DOCUMENTS DE TREBALL DE LA DIVISIÓ DE CIÈNCIES JURÍDIQUES, ECONÒMIQUES I SOCIALS

COL·LECCIÓ D'ECONOMIA

Linear least squares estimation of the first order moving average parameter

Emili Valdero

Departament de Matemàtica Econòmica, Financera i Actuarial
Facultat d'Econòmiques - Universitat de Barcelona
Avda. Diagonal 690, 08034 - Barcelona
Tel: 934021953, Fax: 934021953
e-mail: evaldero@eco.ub.es

Juny 2002

ABSTRACT. We propose an iterative procedure to minimize the sum of squares function which avoids the nonlinear nature of estimating the first order moving average parameter and provides a closed form of the estimator. The asymptotic properties of the method are discussed and the consistency of the linear least squares estimator is proved for the invertible case. We perform various Monte Carlo experiments in order to compare the sample properties of the linear least squares estimator with its nonlinear counterpart for the conditional and unconditional cases. Some examples are also discussed.

Keywords: *Moving average processes, Invertible models, Consistent estimator, Nonlinear optimization, Monte Carlo simulation in time series.*

JEL Classification Code: C22.

RESUM. En aquest document de treball es proposa un procediment iteratiu per minimitzar la suma de quadrats dels errors que evita la naturalesa no lineal de l'estimació del paràmetre del model mitjana mòbil de primer ordre i proporciona una expressió de l'estimador en forma tancada. A continuació es discuteixen les propietats asimptòtiques del mètode i es demostra la consistència de l'estimador per mínims quadrats lineals per a valors del paràmetre dins l'interval obert $(-1, 1)$. També es duen a terme diversos experiments de Monte Carlo per tal de comparar les propietats mostrals de l'estimador per mínims quadrats lineals amb el seu homòleg no lineal pel cas condicional i pel no condicional. Finalment, es discuteixen alguns exemples.

Paraules clau: *Processos mitjana mòbil, Models invertibles, Estimador consistent, Optimització no lineal, Simulació de Monte Carlo en sèries temporals.*

Codi de Classificació JEL: C22.

1. Introduction

The main discussion of this article will center around the first order moving average model, $\mathrm{MA}(1)$ for short, defined as

$$x_t = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad t = 1, \ldots, n \tag{1}$$

where $|\theta| \leq 1$ and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and the same variance σ_ε^2.

The model (1) is said to be invertible if $|\theta| < 1$. When $|\theta| = 1$, the model (1) is noninvertible. The autocorrelation function of the process (1) is given by

$$\rho_k = \begin{cases} \dfrac{\theta}{1+\theta^2}, & \text{if } k = 1 \\ 0, & \text{if } k > 1, \end{cases} \tag{2}$$

and $|\rho_1| \leq 1/2$.

Replacement of θ by $1/\theta$ yields a process with identical autocorrelation function. Therefore, for every noninvertible model with $|\theta| > 1$, there is an equivalent invertible model. This lack of global identification is conventionally removed by imposing the condition $|\theta| \leq 1$ to the process (1).

A brief summary of the nonlinear least squares estimation of $\mathrm{MA}(1)$ processes is given in Section 2. In Section 3 we derive a linear method to minimize the sum of squares function. In Section 4 we prove for invertible models the asymptotic equivalence of the linear procedure to the nonlinear least squares and in Section 5 we demonstrate its consistency for the invertible case $|\theta| < 1$. We perform various Monte Carlo experiments in order to compare the sample properties of the linear and nonlinear least squares estimators for the conditional and unconditional cases in Section 6 and in Section 7 we discuss some examples by applying the least squares and the maximum likelihood methods to two sets of real data. Finally, in the last section we summarize the main conclusions that are drawn from the empirical evidence.

2. Nonlinear least squares estimation

The least squares estimator of the first order moving average parameter θ is a well-known nonlinear estimation problem. Box and Jenkins (1976, Ch. 7) concentrate on two least squares procedures to estimate moving average models given a sample of n observations $x_1, x_2, \ldots, x_n$ of the process (1) that differ in their treatment of the pre-sample residual ε_0.

The conditional least squares estimator of the moving average parameter θ is determined by setting the unknown pre-sample error value, ε_0, to zero and minimizing the conditional sum of squares function of the model (1) given by

$$\mathbf{S}^*(\theta) = \sum_{t=1}^{n} [\varepsilon_t]^2, \tag{3}$$

where n is the sample size and $[\varepsilon_t]$ denotes the expectation of ε_t conditional on θ and the sample values $x_1, x_2, \ldots, x_n$.

The unconditional least squares estimator of the coefficient parameter θ is found using the 'backforecasting' technique developed by Box and Jenkins (1976, App. A7.4) to compute the pre-sample expectation error $[\varepsilon_0]$, and then minimizing the unconditional sum of squares given by

$$\mathbf{S}(\theta) = \sum_{t=0}^{n} [\varepsilon_t]^2. \tag{4}$$

Kang (1975) examined the features of the unconditional likelihood and sum of squares functions. She showed that the unconditional sum of squares function has the undesirable property that is decreasing as the boundary of the invertibility region is crossed, so that the minimum may be at $\theta = \pm 1$. This implies that the unconditional maximum likelihood and least squares estimators of the parameter θ have a positive probability of being equal to ± 1 when the true value is $|\theta| < 1$. This phenomenon is known in the econometric literature as 'pile-up effect' and has been extensively analyzed by Ansley and Newbold (1980), Cryer and Ledolter

(1981) , Davidson (1981a, 1981b), Sargan and Bhargava (1983), Anderson and Takemura (1986) and and Anderson and Mentz (1993).

For practical purposes, minimization either of $S^{*}(\theta)$ or $S(\theta)$ may be carried out by the Gauss-Newton algorithm or other suitable procedures using an initial estimator of θ as a starting point of the nonlinear iterative optimization. The estimator of θ which is likely to be employed at a preliminary stage of nonlinear iterative methods is the moment estimate based on the relationship between θ and ρ_1 given by (2). Another initial estimator is the one obtained using the innovations algorithm proposed by Brockwell and Davis (1991, Sec. 8.3) , which is more efficient than the moment estimator, [see also Brockwell and Davis (1988)]. Finally, the estimator based on the autoregressive representation of the process proposed by Galbraith and Zinde-Walsh (1994) can also be used as a preliminary estimator. This last estimator has the advantages respect to the Brockwell-Davis one that is noniterative and extremely simple to compute and that has the same asymptotic variance equal to $1/n$.

Fuller (1976, Sec. 8.3) considered an estimation procedure based on the Gauss-Newton method. Note that we can write the model (1) as

$$\varepsilon_t\left(x;\theta,\varepsilon_0\right) = x_t - \theta\varepsilon_{t-1}\left(x;\theta,\varepsilon_0\right), \quad t = 1, 2, \ldots, n, \tag{5}$$

where the notation $\varepsilon_t\left(x;\theta,\varepsilon_0\right)$ is used to emphasize the fact that the ε_t depend on the observations, x_t, on the pre-sample error value, ε_0, and on the parameter θ. The random variable ε_0 is unknown and is independent of $\{\varepsilon_{t,}\ t \geq 1\}$. Only in obtaining the estimator of θ we treat ε_0 as a fixed unknown parameter.

Let $W_t\left(x;\hat{\theta}^i,\hat{\varepsilon}_0^i\right)$ and $U_t\left(x;\hat{\theta}^i,\hat{\varepsilon}_0^i\right)$ denote the negatives of the partial derivatives of $\varepsilon_t\left(x;\theta,\ \varepsilon_0\right)$ with respect to θ and ε_0 evaluated at $\theta = \hat{\theta}^i$ and $\varepsilon_0 = \hat{\varepsilon}_0^i$, where $\hat{\theta}^i$ and $\hat{\varepsilon}_0^i$ are the estimators of θ and ε_0 obtained in the ith iteration of the method, being $\hat{\theta}^0$ and $\hat{\varepsilon}_0^0$ some initial estimates of θ and ε_0. The ith step Gauss-Newton correction $\Delta\hat{\theta}^i$ for θ is the coefficient of $W_t\left(x;\hat{\theta}^{i-1},\ \hat{\varepsilon}_0^{i-1}\right)$ in the regression of $\varepsilon_t\left(x;\hat{\theta}^{i-1},\hat{\varepsilon}_0^{i-1}\right)$ on $W_t\left(x;\hat{\theta}^{i-1},\hat{\varepsilon}_0^{i-1}\right)$ and $U_t\left(x;\hat{\theta}^{i-1},\hat{\varepsilon}_0^{i-1}\right)$, where

$$W_t\left(x;\hat{\theta}^i,\hat{\varepsilon}_0^i\right)=\begin{cases}\hat{\varepsilon}_0^i, & t=1,\\ \sum_{j=1}^{t-1} j\left(-\hat{\theta}^i\right)^{j-1}x_{t-j}+t\left(-\hat{\theta}^i\right)^{t-1}\hat{\varepsilon}_0^i, & t=2,\ldots,n,\end{cases} \tag{6}$$

and

$$U_t\left(x;\hat{\theta}^i,\hat{\varepsilon}_0^i\right)=-\left(-\hat{\theta}^i\right)^t, \qquad t=1,\ldots,n. \tag{7}$$

The improved estimator of θ in the ith iteration is then

$$\hat{\theta}^i=\hat{\theta}^{i-1}+\Delta\hat{\theta}^i, \tag{8}$$

and the iterations are halted when the desired convergence is achieved.

For the conditional case we have that $\varepsilon_0=0$ and thereby, the computation of the ith step Gauss-Newton correction is found simply regressing $\varepsilon_t\left(x;\hat{\theta}^{i-1},0\right)$ on $W_t\left(x;\hat{\theta}^{i-1},0\right)$. Hence, the estimator of $\Delta\hat{\theta}^i$ is given by

$$\Delta\hat{\theta}^i=\frac{\sum_{t=1}^n \varepsilon_t\left(x;\hat{\theta}^{i-1},0\right)W_t\left(x;\hat{\theta}^{i-1},0\right)}{\sum_{t=1}^n\left[W_t\left(x;\hat{\theta}^{i-1},0\right)\right]^2}. \tag{9}$$

MacPherson and Fuller (1983) proved the consistency of the least squares estimator (8) for the parameter θ in the closed interval $[-1,1]$.

3. Linear least squares estimation

In this section we derive an alternative procedure to minimize the sum of squares function which avoids the nonlinear nature of estimating the parameter θ and allows us to express the estimator in explicit form. The method is based on a linear approximation to the minimization of the residual sum of squares. The estimator can be obtained iteratively by computing in each stage the expectations of the errors and its first derivatives.

This linear approximation to the least squares estimators, that we shall call linear least squares for short, is the only estimation procedure of the parameter θ that provides a closed form of the estimator.

We derive the method only for the unconditional case, being the same for the conditional case putting $[\varepsilon_0] = 0$.

Consider the derivative of the unconditional sum of squares (4) given by

$$\frac{\partial \mathbf{S}(\theta)}{\partial \theta} = 2 \sum_{t=0}^{n} \left([\varepsilon_t] \, \frac{\partial [\varepsilon_t]}{\partial \theta} \right),$$

and then, we can write that

$$2 \sum_{t=0}^{n-1} \left([\varepsilon_t] \, \frac{\partial [\varepsilon_t]}{\partial \theta} \right) = \frac{\partial \mathbf{S}(\theta)}{\partial \theta} - 2 [\varepsilon_n] \, \frac{\partial [\varepsilon_n]}{\partial \theta} \,. \tag{10}$$

>From the MA(1) process given by (1), we can obtain the expectations of the errors by recursive calculation through the formula

$$[\varepsilon_t] = x_t - \theta [\varepsilon_{t-1}] \,, \qquad t = 0, 1, \ldots, n, \tag{11}$$

where $[\varepsilon_{-1}] = 0$ and $[\varepsilon_0] = x_0$ is generated by backforecasting of the series. Hence, we can rewrite the unconditional sum of squares (4) as

$$S(\theta) = x_0^2 + (x_1 - \theta [\varepsilon_0])^2 + (x_2 - \theta [\varepsilon_1])^2 + \cdots + (x_n - \theta [\varepsilon_{n-1}])^2 \,. \tag{12}$$

Differentiating (12) with respect to θ we obtain

$$\begin{aligned} \frac{\partial \mathbf{S}(\theta)}{\partial \theta} &= -2 (x_1 - \theta [\varepsilon_0]) \left([\varepsilon_0] + \theta \frac{\partial [\varepsilon_0]}{\partial \theta} \right) - 2 (x_2 - \theta [\varepsilon_1]) \left([\varepsilon_1] + \theta \frac{\partial [\varepsilon_1]}{\partial \theta} \right) \\ &\quad - \cdots - 2 (x_n - \theta [\varepsilon_{n-1}]) \left([\varepsilon_{n-1}] + \theta \frac{\partial [\varepsilon_{n-1}]}{\partial \theta} \right) \\ &= -2 \sum_{t=0}^{n-1} x_{t+1} [\varepsilon_t] - 2\theta \sum_{t=0}^{n-1} \left(x_{t+1} \frac{\partial [\varepsilon_t]}{\partial \theta} \right) + 2\theta \sum_{t=0}^{n-1} [\varepsilon_t]^2 \\ &\quad + 2\theta^2 \sum_{t=0}^{n-1} \left([\varepsilon_t] \, \frac{\partial [\varepsilon_t]}{\partial \theta} \right). \end{aligned} \tag{13}$$

If we substitute the expression (10) in the last term of (13) we can write that

$$\begin{aligned}\frac{\partial \mathbf{S}(\theta)}{\partial\theta} &= -2\sum_{t=0}^{n-1} x_{t+1}\left[\varepsilon_t\right] - 2\theta\sum_{t=0}^{n-1}\left(x_{t+1}\frac{\partial\left[\varepsilon_t\right]}{\partial\theta}\right) + 2\theta\sum_{t=0}^{n-1}\left[\varepsilon_t\right]^2 \\ &\quad +\theta^2\left(\frac{\partial \mathbf{S}(\theta)}{\partial\theta} - 2\left[\varepsilon_n\right]\frac{\partial\left[\varepsilon_n\right]}{\partial\theta}\right), \end{aligned} \tag{14}$$

and operating we have that

$$\begin{aligned}\frac{\partial \mathbf{S}(\theta)}{\partial\theta} &= \frac{2}{1-\theta^2}\left[-\sum_{t=0}^{n-1} x_{t+1}\left[\varepsilon_t\right] - \theta\sum_{t=0}^{n-1}\left(x_{t+1}\frac{\partial\left[\varepsilon_t\right]}{\partial\theta}\right)\right. \\ &\quad \left. +\theta\sum_{t=0}^{n-1}\left[\varepsilon_t\right]^2 - \theta^2\left[\varepsilon_n\right]\frac{\partial\left[\varepsilon_n\right]}{\partial\theta}\right]. \end{aligned} \tag{15}$$

The unconditional least squares estimator of the parameter θ is the one which minimizes the unconditional sum of squares function $S(\theta)$. Then, equating the above expression to zero we arrive to the following second degree equation of the parameter θ,

$$\left[\varepsilon_n\right]\frac{\partial\left[\varepsilon_n\right]}{\partial\theta}\theta^2 + \left[\sum_{t=0}^{n-1}\left(x_{t+1}\frac{\partial\left[\varepsilon_t\right]}{\partial\theta}\right) - \sum_{t=0}^{n-1}\left[\varepsilon_t\right]^2\right]\theta + \sum_{t=0}^{n-1} x_{t+1}\left[\varepsilon_t\right] = 0, \tag{16}$$

and finally, we obtain that

$$\theta = \frac{\sum_{t=0}^{n-1} x_{t+1}\left[\varepsilon_t\right]}{\sum_{t=0}^{n-1}\left[\varepsilon_t\right]^2 - \sum_{t=0}^{n-1}\left(x_{t+1}\frac{\partial\left[\varepsilon_t\right]}{\partial\theta}\right)} + \mathbf{Q}\,\theta^2, \tag{17}$$

where the term Q associated to θ^2 in (17) is given by

$$\mathbf{Q} = \frac{\left[\varepsilon_n\right]\dfrac{\partial\left[\varepsilon_n\right]}{\partial\theta}}{\sum_{t=0}^{n-1}\left[\varepsilon_t\right]^2 - \sum_{t=0}^{n-1}\left(x_{t+1}\frac{\partial\left[\varepsilon_t\right]}{\partial\theta}\right)}. \tag{18}$$

The linear approximation to the least squares estimator consist in neglecting the term Q in (17). By this way, the minimization of the sum squares function yields as a result an explicit expression of the parameter estimator. Thus, the unconditional linear least squares estimator is given by

$$\hat{\theta} = \frac{\sum_{t=0}^{n-1} x_{t+1}\left[\varepsilon_t\right]}{\sum_{t=0}^{n-1}\left[\varepsilon_t\right]^2 - \sum_{t=0}^{n-1}\left(x_{t+1}\frac{\partial\left[\varepsilon_t\right]}{\partial\theta}\right)}. \tag{19}$$

In the following section we shall prove a lemma which establish that $Q \to 0$ as $n \to \infty$ and therefore, assures the asymptotic equivalence of the linear estimator (19) to the one obtained by the nonlinear minimization of the sum of squares function (4).

To obtain the linear least squares estimator of θ we have to apply the formula (19) in successive iterations, recalculating in each iteration the values of the expectations of the errors $[\varepsilon_t]$ and its first derivatives $\partial\left[\varepsilon_t\right]/\partial\theta$ for $t = 0, 1, \ldots, n-1$, and using in the first iteration as initial values those obtained with a preliminary estimator of θ.

4. Asymptotic equivalence to the nonlinear least squares estimator

We now prove the following lemma.

Lemma Let the MA(1) model given by (1) with $|\theta| < 1$, then, the term Q associated to θ^2 in (17) given by the expression (18) is at most of order in probability n^{-1}, that is,

$$\mathbf{Q} = O_p\left(n^{-1}\right).$$

Proof. If we make the change of notation

$$\left.\begin{aligned} R_1 &= \sum_{t=0}^{n-1}\left(x_{t+1}\frac{\partial\left[\varepsilon_t\right]}{\partial\theta}\right)\\ R_2 &= \sum_{t=0}^{n-1}\left[\varepsilon_t\right]^2\\ Q_n &= \left[\varepsilon_n\right]\frac{\partial\left[\varepsilon_n\right]}{\partial\theta}\end{aligned}\right\}, \tag{20}$$

we can rewrite the term Q given by (18) as

$$\mathbf{Q} = \frac{Q_n}{R_2 - R_1}. \tag{21}$$

First of all we shall prove that the numerator Q_n in (21) is $O_p(1)$. If we compute the derivatives of the expectations of the errors given by (11) we obtain the recurrence expressions

$$\frac{\partial [\varepsilon_t]}{\partial \theta} = \begin{cases} 0, & t = 0, \\ -\theta \dfrac{\partial [\varepsilon_{t-1}]}{\partial \theta} - [\varepsilon_{t-1}], & t = 1, \ldots, n. \end{cases} \tag{22}$$

By successive substitution of $\partial [\varepsilon_1] / \partial \theta$ into $\partial [\varepsilon_2] / \partial \theta$ and so on until substituting $\partial [\varepsilon_{n-1}] / \partial \theta$ into $\partial [\varepsilon_n] / \partial \theta$ we finally find that

$$\begin{aligned} \frac{\partial [\varepsilon_n]}{\partial \theta} &= -\left[[\varepsilon_{n-1}] - \theta [\varepsilon_{n-2}] + \theta^2 [\varepsilon_{n-3}] + \cdots + (-\theta)^{n-2} [\varepsilon_1] + (-\theta)^{n-1} [\varepsilon_0] \right] \\ &= -\sum_{j=0}^{n-1} (-\theta)^{n-j-1} [\varepsilon_j]. \end{aligned}$$

Clearly, $\partial [\varepsilon_n] / \partial \theta$ is a linear combination of independent and identically distributed expectation errors in different time periods with weights decreasing geometrically with alternate sign. Then, the numerator Q_n in (21) is given by

$$Q_n = [\varepsilon_n] \frac{\partial [\varepsilon_n]}{\partial \theta} \tag{23}$$

$$= -[\varepsilon_n] \sum_{j=0}^{n-1} (-\theta)^{n-j-1} [\varepsilon_j], \tag{24}$$

so, for every $\alpha > 0$ exist a positive real number M_α such that

$$\Pr [|Q_n| \geq M_\alpha] \leq \alpha, \tag{25}$$

for all n and then Q_n is $O_p(1)$.

Next, we shall prove that the denominator in (21) is $O_p(n)$. The denominator of the term Q is given by the difference $R_2 - R_1$. We can rewrite the component R_1 given by (20) as

$$R_1 = \sum_{t=0}^{n-1} \left[(\theta [\varepsilon_t] + [\varepsilon_{t+1}]) \frac{\partial [\varepsilon_t]}{\partial \theta} \right]$$

$$= \theta \sum_{t=0}^{n-1} \left([\varepsilon_t] \frac{\partial [\varepsilon_t]}{\partial \theta} \right) + \sum_{t=0}^{n-1} \left([\varepsilon_{t+1}] \frac{\partial [\varepsilon_t]}{\partial \theta} \right). \tag{26}$$

Making the following change of notation

$$Q_t = [\varepsilon_t] \frac{\partial [\varepsilon_t]}{\partial \theta} \tag{27}$$

$$Q_t^* = [\varepsilon_{t+1}] \frac{\partial [\varepsilon_t]}{\partial \theta}, \tag{28}$$

we can express the R_1 term given by (26) as

$$R_1 = \theta \sum_{t=0}^{n-1} Q_t + \sum_{t=0}^{n-1} Q_t^*. \tag{29}$$

We can note that the structure of the expressions (27) and (28) is very similar to that of the numerator Q_n given by (23). Therefore, using (24) we can rewrite the first summation on the right hand side of (29) as

$$\sum_{t=0}^{n-1} Q_t = -\sum_{t=0}^{n-1} \left([\varepsilon_t] \sum_{j=0}^{t-1} (-\theta)^{t-j-1} [\varepsilon_j] \right). \tag{30}$$

Bearing in mind that $[\varepsilon_t] = 0$ for $t < 0$, we can develope the right hand side of (30) for every time period $t = 0, 1, \ldots, n-1$, in the following way,

$$\left.\begin{array}{rl} Q_0 = & 0 \\ Q_1 = & -[\varepsilon_0][\varepsilon_1] \\ Q_2 = & -[\varepsilon_1][\varepsilon_2] + \theta [\varepsilon_0][\varepsilon_2] \\ Q_3 = & -[\varepsilon_2][\varepsilon_3] + \theta [\varepsilon_1][\varepsilon_3] - \theta^2 [\varepsilon_0][\varepsilon_3] \\ \vdots & \qquad \vdots \\ Q_{n-1} = & -[\varepsilon_{n-2}][\varepsilon_{n-1}] + \theta [\varepsilon_{n-3}][\varepsilon_{n-1}] + \cdots + (-1)^{n-1} \theta^{n-2} [\varepsilon_0][\varepsilon_{n-1}] \end{array}\right\}. \tag{31}$$

Hence, (30) becomes

$$\begin{aligned} \sum_{t=0}^{n-1} Q_t &= -\sum_{t=0}^{n-2} [\varepsilon_t][\varepsilon_{t+1}] + \theta \sum_{t=0}^{n-3} [\varepsilon_t][\varepsilon_{t+2}] - \theta^2 \sum_{t=0}^{n-4} [\varepsilon_t][\varepsilon_{t+3}] \\ &\quad + \cdots + (-1)^{n-1} \theta^{n-2} [\varepsilon_0][\varepsilon_{n-1}], \end{aligned} \tag{32}$$

and multiplying both sides of (32) by $1/n$ and taking limits as $n \rightarrow \infty$, each summation term on the right hand side of (32) converges in probability to its respective autocovariance $\gamma_{\varepsilon}(k) = cov(\varepsilon_t, \varepsilon_{t+k})$, and then, we have

$$\lim_{n\rightarrow\infty} \frac{1}{n} \sum_{t=0}^{n-1} Q_t = \sum_{k=1}^{n-1} (-1)^k \theta^{k-1} \gamma_{\varepsilon}(k). \tag{33}$$

Since $\gamma_{\varepsilon}(k) = 0$, for $k \geq 1$, we have from (33) that $n^{-1} \sum_{t=0}^{n-1} Q_t \rightarrow 0$ as $n \rightarrow \infty$, and consequently $\sum_{t=0}^{n-1} Q_t$ is $O_p(n)$.

The argument for the second term in (29) given by $\sum_{t=0}^{n-1} Q_t^*$ is completely analogous to that for $\sum_{t=0}^{n-1} Q_t$ and is therefore omitted. So, we have that $\sum_{t=0}^{n-1} Q_t^*$ is also $O_p(n)$ and in consequence, the component R_1 of the denominator in (21) is $O_p(n)$.

The term R_2 of the denominator in (21) is given by

$$R_2 = \sum_{t=0}^{n-1} [\varepsilon_t]^2. \tag{34}$$

Multiplying both sides of (34) by $1/n$ and taking limits as $n \rightarrow \infty$, we can see that the right hand side of (34) converges in probability to the variance of the process $\{\varepsilon_t\}$, and hence, the component R_2 of the denominator in (21) is $O_p(n)$.

Therefore, we have proved that $Q_n = O_p(1)$, $R_1 = O_p(n)$ and $R_2 = O_p(n)$, and using (21) we have that the term Q associated to θ^2 in (17) given by the expression (18) is $O_p\left(n^{-1}\right)$. $\square$

Using the result in the previous lemma, we have that $Q \rightarrow 0$ as $n \rightarrow \infty$ for $|\theta| < 1$. This guarantees the asymptotic equivalence of the linear least squares estimator given by (19) to that obtained by minimizing the sum of squares function using the Gauss-Newton algorithm or other nonlinear techniques. When the MA(1) process is noninvertible however, the linear estimator is not asymptotically equivalent to the nonlinear one because the numerator Q_n in (21) given by (24) is not bounded in probability.

5. Consistency of the linear least squares estimator

We are now in position to prove the consistency of the linear least squares estimator given by (19).

Theorem Let x_t satisfy the MA(1) model

$$x_t = \varepsilon_t + \theta \varepsilon_{t-1}, \qquad t = 1, \ldots, n$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and the same variance σ_ε^2. Then, the linear least squares estimator $\hat{\theta}$ given by (19) is a consistent estimator of θ for $|\theta| < 1$.

Proof. The denominator in (19) is the difference between the terms R_2 and R_1 given by expression (20). Using the results of the previous lemma, we have that $n^{-1}R_2 \to \sigma_\varepsilon^2$ and $n^{-1}R_1 \to 0$ as $n \to \infty$.

On the other hand, we have that the numerator in (19) can be rewritten as

$$\begin{aligned} \sum_{t=0}^{n-1} x_{t+1} [\varepsilon_t] &= \sum_{t=0}^{n-1} ([\varepsilon_{t+1}] + \theta [\varepsilon_t]) [\varepsilon_t] \\ &= \sum_{t=0}^{n-1} [\varepsilon_{t+1}] [\varepsilon_t] + \theta \sum_{t=0}^{n-1} [\varepsilon_t]^2, \end{aligned} \tag{35}$$

and multiplying both sides of (35) by $1/n$ and taking limits as $n \to \infty$, the two summations on the right hand side converge in probability to the autocovariance at lag one and the variance of the process $\{\varepsilon_t\}$ respectively. Hence,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} x_{t+1} [\varepsilon_t] = \theta \sigma_\varepsilon^2. \tag{36}$$

Therefore, taking the limit as $n \to \infty$ of the linear least squares estimator $\hat{\theta}$ given by (19) we come to the conclusion that $\hat{\theta}$ is a consistent estimator of θ for $|\theta| < 1$. $\square$

6. Simulation results

This section reports the Monte Carlo simulation results obtained for the MA(1) model given by (1) for the invertible case $|\theta| < 1$. Since the parameter space is a closed interval for this model, we can attempt to determine the complete properties of the estimators by simulation, and 19 equally spaced experiments were conducted for θ varying from -0.9 to 0.9 by steps of 0.1. With respect to sample size we have chosen $n = 30$ and $n = 100$ as in Nelson (1974).

To generate the time series, the pseudo-random number generator from the G05DDF NAG routine has been employed producing standard normally distributed variables $\{\varepsilon_t\}$. When a run was generated, the first 100 numbers were discarded to avoid possible start-up problems. For each choice of parameter and sample size, $5,000$ replications were done. All computations were performed on a Sun/Ultra2 computer at the University of Barcelona. Double precision arithmetic within FORTRAN 77 programs was used throughout the study.

The linear least squares (LLS) and the Fuller-Gauss-Newton (FGN) estimators were computed using formulas (19) and (8) respectively. The ith step Gauss-Newton correction $\Delta\hat{\theta}^{i}$ was calculated using formula (9) for the conditional case while the regression of $\varepsilon_t\left(x;\hat{\theta}^{i-1},\hat{\varepsilon}_0^{i-1}\right)$ on $W_t\left(x;\ \hat{\theta}^{i-1},\ \ \hat{\varepsilon}_0^{i-1}\right)$ and $U_t\left(x;\hat{\theta}^{i-1},\hat{\varepsilon}_0^{i-1}\right)$ used to determine $\Delta\hat{\theta}^{i}$ for the unconditional case was performed using the G02DAF NAG routine.

For both methods, the Galbraith/Zinde-Walsh preliminary estimator computed using an autoregressive approximation of order 15 was used as a starting value of the iteration procedure.

For the FGN method the estimates were constrained to lie in the unit interval only after the last iteration, so if at the end of the procedure there was no local minimum inside the region $[-1, 1]$ we defined the estimate to be on the boundary.For the LLS method we imposed that the estimator had to be less than one in absolute value at the end of each iteration since the consistency of this procedure requires that the invertibility of the process be met at each stage as it is shown in the Theorem of Section 5. To overcome this problem we checked up the estimator value at the end of each iteration and reset any estimation which has moved outside the invertibility region to a value inside it. Thus, at any iteration when an estimator of θ is obtained outside the invertible boundary we simply set $\hat{\theta} = \pm 0.9999$.

In the implementation the number of iterations for both methods was limited to a maximum of $1,000$ and the procedure was regarded as having converged if the absolute value of the difference between two consecutive values fell below 10^{-4}. When the desired convergence was not achieved after $1,000$ iterations the procedure was stopped and the estimation declared nonfeasible. This occurred in a few cases in experiments with values of $|\theta| \geq 0.8$.

Not only have we computed for all the estimators the usual simulation statistics as bias, standard error and mean squared error but also the percentage times the parameter estimator falls in the interval $0.99 \leq |\hat{\theta}| \leq 1$ as an indicator of the magnitude of the 'pileup effect' described in Section 2.

In Table I, the simple average over values of θ of the simulation mean squared errors provides a uniformly weighted expected risk measure of overall estimator performance. Entries in Table I are defined as $(1/19)\sum_{i=1}^{19}\mathrm{MSE}(\theta_i)$, where $\mathrm{MSE}(\theta_i)$ is the simulation mean squared error obtained in an experiment with $\theta = \theta_i$ and $\theta_i = 0.1(i-1) - 0.9$, $i = 1, \ldots, 19$. It will be observed that the conditional LLS estimator is, by a very little margin, the best estimator on this criterion in the smaller sample size, and the unconditional estimators performs relatively poorly which supports the findings of Nelson (1974), Dent and Min (1978) and Ansley and Newbold (1980). For $n = 100$, the linear estimators behave very similarly than the nonlinear ones in terms of the mean squared error, which agrees with the asymptotic equivalence between them proved in Section 4.

TABLE I
Average of the simulation MSE

	Conditional		Unconditional	
Sample size	LLS	FGN	LLS	FGN
$n = 30$	0.03598	0.03612	0.04564	0.04503
$n = 100$	0.00824	0.00820	0.00894	0.00878

Table II gives the results for the conditional case and $n = 30$. Except for values near the invertibility boundary, the conditional LLS estimators seem to be better than the nonlinear ones due to their smallest variance. This is specially significant for θ in the interval $[-0.6, 0.6]$ where the conditional LLS estimator has less bias and standard error than the FGN estimator.

TABLE II
Simulation statistics for the conditional case and $n = 30$

	Bias		Standard error		Mean squared error		% times $0.99 \leq \lvert\hat{\theta}\rvert \leq 1$	
θ	LLS	FGN	LLS	FGN	LLS	FGN	LLS	FGN
-0.9	0.0717	0.0721	0.1362	0.1315	0.0237	0.0225	17.7	11.7
-0.8	0.0295	0.0282	0.1483	0.1460	0.0229	0.0221	11.0	8.0
-0.7	0.0072	0.0050	0.1643	0.1632	0.0270	0.0267	6.3	4.4
-0.6	-0.0047	-0.0063	0.1797	0.1796	0.0323	0.0323	3.7	2.6
-0.5	-0.0094	-0.0112	0.1913	0.1927	0.0367	0.0373	1.7	1.6
-0.4	-0.0102	-0.0125	0.2024	0.2050	0.0411	0.0422	0.8	0.7
-0.3	-0.0088	-0.0107	0.2090	0.2114	0.0438	0.0448	0.4	0.4
-0.2	-0.0061	-0.0068	0.2126	0.2138	0.0452	0.0457	0.2	0.1
-0.1	-0.0028	-0.0030	0.2146	0.2159	0.0461	0.0466	0.1	0.0
0.0	0.0007	0.0005	0.2146	0.2168	0.0461	0.0470	0.1	0.0
0.1	0.0035	0.0037	0.2142	0.2161	0.0459	0.0467	0.1	0.1
0.2	0.0072	0.0072	0.2137	0.2141	0.0457	0.0459	0.2	0.1
0.3	0.0095	0.0098	0.2080	0.2093	0.0434	0.0439	0.4	0.3
0.4	0.0106	0.0111	0.2012	0.2032	0.0406	0.0414	0.8	0.6
0.5	0.0089	0.0096	0.1912	0.1918	0.0366	0.0369	1.4	1.2
0.6	0.0029	0.0059	0.1780	0.1792	0.0317	0.0321	2.9	2.1
0.7	-0.0079	-0.0040	0.1652	0.1637	0.0274	0.0268	5.7	4.4
0.8	-0.0304	-0.0268	0.1493	0.1470	0.0232	0.0223	10.2	8.0
0.9	-0.0728	-0.0708	0.1381	0.1338	0.0244	0.0229	16.5	11.7

In Table III we present the results for the unconditional case and $n = 30$. A significant 'pile-up effect' was found in all the models with $|\theta| \geq 0.6$. Comparing the results of Tables II and III we can see that the conditional estimators have always less mean squared error than the unconditional ones except for the case $\theta = \pm 0.9$ due to the presence of a strong 'pile-up effect' of about 70 percent of the estimates.

TABLE III
Simulation statistics for the unconditional case and $n = 30$

	Bias		Standard error		Mean squared error		% times $0.99 \leq \lvert\hat{\theta}\rvert \leq 1$	
θ	LLS	FGN	LLS	FGN	LLS	FGN	LLS	FGN
-0.9	-0.0438	-0.0489	0.1160	0.1055	0.0154	0.0135	72.8	69.4
-0.8	-0.0724	-0.0761	0.1572	0.1521	0.0300	0.0289	49.7	47.0
-0.7	-0.0756	-0.0763	0.1915	0.1896	0.0424	0.0418	30.8	28.1
-0.6	-0.0669	-0.0659	0.2145	0.2108	0.0505	0.0488	17.5	15.5
-0.5	-0.0550	-0.0550	0.2242	0.2238	0.0533	0.0531	9.3	8.6
-0.4	-0.0427	-0.0418	0.2307	0.2297	0.0550	0.0545	4.7	4.1
-0.3	-0.0297	-0.0301	0.2328	0.2329	0.0551	0.0551	2.2	2.3
-0.2	-0.0190	-0.0197	0.2342	0.2347	0.0552	0.0555	1.2	1.3
-0.1	-0.0093	-0.0093	0.2338	0.2343	0.0547	0.0550	0.6	0.7
0.0	-0.0004	-0.0004	0.2333	0.2320	0.0544	0.0538	0.6	0.4
0.1	0.0088	0.0079	0.2303	0.2303	0.0531	0.0531	0.5	0.4
0.2	0.0189	0.0173	0.2342	0.2324	0.0552	0.0543	1.1	0.8
0.3	0.0290	0.0286	0.2314	0.2314	0.0544	0.0544	2.1	1.7
0.4	0.0403	0.0393	0.2276	0.2270	0.0534	0.0531	4.1	3.5
0.5	0.0507	0.0505	0.2200	0.2200	0.0510	0.0510	7.9	7.0
0.6	0.0606	0.0611	0.2106	0.2067	0.0480	0.0465	14.8	13.4
0.7	0.0717	0.0752	0.1859	0.1869	0.0397	0.0406	27.6	26.6
0.8	0.0662	0.0760	0.1639	0.1518	0.0312	0.0288	46.7	46.2
0.9	0.0390	0.0511	0.1181	0.1056	0.0155	0.0138	69.3	70.0

Corresponding results for the conditional and unconditional cases and $n = 100$ appear in Tables IV and V respectively. For this sample size, the LLS and FGN estimators have a very similar behavior as we can expect from the asymptotic theory developed in Section 4. For this sample size, the 'pile-up' effect for the unconditional case is only significant for models with $|\theta| \geq 0.8$ and for $\theta = \pm 0.9$ about one third of the estimates lies in the interval $0.99 \leq |\hat{\theta}| \leq 1$.

TABLE IV
Simulation statistics for the conditional case and $n = 100$

	Bias		Standard error		Mean squared error		% times $0.99 \leq \|\hat{\theta}\| \leq 1$	
θ	LLS	FGN	LLS	FGN	LLS	FGN	LLS	FGN
-0.9	0.0213	0.0213	0.0637	0.0609	0.0045	0.0042	6.9	3.5
-0.8	0.0035	0.0037	0.0711	0.0696	0.0051	0.0049	1.3	0.7
-0.7	-0.0026	-0.0023	0.0794	0.0783	0.0063	0.0061	0.2	0.1
-0.6	-0.0047	-0.0047	0.0865	0.0863	0.0075	0.0075	0.0	0.0
-0.5	-0.0054	-0.0055	0.0925	0.0926	0.0086	0.0086	0.0	0.0
-0.4	-0.0054	-0.0054	0.0972	0.0972	0.0095	0.0095	0.0	0.0
-0.3	-0.0049	-0.0049	0.1008	0.1008	0.0102	0.0102	0.0	0.0
-0.2	-0.0041	-0.0041	0.1033	0.1033	0.0107	0.0107	0.0	0.0
-0.1	-0.0030	-0.0030	0.1047	0.1047	0.0110	0.0110	0.0	0.0
0.0	-0.0018	-0.0018	0.1052	0.1052	0.0111	0.0111	0.0	0.0
0.1	-0.0006	-0.0006	0.1048	0.1048	0.0110	0.0110	0.0	0.0
0.2	0.0006	0.0006	0.1033	0.1033	0.0107	0.0107	0.0	0.0
0.3	0.0016	0.0016	0.1007	0.1007	0.0101	0.0101	0.0	0.0
0.4	0.0025	0.0024	0.0971	0.0970	0.0094	0.0094	0.0	0.0
0.5	0.0028	0.0027	0.0921	0.0919	0.0085	0.0085	0.0	0.0
0.6	0.0023	0.0022	0.0853	0.0853	0.0073	0.0073	0.0	0.0
0.7	0.0005	0.0002	0.0780	0.0773	0.0061	0.0060	0.2	0.1
0.8	-0.0056	-0.0055	0.0696	0.0686	0.0049	0.0047	1.1	0.6
0.9	-0.0225	-0.0225	0.0626	0.0602	0.0044	0.0041	6.1	3.4

TABLE V

Simulation statistics for the unconditional case and $n = 100$

	Bias		Standard error		Mean squared error		% times $0.99 \le \lvert\hat{\theta}\rvert \le 1$	
θ	LLS	FGN	LLS	FGN	LLS	FGN	LLS	FGN
-0.9	-0.0307	-0.0313	0.0615	0.0592	0.0047	0.0045	34.3	31.3
-0.8	-0.0253	-0.0251	0.0818	0.0770	0.0073	0.0066	7.8	5.6
-0.7	-0.0194	-0.0187	0.0845	0.0827	0.0075	0.0072	1.2	0.7
-0.6	-0.0156	-0.0157	0.0894	0.0894	0.0082	0.0082	0.1	0.1
-0.5	-0.0133	-0.0133	0.0953	0.0951	0.0093	0.0092	0.1	0.0
-0.4	-0.0110	-0.0110	0.0993	0.0991	0.0100	0.0099	0.0	0.0
-0.3	-0.0087	-0.0087	0.1024	0.1023	0.0106	0.0105	0.0	0.0
-0.2	-0.0065	-0.0065	0.1047	0.1047	0.0110	0.0110	0.0	0.0
-0.1	-0.0042	-0.0042	0.1060	0.1060	0.0113	0.0113	0.0	0.0
0.0	-0.0018	-0.0018	0.1065	0.1065	0.0113	0.0113	0.0	0.0
0.1	0.0005	0.0005	0.1061	0.1061	0.0113	0.0113	0.0	0.0
0.2	0.0029	0.0029	0.1048	0.1048	0.0110	0.0110	0.0	0.0
0.3	0.0053	0.0052	0.1024	0.1024	0.0105	0.0105	0.0	0.0
0.4	0.0077	0.0077	0.0989	0.0988	0.0098	0.0098	0.0	0.0
0.5	0.0103	0.0101	0.0943	0.0939	0.0090	0.0089	0.0	0.0
0.6	0.0130	0.0128	0.0887	0.0880	0.0080	0.0079	0.2	0.1
0.7	0.0165	0.0159	0.0825	0.0811	0.0071	0.0068	0.9	0.4
0.8	0.0229	0.0215	0.0779	0.0752	0.0066	0.0061	6.5	4.6
0.9	0.0288	0.0292	0.0608	0.0591	0.0045	0.0043	31.7	29.4

7. Empirical examples

In this section we compare the performance of the LLS method with the FGN procedure and the maximum likelihood estimator by applying it to the daily IBM common stock closing prices data of Box and Jenkins (1976, p. 526) and to the yearly U.S. tobacco production data of Wei (1990, p. 449). We computed the conditional LLS and FGN estimators using FORTRAN 77 programs as in the previous section. The Galbraith/Zinde-Walsh preliminary estimator calculated using an autoregressive approximation of order 15 was used as a starting value of the iteration procedure. The exact maximum likelihood (EML) estimator was computed using Mélard's algorithm [see Mélard (1984)] with the help of the SPSS statistical package.

The asymptotic variance of the least squares estimator $\hat{\theta}$ is given by [see Box and Jenkins (1976, p. 227)],

$$\operatorname{var}(\hat{\theta}) \simeq 2\sigma_{\varepsilon}^{2}\left[E\left(\frac{\partial^{2}\mathbf{S}^{*}(\theta)}{\partial\theta^{2}}\right)\right]^{-1}, \tag{37}$$

where $S^{*}(\theta)$ is the conditional sum of squares function. Furthermore, for large samples, we can approximate the expected value of the second partial derivative in (37) by the sample values actually observed, and hence, the estimated asymptotic variance of the estimator is

$$\operatorname{var}(\hat{\theta}) \simeq \hat{\sigma}_{\varepsilon}^{2}\left[\sum_{t=1}^{n}\left(\frac{\partial\left[\varepsilon_{t}\right]}{\partial\theta}\right)^{2}+\sum_{t=1}^{n}\left(\left[\varepsilon_{t}\right]\frac{\partial^{2}\left[\varepsilon_{t}\right]}{\partial\theta^{2}}\right)\right]^{-1}, \tag{38}$$

where $\hat{\sigma}_{\varepsilon}^{2}=S^{*}(\theta)\,/n$.

For each example, we have computed the parameter estimator $\hat{\theta}$, its asymptotic standard error, using the expression (38) for the LLS and FGN methods, the estimated residual variance $\hat{\sigma}_{\varepsilon}^{2}$ and the Akaike information criterion (AIC) as a standard tool of model selection.

Example 7.1. Let $\{Y_t,\, t=1,\ldots,369\}$ denote the daily IBM common stock closing prices data from 17th May 1961 to 2nd November 1962. As in the analysis of Box and Jenkins (1976, Table 6.4, series B), the data was transformed by taking first differences to produce a new stationary series with a sample autocorrelation function which suggests a moving average model of order one. Then, we can write the model as

$$\nabla Y_t=\varepsilon_t+\theta\varepsilon_{t-1}, \qquad t=1,\ldots,368, \tag{39}$$

where the operator ∇ is the first difference operator and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and the same variance σ_ε^2.

The preliminary Galbraith/Zinde-Walsh estimator took the value $\hat{\theta} = 0.0888$. The results of the estimation of the model (39) are shown in Table VI.

TABLE VI
Estimation results for the daily IBM
common stock closing prices data

Method	$\hat{\theta}$	s.e$(\hat{\theta})$	$\hat{\sigma}_\varepsilon^2$	AIC
LLS	0.08658	0.05130	52.21903	1457.605
FGN	0.08657	0.05130	52.21903	1457.605
EML	0.08636	0.05203	52.36119	1458.605

Example 7.2. Let $\{Z_t,\ t = 1, \ldots, 114\}$ denote the yearly U.S. tobacco production data from 1871 to 1984 in millions of pounds. Following the analysis of Wei (1990, Table 7.3, series W6), the data was first transformed by taking natural logarithms and then applying the difference operator to generate a new stationary series with a sample autocorrelation function which indicates a moving average model of order one. Therefore, we can write the model as

$$\nabla \ln Z_t = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad t = 1, \ldots, 113, \tag{40}$$

where the operator ∇ is the first difference operator and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and the same variance σ_ε^2.

The preliminary Galbraith/Zinde-Walsh estimator took the value $\hat{\theta} = -0.7038$. The results of the estimation of the model (40) are displayed in Table VII.

TABLE VII
Estimation results for the yearly
U.S. tobacco production data

Method	$\hat{\theta}$	s.e$(\hat{\theta})$	$\hat{\sigma}_\varepsilon^2$	AIC
LLS	-0.60027	0.06750	0.02785	-402.653
FGN	-0.60009	0.06754	0.02785	-402.653
EML	-0.60558	0.07620	0.02798	-402.137

For each model, the two sets of least squares estimates are virtually the same and have identical AIC. This is not surprising, as series containing more than a hundred observations may be expected to follow large sample theory. Nevertheless, in the U.S. tobacco production model the conditional LLS estimator has, by a very small margin, less standard error than the others. It's also important to note that the EML estimator has, in both models, greater standard error and AIC than the least squares estimates, which supports the findings of Nelson (1974) and Dent and Min (1978) that the conditional least squares estimator may be marginally superior to the exact maximum likelihood one.

8. Concluding remarks

Four lessons seem to emerge from the simulation results; first of all, the linear and the nonlinear least squares estimators have the same asymptotic properties as we can see in the experiment performed for $n = 100$ and in the examples examined in Section 7.

Second, at the least there seems to be no justification for the computation of the pre-sample error ε_0 because the conditional estimators performs better than the unconditional ones for both sample sizes. This supports the findings of Dent and Min (1978) and Davidson (1981a, 1981b).

Third, the results of the experiment for $n = 30$ suggest that in small samples the conditional LLS estimator performs globally as well as the Fuller-Gauss-Newton one in terms of the mean squared error and is clearly superior when the true parameter value lies in the range $[-0.6, 0.6]$.

Finally, a 'pile-up' effect of a substantial magnitude has been detected for the unconditional LLS and FGN estimators, being higher as the sample size decreases and the invertibility boundary is approached. For both sample sizes and $\theta = \pm 0.9$ a high proportion of replications produce a boundary minimum for the unconditional estimators generating a simulation mean squared error close to zero, while they are points of maximal bias for the conditional estimators. This supports the findings of Ansley and Newbold (1980) and Davidson (1981a, 1981b).

Acknowledgements

We are very grateful to Ruey Tsay and Philip Rothman for valuable comments and suggestions on the first draft of this paper and to Antoni Vidiella for his FORTRAN 77 programming support.

References

Anderson, T.W. and Mentz, R.P. (1993) . "A note on maximum likelihood estimation in the first-order Gaussian moving average model", Statistics and Probability Letters, 16, 205-211.

Anderson, T.W. and Takemura, A. (1986) , "Why do noninvertible estimated moving averages occur?", Journal of Time Series Analysis, 7, 235-254.

Ansley, C.F. and Newbold, P. (1980) . "Finite sample properties of estimators for autoregressive moving average models", Journal of Econometrics, 13, 159-183.

Box, G.E.P. and Jenkins, G.M. (1976) , Time Series Analysis: Forecasting and Control, (2nd. edn), Holden-Day, San Francisco.

Brockwell, P.J. and Davis, R. A. (1991) , Time Series: Theory and Methods, (2nd. edn), Springer-Verlag, New York.

Brockwell, P.J. and Davis, R. A. (1988) , ”Simple consistent estimation of the coefficients of a linear filter”, Stochastic Processes and their Applications, 28, 47-59.

Cryer, J.D. and Ledolter, J. (1981) , ”Small-sample properties of the maximum likelihood estimator in the first-order moving average model”, Biometrika, 68, 691-694.

Davidson, J.E.H. (1981a) , ”Small sample properties of estimators of the moving average process”, In E.G. Charatsis, ed., Proceedings of the Econometric Society European Meeting 1979: Selected econometric papers in memory of Stephan Valavanis, North-Holland, Amsterdam, 27-62.

Davidson, J.E.H. (1981b) , ”Problems with the estimation of moving average processes”, Journal of Econometrics, 16, 295-310.

Dent, W. T. and Min, A. S. (1978) , ”A Monte Carlo study of autoregressive integrated moving average processes”, Journal of Econometrics, 7, 23-55.

Fuller, W.A. (1976) , Introduction to Statistical Time Series, Wiley, New York.

Kang, K.M. (1975) , ”A comparison of estimators for moving average processes”, Unpublished Paper, Australian Bureau of Statistics.

Galbraith, J.W. and Zinde-Walsh, V. (1994) , ”A simple noniterative estimator for moving average models”, Biometrika, 81, 143-155.

MacPherson, B.D. and Fuller, W. A. (1983) , ”Consistency of the least squares estimator of the first order moving average parameter”, Annals of Statistics, 11, 326-329.

Mélard, G. (1984) , ”A fast algorithm for the exact likelihood of autoregressive-moving average models”, Applied Statistics, 33, 104-114.

Nelson, C.R. (1974) , ”The first order moving average process: Identification, estimation and prediction”, Journal of Econometrics, 2, 121-141.

Sargan, J.D. and Bhargava, A. (1983) , ”Maximum likelihood estimation of regression models with first-order moving average errors when the root lies on the unit circle”, Econometrica, 51, 799-820.

Wei, W.W.S. (1990) . Time Series Analysis. Univariate and Multivariate Methods, Addison-Wesley, Redwood City, California.