i)-T\sum_i s_i/(1+xh_i)^2=0. \tag{8}$$

⟨Proof⟩ Observing that, for $x\neq 0$,

$$\sum_i h_i/(1+xh_i)=(1/x)\sum_i[1-1/(1+xh_i)]=T/x-(1/x)\sum_i 1/(1+xh_i),$$

and

$$\sum_i s_ih_i/(1+xh_i)^2=(1/x)\sum_i[s_i(1+xh_i)-s_i]/(1+xh_i)^2$$
$$=(1/x)\sum_i s_i/(1+xh_i)-(1/x)\sum_i s_i/(1+xh_i)^2,$$

it is easy to see that solving (7) for x is equivalent to solve

$$(1/xT)\sum_i s_i/(1+xh_i)[T-\sum_i 1/(1+xh_i)]$$
$$=(1/x)[\sum_i s_i/(1+xh_i)-\sum_i s_i(1+xh_i)^2],$$

or

$$\sum_i s_i/(1+xh_i)-\left(\frac{1}{T}\right)\sum_i s_i/(1+xh_i)\sum_i 1/(1+xh_i)$$
$$=\sum_i s_i/(1+xh_i)-\sum_i s_i/(1+xh_i)^2,$$

which proves the lemma.

3. Numerical Computation of MLE

The corresponding matrix form of our model (1) becomes

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{b}+\boldsymbol{w},\quad L(\boldsymbol{w})\equiv N(\boldsymbol{o},\ \boldsymbol{G}), \tag{9}$$

where $\boldsymbol{G}=v(\boldsymbol{I}+r\boldsymbol{M})$, $\boldsymbol{M}$ is the matrix defined as in lemma 1, $v=\mathrm{var}(y_t)$, and $r=$correlation $(y_t,\ y_{t+1})$. To obtain MLE for $(v,\ r,\ \boldsymbol{b})$ numerically we first perform an orthogonal transformation,

$$\boldsymbol{y}^*=\boldsymbol{Q}\boldsymbol{y},\quad \boldsymbol{X}^*=\boldsymbol{Q}\boldsymbol{X},\ \text{and}\ \boldsymbol{w}^*=\boldsymbol{Q}\boldsymbol{w},$$

where Q is the matrix defined as in lemma 1. Our transformed model is then by lemma 1,

$$y^*=X^* b+w^*,\ \ L(w^*)\equiv N[o,\ D\{v(1+rh_i)\}],$$

where $h_i=2\cos[i\pi/(T+1)]$, $i=1,\cdots\cdots,T$, $v>o$ and $|r|<h_1^{-1}$.

The corresponding joint density function is then

$$f(v,\ r,\ b)\propto\sum_{i=1}^{T}v(1+rh_i)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\ D\left\{\frac{1}{v(1+rh_i)}\right\}S(b)\right],$$

where $S(b)=(y^*-X^*b)(y^*-X^*b)'$. To minimize $-2\log f$ with respect to $(v,\ r,\ b)$, we obtain the likelihood equations;

$$(T/v)-(1/v^2)\sum_i s_i(b)/(1+rh_i)=0, \tag{10}$$

$$\sum_i h_i/(1+rh_i)-(1/v)\sum_i s_i(b)h_i/(1+rh_i)^2=0, \tag{11}$$

$$-2X^{*\prime}D[1/v(1+rh_i)](y^*-X^*b)=0, \tag{12}$$

where $s_i(b)=[S(b)]_{ii}$. Combining (10) and (11), we obtain

$$\sum_i s_i(b)/(1+rh_i)\sum_i h_i/(1+rh_i)-T\sum_i s_i(b)h_i/(1+rh_i)^2=0. \tag{13}$$

From the result of lemma 2, solving (13) for r is equivalent to solve

$$g(r)=\sum_i s_i(b)/(1+rh_i)\sum_i 1/(1+rh_i)-T\sum_i s_i(b)/(1+rh_i)^2=0. \tag{14}$$

For fixed b, let

$$w_i(r)=[1/(1+rh_i)]/\sum_j 1/(1+rh_j),\ \text{and}\ a_i(r)=s_i/(1+rh_i),$$

then solving (14) for r is equivalent to solve

$$\sum_{i=1}^{T}w_i(r)a_i(r)=\frac{1}{T}\sum_{i=1}^{T}a_i(r),\ \ \sum_i w_i(r)=1. \tag{15}$$

As r approaches from the left to $-1/h_1$, the values $w_1(r)$ and $a_1(r)$ become tremendously larger compared to other $w_i(r)$'s and $a_i(r)$'s respectively so that for such r we have $g(r)>0$. Similarly as r approaches from right to $+1/h_1$, $w_T(r)a_T(r)$ dominates. I. e., $g(r)>0$ at the two end points in the open interval $(-1/h_1,\ +1/h_1)$. Observing that $g(o)=0$ and

$$\left.\frac{dg(r)}{dr}\right|_{r=0}=T\sum s_ih_i, \tag{16}$$

which is zero with probability zero, we know that the existence of $r\neq0$ satisfying (14) is trivial. At this moment the uniqueness part is unfortunately not proved. However, for the 40 sets of generated data we observe that $g(r)$ is monotonically decreasing on $(-1/h_1,\ L]$ and increasing on $[L,\ +1/h_1)$, where $L<0$ is the minimum of g, and that the solution is unique.

3-1. Computational Method

Our computer program to solve the system of likelihood equations, (14), (10) and (12), operates according to the following rule: At the k-th iteration, if $\sum s_i(\boldsymbol{b}_k)h_i>0$ then we trace to find the negative root, say r_k, of the funcion g. The value of v_k is then uniquely determined by (10). For such (r_k, v_k), we obtain the unique $\boldsymbol{b}_{k+1}$ from (12). If $\sum s_i(\boldsymbol{b}_k)h_i<0$ then we take similar procedure for the positive r_k. This iterated procedure is terminated at the k-th stage if

$$\max\left[\max_{1\le i\le p}\frac{|b_{ik}-b_{ik-1}|}{|b_{ik-1}|},\ \frac{|r_k-r_{k-1}|}{|r_{k-1}|},\ \frac{|v_k-v_{k-1}|}{|v_{k-1}|}\right]\le\epsilon, \tag{17}$$

where ϵ is preassigned small positive number (we take $\epsilon=10^{-8}$). For all the 40 cases of generated data we find the number of iteration required to satisfy (17) is very small.

3-2. Comparison in the Special Case.

Let $\hat{r}$ is MLE for r. If $|\hat{r}|<1/2$ then the invariant property of MLE implies that

$$\hat{r}=\frac{d}{1+\hat{d}^2}, \tag{18}$$

where $\hat{d}$ is MLE for d. From (18) we obtain

$$\hat{d}=\left[1-(1-4\hat{r}^2)^{\frac{1}{2}}\right]/2\hat{r}. \tag{19}$$

In the case of $\boldsymbol{b}=\boldsymbol{0}$, we compare $\hat{d}$, the MLE of d obtained from (19) and the efficient estimate Durbin (1959) proposed, namely,

$$\hat{d}=-\sum_{i=0}^{k-1}\hat{l}_i\hat{l}_{i+1}/\sum_{i=0}^{k}\hat{l}^2{}_1,\ 2k<T,$$

where $[\hat{l}_i]$ is the sequence of least square estimates of coefficient when our model is approximated with finite autoregressive representation, i.e., $[\hat{l}_i]$ minimizes

$$\sum_{t=k+1}^{T}(y_t-l_1y_{t-1}-\cdots\cdots-l_ky_{t-k})^2.$$

The numerical comparison is presented in the Appendix for a set of generated data.

APPENDIX

(1) The random created data for $T=43$ according to the model,

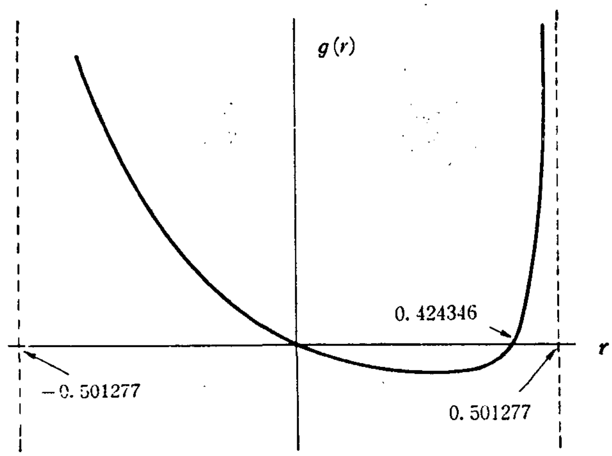
$$y_t=e_t+de_{t-1},$$

where e_t were drawn from $N(0,\ 36)$ and $d=0.5$ are:

0.99	2.20	−4.56	−1.78	−0.35	4.12	−6.68	1.19	3.50	5.94
−2.66	2.44	3.89	−7.20	0.05	5.73	0.41	−4.29	−18.05	−9.87
0.75	−6.24	−7.34	4.75	0.84	−2.23	0.25	5.23	11.44	19.47
11.67	1.30	5.83	4.74	10.53	−0.42	−10.67	−3.42	−8.40	−12.34
−5.91	−4.84	6.54							

The estimates are:

$\hat{r}=0.424346$, $\hat{v}=34.42455$, $\hat{d}=0.555103$ (MLE), $\hat{d}=0.417507$ (Durbin).



(2) The following data is the demand for money series.

31.0	38.7	36.7	39.5	39.4	44.9	41.6	43.0	45.1	48.1
48.6	50.7	55.7	61.2	64.7	68.3	71.4	78.6	84.2	91.6
101.5	109.2	121.0	131.5	136.9	151.2	174.1	177.6	192.1	226.1
251.4	236.5	250.9	275.0	306.5	296.9	300.1	342.6	360.9	357.3
380.5	463.8								

The estimates (when $\boldsymbol{b}=\boldsymbol{0}$) are:

$$\hat{r}=0.500936, \text{ and } \hat{v}=28484.611$$

(3) The created data for $T=11$ according to the model,

$$y_t=b_0+b_1X_t+e_t+de_{t-1},$$

where e_t were drawn from $N(0,\ 0.1)$, $b_0=0.3$, $b_1=0.4$, $d=-0.5$ and given $[X_t]$ are as follow:

X_t	1.17	1.24	1.37	1.47	1.54	1.63	1.76	1.92	2.14	2.31	2.48
Y_t	0.83	0.77	0.89	1.00	1.08	1.04	1.14	1.19	1.34	1.34	1.37

The estimates are:

stage	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{v}$
0	0.0	0.0	0.509127	0.813577
1	0.265984	0.491840	0.323700	0.002480
2	0.302694	0.454691	0.255736	0.001948
3	0.299193	0.456637	0.254116	0.001946
4 (final)	0.299110	0.456685	0.254109	0.001945

REFERENCES

1. Durbin, J. [1959], "Efficient estimation of parameters in moving average models," *Biometrika* 46, 306.
2. Peckham, G. [1970], "A new method for minimizing a sum of squares without calculating gradients," *The Computer Journal*, Vol. 13, No. 4, 418.
3. Shaman, P. [1969], "On the inverse of the covariance matrix of a first order moving average," *Biometrika* 56, 509.

SUMMARY

In the first-order moving average model, we present the exact likelihood equations as function of variance, correlation and parameters of coefficients in the orthogonally transformed model. Existence of maximum likelihood estimates for these unknowns are studied and a computational method is provided. (Because of the limited space we do not present the computer program which is written in FORTRAN.) 40 sets of generated data and economic data are used to demonstrate, and few of them are presented in the Appendix. A numerical comparison of MLE with the efficient estimate proposed by Durbin is presented in the particular case.