***S**tatistica & **A**pplicazioni*
Vol. IV, n. 1, 2006
pp. 3-22

A Nonlinear Least Squares Solution of a Minimax Estimation Problem for Stationary Time Series

Marica Manisera[§]

Aride Mazzali[§]

Summary: *The parameter estimation problem for the ARMA models is considered under a minimax approach: when the objective is to minimize the largest deviations in time series, it is useful to define a generalized version of a minimax estimator, minimizing the sum of the r-th powers of the k<n largest absolute deviations. The solution to that problem can be obtained by a searching procedure based on a grid of values on the admissible parametric space. The aim of this paper is propose a new procedure (called FINLS) to solve that minimax problem when a quadratic loss function is considered. FINLS approximates the searching procedure by a modified NLS method, with a filter matrix W selecting the largest absolute deviations to be included in the computation. Some applications account for the validity of the proposed estimation process. The results of a simulative study empirically show that the asymptotic properties of the NLS estimator could be extended to the proposed estimator. We also propose an asymptotic estimator for the covariance matrix of the FINLS estimated parameters.*

Keywords: *stationary time series, minimax, nonlinear least squares, filtered input.*

1. Introduction

The parameters of invertible and stationary ARMA(p,q) models are usually estimated by the maximum likelihood (ML) method. Since the likelihood function can sometimes be difficult to derive, for suitably long series the

[§] Department of Quantitative Methods – University of Brescia – C.da S. Chiara, 50, 25122 Brescia (e-mail: mazzali@eco.unibs.it, manisera@eco.unibs.it).

This paper is the outcome of the assiduous collaboration of the authors, who cooperated jointly. However, the sections 1, 3, 5, 6.1 *can be referred to Aride Mazzali and the sections* 2, 4, 6.2, 7, 8 *to Marica Manisera*.

parameter estimates are obtained by approximate procedures minimizing, with respect to the unknown parameters, a sum of squares function. Such procedures usually provide very close approximations to the ML estimates (Box & Jenkins, 1970). Those estimates are known as *unconditional* or *conditional least squares estimates*, whether the initial values are provided by a preliminary back-calculation (i.e., the back-forecasting) or not. Since for the ARMA(p,q) models with $q > 0$ the terms involved in the sum of squares are always nonlinear functions of the parameters, iterative algorithms must be used to estimate the parameters, for example the *nonlinear least squares* (NLS) *method*, which iteratively applies the linear least squares method.

In many real situations it is important to have a model that adequately represents the series but, at the same time, accounts for a set of $k<n$ largest absolute deviations[1]; such a model can be obtained if the parameters are estimated by minimizing the sum of squares or, more generally, of r-th powers of the absolute value (*a.v.*) of a certain number $k<n$ of largest errors instead of the sum of all n squared errors. This aim can be achieved by means of a possible generalization of the minimax method.

In the present paper the generalized version of the minimax method for the estimation of the parameters of an ARMA(p,q) model (Mazzali & Manisera, 2005) minimizing the maximum of the sum of $k<n$ squared deviations is considered, with a twofold aim: firstly, to show that the minimax problem can be viewed as a particular case of the nonlinear least squares (NLS) problem, with filtered (or selected) input, and secondly, to highlight the statistical asymptotic properties of the resulting estimator.

In the second section the problems related to the parameter estimation in stationary and invertible time series are briefly recalled. The third section focuses on a possible generalization of the classical minimax estimation considering a quadratic loss function; the solution is found by a searching procedure. We propose (in the fourth section) an alternative solution (called FINLS) based on the NLS method with selection of the inputs and (in the fifth section) an index of stability for the selection matrix useful to identify that solution. Some applications accounting for the validity of the proposed estimation process are described in the sixth section. Section seven illustrates a simulative study on the asymptotic statistical properties of the resulting estimates. We also propose an asymptotic estimator for the covariance matrix of the FINLS estimates. The conclusions are in section eight.

---

[1] The k largest absolute deviations are not outliers, but are the extreme values once discarded the outliers.

2. Parameter estimation of ARMA models

In this paper, we consider a univariate Gaussian time series z_t following the autoregressive moving average (ARMA) model:

$$\phi(B)(z_t - \mu) = \theta(B)a_t \tag{1}$$

where $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - ... - \theta_q B^q$ are polynomials in B of degrees p and q, respectively, B is the backshift operator such that $Bz_t = z_{t-1}$, $\mu = E(z_t)$, and a_t is a sequence of independent Gaussian noise with mean 0 and variance σ_a^2. In the model (1) we assume that all of the zeros of $\phi(B)$ and $\theta(B)$ are outside the unit circle and that $\phi(B)$ and $\theta(B)$ have no common factors.

The ARMA model (1) can also be written as:

$$a_t = (z_t - \mu) - \phi_1(z_{t-1} - \mu) - ... - \phi_p(z_{t-p} - \mu) + \theta_1 a_{t-1} + ... + \theta_q a_{t-q}\ .$$

Because of the assumptions on the a_t's, the likelihood of the parameter vector $\boldsymbol{\xi} = (\boldsymbol{\beta}, \sigma_a^2)$, with $\boldsymbol{\beta} = [\phi_1, ..., \phi_p, \theta_1, ..., \theta_q]$, can be written as:

$$L(\boldsymbol{\xi} \mid \mathbf{z}) = h(\boldsymbol{\xi}) \exp\left[-\frac{1}{2\sigma_a^2} S(\boldsymbol{\beta}) \right] \tag{2}$$

where $\mathbf{z}$ is the observed time series of n values, h is a function of the parameters $\boldsymbol{\xi}$ and

$$S(\boldsymbol{\beta}) = \sum_{t=1-g}^{n} [a_t \mid \mathbf{z}]^2\ , \tag{3}$$

g=max(p,q); $[a_t \mid \mathbf{z}] = E[a_t \mid \mathbf{z}]$ denotes the expectation of a_t conditional on $\mathbf{z}$ and on $\boldsymbol{\beta}$:

$$[a_t \mid \mathbf{z}] = \begin{cases} a_t & \text{for} \quad t = 1,2,...,n \\ E(a_t \mid \mathbf{z}, \boldsymbol{\beta}) & \text{for} \quad t < 1 \end{cases};$$

$[a_t \mid \mathbf{z}]$ is a linear function of the initial unobservable values $\mathbf{a}^*$ e $\mathbf{z}^*$ that are needed for the evaluation of a_t for t<1.

Maximum likelihood estimates of the parameter vector $\boldsymbol{\xi}$ for the model (1) can be obtained by *maximizing* the function (2). In general, such a maximization is difficult; for this reason, many approximations have been considered in the literature. A commonly used approximation is obtained by ignoring the function $h(\boldsymbol{\xi})$ and minimizing $S(\boldsymbol{\beta})$. The resulting estimates are called least squares (LS) estimates.

Even in the minimization of $S(\boldsymbol{\beta})$ there are some difficulties: the initial expectations can be hard to evaluate and furthermore $S(\boldsymbol{\beta})$ can be a complicated function of the parameters.

Computationally simpler estimates can be obtained by minimizing the conditional sum of squares $S_C(\boldsymbol{\beta}) = \sum_{t=g+1}^{n}[a_t]^2 = \sum_{t=g+1}^{n} a_t^2$, where the starting values are set equal to zero. The resulting estimates are called *conditional least squares* (*CLS*) estimates.

Suppose that $\{z_t\}$, t=1,…,n, is a realization of a stationary and invertible Gaussian ARMA(p,q) process. Once identified p and q, for each choice of the initial values a_i^*, z_i^* and of the parameters $\boldsymbol{\beta}$, it is possible to obtain the succession of the a_t's. Since they depend on the parameters and on the initial values, they are denoted by $\hat{a}_t(\boldsymbol{\beta}; a_i^*, z_i^*)$ or just by $\hat{a}_t$; for $t = g+1, g+2, ..., n$, we have

$$\hat{a}_t(\boldsymbol{\beta}; a_i^*, z_i^*) = \hat{a}_t =$$
$$= (z_t - \mu) - \phi_1(z_{t-1} - \mu) - ... - \phi_p(z_{t-p} - \mu) + \theta_1 a_{t-1} + ... + \theta_q a_{t-q}$$

The *CLS* estimates are simply obtained by computing, for each choice of $\boldsymbol{\beta}$ on a given grid of values in the admissible region, the sum of squared $\hat{a}_t$'s and then identifying the vector associated with the minimum of that sum. Formally:

$$\hat{\boldsymbol{\beta}} = \arg\{\min S_C(\boldsymbol{\beta})\} = \arg\{\min_{\boldsymbol{\beta}} \sum_{t=g+1}^{n} \hat{a}_t(\boldsymbol{\beta}; a_i^*, z_i^*)^2\}. \qquad (4)$$

Analogously, the unconditional least squares estimates can be obtained minimizing the sum of squares (3).

When the number of parameters of the ARMA model is not small, the searching procedure becomes somewhat difficult. Special simplifications occur in obtaining LS estimates for the autoregressive process. But if the *ARMA* model (1) contains moving average terms, then the conditional and unconditional sums of squares are not quadratic functions of the parameters (Box & Jenkins, 1970). This is because the term a_t (or $[a_t]$) is nonlinear in the parameters. Hence nonlinear least squares procedures must be used to minimize $S(\boldsymbol{\beta})$ or $S_C(\boldsymbol{\beta})$.

The NLS estimation method considers a_t as a function of $\boldsymbol{\beta}$ and linearizes this function at the point $\hat{a}_{t,0}$ (corresponding to some guessed set of parameter values $\boldsymbol{\beta}_0 = (\beta_{1,0}, \beta_{2,0}, ..., \beta_{s,0})$), $s=p+q$, in a first Taylor series expansion:

$$\hat{a}_t = \hat{a}_{t,0} + (\beta_1 - \beta_{1,0}) \left.\frac{\partial \hat{a}_t}{\partial \beta_1}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + ... + (\beta_s - \beta_{s,0}) \left.\frac{\partial \hat{a}_t}{\partial \beta_s}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} .$$

Considering this equation for t = 1, 2, … the linear representation in matrix form is obtained:

$$\hat{\mathbf{a}}_0 = \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \hat{\mathbf{a}}; \tag{5}$$

$\mathbf{X}=[x_{it}]$ is a $(n\text{-}p\text{-}q)$x$(p+q)$ matrix where $x_{it} = -\left.\frac{\partial \hat{a}_t}{\partial \beta_i}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$, $\hat{\mathbf{a}}_0$, and $\hat{\mathbf{a}}$ are column vectors with $n\text{-}p\text{-}q$ elements. The negative of the derivatives are numerically computed as follows (Box & Jenkins, 1970, p. 233):

$$x_{it} = \frac{\{(\hat{a}_t \mid \mathbf{z}, \beta_{1,0}, ..., \beta_{i,0}, ..., \beta_{s,0}) - (\hat{a}_t \mid \mathbf{z}, \beta_{1,0}, ..., \beta_{i,0} + \delta_i, ..., \beta_{s,0})\}}{\delta_i} .$$

NLS estimates of the ARMA(p,q) model can be obtained iteratively:

$$\hat{\boldsymbol{\beta}}_{r+1} = \hat{\boldsymbol{\beta}}_r - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{a}}$$

with r = 1, 2, ….

Once obtained the parameter estimates, the variance is estimated by:

$$\hat{\sigma}_a^2 = \frac{1}{n-p-q-1} \sum_{t=g+1}^{n} \hat{a}_t^2 \tag{6}$$

and the estimated asymptotic covariance matrix for the parameters is:

$$Var(\hat{\boldsymbol{\beta}}) = \hat{\sigma}_a^2 (\mathbf{X}'\mathbf{X})^{-1} \tag{7}$$

where the matrix $\mathbf{X}'\mathbf{X}$ is the sum of squares and cross products computed at the last iteration. With Gaussian noise, the estimator is an approximate ML estimator conditioned on the initial observations.

3. Minimax estimates with quadratic loss function

In the financial, environmental, pollution, and quality control fields, the *smallest* fluctuations can be often considered negligible and important damages are only due to the *largest* fluctuations, exceeding a certain level L. In those situations it is useful to define a minimax estimator with a loss function considering only the selection of data providing the k largest absolute deviations.

Formally, the estimator is obtained as the solution of the following minimax problem:

$$\tilde{\boldsymbol{\beta}} = \arg\left\{ \min_{\boldsymbol{\beta}} \left[\max_{a_t} \sum_{t=1}^{k} |\hat{a}_t(\boldsymbol{\beta})|^r \right] \right\} \tag{8}$$

where $\hat{a}_t(\boldsymbol{\beta}) = z_t - \hat{z}_t(\boldsymbol{\beta})$. Because of the identity:

$$\max \sum_{t=1}^{k} |\hat{a}_t(\boldsymbol{\beta})|^r = \sum_{t=1}^{k} \max |\hat{a}_t(\boldsymbol{\beta})|^r$$

(8) can be written as

$$\tilde{\boldsymbol{\beta}} = \arg\left\{ \min_{\boldsymbol{\beta}} \left[\sum_{t=1}^{k} \max |\hat{a}_t(\boldsymbol{\beta})|^r \right] \right\}$$

Since this method is a possible generalization of the standard minimax estimation method, it has been proposed in Mazzali and Manisera (2005) with the name of generalized minimax method gmM_k, with k being the number of deviations included in the objective function. In fact, it coincides with the standard minimax method when k=1 e r=1, with the minimum absolute deviations method when k=n e r=1, and with the LS method when k=n e r=2.

This work focuses on the case of $1<k<n$ (k is usually a fixed percentage of n) and r=2. In this case, the method corresponds to the (conditional or

unconditional) LS method (see (4)) applied only to that particular subset of observations corresponding to the k largest deviations (given $\boldsymbol{\beta}$). In this case the problem can be written as

$$\tilde{\boldsymbol{\beta}} = \arg\left\{\min_{\boldsymbol{\beta}} S_W(\boldsymbol{\beta})\right\} = \arg\left\{\min_{\boldsymbol{\beta}}\left[\sum_{t=1}^{n}[\hat{a}_t(\boldsymbol{\beta})]^2 w_t\right]\right\} \tag{9}$$

where w_t equals 1 when the deviation belongs to the set of the k largest absolute deviations and equals zero otherwise. We could define a nxn selection (filter) matrix $\mathbf{W}$=diag(w_t) as a diagonal matrix whose generic element is given by w_t. In this sense, $\tilde{\boldsymbol{\beta}}$ can also be interpreted as a special version of the Weighted Nonlinear Least Squares (WNLS) estimator (a.o., Carroll & Ruppert, 1988) with weights assuming only values 0 and 1. Only the k deviations with weight equal to 1 are considered in S; the other deviations are discarded from the computation. It is important to say that the weights can be defined only after computing the deviations, which in their turn depend on the parameters to be estimated.

The estimator (9) can be obtained in two different ways: (1) by a searching algorithm, looking for the minimum of $S_W(\boldsymbol{\beta})$ when parameters vary on a grid within the admissible region for $\boldsymbol{\beta}$ (with or without back-forecasting); (2) by taking advantage of the recursive relation defining the NLS method, as shown in the next section.

The searching estimation method (referred as to the GRID method) requires the following steps:

GRID *algorithm*

step 1: Fix a value for $\boldsymbol{\beta}$ on the grid, which has been previously defined within the admissible region for the parameters, with reference to the process being studied;

step 2: Compute the errors $\hat{a}_t(\boldsymbol{\beta}) = z_t - \hat{z}_t(\boldsymbol{\beta})$ as a function of the $\boldsymbol{\beta}$ fixed in step 1;

step 3: Identify the k largest absolute deviations and compute the corresponding sum of squares;

step 4: Repeat steps 1, 2, and 3 for every point on the grid;

step 5: Identify the *minimax* estimate $\tilde{\boldsymbol{\beta}}$ as the value of $\boldsymbol{\beta}$ corresponding to the minimum sum of squares.

4. A Nonlinear Least Squares solution with Filtered Input

A different procedure to solve the minimax problem (9) with considerable practical advantages is now considered.

On the basis of (9), it seems to be obvious to solve our estimation problem by looking for the Weighted Nonlinear Least Squares (WNLS) estimator with stochastic weights (depending on the estimated parameters).

Considering the linear representation (5) of the ARMA(p,q) model, the following WNLS iterative solution can be obtained:

$$\hat{\boldsymbol{\beta}}_{r+1} = \hat{\boldsymbol{\beta}}_r - (\mathbf{X}'\mathbf{W}_{r+1}'\mathbf{W}_{r+1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_{r+1}'\mathbf{W}_{r+1}\hat{\mathbf{a}},$$

which reduces to

$$\hat{\boldsymbol{\beta}}_{r+1} = \hat{\boldsymbol{\beta}}_r - (\mathbf{X}'\mathbf{W}_{r+1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_{r+1}\hat{\mathbf{a}} \tag{10}$$

because $\mathbf{W}_{r+1}$ is symmetric and idempotent. ($\mathbf{W}_{r+1}$ is computed at the r-th iteration on the basis of $\hat{\boldsymbol{\beta}}_r$). This could be only a theoretical solution to the problem: in practice, due to the singularity of $\mathbf{X}'\mathbf{W}_{r+1}\mathbf{X}$ when k is small, the WNLS solution (10) appears to be acceptable only for k sufficiently large ($k>n/2$, analogously to the Nonlinear Least Trimmed Squares estimator, see for example Stromberg & Ruppert, 1992).

In order to get a general solution valid for every k, we propose an approximate solution to the problem, which is very close to the "exact" one, identified by the GRID method. This closeness can be easily verified by comparing the estimates with the GRID ones.

We consider the linear representation (5) of the ARMA(p,q) model and impose to the response $\tilde{\mathbf{a}}_0$ the selection defined by the diagonal matrix $\mathbf{W}$ identifying the k largest absolute deviations. In other words, we consider the linear model

$$\mathbf{W}\tilde{\mathbf{a}}_0 = \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \tilde{\mathbf{a}} \tag{11}$$

representing the filtered response as a function of the regressors $\mathbf{X}$ given by the negative of the derivatives of $\mathbf{a}$ with respect to the parameters.

Starting from (11), the recursive formula

$$\tilde{\boldsymbol{\beta}}_{r+1} = \tilde{\boldsymbol{\beta}}_r - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_{r+1}\tilde{\mathbf{a}} \tag{12}$$

can be obtained. When the algorithm converges, the desired estimates can be achieved from (12) (with r+1 indicating the last iteration).

(12) can be viewed as a special version of the NLS method that can be called Filtered Input Nonlinear Least Squares (FINLS). The FINLS solution identifies a NLS solution combined with the selection of the k largest absolute (or squared) deviations as input of the solution itself. The very important difference between the WNLS method and the proposed FINLS method is that while in (10) both the input and the transfer function of the system are filtered, in (12) only the input is filtered while the transfer function keeps the characteristics of the NLS system: even when some inputs are imposed to be zero by the filter, all the n deviations are considered in the computation of $\mathbf{X}$ and, therefore, in the transfer function. In this sense the final FINLS solution can be considered as the solution of an "adaptive" system: the computation of the solution takes into account the selection of the largest absolute (squared) deviations, whose sum has to be minimized, but it is also based on the whole set of deviations. This is important to obtain a model still representative of the series, but based on the leading principle of the generalized minimax, focused on the minimization of the k largest (absolute) deviations.

FINLS gives an approximate solution because the filter matrix $\mathbf{W}$ reset to zero the elements in $\tilde{\mathbf{a}}$ corresponding to the n-k smallest absolute deviations. It is worthy to note that the WNLS solution is also approximate in this context: the n-k deviations with zero weights are not discarded; the corresponding elements in $\hat{\mathbf{a}}$ and in $\mathbf{X}$ are put equal to zero by $\mathbf{W}$.

The FINLS method appears to be an hybrid strategy with the advantage to have always a solution that is very close to the "exact" one. Although it could be a sub-optimal solution, it is asymptotically a NLS solution, as will be shown in the followings. The algorithm[2] can be described as follows:

FINLS *algorithm*

step 1: Start with a preliminary estimate $\tilde{\boldsymbol{\beta}}_0$ (usually the ML or NLS estimate);

step 2: Generate the residuals $\tilde{\mathbf{a}}_0$;

step 3: Form the selection matrix $\mathbf{W}_1$;

step 4: Obtain the FINLS estimate $\tilde{\boldsymbol{\beta}}_1$ as in (12), using the filter $\mathbf{W}_1$;

step 5: Update the preliminary estimator by setting $\tilde{\boldsymbol{\beta}}_0 = \tilde{\boldsymbol{\beta}}_1$, generate the new n residuals and update the filter;

step 6: Repeat steps 4 and 5 until convergence.

[2] The FINLS algorithm adjusts the initial (ML or NLS) estimates: the usual procedures of model identification (Box & Jenkins, 1970) are still valid.

5. An Index of Stability for the Filter Matrix W

When computing FINLS estimates (12) and WNLS estimates (10), some important considerations should be taken into account.

The filter or selection or weight matrix **W** is defined after having obtained the deviations $\tilde{\mathbf{a}}$; therefore, depending on the chosen initial values, the structure of **W** changes, according to the changes in the values of the largest absolute deviations, due to the changes in the parameters. If the chosen initial values for the parameters $\boldsymbol{\beta}_0$ are close to the optimal values – for example, the ML or the NLS estimates – the filter matrix rapidly converges to a stable configuration and the set of observations entering the optimization process becomes fixed: $\mathbf{W}=\mathbf{W}_r=\mathbf{W}_{r+1}=\mathbf{W}_{r+2}$; once achieved that stable configuration, **W** is not computed anymore and the convergence of the estimates can be reached more easily. In fact, a necessary condition for the algorithm to be convergent is that **W** becomes stable and assumes a fixed configuration.

It is clear that if the structure of the filter matrix **W** continuously changes, new and different observations enter the computation, and this hinders the convergence of the estimates. On the contrary, the estimation process converges if the filter matrix is stable and, obviously, the required conditions (given in Jenrich, 1969 and discussed, among others, by Wu, 1981 and Seber & Wild, 1989, 563-564) for the existence of the NLS estimator are satisfied.

In order to study the stability of the filter matrix, we propose the following empirical index $J_{r,r-1}$:

$$J_{r,r-1} = \frac{1}{2}\sum_i | w_{ii}^{(r)} - w_{ii}^{(r-1)} |$$

where $w_{ii}^{(r)}$ is the *i*-th diagonal element of the filter matrix $\mathbf{W}_r$ obtained at the *r*-th iteration.

$J_{r,r-1}$ gives the number of changes in the position of the nonzero elements of the filter matrix. $J_{r,r-1}$ equals zero when **W** has reached stability: from that moment onwards the deviations entering the computation are always the same and the FINLS procedure becomes a NLS procedure. When $J_{r,r-1}=h$, with h =1, 2, 3, … there are h modified positions, i.e., h new deviations come into the process at *r*-th iteration substituting deviations included in the previous (r–1)-th iteration[3].

[3] In practice, in some computation programs, in the presence of ties on the values of the deviations, it can happen that $J_{r,r-1}$ is not always zero but equals (sometimes cyclically) very small values (0, 1 or 2): this means that the computational process can sometimes show one or two modified positions from one iteration to the next

Adopting this point of view, the iterative estimation procedure can be considered as composed of two stages: in the first one a fixed configuration for the filter matrix is obtained; in the second one the final convergence of the estimates is achieved.

6. Some Applications

In the present section some applications, accounting for the validity of the proposed estimation process, are described. Two datasets were used: (1) the first one refers to the real time series of the closing quotes of Edison shares and (2) the second one is a time series simulated from an *ARMA*(1,2) process.

6.1 *The Edison time series*

We consider the time series from 11.06.1998 to 31.05.2000 (n=501) of the closing quotes of Edison shares (Milan Stock Exchange). In a previous work (Mazzali & Manisera, 2006) we fitted an ARMA(0,1) model to the series y_t of the relative price changes under the hypothesis of homoscedasticity of the errors and we computed the NLS and the GRID estimates with k=50,100. The results are displayed in Table 1, which also shows the FINLS estimates.

Table 1. NLS, GRID *and* FINLS *estimates of* μ *and* θ_1 *for* k=50,100

Method		μ *estimates*	θ_1 *estimates*
NLS		.004	.240
GRID	k = 50	.150	.340
	k = 100	.200	.280
FINLS	k = 50	.101	.336
	k = 100	.141	.313

Table 2 displays the partial sums $S(j)$ of the squared residuals of the first j largest deviations for the three models when the focus is on the solutions obtained for k=50. As expected, the $S(j)$ associated with the model estimated by GRID and FINLS are clearly smaller for j<50; but this also holds for some j>50; obviously, if k=n=500 then the sum of squared NLS residuals is the smaller one. This is also evident in Figure 1, representing the the differences between the sums $S_{NLS}(j)$ and $S_{GRID}(j)$ and between the sums

one. Even in this situation, after some investigations, it is possible to verify if **W** has reached a substantial stability.

$S_{NLS}(j)$ and $S_{FINLS}(j)$ corresponding to the models obtained by NLS, GRID (k=50) and FINLS (k=50) estimates.

Table 2. *Sums S*(*j*) *of the first j largest squared estimated residuals for some values of j*

$S(j)$	NLS *model*	GRID *model* (*k=50*)	FINLS *model* (*k=50*)
$S(1)$	55.57	52.76	53.34
$S(10)$	327.52	321.06	320.90
$S(30)$	720.47	713.99	712.71
$S(50)$	1003.17*	989.15*	989.41*
$S(100)$	1430.23	1423.22	1423.91
$S(158)$	1714.67	1715.02	1714.66
$S(159)$	1718.15	1718.95	1718.52
$S(500)$	2069.05	2093.71	2089.33

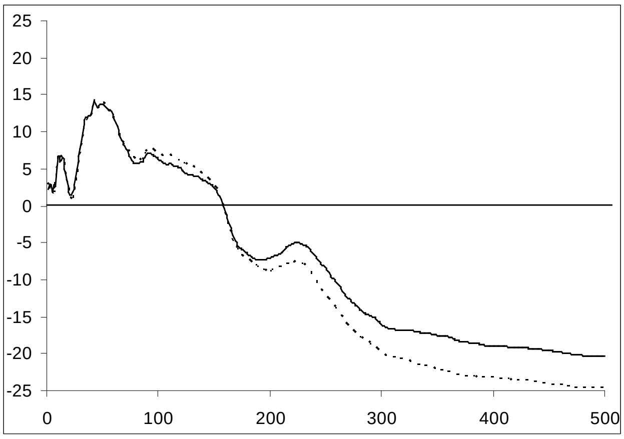

Figure 1. *Differences between the sums* $S_{NLS}(j)$ *and* $S_{GRID}(j)$ (*dashed line*) *and between the sums* $S_{NLS}(j)$ *and* $S_{FINLS}(j)$ (*black line*) *of the first j largest squared estimated residuals of models obtained by NLS*, *GRID* (k=50) *and FINLS* (k=50) *estimates*

This conclusion indicates that if a cost function penalizing the largest deviations (e.g., the deviations larger than a level L) is adopted, then the model obtained by minimizing the largest $k<n$ absolute (squared) deviations can be preferred.

If we consider larger values of k ($k>n/2$), then we can compare FINLS and WNLS methods. Table 3 shows that the FINLS and the WNLS estimates for k=300, k=350, k=400 coincide.

Table 3. FINLS *and* WNLS *estimates for* k=300, k=350, *and* k=400

k	*Parameters*	WNLS	FINLS
300	μ	.0488	.0488
	θ_1	-.2510	-.2510
350	μ	.0300	.0300
	θ_1	-.2509	-.2509
400	μ	.0343	.0343
	θ_1	-.2428	-.2428

6.2 *The simulated ARMA(1,2) time series*

The validity of the FINLS method has also been verified on a time series simulated from an *ARMA*(1,2) process z_t=.75z_{t-1}+a_t–.45a_{t-1}+.20a_{t-2}, with a_t~N(0,1) and n=500. The aim is not to obtain the true parameter values; rather, we want to show that, unlike the GRID estimates, the FINLS estimates are easy to obtain also when the number of parameters is not small. Figure 2 represents the resulting series and Table 4 displays the NLS and the FINLS (with k=50,100) estimates.

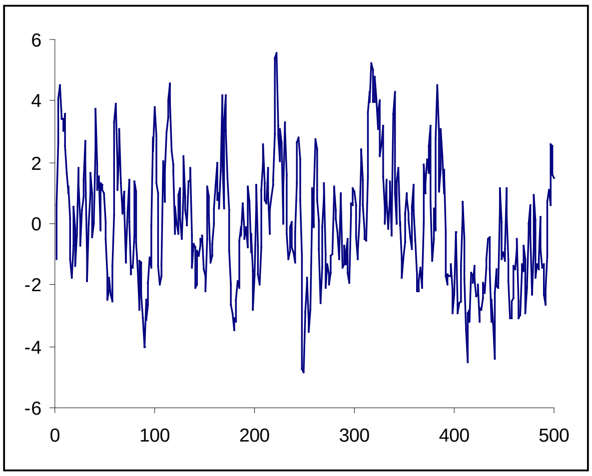

Figure 2. *The simulated* ARMA(1,2) *series from:* z_t=.75z_{t-1}+a_t–.45a_{t-1}+.20a_{t-2}; a_t=WN(0,1); (n=500)

Table 4. NLS *and* FINLS (k=50,100) *estimates for the parameters of the simulated* ARMA(1,2) *series* (n=500)

	Parameter values			
Method	$\mu = 0$	ϕ_1=.75	θ_1=-.45	θ_2=.25
NLS	.0254	.8043	-.4254	.2432
FINLS (k=50)	.7689	.7674	-.4799	.1948
FINLS (k=100)	.4862	.7591	-.4942	.1800

With reference to the NLS models, using (6) and (7) we are now able to compute the (approximate) asymptotic estimates for the error variance and for the covariance matrix of the parameters:

$$\hat{\sigma}_a^2 = \frac{1}{496} \cdot 471.0802 = 0.9497$$

$$Var(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 0.0695 & 0.0000 & 0.0000 & 0.0000 \\ & 0.0018 & 0.0021 & 0.0019 \\ & & 0.0043 & 0.0032 \\ & & & 0.0040 \end{bmatrix}$$

The models estimated by the minimax approach (FINLS) is still representative of the original series, as shown by the ACF and PACF functions of the residuals (not displayed here). The conclusion that both the fitted models are appropriate is confirmed by the Portmanteau test, being equal to 20.930 (p-value .401) and 20.710 (p-value .414) with reference to the first 20 autocorrelations (as suggested by Box & Jenkins, 1970, p. 290-291) for the FINLS model with k=50 and k=100 respectively.

The main difference between the two models estimated by FINLS with k=50 and k=100 regards the estimate for the mean. As expected, the mean estimate moves towards the true mean value when the number of the largest absolute deviations entering the sum to be minimized is larger. In fact, the smaller is k, the more the FINLS model is different from the NLS one, which takes into account all the n observations.

As in the previous application, we computed the partial sums $S_{\text{NLS}}(j)$ and $S_{\text{FINLS}}(j)$ of the of the first j largest squared estimated residuals for the three models NLS, FINLS (k=50), and FINLS (k=100) respectively. Table 5 displays those sums for some values of j. As expected, $S_{\text{NLSFI}}(j)<S_{\text{NLS}}(j)$ for $j=k$ but also for some larger values. On the contrary, when $j=n$=500, then $S_{\text{NLSFI}}(j)>S_{\text{NLS}}(j)$.

Table 5. *Sums S(j) of the first j largest squared estimated residuals for some values of j*

j	$S_{NLS}(j)$	$S_{FINLS}(j)$ (k=50)	$S_{FINLS}(j)$ (k=100)
20	124.4843	118.5947	120.3228
30	160.0241	155.8741	156.6415
50	215.0426	212.0845	212.4984
100	308.2035	309.2909	307.7385
500	471.0802	481.0588	475.9161

The same conclusions can be drawn from Figure 3, which displays the difference between the partial sums $S(j)$ of the first j largest squared estimated residuals for the NLS model and the FINLS models obtained with k=50 (dotted line) and k=100 (black line). The $S(j)$ associated to the FINLS models are clearly smaller for $j \leq k$. It is worthy to note that this is true also for some $j>k$ (up to j=83 for the FINLS model with k=50 and up to j=122 for the FINLS model with k=100); obviously, if $j=n$=500 then $S_{NLS}(j)<S_{FINLS}(j)$.

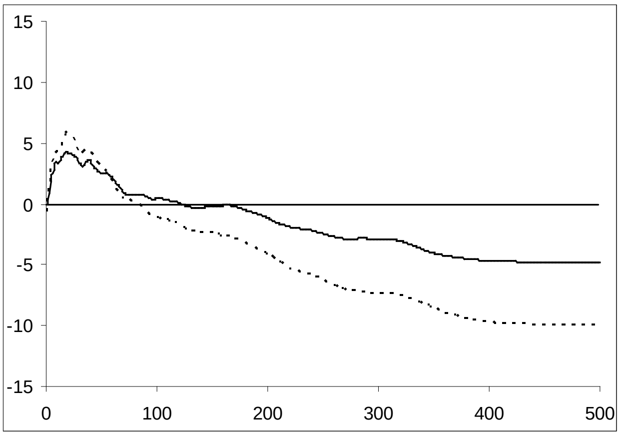

Figure 3. *Differences between the sums* $S_{NLS}(j)$ *and* $S_{FINLS}(j)$ *of the first j largest squared estimated residuals of models obtained by* NLS *and* FINLS (k=50 *dotted line,* k=100 *black line*)

Also in the present application we can compare the FINLS and the WNLS estimates for k>250. Again, the estimates obtained by the two methods are nearly the same (Table 6).

Table 6. FINLS *and* WNLS *estimates for k*=300, *k*=350, *and k*=400

k	*Parameters*	WNLS	FINLS
300	μ	.1342	.1130
	ϕ_1	.7914	.7930
	θ_1	-.4308	-.4297
	θ_2	.2446	.2440
350	μ	.0686	.0567
	ϕ_1	.7901	.7933
	θ_1	-.4340	-.4295
	θ_2	.2341	.2399
400	μ	.0486	.0485
	ϕ_1	.8037	.8038
	θ_1	-.4230	-.4227
	θ_2	.2474	.2475

7. Asymptotic properties of the estimator

In the previous sections it was shown that the proposed minimax estimator can computationally be obtained by the algorithm we called FINLS, that can be viewed as a variant of the NLS method.

In general, the NLS method implies the iterative application of linear least squares, i.e. the linearization of the nonlinear model. Unlike in the linear model, when the desirable properties of the LS estimator hold for finite samples, in the nonlinear model the NLS estimator do not enjoy any tractable finite sample optimality property (Wu, 1981) but only some asymptotic properties are valid. For example, the NLS estimator will be nearly always biased (Box, 1971). This is because the LS estimator for the nonlinear model is calculated iteratively using a linear approximation that is only asymptotic[4]. Therefore, also our FINLS estimator should generally enjoy only asymptotic properties.

In the study of the properties of the FINLS estimator three observations seem to be relevant.

The first one refers to the behaviour of the estimates when *k* tends to *n*, which remains constant. We already observed that for *k*=*n* the FINLS and the NLS estimators coincide and, therefore, the asymptotic properties of the

[4] The asymptotic properties are valid only when the linear approximation is a satisfactory approximation of the nonlinear function. Some authors (a.o., Box, 1971; see Seber & Wild, pp. 181/182-187) proposed measures of nonlinearity helping the evaluation of the goodness of the linear approximation.

NLS estimator (a.o., Wu, 1981; Amemiya, 1983; Judge et al., 1985; Sen, Srivastava, 1990) can be extended to the FINLS estimator.

The second observation refers to a more complex question: what happens to the FINLS estimates when n diverges and k is a fixed percentage of n? In order to answer this question, we simulated 1000 ARMA(0,1) series (with parameters $\mu = 0$ and θ_1=.45) with different dimensions: n=500, n=2000, n=5000, and n=10000, and we computed the FINLS estimates considering k=.10n and then k=.20n in all four cases.

The obtained estimates are displayed in Table 7.

Table 7. NLS *and* FINLS (*with* k=.10n *and* k=.20n) *estimates resulting from simulations of* 1000 ARMA(0,1) *series* ($\mu = 0$ *and* θ_1=.45) *with* n=500, 2000, 5000, 10000

	NLS		FINLS (k=.10n)		FINLS (k=.20n)	
n	$E(\hat{\mu})$	$E(\hat{\theta}_1)$	$E(\tilde{\mu})$	$E(\tilde{\theta}_1)$	$E(\tilde{\mu})$	$E(\tilde{\theta}_1)$
500	.0006 (.0249)	.4522 (.0387)	.0014 (.0438)	.4452 (.1100)	.0009 (.0321)	.4494 (.0605)
2000	.0001 (.0122)	.4501 (.0196)	-.0000 (.0198)	.4484 (.0326)	-.0000 (.0154)	.4496 (.0246)
5000	-.0000 (.0075)	.4504 (.0121)	-.0001 (.0120)	.4497 (.0205)	-.0002 (.0097)	.4501 (.0156)
10000	-.0001 (.0057)	.4502 (.0091)	.0003 (.0091)	.4505 (.0137)	.0003 (.0072)	.4500 (.0111)

As n increases, the results shown in Table 7 seem to converge towards the true parameter values and the associated variability gradually decreases. In fact, as n diverges, k also increases, and consequently the estimates are based on a larger amount of information. Obvioulsy, this trend is more emphasized when k=.20n, according to the first observation in this section. We may argue that these results empirically show that the FINLS estimator may be considered asymptotically unbiased and consistent (we adopt here the weak definition of consistency, see for example Mood, Graybill & Boes, 1974). In conclusion, the model estimated to take into account the largest absolute deviations results in substantially different parameter estimates when the series size is relatively small, while it tends towards the true model when the series size diverges.

The last consideration refers to the behaviour of the FINLS solution when $k>n/2$, which substantially equals the WNLS solution. The properties associated to the WNLS algorithm (White, 1980; Amemiya, 1983, p. 359) show that, under the adopted conditions, the WNLS estimates and therefore the FINLS estimates seem to be consistent.

On the basis of the previous considerations, we can propose an estimator for the covariance matrix of the parameter estimates. Let $\tilde{\sigma}_u^2$ the variance of $\tilde{\mathbf{u}} = \mathbf{W}\tilde{\mathbf{a}}$, where $\mathbf{W}$ is the filter matrix obtained at the last iteration of the FINLS algorithm; the nonzero elements in $\tilde{\mathbf{u}}$ are the largest absolute values in $\tilde{\mathbf{a}}$. We have:

$$\tilde{\sigma}_u^2 = \frac{1}{k-p-q-1}\sum_{i=1}^{n}\tilde{u}_i^2 = \frac{1}{k-p-q-1}\sum_{i=1}^{k}\tilde{a}_{(i)}^2$$

with $\tilde{a}_{(i)}$ being the i-th value in the decreasing ordering of absolute residuals $|\tilde{a}|_{(1)} > |\tilde{a}|_{(2)} > \dots$. Replacing $\mathbf{W}\tilde{\mathbf{a}}$ by $\tilde{\mathbf{u}}$ in the (12), we find:

$$Var(\tilde{\boldsymbol{\beta}}) = \tilde{\sigma}_u^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

Clearly, when $k \to n$, $\tilde{\sigma}_u^2 \to \hat{\sigma}_a^2$.

For example, with reference to the FINLS (k=100) model estimated in the section 6.2, the covariance matrix of $\tilde{\boldsymbol{\beta}}$ results:

$$Var(\tilde{\boldsymbol{\beta}}) = 3.2072 \cdot (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.2050 & 0.0090 & 0.0064 & 0.0057 \\ & 0.0076 & 0.0086 & 0.0080 \\ & & 0.0163 & 0.0131 \\ & & & 0.0151 \end{bmatrix}$$

8. Conclusions

In the present article, after a brief recall of the parameter estimation problem for stationary and invertible ARMA models, we considered a special version of a *minimax* estimator with a quadratic loss function, which implies the minimization of the sum of the squares of a certain number of $k<n$ largest absolute deviations. The solution can be obtained by a searching procedure based on a grid of values defined on the admissible parametric space, as introduced in a previous work (Mazzali & Manisera, 2005).

The searching procedure has been here considered again and, when the focus is on a quadratic loss, it can be approximated by a variant of the NLS problem, which we called FINLS. This is a NLS solution with a filter or selection matrix **W** selecting the largest absolute deviations to be included in the computation. The proposed index of stability is useful to examine the stability of matrix **W** and, consequently, to simplify the algorithm. Since we showed that as the iterative process converges, the selection matrix **W** becomes stable, and that as the series size increases, the FINLS estimates empirically converge to the NLS estimates, we argue that it is possible to extend the asymptotic properties of the NLS estimates to the proposed estimates. Some applications account for the validity of the proposed estimation process.

Moreover, when $k>n/2$, the FINLS results substantially equal the WNLS results. This further confirms the validity of the algorithm and supports the choice to filter the inputs in the proposed way. We found that the FINLS and the WNLS solutions are both approximations to the true solution, better identified by the GRID algorithm. While the WNLS solution can be computed for large values of k and the GRID solution is easy to compute only when p and q are small, the FINLS solution is feasible, easy and fast for any k and any p,q. It would be interesting to compare the three solutions (GRID, WNLS, and FINLS) in order to evaluate the discrepancies in the different levels of approximation.

The usefulness of the proposed model is clear when the loss function penalizes only or mostly the largest deviations (larger than a level L). In fact, in those situations the FINLS model can be preferred to the optimal model (estimated by ML or NLS) because the parametric structure of the model assures a better performance (smaller absolute forecast errors) just on the largest deviations.

The generalized version of the minimax estimator proposed in the third section can also be studied when the function to be maximinimized is not quadratic but expressed in absolute values (in the L_1 metric); in other words, when $1<k<n$ ed $r=1$. Moreover, it could be interesting to study the minimax estimator also when the loss function is asymmetric, in the sense that each deviation is associated to a different cost according to its sign. This will be the subject of another research.

References

Amemiya T. (1983). Nonlinear regression models. In Z. Griliches and M. D. Intriligator (Eds.). *Handbook of econometrics*. North-Holland, Amsterdam.

Box E. P., Jenkins G. M. (1970). *Time series analysis forecasting and control*. Holden Day, S. Francisco.

Box M. J. (1971). Bias in nonlinear estimation. *Journal of the Royal Statistical Society, Series B,* **26**, 211-252.

Carroll R. J., Ruppert D. (1988). *Transformation and weighting in regression*. Chapman and Hall, New York.

Jennrich R. I. (1969). Asymptotic properties of nonlinear least squares estimation. *Annals of Mathematical Statistics*, **40**, 633-643.

Judge G. G., Griffiths W. E., Hill R. C., Lutkepohl H., Lee T. (1985). *The Theory and practice of econometrics*. Wiley, New York.

Mazzali A., Manisera M. (2005). Stime minimax per serie storiche stazionarie. *Rapporti di ricerca del Dipartimento Metodi Quantitativi, Università degli Studi di Brescia*, n. 258.

Mazzali A., Manisera M. (2006). Minimax estimates for stationary time series. In Atti del Convegno SER2006, *Convegno Nazionale delle Ricerche sulle Serie Temporali*, Università di Roma TorVergata, Aprile 2006.

Mood A. M., Graybill F. A., Boes D. C. (1974). *Introduction to the theory of statistics*. McGraw Hill, New York.

Seber G. A. F., Wild C. J. (1989). *Nonlinear regression*. Wiley, New York.

Sen A. K., Srivastava M. (1990). *Regression analysis. Theory, methods, and applications*. Springer, New York.

Stromberg A. J., Ruppert D. (1992). Breakdown in nonlinear regression. *Journal of American Statistical Association*, **87**, 991-997.

White H. (1980). Nonlinear regression on cross-section data. *Econometrica*, **48**, 721-746.

Wu C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, **9**, 3, 501-513.