T E C H N I C A L
R E P O R T

0566

ON-LINE ESTIMATION OF THE PARAMETERS OF A TIME SERIES MODELS, WITH APPLICATIONS

OUAKASSE A., and G. MELARD

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

On-line estimation of the parameters of a time series models, with applications

A. Ouakasse and G. Mélard
Université Libre de Bruxelles. *

November 23, 2005.

Abstract

Recursive estimation methods for time series models usually make use of recurrences for the vector of parameters, the model error and its derivatives with respect to the parameters, plus a recurrence for the Hessian of the model error. An alternative method is proposed in the case of an ARMA model, where the Hessian is not updated but is replaced, at each time, by the inverse of the Fisher information matrix evaluated at the current parameter. The asymptotic properties, consistency and asymptotic normality, of the new estimator are obtained. Monte Carlo experiments indicate that the estimates may converge faster to the true values of the parameters than when the Hessian is updated. The paper is illustrated by an example on forecasting the speed of wind.
Keywords: time series, ARMA processes, recursive estimation, on-line estimation, Fisher information matrix.

1 Introduction

The development of estimation methods of the parameters of statistical and econometric models was influenced by the availability of more powerful computers. Numerical calculations are lighter and faster with the increased speed of computers, and bigger data bases can be used. For non-linear models, it is generally not possible to find the estimator analytically so numerical optimisation procedures are applied to obtain the maximum likelihood or even the least squares estimator. These procedures are iterative and make use of all the data at each iteration. They are called off-line because they are applied when all the

*ISRO CP 210, Campus Plaine ULB, Bd du Triomphe, B-1050 Bruxelles. G. Mélard is also with ECARES, Université Libre de Bruxelles. Phone: +32-2-6505890, fax: +32-2-6505899, e-mail: gmelard@ulb.ac.be.
This paper has benefited from an IAP-network in Statistics, contract P5/24, financially supported by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC). We thank Tractebel which kindly provided the data used in Section 5, and Nicolas Standaert, Denis Devleeschouwer and Bogdan Petrescu who performed an off-line analysis of these data.

data are available. Each time we have a new observation the whole estimation procedure has to be repeated. This is not a problem with quarterly or monthly data but availability of large capacity memory also implies that much more data are stored and more frequently. Instead of collecting data at a yearly, quarterly or monthly level, data are more and more collected in real time, starting with financial markets. Also, new fields of applications have appeared, like mobile telecommunications or fluid flow management, where quick automated decisions are required.

When the interval of time between two observations is very short, working with past, off-line, methods become inefficient if all data need to be used at high frequency rates and doing huge computations, because of the expensive calculation power needed as well as the memory space. Instead of being used by humans on their desks, the work should be done "on the spot" by computer systems and in an automated way. This raises new problems which are not yet entirely solved like model identification, and detection and treatment of outliers. The idea is to use on-line or recursive methods. They make use of a very small subset of data at each time. These methods appeared first in linear models (Plackett, 1950, who referred to Gauss) when computation was a major annoyance. In statistics they reappeared later (Brown *et al.*, 1975) as a way to check the stability of model specification with respect to time. In the discussion of that paper, the influence of Kalman (1960) became clear. Recursive methods became particularly interesting in the context of time series models, see Young (1985). These methods were indeed developed mainly in engineering, under the name of Recursive Identification, for data available on-line in telecommunications, transmissions, management of fluids, etc. For some recent contributions to recursive estimation methods, see Guo (1994), Kushner and Yin (1997), Moulines *et al.* (2004), Subba Rao and Dahlhaus (2004).

Among these recursive methods there is the RML (Recursive Maximum Likelihood) method which was introduced by Söderström (1973), see also Young (1984). We know that, under general conditions, the (off-line) maximum likelihood method gives an estimator which is asymptotically efficient, i.e. it is distributed asymptotically like a normal law whose asymptotic variance-covariance matrix is equal to the Cramér-Rao upper bound. Under certain conditions, Ljung and Söderström (1983) have shown that the RML estimator has the same asymptotic properties as the maximum likelihood estimator. But they noticed that for a finite series, $\{y_1, ..., y_n\}$, the maximum likelihood estimator is always better than the RML estimator. The RML estimator is based on a first order approximation of the Taylor expansion of the sum of squares of the errors. Let β be the vector of parameters of the model. As we will see in Section 2, the estimate at time t, $\widehat{\beta}_t$, makes use of the value at the previous time, $\widehat{\beta}_{t-1}$, but also of a matrix R_t which is an approximation of the Hessian of the sum of squares of errors. A recurrence for the residual is used but also a recurrence for the derivative of the error with respect to the parameters and an updating recurrence for the Hessian.

Mélard (1989) and Zahaf (1999) observed that the latter recurrence, with highly variable successive values of R_t, is often the cause for wild variations in the estimates and proposed a modified RML estimator for ARMA models. While keeping the spirit of the algorithm, instead of the recurrence for the Hessian R_t, Zahaf (1999) proposed to use the evaluation of the asymptotic Fisher information at the current value of the estimator, $\beta = \widehat{\beta}_{t-1}$.

Zahaf (1999) noticed that the asymptotic theory developed by Ljung (1977) and Ljung and Söderström (1983) no longer applies. He outlined an asymptotic theory based on the stochastic approximation of Robbins-Monro following Duflo (1997) but it was not complete. Moreover convergence in law of the estimator rested on a conjecture which was later proved to be wrong. For these reasons, after vain attempts including with the alternative approach of Kushner and Huang (1979), we preferred to adapt the approach of Ljung and Söderström.

In Section 2, we remind the necessary concepts of recursive maximum likelihood (RML) estimation in order to be able to introduce our version at the beginning of Section 3. The remaining of Section 3 is devoted to the main theorems in order to establish consistency and asymptotic normality of the new estimator. In Section 4, we show small samples results obtained by Monte Carlo simulations. They indicate that the new estimator is an improvement over the classical RML estimator. Section 5 will present an example of wind forecasting.

2 Recursive maximum likelihood estimation

Let us first describe the RML estimator before introducing how we have modified it. The algorithm for that estimator is derived from the off-line maximum likelihood estimator, see Ljung (1978) and Aström (1980). We assume for simplicity that the observations $\{y_t; t = 1, ..., N\}$ follow a univariate ARMA(p, q) model defined by the equation:

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - ... - \phi_p y_{t-p} = e_t - \theta_1\, e_{t-1} - \theta_2\, e_{t-2} - ... - \theta_q\, e_{t-q}, \quad (1)$$

where the roots of the autoregressive and moving average polynomials $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_p B^q$ are outside of the unit circle and e_t's are i.i.d. random variables with $E(e_t) = 0$ and $E(e_t^2) = \sigma_e^2$. Let $\beta = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)^T$ be the vector of the parameters of interest, where ^{T} denotes transposition, and let β^* be the true value of β . The estimator at time t will be denoted $\hat{\beta}_t = \left(\widehat{\phi}_{1,t}, ..., \widehat{\phi}_{p,t}, \widehat{\theta}_{1,t}, ..., \widehat{\theta}_{q,t}\right)^T$. For a given β, the forecast $\hat{y}_{t|t-1}(\beta)$ for time t can be computed at time $t-1$, provided we replace the true errors e_s, $s < t$, by the residuals $\varepsilon_s(\beta) = y_s - \hat{y}_{s|s-1}(\beta)$, computed by recurrence. This requires suitable initial values whose effect can be neglected because of the assumption on the polynomials. In off-line estimation, under the Gaussian assumption on e_t's, the maximum likelihood estimator is obtained by minimizing the sum of squares of the residuals

$$V_N(\beta) = \frac{1}{2} \sum_{t=1}^{N} \varepsilon_t^2(\beta). \quad (2)$$

Example 1

Specific parts will be illustrated with the ARMA(1,1) model defined by

$$y_t - \phi y_{t-1} = e_t - \theta e_{t-1}, \tag{3}$$

with $\beta^T = (\phi, \theta)$. Note that (3) implies

$$\hat{y}_{t|t-1}(\beta) = \phi y_{t-1} - \theta \varepsilon_{t-1}(\beta) \tag{4}$$

and

$$y_t - \hat{y}_{t|t-1}(\beta) = y_t - \phi y_{t-1} + \theta(y_{t-1} - \hat{y}_{t-1|t-2}(\beta)), \tag{5}$$

where the starting value $\hat{y}_{1|0}(\beta)$ can be taken equal to 0. Indeed the effect of a starting value decreases like $|\theta|^{t-1}$, and the assumption made implies that $|\theta| < 1$. This recurrence allows computing $\varepsilon_t(\beta)$.

For ARMA models, $V_N(\beta)$ is a non-linear function of β, so $V_N(\beta)$ cannot be minimized analytically but well using numerical procedures, requiring many iterations on basis of the data from $t = 1$ to $t = N$. An on-line or recursive algorithm requires a vector of fixed size, preferably small with respect to N. Therefore we want an approximation of the off-line maximum likelihood estimator $\hat{\beta}_N$ that can be obtained by recurrences.

Given $\hat{\beta}_{t-1}$, we want to obtain $\hat{\beta}_t$ which is close to the minimum of $V_t(\beta)$. By a Taylor expansion of $V_t(\beta)$ around $\hat{\beta}_{t-1}$ limited to the second order we obtain

$$\begin{aligned} V_t(\beta) \simeq V_t(\hat{\beta}_{t-1}) &+ \left(\frac{\partial V_t(\beta)}{\partial \beta^T}\right)_{\beta=\hat{\beta}_{t-1}} [\beta - \hat{\beta}_{t-1}] \\ &+ \frac{1}{2}[\beta - \hat{\beta}_{t-1}]^T \left(\frac{\partial^2 V_t(\beta)}{\partial \beta\, \partial \beta^T}\right)_{\beta=\hat{\beta}_{t-1}} [\beta - \hat{\beta}_{t-1}]. \end{aligned} \tag{6}$$

Minimizing the right hand side with respect to β leads to

$$\hat{\beta}_t = \hat{\beta}_{t-1} - \left(\frac{\partial^2 V_t(\beta)}{\partial \beta\, \partial \beta^T}\right)^{-1}_{\beta=\hat{\beta}_{t-1}} \left(\frac{\partial V_t(\beta)}{\partial \beta^T}\right)^T_{\beta=\hat{\beta}_{t-1}}. \tag{7}$$

Denoting $\psi_t(\beta) = -\left[\partial \varepsilon_t(\beta)/\partial \beta^T\right]^T$, the opposite of the derivative of $\varepsilon_t(\beta)$ with respect to β, we have

$$\left[\frac{\partial V_t(\beta)}{\partial \beta^T}\right]^T = -\sum_{k=1}^{t} \psi_k(\beta)\varepsilon_k(\beta) = \left[\frac{\partial V_{t-1}(\beta)}{\partial \beta^T}\right]^T - \psi_t(\beta)\varepsilon_t(\beta), \tag{8}$$

and a further differentiation yields the Hessian:

$$\frac{\partial^2 V_t(\beta)}{\partial \beta\, \partial \beta^T} = \frac{\partial^2 V_{t-1}(\beta)}{\partial \beta\, \partial \beta^T} + \psi_t(\beta)\psi_t^T(\beta) + \frac{\partial^2 \varepsilon_t(\beta)}{\partial \beta\, \partial \beta^T}\varepsilon_t(\beta). \tag{9}$$

In order to evaluate (7), the following approximations are made.

1. We assume that $\hat{\beta}_t$ is close to $\hat{\beta}_{t-1}$, a quite reasonable approximation for large t, justifying (6) and

$$\left(\frac{\partial^2 V_t(\beta)}{\partial\beta\,\partial\beta^T}\right)_{\beta=\hat{\beta}_t} \simeq \left(\frac{\partial^2 V_t(\beta)}{\partial\beta\,\partial\beta^T}\right)_{\beta=\hat{\beta}_{t-1}}. \tag{10}$$

2. We proceed as if $\hat{\beta}_{t-1}$ were optimal at time $t-1$, i.e.

$$\left(\frac{\partial\, V_{t-1}(\beta)}{\partial\beta^T}\right)_{\beta=\hat{\beta}_{t-1}} \simeq 0. \tag{11}$$

3. Since, for β close to β^*, $\{\varepsilon_t(\beta)\}$ will almost behave like a white noise process, i.e. $\varepsilon_t(\beta)$ will have a mean close to 0 and be nearly independent from the observations and residuals before time t, allowing to neglect the last term of (9).

Then, inserting (10) in (9) evaluated at $\beta = \hat{\beta}_{t-1}$, we have an approximation of the Hessian, $\overline{R}_t$, which can be computed recursively by

$$\overline{R}_t = \overline{R}_{t-1} + \psi_t(\hat{\beta}_{t-1})\psi_t^T(\hat{\beta}_{t-1}). \tag{12}$$

Insertion of (11) in (8) evaluated at $\beta = \hat{\beta}_{t-1}$, yields

$$\left(\frac{\partial\, V_t(\beta)}{\partial\beta^T}\right)^T_{\beta=\hat{\beta}_{t-1}} = -\psi_t(\hat{\beta}_{t-1})\varepsilon_t(\hat{\beta}_{t-1}).$$

Using the approximation $\overline{R}_t$ in (7), we have

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \overline{R}_t^{-1}\psi_t(\hat{\beta}_{t-1})\varepsilon_t(\hat{\beta}_{t-1}). \tag{13}$$

Denoting $tR_t = \overline{R}_t$ we have the two equations

$$\begin{cases} R_t = R_{t-1} + \frac{1}{t}\{\psi_t(\hat{\beta}_{t-1})\psi_t^T(\hat{\beta}_{t-1}) - R_{t-1}\} \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \frac{1}{t}R_t^{-1}\psi_t(\hat{\beta}_{t-1})\varepsilon_t(\hat{\beta}_{t-1}). \end{cases} \tag{14}$$

There remains to derive equations for computing $\varepsilon_t(\hat{\beta}_{t-1})$ and $\psi_t(\hat{\beta}_{t-1})$. Let us first look at the ARMA(1,1) example (3).

Example 2

We have $\psi_t^T(\beta) = \partial\hat{y}_{t|t-1}(\beta)/\partial\beta^T$ and differentiation of $\hat{y}_{t|t-1}(\beta) - \theta\hat{y}_{t-1|t-2}(\beta) = (\phi-\theta)y_{t-1}$, which is also deduced from (3), gives the two equations:

$$\frac{\partial\hat{y}_{t|t-1}(\beta)}{\partial\phi} - \theta\frac{\partial\hat{y}_{t-1|t-2}(\beta)}{\partial\phi} = y_{t-1}, \tag{15}$$

$$\frac{\partial\hat{y}_{t|t-1}(\beta)}{\partial\theta} - \hat{y}_{t-1|t-2}(\beta) - \theta\frac{\partial\hat{y}_{t-1|t-2}(\beta)}{\partial\theta} = -y_{t-1}. \tag{16}$$

The latter can also be written

$$\frac{\partial \hat{y}_{t|t-1}(\beta)}{\partial \theta} - \theta \frac{\partial \hat{y}_{t-1|t-2}(\beta)}{\partial \theta} = -\varepsilon_{t-1}(\beta). \tag{17}$$

Grouping (15) and (17) gives

$$\psi_t(\beta) - \theta \psi_{t-1}(\beta) = \begin{pmatrix} y_{t-1} \\ -\varepsilon_{t-1}(\beta) \end{pmatrix}. \tag{18}$$

We can compute $\varepsilon_t(\hat{\beta}_{t-1})$ and $\psi_t(\hat{\beta}_{t-1})$ by using equations like (4) and (18) but this requires all the observations y_s, $s = 1, ..., t-1$. Let us derive approximations of $\varepsilon_t(\hat{\beta}_{t-1})$ and $\psi_t(\hat{\beta}_{t-1})$ that can be computed by recurrence using additional approximations. A natural approximation consists in using only the current estimator and $\max(p, q)$ previous values of ε, y and ψ as initial values.

Example 3

In the case of (3), $\varepsilon_t(\hat{\beta}_{t-1})$ is approached by ε_t, computed by

$$\varepsilon_t = y_t - \hat{y}_{t|t-1} = y_t - \widehat{\phi}_{t-1} y_{t-1} + \widehat{\theta}_{t-1}(y_{t-1} - \hat{y}_{t-1|t-2}).$$

Let us introduce $\varphi_{t-1}^T = (y_{t-1}, -\varepsilon_{t-1})$. Using (4), we can write

$$\varepsilon_t = y_t - \widehat{\beta}_{t-1}^T \varphi_{t-1}. \tag{19}$$

Similarly, (18) leads to a natural approximation ψ_t of $\psi_t(\hat{\beta}_{t-1})$

$$\psi_t = \widehat{\theta}_{t-1} \psi_{t-1} + \varphi_{t-1}. \tag{20}$$

At time t we only need to know φ_{t-1}, ψ_{t-1} et $\hat{\beta}_{t-1}$. Adding these equations to those of (14) and performing substitutions, we obtain the system

$$\begin{cases} \psi_t = \hat{\theta}_{t-1} \psi_{t-1} + \varphi_{t-1}, \\ \overline{R}_t = \overline{R}_{t-1} + \psi_t \psi_t^T, \\ \varepsilon_t = y_t - \hat{\beta}_{t-1}^T \varphi_{t-1}, \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \overline{R}_t^{-1} \psi_t \varepsilon_t. \end{cases} \tag{21}$$

Let us now go back to the general case (1). To improve the behaviour of the algorithm, we replace the factor $1/t$ by a sequence γ_t of positive scalars decreasing to 0 such that $\sum \gamma_t$ is divergent. If we now denote $\varphi_t^T = (y_t, \cdots, y_{t-p+1}, \ -\varepsilon_t, \cdots, -\varepsilon_{t-q+1})$, with a due generalisation of (20), the RML algorithm can now be written:

$$\begin{cases} \psi_t = \sum_{k=1}^q \hat{\theta}_{k,t-1} \psi_{t-k} + \varphi_{t-1}, \\ R_t = R_{t-1} + \gamma_t(\psi_t \psi_t^T - R_{t-1}), \\ \varepsilon_t = y_t - \hat{\beta}_{t-1}^T \varphi_{t-1}, \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \gamma_t R_t^{-1} \psi_t \varepsilon_t. \end{cases} \tag{22}$$

3 Estimation by the RML$_{MZ}$ method

Let us now consider a modification of the method of Section 2 called the RML$_{MZ}$ method. From a theoretical point of view, under some assumptions, the RML algorithm (22) provides a consistent estimator with a rate of convergence $\sqrt{t}$. However, Mélard (1989) and Zahaf (1999) have observed huge variations of R_t with respect to time, which produce disturbances in the RML estimator. While keeping the recursive nature of the algorithm, they have tried to improve its accuracy by replacing the central recurrence (12) for the Hessian $\partial^2 V(\beta) \,/\, \partial \beta \, \partial \beta^T$, by the computation of its expectation at the current value of the estimator. Indeed, $\sigma_e^2 R_t^{-1}$ is an approximation of the asymptotic covariance matrix $\Gamma(\beta^*)$ of the maximum likelihood estimator. But, β^* being unknown, they suggest to replace $\Gamma(\beta^*)$ by the asymptotic covariance matrix evaluated at the last value of the estimator, $\Gamma(\hat{\beta}_{t-1})$. If $\hat{\beta}_t$ converges to β^*, which will be shown later, then $\Gamma(\hat{\beta}_{t-1})$ converges to $\Gamma(\beta^*)$. Moreover, $\Gamma(\hat{\beta}_{t-1})$ is the inverse $F^{-1}(\hat{\beta}_{t-1})$ of the Fisher information matrix $F(\beta)$ computed at $\beta = \hat{\beta}_{t-1}$. At each time, we will compute $\sigma_e^2 F(\hat{\beta}_{t-1})$ and then its inverse $\sigma_e^{-2} F^{-1}(\hat{\beta}_{t-1})$ which will replace R_t^{-1} in (22). For a given σ_e^2, the algorithm is written:

$$\begin{cases} \psi_t = \sum_{k=1}^{q} \hat{\theta}_{k,\,t-1} \psi_{t-k} + \varphi_{t-1}, \\ \varepsilon_t = y_t - \hat{\beta}_{t-1}^T \varphi_{t-1}, \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \gamma_t \sigma_e^{-2} \, F^{-1}(\widehat{\beta}_{t-1}) \psi_t \varepsilon_t, \end{cases} \tag{23}$$

where φ_t is like before. Therefore the recurrence for R_t in (22) will no longer be needed. Note that

$$F(\beta) = \sigma_e^{-2} E\{\psi_t(\beta) \psi_t^T(\beta)\}, \tag{24}$$

where

$$\psi_t(\beta) = \sum_{k=1}^{q} \theta_k \, \psi_{t-k}(\beta) + \varphi_t^1,$$

and $\varphi_t^1 = (y_t, \cdots, y_{t-p+1}, -e_t \cdots, -e_{t-q+1})$. Note also that $F(\beta)$ doesn't depend on t. For simple models, an analytic expression does exist for $F^{-1}(\beta)$, see Box *et al.* (1994). Otherwise, there are simple algorithms for computing $F(\beta)$, see e.g. Klein and Mélard (1989).

But σ_e^2 is generally unknown so the algorithm (23) is modified as follows

$$\psi_t = \sum_{k=1}^{q} \hat{\theta}_{k,\,t-1} \psi_{t-k} + \varphi_{t-1}, \tag{25}$$

$$\widehat{\sigma}_t^2 = \widehat{\sigma}_{t-1}^2 + \gamma_t (\varepsilon_{t-1}^2 - \widehat{\sigma}_{t-1}^2), \tag{26}$$

$$\varepsilon_t = y_t - \hat{\beta}_{t-1}^T \varphi_{t-1}, \tag{27}$$

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \gamma_t \widehat{\sigma}_t^{-2} \, F^{-1}(\widehat{\beta}_{t-1}) \psi_t \varepsilon_t. \tag{28}$$

We will now study the statistical properties of the RML$_{MZ}$ recursive estimator before discussing small sample results obtained by Monte Carlo simulation, including a comparison with the original RML algorithm, in Section 5.

3.1 Almost sure convergence

Zahaf (1999) has used results from Duflo (1997) about Robbins-Monro stochastic approximation in order to obtain asymptotic properties for a Newton approximation to the RML$_{MZ}$ estimator, called the RML$_{NE}$ estimator. The algorithm has the form $\widehat{\beta}_{t+1} = \widehat{\beta}_t + \gamma_t Y_{t+1}$, where the conditional expectation of Y_{t+1} given the past information fulfils $E[Y_{t+1}/F_t]$ is a measurable function of $\widehat{\beta}_t$. But here $E[Y_{t+1}/F_t]$ depends on both $\widehat{\beta}_t$ and t, and it is even difficult to deduce convergence of the RML$_{MZ}$ estimator from its Newton version.

The theory contained in Ljung and Söderström (1983) is based on writing the algorithm under the following form

$$\begin{cases} h_t = A(\widehat{x}_{t-1})\, h_{t-1} + B(\widehat{x}_{t-1})\, z_t, \\ \widehat{x}_t = \widehat{x}_{t-1} + \gamma_t\, Q(t, \widehat{x}_{t-1}, h_t), \end{cases} \tag{29}$$

where $A(.)$, $B(.)$, and $Q(.,.,.)$ are functions, γ_t is like in Section 2 and z_t makes use of the data. Like in Ljung (1977), the idea is to associate an ordinary differential equation (ODE) to the algorithm and obtain the attraction domain of an invariant set of that ODE. For the original RML estimator, $\widehat{x}_t = (\widehat{\beta}_t^T, \mathrm{vec}(R_t)^T)^T$ and it appears that $A(.)$ and $B(.)$ depend only on $\widehat{\beta}_t$. Here we have to consider the same but where $\widehat{x}_t = (\widehat{\beta}_t^T, \widehat{\sigma}_{t+1}^2)^T$

$$\begin{cases} h_t = A(\widehat{\beta}_{t-1})\, h_{t-1} + B(\widehat{\beta}_{t-1})\, z_t, \\ \begin{pmatrix} \widehat{\beta}_t \\ \widehat{\sigma}_{t+1}^2 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_{t-1} \\ \widehat{\sigma}_t^2 \end{pmatrix} + \frac{1}{t}\, Q(t, \widehat{x}_{t-1}, h_t), \end{cases} \tag{30}$$

where h_t is $\{q(p+q+1\} \times 1$, $Q(t,x,h)$ is $(p+q+1) \times 1$, $x = (\beta^T, \sigma^2)^T$, and

$$h_t = (\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q+1}, \psi_t^T, \psi_{t-1}^T, ..., \psi_{t-q+1}^T)^T, \quad z_t = (y_t, \cdots, y_{t-p})^T, \tag{31}$$

$$Q(t,x,h) = \left[\sigma^{-2}\left\{F^{-1}(\beta)(h_{q+1}, h_{q+2}, ..., h_{2q+p})^T\right\}^T h_1, h_1^2 - \sigma^2\right]^T, \tag{32}$$

so that h_1 represents ε_t, $(h_{q+1}, h_{q+2}, ..., h_{2q+p})^T$ represents ψ_t, and

$$Q(t, \widehat{x}_{t-1}, h_t) = \left[\widehat{\sigma}_t^{-2}\left(F_t^{-1}(\widehat{\beta}_{t-1})\psi_t\right)^T \varepsilon_t, \varepsilon_t^2 - \widehat{\sigma}_t^2\right]^T.$$

Notice that R_t, obtained by the Fisher information matrix evaluated at $\beta = \widehat{\beta}_{t-1}$, appears in the second term of the right hand side of the second equation of (30), making derivations very different from Ljung and Söderström (1983). Their theory cannot be applied directly for the RML$_{MZ}$ algorithm. However, the first equation of (30) still holds with the same choice for the matrices A and B as in (29). For an ARMA(p,q) model, it can be seen that

$$\det\left(A(\beta) - \lambda I\right) = (-1)^{\frac{q(q-1)(p+2q-1)}{2}}\left(-\left(\lambda^q - \lambda^{q-1}\theta_1 - \lambda^{q-2}\theta_2 - ... - \theta_q\right)\right)^{p+q+1}. \tag{33}$$

To show convergence of the algorithm to the optimal value, we make two assumptions; the first one is about the true value of the vector of parameters β^* and the second one is about the data.

Assumption 1: (on the model) The autoregressive and moving average polynomials have no common root and their roots are all outside of the unit circle (satisfying the causality or stationarity condition and the invertibility condition of the process).

Assumption 2: (on the data) The sequence of observations $\{ y_t \}$ has a uniform upper bound in absolute value: there exists a constant M, independent of t, such that $\forall t$,

$$|\, y_t \,| < M.$$

The latter assumption seems to be a convenient and ubiquitous assumption in this context. Despite that assumption, the proof is very technical so most of the details will be given in Appendix 1.

Let $D_S = \{(\beta^T, \sigma^2) \in \mathbb{R}^{p+q+1} \diagup$ the eigenvalues of $A(\beta)$ are in the unit circle$\}$, hence $D_S = \{(\beta^T, \sigma^2) \in \mathbb{R}^{p+q+1} \diagup$ the roots of the moving average polynomial are outside of the unit circle$\}$. Because of the Fisher information matrix, the definition of the set D_R is also different:
$D_B = \{\beta \in \mathbb{R}^{p+q} \diagup$ the roots of the autoregressive and moving average polynomials are outside of the unit circle, $F(\beta)$ is invertible, $\left\|F^{-1}(\beta)\right\| < k$ for some constant $k > 0$ large enough$\}$
$D_R = \{(\beta^T, \sigma^2) \in \mathbb{R}^{p+q+1} \diagup\ \beta^T \in D_B$ and $\sigma^2 > \delta,$ for some constant $\delta > 0$ small enough$\}$

We will make use of Theorem 1 and Theorem 4 of Ljung (1977). Here is the third subset of his conditions, denoted by **C**, without **C7** which is not needed:

C1: $Q(t, x, h)$ is Lipschitz continuous in x et h :

$$\|Q(t, x_1, h_1) - Q(t, x_2, h_2)\| < \mathcal{K}_1(x, h, \rho, \upsilon) \{\|x_1 - x_2\| + \|h_1 - h_2\|\}$$

for $x_i \in \mathcal{B}(x, \rho)$, an open ball of centre x and diameter ρ, for $\rho = \rho(x) > 0$, where $x \in D_R,\ h_i \in \mathcal{B}(h, \upsilon)$ for $\upsilon \geq 0$;

C2: Matrices $A(.)$ and $B(.)$ are Lipschitz continuous functions over D_R.

C3: $f(\overline{x}) = \lim\limits_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} Q(k, \overline{x}, \overline{h}_k(\overline{x}))$ does exist for all $\overline{x} \in D_R$.

C4: For all $\overline{x} \in D_R$, $0 < \lambda < 1$ and $c < \infty$, the random variable $k_v(t, \overline{x}, \lambda, c)$ defined by

$$\begin{aligned} k_v(t, \overline{x}, \lambda, c) = k_v(t-1, \overline{x}, \lambda, c) + \gamma_t\, [&\mathcal{K}_1(\overline{x}, h, \rho(\overline{x}), v(t, \lambda, c))(1 + v(t, \lambda, c)) \\ &- k_v(t-1, \overline{x}, \lambda, c)] \end{aligned}$$

with $k_v(0, \overline{x}, \lambda, c) = 0$ and $v(t, \lambda, c) = c \sum_{k=1}^{t} \lambda^{t-k} |z(k)|$, converges to a finite limit when $t \to \infty$.

C5: $\sum_{t=1}^{\infty} \gamma_t = \infty$;

C6: $\lim\limits_{t\to\infty} \gamma_t = 0$.

According to Ljung (1977), these conditions are used in the deterministic case, but the results are valid with probability 1 as far as z_t is such that the conditions **C3** and **C4** are satisfied with probability 1.

Theorem 1. *Under Assumptions 1 and 2, conditions* **C1-C6** *of Theorem 4 of Ljung (1977) are satisfied.*

The proof is given in Appendix 1.

According to Theorem 4 of Ljung (1977) and given (53) in the proof of Theorem 1, we have to analyse the following ODE

$$\frac{\partial x(t)}{\partial t} = \frac{\partial \left(\beta^T(t), \sigma^2(t)\right)^T}{\partial t} = H\left(\beta(t), \sigma^2(t)\right),$$

where

$$H\left(\beta, \sigma^2\right) = \left[\sigma^{-2}\{F^{-1}(\beta)E(\psi(\beta)\varepsilon(\beta))\}^T, E\{\varepsilon^2(\beta)\} - \sigma^2\right]^T.$$

Letting $f(\beta) = E\{\psi(\beta)\varepsilon(\beta)\}$ et $V(\beta) = E\{\varepsilon^2(\beta)\}$, the ODE can be put under the form

$$\frac{\partial \beta(t)}{\partial t} = \sigma^{-2}(t)F^{-1}(\beta(t))f(\beta(t)), \tag{34}$$

$$\frac{\partial \sigma^2(t)}{\partial t} = V(\beta(t)) - \sigma^2(t). \tag{35}$$

We need to check some assumptions on that differential equation. We have

$$V(\beta(t)) = E\{\varepsilon^2(\beta(t))\} \geqslant \sigma_e^2 > 0,$$

and

$$\begin{aligned}\frac{\partial V(\beta(t))}{\partial t} &= \frac{\partial V(\beta(t))}{\partial \beta^T(t)} \frac{\partial \beta(t)}{\partial t} \\ &= -2f(\beta(t))^T \sigma^{-2}(t)F^{-1}(\beta(t))f(\beta(t)) \leqslant 0,\end{aligned}$$

since $F^{-1}(\beta(t))$ is a symmetric positive definite matrix in D_R and $\sigma^{-2}(t)$ is positive. Let $\dot{V}(\beta) = \partial V(\beta)/\partial \beta^T$. We know by Lemma 1.3 (see Appendix 1) that an invariant set of the ODE is $E = \{(\beta^T, \sigma^2) \in D_R/\dot{V}(\beta) = 0\} = \left\{(\beta^T, \sigma^2) \in D_R/\ f(\beta) = 0\right\} = \{\beta^*\} \times \mathbb{R}^+$.

By Lemma 1.4 (see Appendix 1), there is a solution of the ODE (34 - 35) over some interval $[t_0, t_1]$ and, like in Ljung and Söderström (1983), we can find a part D_2 of the attraction domain for $E = \{\beta^*\} \times \mathbb{R}^+$. Let

$$c(\beta^*) = \sup_{D \in K_D} \inf_{x \in Fr(D)} E(V(x)),$$

where K_D is a set of connex parts of D_B containing β^*.

Let $D_2 = \left\{(\beta^T, \sigma^2) \in D_R/V(\beta) \leqslant c(\beta^*) - \varrho\right\}$ with a very small positive constant ϱ, and $c(\beta^*)$ is the largest possible value such that the set D_2 is the broadest set of the form $\{V(\beta) < c(\beta^*)\}$ strictly included in D_R. The set D_2 is included in the attraction domain of the invariant set $\{\beta^*\} \times \mathbb{R}^+$ because it fulfils the conditions of Lemma 1.6. Indeed, like in Lemma 1.5, let $(\beta(t_0) = \beta_0, \sigma^2(t_0) = \sigma_0^2) \in D_2$ since $V(\beta(t))$ is decreasing in t then for all $t > t_0$, $V(\beta(t)) < V(\beta(t_0)) \leqslant c(\beta^*) - \varrho$. Hence $\forall t \in [t_0, t_1]$, $(\beta(t), \sigma^2(t)) \in D_2$ and by Lemma 1.4, $\forall t > t_0$, $(\beta(t), \sigma^2(t)) \in D_2$. Then, we can apply Lemma 1.7 which summarises Theorems 1 and 4 of Ljung (1977), by letting $D_1 = D_B$. This can be summarized by the following theorem.

Theorem 2. *Under Assumptions 1 and 2, let the recursive* RML_{MZ} *estimator* (25-28) *be replaced by the following recurrences*

$$\widehat{\beta}_t = \left[\hat{\beta}_{t-1} + \frac{1}{t}\sigma_t^{-2} F^{-1}(\widehat{\beta}_{t-1})\psi_t \varepsilon_t\right]_{D_B, D_2}, \tag{36}$$

where

$$[z]_{D_B,D_2} = \begin{cases} z & \text{if } z \in D_B \\ \text{a point in } D_2 & \text{if } z \notin D_B \end{cases}$$

and

$$\begin{cases} \varepsilon_t = y_t - \hat{\beta}_{t-1}^T \varphi_{t-1}, & \\ \psi_t = \sum_{k=1}^q \hat{\theta}_{k,t-1}\psi_{t-k} + \varphi_{t-1}, & \text{if } \widehat{\beta}_{t-1} \in D_B, \\ (\varepsilon_t, \psi_t)^T \text{ a point in } K & \text{if } \widehat{\beta}_{t-1} \notin D_B, \end{cases}$$

where K is a compact subset of $\mathbb{R}^{p+q+1}$ defined in advance. Then $\widehat{\beta}_t$ converges to β^* almost surely when $t \to \infty$.

Example 4

Let the ARMA(1,1) model defined by (3). Let $\beta^* = (\phi^*, \theta^*)^T$ and assume that $\phi^* \neq \theta^*$. We know that

$$\begin{aligned} F^{-1}(\beta) &= \left\{ E\left[\frac{\partial \varepsilon_t(\beta)}{\partial \beta}\frac{\partial \varepsilon_t(\beta)}{\partial \beta^T}\right]\right\}^{-1} \\ &= \frac{1-\phi\theta}{(\phi-\theta)^2}\left[\begin{array}{cc} (1-\phi^2)(1-\phi\theta) & (1-\phi^2)(1-\theta^2) \\ (1-\phi^2)(1-\theta^2) & (1-\theta^2)(1-\phi\theta) \end{array}\right]. \end{aligned}$$

The RML_{MZ} algorithm can be written

$$\begin{cases} h_t = \begin{pmatrix} \widehat{\theta}_{t-1} & 0 & 0 \\ 0 & \widehat{\theta}_{t-1} & 0 \\ -1 & 0 & \widehat{\theta}_{t-1} \end{pmatrix} h_{t-1} + \begin{pmatrix} 1 & -\widehat{\phi}_{t-1} \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \\ \begin{pmatrix} \widehat{\beta}_t \\ \widehat{\sigma}_{t+1}^2 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_{t-1} \\ \widehat{\sigma}_t^2 \end{pmatrix} + \frac{1}{t}Q(t, \widehat{\beta}_{t-1}, \widehat{\sigma}_t^2, h_t), \end{cases}$$

with $Q(t, \widehat{\beta}_{t-1}, \widehat{\sigma}_t^{-2}, h_t) = \left(\widehat{\sigma}_t^{-2}\, \psi_t^T F^{-1}(\widehat{\beta}_{t-1})\varepsilon_t, \varepsilon_t^2 - \widehat{\sigma}_t^{-2}\right)^T$ and $h_t^T = (\varepsilon_t, \psi_t^T)$. Hence

$$A(\beta) = \begin{pmatrix} \theta & 0 & 0 \\ 0 & \theta & 0 \\ -1 & 0 & \theta \end{pmatrix}, \quad \det(A(\beta) - \lambda I) = (\theta - \lambda)^3.$$

Let $U = \{\beta = (\phi, \theta) \in]-1,1[\times]-1,1[\}$. For that model, we have
$D_S = \left\{(\beta^T, \sigma^2) \in \mathbb{R}^3 \diagup \theta \in]-1,1[\ \right\}, D_B = \left\{\beta = (\phi,\theta) \in U \diagup \left\|F^{-1}(\beta)\right\| < k\right\}$,
$D_R = \left\{\beta \in D_B,\ \sigma^2 \in \mathbb{R} \diagup \sigma^2 > \delta\right\}$,
hence $D_R \subset D_S$ and $D_R = (U \backslash \{(\phi,\theta) \in U \diagup |\phi - \theta| > \kappa\}) \times \{\sigma^2 > \delta\}$, where κ is a very small positive real number.

Let us compute $E(\varepsilon_t^2(\beta))$. We have $\varepsilon_t = \Theta_1^{-1}(B)\Phi_1(B)y_t = \Theta_1^{-1}(B)\Phi_1(B)$ $\Phi_1^{*-1}(B)\Theta_1^*(B)e_t$. Let $\phi(\omega)$ the spectral density of ε_t. We have

$$\phi(\omega) = \frac{1}{2\pi}\left|\Theta_1(e^{i\omega})\right|^{-2}\left|\Phi_1(e^{i\omega})\right|\left|\Phi_1^*(e^{i\omega})\right|^{-2}\left|\Theta_1^*(e^{i\omega})\right|$$

hence

$$\begin{aligned}
E(\varepsilon_t^2(\beta)) &= \int_{-\pi}^{\pi}\phi(\omega)d\omega \\
&= \frac{1}{2\pi i}\oint \Theta_1^{-1}(z)\Theta_1^{-1}(1/z)\Phi_1(z)\Phi_1(1/z)\Phi_1^{*-1}(z)\Phi_1^{*-1}(1/z)\Theta_1^*(z)\Theta_1^*(1/z)\frac{dz}{z} \\
&= \frac{1}{2\pi i}\oint \frac{(1-\phi z)(z-\phi)(1-\theta^* z)(z-\theta^*)}{(1-\theta z)(z-\theta)(1-\phi^* z)(z-\phi^*)}\frac{dz}{z}.
\end{aligned}$$

It is obvious that when θ comes close to 1 or -1, $E(\varepsilon_t^2(\beta))$ converges to infinity except when $\phi = \theta$. For all β such that $\phi = \theta$,

$$E(\varepsilon_t^2(\beta)) = \frac{(1-\theta^*\phi^*)(\phi^* - \theta^*)}{(1-\phi^{*2})} + \frac{\theta^*}{\phi^*},$$

and the Fisher information matrix is not invertible. Let us consider the ODE

$$\begin{aligned}
\dot{\beta} &= \sigma^{-2}(t)f(\beta) \\
\dot{\sigma}^2(t) &= V(\beta) - \sigma^2(t)
\end{aligned}$$

where

$$f(\beta) = F^{-1}(\beta)E\left[\varepsilon_t(\beta)\psi_t(\beta)\right] \text{ and } V(\beta) = E\{\varepsilon_t^2(\beta)\}.$$

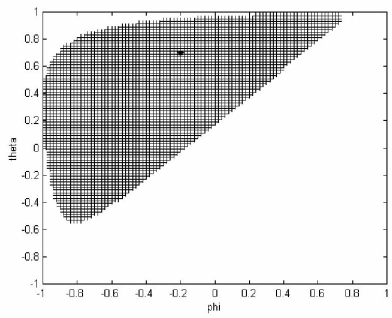
Let

$$c(\phi^*, \theta^*) = \sup_{D \in K_D}\inf_{x \in Fr(D)} E(V(x)),$$

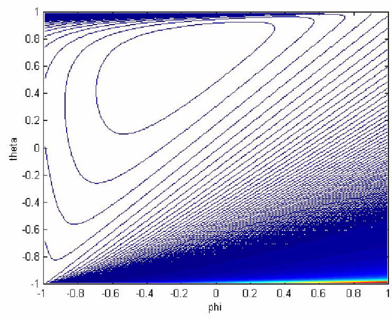
where K_D is a set of connex parts of D_B containing β^*.

Let $D_2 = \{V(\beta) \leqslant c(\beta^*) - \varrho\}$ with a very small positive ϱ. If we know a value in D_2, we can use it for estimation by the algorithm of Theorem 4 of Ljung (1977). In that case we have almost sure convergence.

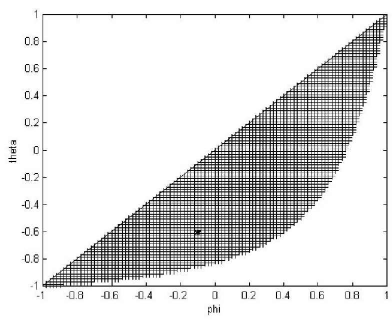
In the cases shown in Figure 1, the part D_2 where we can project the estimator to achieve convergence is the crossed surface.



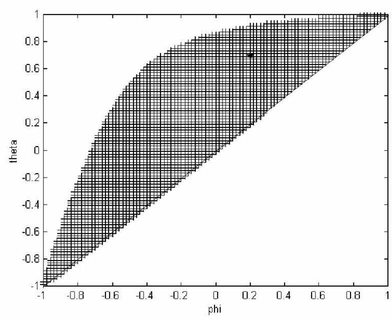
(a) $(\phi^*, \theta^*) = (-0.2, 0.7)$



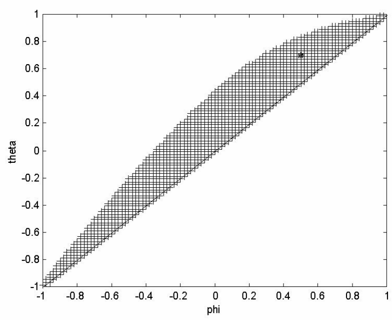
(b) $(\phi^*, \theta^*) = (-0.2, 0.7)$



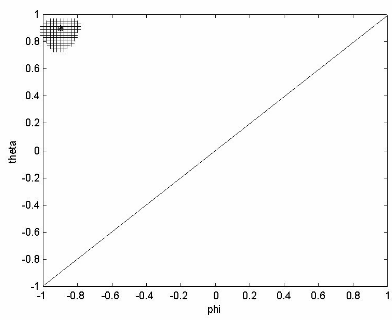
(c) $(\phi^*, \theta^*) = (-0.1, -0.6)$



(d) $(\phi^*, \theta^*) = (0.2, 0.7)$



(e) $(\phi^*, \theta^*) = (0.5, 0.7)$



(f) $(\phi^*, \theta^*) = (-0.9, 0.9)$

Figure 1. For several ARMA(1,1) processes characterized by (ϕ^*, θ^*) values of ρ, part D_2 is shown where we can project $\hat{\beta}_t$ in order to achieve convergence. For graph (b), level curves for several values of $V(\beta)$ are shown instead.

As said before, the admissible region is the square U except the diagonal joining the points $(-1,-1)$ and $(1,1)$ so it is composed of two half squares. For each of the six cases $V(\beta)$ has a unique minimum located in one of the half squares. The part D_2, shown in Figure 1 except in case (b), is always in the half square where the minimum is located. If the initial value of the ODE is in that half square, and better in D_2, the solutions will turn towards that minimum when t goes to infinity, hence also the estimator $(\widehat{\phi}_t, \widehat{\theta}_t)$. Convergence is faster in D_2. If the initial value is in the other half square, the solutions of the ODE will turn towards the frontier formed by the diagonal joining the points $(-1,-1)$ and $(1,1)$ but will stop before reaching it. Similarly, the estimator $(\widehat{\phi}_t, \widehat{\theta}_t)$ will turn towards the minimum but will have to jump over the diagonal since we may not have $\widehat{\phi}_t = \widehat{\theta}_t$, because the Fisher information matrix is not invertible there. Convergence will also be slower than in the other half square and much slower than in D_2. This is well illustrated by (b) which shows the contour levels of $V(\beta)$ for the same parameter values as (a). The level corresponding to D_2 can be seen and even smaller areas where convergence will be still faster. The other levels are higher in the upper half square and much higher in the lower half square.

Figure 1 shows that D_2 is sometimes lenticular, like in (c, d, e) but not always. Its size depends on the true values of the parameters and is smaller when they are close one from the other. Case (c) shows a situation where the point corresponding to the true values of the parameters is in the lower half square. The surface of D_2 is small when the point is close to the boundary, like in (e) and (f).

3.2 Convergence in law

Fabian (1968) has studied asymptotic normality of the algorithm

$$\widetilde{\beta}_{t+1} = (I - t^{-\alpha}\Gamma_t)\widetilde{\beta}_t + t^{-(\alpha+\delta)/2}\Phi_t V_t + t^{-\alpha-\delta/2}T_t,$$

where $\widetilde{\beta}_t = \widehat{\beta}_t - \beta^*$ in our case, Γ_t, Φ_t are matrices, V_t and T_t are vectors, by letting conditions on the components of that algorithm. He has shown that $t^{\delta/2}\widetilde{\beta}_t$ converges in law to the normal distribution. In our case, we let $\alpha = \delta = 1$, $\Gamma_t = 0$, $T_t = 0$, $\Phi_t V_t = F^{-1}(\widehat{\beta}_t)\psi_t\varepsilon_t$, but one of the conditions of Fabian (1968) is that Γ_t is definite positive. Ljung *et al.* (1992) have studied a special case of that algorithm by letting $T_t = T$, $\alpha = 1$ and $\Phi_n = I$. They have shown convergence in law under other conditions. We have tried to verify the conditions of Kushner and Huang (1979) which are more general but they are not satisfied for the RML_{MZ} estimator. Zahaf (1999) has tried to show convergence in law of the RML_{MZ} estimator by using a theorem from Duflo (1997, p. 52). In Zahaf (1999) there is an unproved conjecture which is only valid in some special cases and it is supposed that $F^{-1}(\beta)\psi_t(\beta)\psi_t^T(\beta)$ is positive definite which is not true in general despite it is a product of two positive definite matrices. Therefore, we have preferred to adapt the approach of Ljung and Söderström (1983).

Theorem 3. *Consider an ARMA model defined by (1) and the algorithm (25-28), according to the conditions of Section 3.1. Then,* $\sqrt{t}(\hat{\beta}_t - \beta^*)$ *converges in law to a normal distribution* $N(0, F^{-1}(\beta^*))$ *when* $t \to \infty$.

Proof of Theorem 3. We know that the algorithm (25-28) can be written under the form (30). We have already shown in Theorem 2 that, under some assumptions and a mechanism of projection, the estimator converges almost surely to the true value of the parameter.

Let $\varepsilon_0 = \varepsilon_0(\beta^*) = e_0 = 0$, and consider $t \geqslant 1$. Denote $\overline{\sigma}_t = t\widehat{\sigma}_t^2$. Using (26), we have $\overline{\sigma}_t = \overline{\sigma}_{t-1} + \varepsilon_{t-1}^2$. Define $\sigma_t^2(\beta^*) = \sigma_{t-1}^2(\beta^*) + \frac{1}{t}\left(\varepsilon_{t-1}^2(\beta^*) - \sigma_{t-1}^2(\beta^*)\right)$. Hence

$$\widehat{\sigma}_t^2 = \frac{1}{t}\sum_{k=1}^{t}\varepsilon_{k-1}^2, \qquad \sigma_t^2(\beta^*) = \frac{1}{t}\sum_{k=1}^{t}\varepsilon_{k-1}^2(\beta^*) = \frac{1}{t}\sum_{k=1}^{t}e_{k-1}^2 \tag{37}$$

because $\forall k > 0, \varepsilon_k(\beta^*) = e_k$. Define also

$$\psi_t(\beta^*) = \sum_{k=1}^{q}\theta_k^*\,\psi_{t-k}(\beta^*) + \varphi_{t-1}^1, \tag{38}$$

$$R_t(\beta^*) = R_{t-1}(\beta^*) + \frac{1}{t}\left(\psi_t(\beta^*)\psi_t^T(\beta^*) - R_{t-1}(\beta^*)\right) = \frac{1}{t}\sum_{k=1}^{t}\psi_k(\beta^*)\psi_k^T(\beta^*), \tag{39}$$

where $\varphi_{t-1}^1 = (y_{t-1}, \cdots, y_{t-p}, -e_{t-1} \cdots, -e_{t-q})^T$. Denote

$$\overline{R}_t(\beta^*) = tR_t(\beta^*) = \overline{R}_{t-1}(\beta^*) + \psi_t(\beta^*)\psi_t^T(\beta^*), \tag{40}$$

and let $\widetilde{\beta}_t = \hat{\beta}_t - \beta^*$. From (28), we have

$$\widetilde{\beta}_t = \widetilde{\beta}_{t-1} + \frac{1}{t}\widehat{\sigma}_t^{-2}\,F^{-1}(\widehat{\beta}_{t-1})\psi_t\varepsilon_t. \tag{41}$$

According to Lemma 2.1, $K_t = \overline{\sigma}_t\,\widetilde{\beta}_t$ can be decomposed in a sum of terms. Using that decomposition, we need Lemma 2.12 to show that $\forall\delta > 0, t^{1/2-\delta}\left\|\widetilde{\beta}_t\right\| \to 0$ a.s. when $t \to \infty$. The proof of that Lemma 2.12 makes use of Lemmas 2.2-2.11. We will use that result in Lemmas 2.13 and 2.16.

From (40) and (41), we can write

$$\begin{aligned}\overline{R}_t(\beta^*)\widetilde{\beta}_t &= \overline{R}_t(\beta^*)\widetilde{\beta}_{t-1} + \frac{1}{t}\overline{R}_t(\beta^*)\widehat{\sigma}_t^{-2}\,F^{-1}(\widehat{\beta}_{t-1})\psi_t\varepsilon_t,\\ &= \overline{R}_{t-1}(\beta^*)\widetilde{\beta}_{t-1} + \psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} + R_t(\beta^*)\widehat{\sigma}_t^{-2}\,F^{-1}(\widehat{\beta}_{t-1})\psi_t\varepsilon_t.\end{aligned}$$

But $\psi_t\varepsilon_t$ is equal to

$$\psi_t\left(\varepsilon_t - \varepsilon_t(\hat{\beta}_{t-1})\right) + \left(\psi_t - \psi_t(\hat{\beta}_{t-1})\right)\varepsilon_t(\hat{\beta}_{t-1}) + \psi_t(\hat{\beta}_{t-1})(\varepsilon_t(\hat{\beta}_{t-1}) - e_t) + \psi_t(\hat{\beta}_{t-1})e_t,$$

and, using a Taylor expansion,

$$\varepsilon_t(\hat{\beta}_{t-1}) - e_t = -\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} - \frac{1}{2}\widetilde{\beta}_{t-1}^T\left(\frac{\partial\,\psi_t(\beta)}{\partial\beta^T}\right)_{\beta=\varkappa_t}\widetilde{\beta}_{t-1}, \tag{42}$$

where $\varkappa_t$ is a point between $\hat{\beta}_{t-1}$ and β^*. Letting $U_t = R_t(\beta^*)\widehat{\sigma}_t^{-2}\, F^{-1}(\widehat{\beta}_{t-1})$, we have

$$\begin{aligned}\overline{R}_t(\beta^*)\widetilde{\beta}_t &= \overline{R}_{t-1}(\beta^*)\widetilde{\beta}_{t-1} + \psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} - U_t\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1}\\ &\quad - U_t\left(\psi_t(\hat{\beta}_{t-1}) - \psi_t(\beta^*)\right)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} - \frac{1}{2}U_t\psi_t(\hat{\beta}_{t-1})\widetilde{\beta}_{t-1}^T\left(\frac{\partial\,\psi_t(\beta)}{\partial\beta^T}\right)_{\beta=\varkappa_t}\widetilde{\beta}_{t-1}\\ &\quad + U_t\left(\psi_t - \psi_t(\hat{\beta}_{t-1})\right)\varepsilon_t(\hat{\beta}_{t-1}) + U_t\psi_t\left(\varepsilon_t - \varepsilon_t(\hat{\beta}_{t-1})\right) + U_t\psi_t(\hat{\beta}_{t-1})e_t.\end{aligned}$$

We can write

$$\begin{aligned}(I_{p+q} - U_t)\,\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} &= \left(\widehat{\sigma}_t^2\, F(\beta^*) - R_t(\beta^*)\right)\widehat{\sigma}_t^{-2}\, F^{-1}(\widehat{\beta}_{t-1})\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1}\\ &\quad + \left(F(\widehat{\beta}_{t-1})\ - F(\beta^*)\right)F^{-1}(\widehat{\beta}_{t-1})\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1}.\end{aligned}$$

Letting

$$\begin{aligned}B_{1,t} &= \left(F(\widehat{\beta}_{t-1})\ - F(\beta^*)\right)\ F^{-1}(\widehat{\beta}_{t-1})\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1}\\ &\quad - U_t\left(\psi_t(\hat{\beta}_{t-1}) - \psi_t(\beta^*)\right)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} - \frac{1}{2}U_t\psi_t(\hat{\beta}_{t-1})\widetilde{\beta}_{t-1}^T\left(\frac{\partial\,\psi_t(\beta)}{\partial\beta^T}\right)_{\beta=\eta_t}\widetilde{\beta}_{t-1}.\end{aligned} \tag{43}$$

and

$$B_{2,t} = U_t\left(\psi_t - \psi_t(\hat{\beta}_{t-1})\right)\varepsilon_t(\hat{\beta}_{t-1}) + U_t\psi_t\left(\varepsilon_t - \varepsilon_t(\hat{\beta}_{t-1})\right), \tag{44}$$

we have

$$\begin{aligned}\overline{R}_t(\beta^*)\widetilde{\beta}_t &= \overline{R}_{t-1}(\beta^*)\widetilde{\beta}_{t-1} + \left(\widehat{\sigma}_t^2\, F(\beta^*) - R_t(\beta^*)\right)\widehat{\sigma}_k^{-2}\, F^{-1}(\widehat{\beta}_{t-1})\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1}\\ &\quad + B_{1,t} + B_{2,t} + U_t\psi_t(\hat{\beta}_{t-1})e_t,\end{aligned}$$

hence

$$\begin{aligned}tR_t(\beta^*)\widetilde{\beta}_t &= \sum_{k=1}^{t}\left(\widehat{\sigma}_t^2\, F(\beta^*) - R_k(\beta^*)\right)\widehat{\sigma}_k^{-2}\, F^{-1}(\widehat{\beta}_{k-1})\psi_k(\beta^*)\psi_k^T(\beta^*)\widetilde{\beta}_{k-1} + \sum_{k=1}^{t}B_{1,k}\\ &\quad + \sum_{k=1}^{t}B_{2,k} + \sum_{k=1}^{t}U_k\psi_k(\hat{\beta}_{k-1})e_k.\end{aligned} \tag{45}$$

Hence

$$\sqrt{t}\widetilde{\beta}_t = H_t + L_t + R_t^{-1}(\beta^*)\frac{1}{\sqrt{t}}\sum_{k=1}^{t}R_k(\beta^*)\widehat{\sigma}_k^{-2}\, F^{-1}(\beta^*)\psi_k(\beta^*)e_k, \tag{46}$$

where

$$H_t = R_t^{-1}(\beta^*)\frac{1}{\sqrt{t}}\sum_{k=1}^{t}\left(\sigma_e^2\ F(\beta^*) - R_k(\beta^*)\right)\widehat{\sigma}_k^{-2}\ \ F^{-1}(\widehat{\beta}_{k-1})\psi_k(\beta^*)\psi_k(\beta^*)\widetilde{\beta}_{k-1}$$
$$+ R_t^{-1}(\beta^*)\frac{1}{\sqrt{t}}\sum_{k=1}^{t}\left(\widehat{\sigma}_k^2\ - \sigma_e^2\right) F(\beta^*)\widehat{\sigma}_k^{-2}\ \ F^{-1}(\widehat{\beta}_{k-1})\psi_k(\beta^*)\psi_k(\beta^*)\widetilde{\beta}_{k-1}$$
$$+ R_t^{-1}(\beta^*)\frac{1}{\sqrt{t}}\sum_{k=1}^{t}B_{1,k} + R_t^{-1}(\beta^*)\frac{1}{\sqrt{t}}\sum_{k=1}^{t}B_{2,k}, \tag{47}$$

and

$$L_t = R_t^{-1}(\beta^*)\frac{1}{\sqrt{t}}\sum_{k=1}^{t}R_k(\beta^*)\widehat{\sigma}_k^{-2}\{\ F^{-1}(\widehat{\beta}_{k-1}) - F^{-1}(\beta^*)\}\psi_k(\hat{\beta}_{k-1})e_k$$
$$+ R_t^{-1}(\beta^*)\frac{1}{\sqrt{t}}\sum_{k=1}^{t}R_k(\beta^*)\widehat{\sigma}_k^{-2}\ F^{-1}(\beta^*)\left(\psi_k(\hat{\beta}_{k-1}) - \psi_k(\beta^*)\right)e_k. \tag{48}$$

In Lemma 2.16 we show convergence a. s. to 0 of H_t and L_t.

From (46) and according to Lemmas 2.15 and 2.16, based on Lemmas 2.13 and 2.14, we have that $\sqrt{t}\widetilde{\beta}_t$ converges in law to a normal distribution with mean 0 and variance

$$\begin{aligned} V &= E\left(\psi_1(\beta^*)\psi_1^T(\beta^*)\right)^{-1}\sigma_e^4 F(\beta^*) E\left(\psi_1(\beta^*)\psi_1^T(\beta^*)\right)^{-1} \\ &= \sigma_e^{-2}F^{-1}(\beta^*)\sigma_e^4 F(\beta^*)\sigma_e^{-2}F^{-1}(\beta^*) = F^{-1}(\beta^*), \end{aligned}$$

when $t \to \infty$ since $R_t(\beta^*)$ converges a.s. to $E\left(\psi_1(\beta^*)\psi_1^T(\beta^*)\right)$ which is equal to $\sigma_e^2 F(\beta^*)$ by (24) .

4 Finite sample properties

4.1 Ljung's Toolbox in Matlab

We will compare the results of our algorithm with Ljung (2000) System Identification Toolbox in Matlab version 5.0 (R12), and more specifically function RPEM, using the adm='ff' parameter, i.e. the forgetting factor algorithm which makes use of the algorithm (22). At each iteration, only the elements ε_t, $\psi_t(1)$ and $\psi_t(p_1+1)$ of ψ_t are computed

$$\psi_t(1) = \sum_{k=1}^{q} c_{k\,t}\psi_t(1+k) + y_{t-1}, \tag{49}$$
$$\psi_t(p_1+1) = \sum_{k=1}^{q} c_{k\,t}\psi_t(p_1+1+k) + \varepsilon_{t-1},$$

with $p_1 = \max(p, q)$. After that, a sliding is performed by

$$\psi_{t+1}(2) = \psi_t(1), ..., \psi_{t+1}(p_1) = \psi_t(p_1 - 1),$$
$$\psi_{t+1}(p_1 + 2) = \psi_t(p_1 + 1), ..., \psi_{t+1}(p_1 + q) = \psi_t(p_1 + q - 1).$$

The method in RPEM makes use of a projection using the function 'FSTAB' in Matlab but only for the parameters of the moving average polynomial $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ as follows. Let $\widehat{\theta}_t(B) = 1 - \widehat{\theta}_{1,t} B + ... + \widehat{\theta}_{q,t} B^q$ be the polynomial estimated at time t. The roots should be inside of the unit circle. Therefore, those roots which are outside of the unit circle are inverted and the others are unchanged. Then, the polynomial is computed again. We will not use that procedure in our method.

4.2 Implementation of the RML$_{MZ}$ method

Besides omitting the recurrence for the Hessian, our implementation of the RML$_{MZ}$ estimator for ARMA models is different from that of Ljung in Matlab System Identification toolbox. We will discuss below the effective recurrences used in practice, projection of the parameters, the choice of initial values, the use of forgetting factors, and some information about the program.

Like Ljung, we introduce a second estimate of the forecast error $\overline{\varepsilon}_t$, so that the algorithm (23) becomes

$$\psi_t = \sum_{k=1}^{q} \hat{\theta}_{k,t-1} \psi_{t-k} + \overline{\varphi}_{t-1}, \tag{50}$$

$$\varepsilon_t = y_t - \hat{\beta}_{t-1}^T \overline{\varphi}_{t-1}, \tag{51}$$

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \gamma_t F^{-1}(\hat{\beta}_{t-1}) \psi_t \varepsilon_t, \tag{52}$$

$$\overline{\varepsilon}_t = y_t - \hat{\beta}_t^T \overline{\varphi}_{t-1},$$

where this time $\overline{\varphi}_t^T = (y_t, \cdots, y_{t-p+1}, -\overline{\varepsilon}_t, \cdots, -\overline{\varepsilon}_{t-q+1})$. Note that we don't have made use of the suggestion of Ljung for sliding a reduced subset of ψ_t's.

4.3 Admissibility of β_t

At each time t, we have to check that the estimator $\widehat{\beta}_t$ satisfies conditions analog to those assumed in Section 3. The most important is that the roots of both the AR and MA polynomials be outside of the unit circle. This is done by projecting $\widehat{\beta}_t$ in the unit disk, contrarily to RPEM where only the MA polynomial is treated. As a matter of fact, our implementation offers also the choice of the Ljung and Söderström (1983) projection procedure.

Let us illustrate the case of an AR polynomial. If $\left(\widehat{\phi}_{t,1}, ..., \widehat{\phi}_{t,p}\right)$ is not admissible, let $\rho < 1$, consider instead $\left(\rho\widehat{\phi}_{t,1}, \rho^2\widehat{\phi}_{t,2}, ..., \rho^p\widehat{\phi}_{t,p}\right)$ and iterate until the subset of parameters becomes admissible. That way the roots of $1 + \widehat{\phi}_{t,1} B +$

... + $\widehat{\phi}_{t,p}B$ don't come close to the unit circle like with the procedure of Ljung and Söderström (1983).

4.4 Initial values

To obtain good estimates, starting with appropriate initial values $\widehat{\beta}_0$ is essential. In our RML_{MZ} method, besides satisfying the causality and invertibility conditions, $\widehat{\beta}_0$ should be far enough from the region where the Fisher information matrix is not invertible. For the RML method, an initial matrix R_0 is also needed and Ljung and Söderström (1983) recommend to use $R_0 = 10000\ I$, expressing thereby a large amount of uncertainty. Here, we have only to choose σ_0^2. We have taken $\widehat{\sigma}_0^2 = 10$ or $\widehat{\sigma}_0^2 = 10000$.

4.5 Forgetting factor

We have used a factor γ_t in (22) or (25-28) although this was often taken as $1/t$ in the theory. In practice it should be selected in order to improve convergence. It is often based on the forgetting factor defined by

$$\lambda_t = \frac{\gamma_{t-1}(1-\gamma_t)}{\gamma_t},$$

which corresponds to

$$\gamma_t = \frac{\gamma_{t-1}}{\lambda_t + \gamma_{t-1}}.$$

According to Ljung (1985), using $\gamma_t = 1/t$ (which corresponds to $\lambda_t = 1$, $\forall t$) is justified when the coefficients do not vary with time, which is the case here. More generally, it is recommended to use

$$\lambda_t = \lambda^0 \lambda_{t-1} + (1-\lambda^0),$$

where typically $\lambda^0 \approx 0.95$ or $\lambda^0 \approx 0.99$. Remark that λ_t converges to $\lambda_\infty = 1$ and γ_t converges to $\gamma_\infty = 0$. We have also experimented with a constant forgetting factor.

We will compare our estimator (solid line) with the RML estimator of Ljung as implemented in Matlab (dashed line with a forgetting factor $\lambda = 1$, or dot-dashed line with $\lambda = 0.99$). For our RML_{MZ} algorithm, we have used a different forgetting factor for the variance, denoted with a subscript σ, characterised by $\gamma_{0\sigma} = 1$ or $\lambda_{t\sigma} = 0.9$ with $\lambda^0 = 1$.

4.6 Information about the program

A computer program in Fortran 90 was written in order to experiment with the new method. The computer program is a part of a bigger project described in Ouakasse and Mélard (2005). Indeed the program is able to handle general single input single output (SISO) models. Moreover each polynomial can be factored in a non seasonal polynomial and a seasonal polynomial, a feature which is necessary when dealing with economic or traffic data. These aspects, as well as specific procedures in order to improve the computational efficiency of the method, both in the non-seasonal and seasonal cases, are discussed by Ouakasse *et al.* (2005).

Let us now describe the experiments that follow. Artificial time series were produced in Matlab using simple recurrences and omitting the first 50 observations. They were immediately treated with the RPEM procedure. The series were then exported and treated by the Fortran program.

4.7 ARMA(1,1) model

Let us consider the ARMA(1,1) model with equation (3) with $\phi^* = 0.5$ and $\theta^* = -0.5$, with $\sigma^2 = 1$. We have generated 10000 series of length 1000 for which we have computed the estimates of ϕ and θ, for each time $t = 1, ..., 1000$. The following initial values were used: $\widehat{\sigma}_0^2 = 10$, $\widehat{\phi}_0 = 0.25$, $\widehat{\theta}_0 = -0.25$, $\lambda_0 = 1$, $\gamma_0 = 1$, $\lambda_{0\sigma} = 1$, $\gamma_{0\sigma} = 1$. The averages and standard deviations across the experiments are shown in function of time. For each plot, the true value of the parameter is given. It is even displayed in the plot of the averages as a dotted horizontal line. The averages should be as close as possible of the true value and the standard deviations should be close to 0.

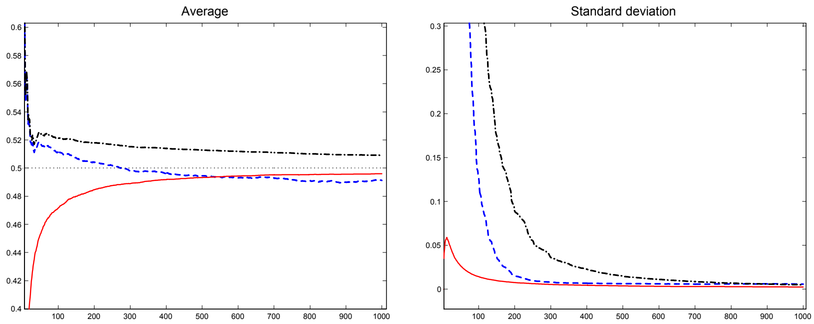


Figure 2. ARMA(1,1) with $\phi^* = 0.5$. Averages (left) and standard deviations (right) over the simulations in function of time for three estimates of ϕ. Solid line: our estimator, dashed line: Ljung/Matlab with $\lambda = 1$, dot-dashed line: same with $\lambda = 0.99$.

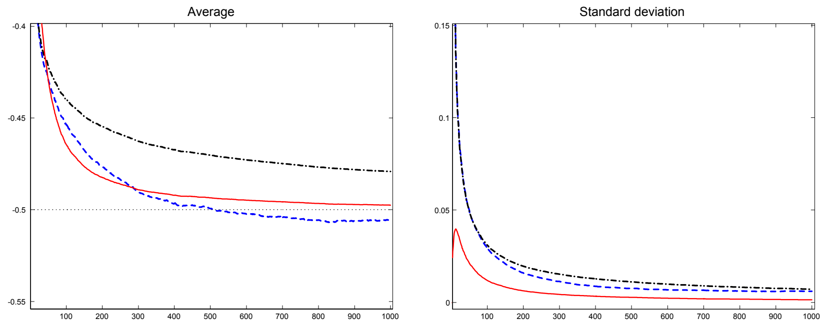


Figure 3. ARMA(1,1) with $\theta^* = -0.5$. Averages (left) and standard deviations (right) over the simulations in function of time for three estimates of θ. Solid line: our estimator, dashed line: Ljung/Matlab with $\lambda = 1$, dot-dashed line: same with $\lambda = 0.99$.

The plots for averages indicate that the new estimator seems to converge faster than the RML estimator. On the plot for standard deviations, we observe that those of the RML estimator decrease more slowly than ours.

4.8 ARMA(2, 2) model

Let us consider the ARMA(2, 2) model with equation

$$(1 + 0.8B + 0.25B^2)y_t = (1 + 1.378B + 0.5B^2)e_t,$$

with $\sigma^2 = 1$. We have generated 10000 series of length 1000 for which we have computed the the estimates of ϕ_1, ϕ_2 and θ_1, θ_2 for each time $t = 1, ..., 1000$. The following initial values were used: $\widehat{\sigma}_0^2 = 10000$, $\widehat{\phi}_{1,0} = 0.5$, $\widehat{\phi}_{2,0} = 0.8$, $\widehat{\theta}_{1,0} = 0.69$, $\widehat{\theta}_{2,0} = 0.14$, $\lambda_0 = 1$, $\gamma_0 = 1$, $\lambda_{0\sigma} = 0.9$, $\gamma_{0\sigma} = 1$. Here are the results that were obtained, presented like for the previous example.

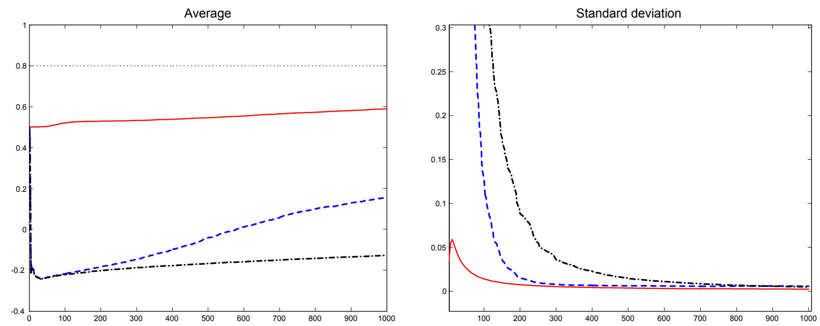


Figure 4. ARMA(2,2) with $\phi_1^* = 0.8$. Averages (left) and standard deviations (right) over the simulations in function of time for three estimates of ϕ_1. Solid line: our estimator, dashed line: Ljung/Matlab with $\lambda = 1$, dot-dashed line: same with $\lambda = 0.99$.

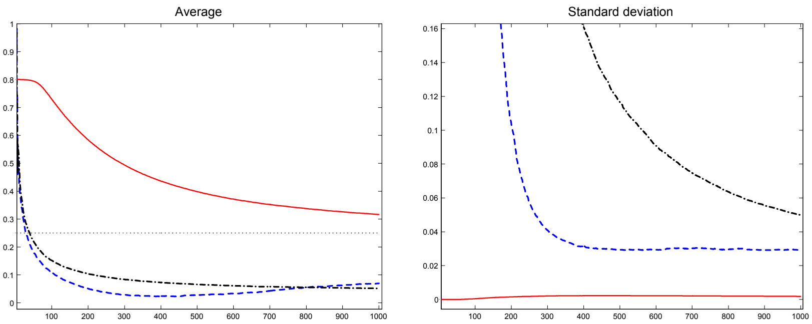


Figure 5. ARMA(2,2) with $\phi_2^* = 0.25$. Averages (left) and standard deviations (right) over the simulations in function of time for three estimates of ϕ_2. Solid line: our estimator, dashed line: Ljung/Matlab with $\lambda = 1$, dot-dashed line: same with $\lambda = 0.99$.

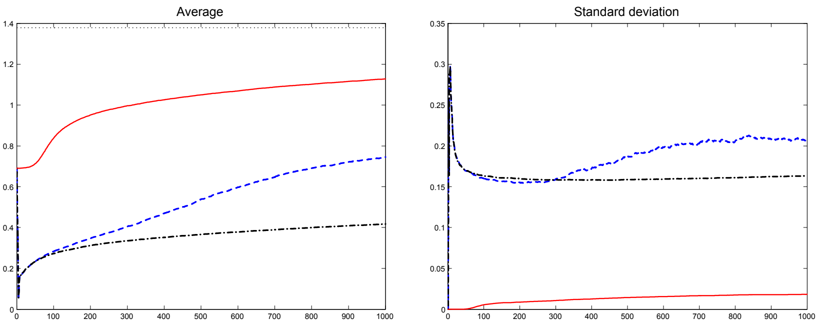


Figure 6. ARMA(2,2) with $\theta_1^* = 1.375$. Averages (left) and standard deviations (right) over the simulations in function of time for three estimates of θ_1. Solid line: our estimator, dashed line: Ljung/Matlab with $\lambda = 1$, dot-dashed line: same with $\lambda = 0.99$.

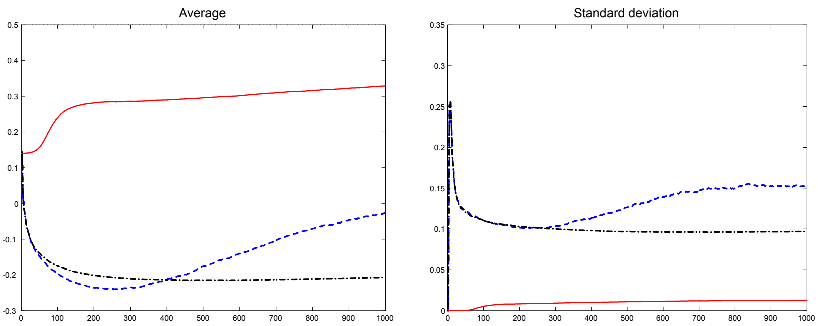


Figure 7. ARMA(2,2) with $\theta_2^* = 0.5$. Averages (left) and standard deviations (right) over the simulations in function of time for three estimates of θ_2. Solid line: our estimator, dashed line: Ljung/Matlab with $\lambda = 1$, dot-dashed line: same with $\lambda = 0.99$.

The graphs show that the averages for our method converge faster than for the RML estimator except for the parameter ϕ_2, and also that the dispersion across simulations is smaller.

5 An example

We will illustrate the procedure on the following example. Windmills produce electricity in a way which is cleaner for the environment than with thermal or nuclear power stations. Electricity is however irregular because it depends of wind irregularity. When the wind is strong, more electricity is produced. Conversely, when the wind is weak, the quantity of electricity is very small. In order to maintain the offer of electricity at the level of demand, it is required to adapt production from traditional power stations in function of the amount of electricity produced by a park of windmills. Response time of a power station can go from a few minutes to several hours according to the technology being used. It is therefore useful to forecast wind speed a few hours in advance. The data come from speed of wind measurements at the top of a windmill. They are available every ten minutes, hence 144 observations per day. We have used about twelve days of measurements, more precisely 1728 observations. The data are shown in Figure 8.

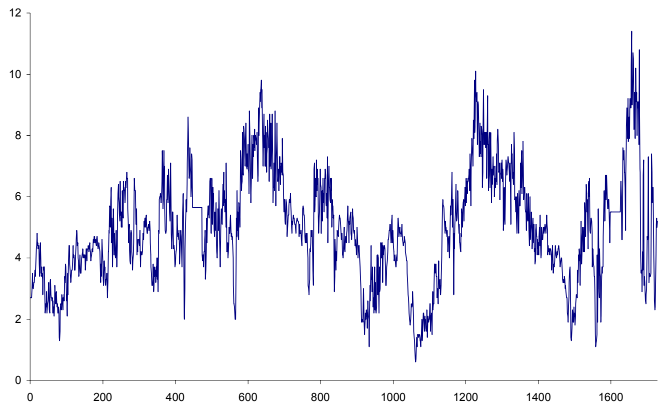


Figure 8. Speed of wind on the top of a windmill. One observation every ten minutes during ten days.

We have specified an ARMA(1,2) model with a constant, described by the equation:

$$(1 - \phi_1 B)(y_t - \mu) = (1 - \theta_1 B - \theta_2 B^2)e_t.$$

Here the vector of parameters is composed of $\beta = (\phi_1, \theta_1, \theta_2, \mu)$. A statistician or an econometrician would probably select a model with a unit root. For that reason, we have used both forgetting factors equal to 1 and an initial value of the variance which is not too large, equal to 500. The estimates are shown in Figure 9.

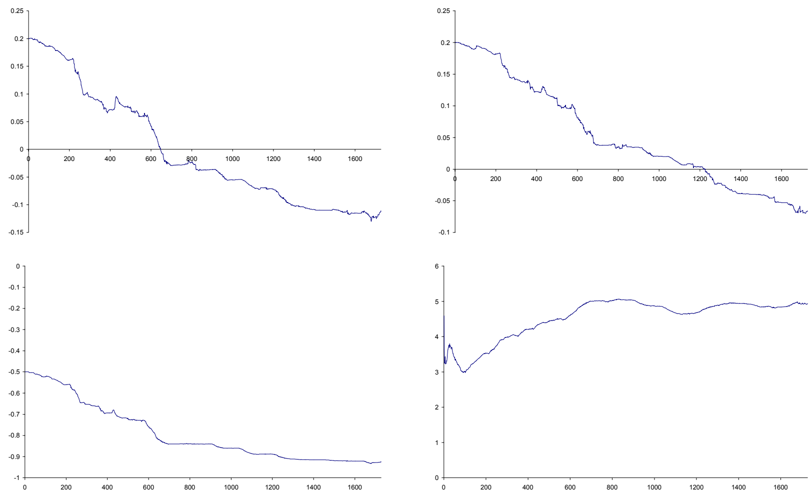


Figure 9. Estimates by the RML_{MZ} method in function of time: ϕ_1 (top left), θ_1 (top right), θ_2 (bottom left), μ (bottom right).

The estimates at the end of the series are $\hat{\beta} = (0.925, 0.112, 0.066, 4.933)$ and the final value of the innovation variance is 0.739. Note that the exact maximum likelihood method (Mélard, 1984) of SPSS (2004) gives the following model using the whole data set:

$$(1 - 0.976B)(y_t - 4.894) = (1 - 0.216B - 0.196B^2)e_t,$$

with an estimate of the innovation variance equal to 0.400.

6 Conclusion

In Section 2, we have recalled the RML method proposed by Ljung (1977) and Ljung and Söderström (1983). That method provides recursive estimates using a system of equations. In one of the equations, the Hessian matrix of the error is updated. An improved RML method called RML_{MZ} is the subject of the present paper. It has been described in Section 3. It is based on using the Fisher information matrix, evaluated at the current value of the estimator, in order to update the estimator, instead of updating the Hessian. The asymptotic statistical properties of the new method have been studied in Subsection 3.1 and 3.2. Under fairly general assumptions, it was proved that the RML_{MZ} estimator is consistent in the almost sure sense and also asymptotically normally distributed. This is done by following Ljung (1977) but the details are very different from those of the Ljung et Söderström (1983) approach. It is based on a result that the mathematical expectation of the errors, $E(\varepsilon^2)$, has an absolute minimum obtained at the true value of the parameter. We have obtained a part of the attraction domain around that minimum of the differential equation associated to the algorithm. Convergence in law makes use of a result of Hannan

(1976) which allows to adapt the Ljung and Söderström (1983) approach to the RML_{MZ} estimator. In Section 4 we have shown Monte Carlo simulations (for some ARMA models and using 10000 series of length 1000) for the comparison between the RML_{MZ} estimator and the original RML method. This suggests that indeed the RML_{MZ} estimator often does converge more quickly in practice. In Section 5 we show an example on real data which shows the usefulness of the new method.

In Ouakasse and Mélard (2005), we will present an extension of these recursive method based on the Fisher information matrix to a wider range of models: single input, single output (SISO) models. In Ouakasse *et al.* (2005), we intend to show how these methods can be adapted to seasonal data in order to cope with the so-called seasonal models of Box *et al.* (1994).

Appendix 1

Here are a few lemmas needed for the proofs in Section 3.1. Invertibility of the Fisher information matrix is satisfied by Assumption 1, given the following lemma.

Lemma 1.1 (Klein and Spreij, 1993). *The Fisher information matrix $F(\beta)$ is invertible if and only if the autoregressive and moving average polynomials have no common root.*

Lemma 1.2 (Harville, 1997, p. 307). *Let F be a matrix function of $\mathbb{R}^m \to \mathbb{R}^{(n,n)}$, let $x \in \mathbb{R}^m$ be a point where matrix F is invertible and continuously differentiable, then*

$$\frac{\partial F^{-1}(x)}{\partial x_i} = -F^{-1}(x)\frac{\partial F(x)}{\partial x_i}F^{-1}(x)$$

Proof of Theorem 1.

We have to prove the conditions **C** for the algorithm (30). Condition **C2** on matrices A and B which are the same as in the RML method is of course valid. Condition **C4** is essentially the same as for the RML method but Ouakasse (2004, pp. 45-46) provides an alternative proof which is more direct than in Ljung and Söderström (1983, p. 175-176). Of course conditions **C5** and **C6** are satisfied since $\gamma_t = 1/t$.

Let us first check condition **C1**.

Suppose $(\beta_1^T, \sigma_1^2)^T$ and $(\beta_2^T, \sigma_2^2)^T$ in a ball $\mathcal{B}\left((\beta^T, \sigma^2)^T, \rho(\beta, \sigma^2)\right)$ with $\rho(\beta, \sigma^2)$ small enough such that (β_1, σ_1^2) and (β_2, σ_2^2) belong to D_R. Let h and h' be two vectors in a ball $\mathcal{B}\left(h^0, v\right)$ of $\mathbb{R}^{q(p+q+1)}$ with $h^0 = (h_1^0, h_2^0, ..., h_{q(p+q+1)}^0)^T$, $h = (h_1, h_2, ..., h_{q(p+q+1)})^T$ and $h' = (h'_1, h'_2, ..., h'_{q(p+q+1)})^T$. Let $k_1^0 = (h_{q+1}^0, h_{q+2}^0, ..., h_{p+2q}^0)$, $k_1 = (h_{q+1}, h_{q+2}, ..., h_{p+2q})$ and $k'_1 = (h'_{q+1}, h'_{q+2}, ..., h'_{p+2q})$. By (32) we have

$$
\begin{aligned}
&Q(t,\beta_1,\sigma_1^2,h)-Q(t,\beta_2,\sigma_2^2,h')\\
&=\left[\left\{\left(h_1-h_1^0\right)\left(k_1^T-k_1'^T\right)+\left(h_1-h_1'\right)\left(k_1'^T-k_1^{0T}\right)\right\}\sigma_1^{-2}F^{-1}(\beta_1)\right.\\
&+\left\{h_1^0\left(k_1^T-k_1'^T\right)+\left(h_1-h_1'\right)k_1^{0T}\right\}\sigma_1^{-2}F^{-1}(\beta_1)\\
&+\left(h_1'-h_1^0\right)\left(k_1'^T-k_1^{0T}\right)\sigma_1^{-2}\left(F^{-1}(\beta_1)-F^{-1}(\beta_2)\right)\\
&+h_1^0\left(k_1'^T-k_1^{0T}\right)\sigma_1^{-2}\left(F^{-1}(\beta_1)-F^{-1}(\beta_2)\right)\\
&+\left(\left(h_1'-h_1^0\right)k_1^{0T}+h_1^0k_1^{0T}\right)\sigma_1^{-2}\left(F^{-1}(\beta_1)-F^{-1}(\beta_2)\right)\\
&+\left(h_1'-h_1^0\right)\left(k_1'^T-k_1^{0T}\right)\left(\sigma_2^2-\sigma_1^2\right)\sigma_1^{-2}\sigma_2^{-2}F^{-1}(\beta_2)\\
&+h_1^0\left(k_1'^T-k_1^{0T}\right)\left(\sigma_2^2-\sigma_1^2\right)\sigma_1^{-2}\sigma_2^{-2}F^{-1}(\beta_2)\\
&+\left(\left(h_1'-h_1^0\right)k_1^{0T}+h_1^0k_1^{0T}\right)\left(\sigma_2^2-\sigma_1^2\right)\sigma_1^{-2}\sigma_2^{-2}F^{-1}(\beta_2),\\
&\left.\left(h_1-h_1^{'}\right)\left(h_1-h^0+h_1^{'}-h^0\right)+2h^0\left(h_1-h_1^{'}\right)+\sigma_2^2-\sigma_1^2\right]^T.
\end{aligned}
$$

Since $\sigma_1^{-2}\leqslant\delta^{-1}$, $\sigma_2^{-2}\leqslant\delta^{-1}$, $\left(h_1-h_1^0\right)\leqslant v$ and $\left(k_1-k_1^0\right)\leqslant v$, there exists a constant C such that

$$
\begin{aligned}
\left\|Q(t,\beta_1,\sigma_1^2,h)-Q(t,\beta_2,\sigma_2^2,h')\right\|\leqslant C\left(\left\|h^0\right\|^2+\left\|h^0\right\|+v^2+v\right)\{\|h-h'\|\\
+\left\|F^{-1}(\beta_1)-F^{-1}(\beta_2)\right\|+\left|\sigma_2^2-\sigma_1^2\right|\}.
\end{aligned}
$$

According to Lemma 1.2, $\partial F^{-1}(\beta)/\partial\beta$ is continuous on the ball $\mathcal{B}\left((\beta^T,\sigma^2)^T,\ \rho(\beta,\sigma^2)\right)$ hence it is bounded, then there exists a constant $C_1>0$ such that $\left\|F^{-1}(\beta_1)\ -F^{-1}(\beta_2)\right\|\leqslant C_1\left\|\beta_1-\beta_2\right\|$, and the preceding expression can be written

$$
\begin{aligned}
\left\|Q(t,\beta_1,\sigma_1^2,h)-Q(t,\beta_2,\sigma_2^2,h')\right\|\leqslant C\left(\left\|h^0\right\|^2+\left\|h^0\right\|+v^2+v+C_1\right)\\
\left(\left\|(\beta_1^T,\sigma_1^2)^T-(\beta_2^T,\sigma_2^2)^T\right\|+\left\|h-h^{'}\right\|\right).
\end{aligned}
$$

Let us now check **C3**. Condition **C3** is basically the same as for the RML method, see Ljung and Söderström (1983, p. 169-170). We have to show that:

A) For all $t,s,\ t\geqslant s$, there exists a random vector $z_s^0(t)$ belonging to the σ-algebra spanned by the $z_i,\ i\leqslant t$, and independent of z_s, such that $E\left\|z_t-z_s^0(t)\right\|^4<C\lambda^{t-s}$, $C<\infty$, $\lambda<1$.

B) The following limit does exist:

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}E\left(Q\left(t,\beta,\sigma^2,h_t(\beta)\right)\right)=H(\beta,\sigma^2).
$$

For the RML case, Ouakasse (2004, pp. 46-49) proves Part A and the proof is still valid for the RML$_{MZ}$ method. Part B is satisfied since

$$
E\left(Q\left(t,\beta,\sigma^2,h_t(\beta)\right)\right)=E\left(\left(\sigma^{-2}F^{-1}(\beta)\psi_t(\beta)\varepsilon_t(\beta)\right)^T,\varepsilon_t^2(\beta)-\sigma^2\right)^T
$$

doesn't depend on β and β^* and doesn't depend on t. Indeed $Q\left(t,\beta,\sigma^2,h_t(\beta)\right) = \left(\psi_t^T(\beta)F^{-1}(\beta)\varepsilon_t(\beta), \varepsilon_t^2(\beta)-\sigma^2\right)^T$ with $\Theta_q(B)\psi_t(\beta) = \varphi_t(\beta)$, $\Theta_q(B)\Phi_p^*(B)\varepsilon_t(\beta) = \Phi_p(B)\Theta_q^*(B)e_t$ and $\varphi_t(\beta) = (y_{t-1},\cdots,y_{t-p},-\varepsilon_{t-1}(\beta),\cdots,-\varepsilon_{t-q}(\beta))$, and we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}E\left(Q\left(t,\beta,\sigma^2,h_t(\beta)\right)\right) = \left[\sigma^{-2}F^{-1}(\beta)E\left(\psi_t(\beta)\varepsilon_t(\beta)\right)^T, E\left(\varepsilon_t^2(\beta)\right)-\sigma^2\right]^T. \tag{53}$$

■

Lemma 1.3 (Äström and Söderström, 1974). *For the ARMA(p,q) model defined by (1), β^* is the unique solution of $E\left[\varepsilon(\beta)\psi(\beta)\right]=0$.*

Lemma 1.4 (Cartan 1967, p. 122). *Let $g(x)$ be a class C_1 function $\omega \longrightarrow \mathbb{R}^n$, ω being an open set of $\mathbb{R}^n$. Let A_ω be a compact of ω. We suppose that any solution of $\dot{x}(t) = g(x(t))$, with the initial condition $x(t_0)=x_0$, defined over $[t_0,t_1]$ is such that $\forall t\in[t_0,t_1], x(t)\in A_\omega$. Then the upper bound of the maximal interval of existence of the ODE is $+\infty$.*

Lemma 1.5 (Rouch and Mawhin 1980, p. 12). *Consider the ODE in Lemma 1.4 where g is a continuous locally Lipschitz function: $g: I\times B_\rho \to \mathbb{R}^n$, $B_\rho = B(0,\rho)\subset\mathbb{R}^n$. Let Γ be a part in $\mathbb{R}^n$ such that $\overline{\Gamma}\subset B_\rho$. Let $V: I\times B_\rho \longrightarrow \mathbb{R}^+$ a function of class C^1, and a, a positive constant. If*

a) $x_0\in\Gamma$, $t_0\in I$,
b) $V(t_0,x_0)<a$,
c) $\forall(t,x)\in I\times Fr(\Gamma), V(t,x)\geqslant a$,
d) $\forall(t,x)\in I\times\Gamma, \dot{V}(t,x)\leqslant 0$,

then the solution of the ODE is such that $\forall t\geqslant t_0$, $x(t)\in\Gamma$.

Consider the following ODE: $\dot{x}=g(x)$, $x(t_0)=x_0$ where $g:\Omega\to\mathbb{R}^n$ is a continuous locally Lipschitz function. Let $\gamma^+(x_0)=\{x(z,x_0), z\geqslant 0\}$ be the trajectory of $x(z,x_0)$.

Lemma 1.6 (Rouch and Mawhin 1980, p. 50). *Let Ψ a compact of Ω, an open set of $\mathbb{R}^n$, and $V:\Omega\longrightarrow\mathbb{R}^+$ a function of class C^1* such that $\forall x\in\Psi$, *$\dot{V}(x)\leqslant 0$. Let $E_\Psi = \left\{x\in\Psi/\ \dot{V}(x)=0\right\}$ and M the largest invariant subset of E. Then for any x_0 such that $\gamma^+(x_0)\subset\Psi$, $x(z,x_0)\underset{z\longrightarrow\infty}{\longrightarrow} M$.*

Lemma 1.7 (Theorems 1 and 4, Ljung, 1977). *Under conditions* **C**, *let us consider the algorithm* (29) *modified as follows*:

$$h_t = \begin{cases} A(\widehat{x}_{t-1})\,h_{t-1} + B(\widehat{x}_{t-1})\,z_t & \text{if } \widehat{x}_{t-1}\in D_1 \\ \text{a point in } D_3 & \text{if } \widehat{x}_{t-1}\notin D_1, \end{cases} \tag{54}$$

$$\widehat{x}_t = \left[\widehat{x}_{t-1} + \gamma_t\,Q(t,\widehat{x}_{t-1},h_t)\right]_{D_1,D_2} \tag{55}$$

where $D_1\subset D_R\subset\mathbb{R}^m$ is a bounded open part containing the compact D_2, D_3 is a compact of $\mathbb{R}^m$, and m is the dimension of $\widehat{x}_t$, and

$$[z]_{D_1,D_2} = \begin{cases} z & \text{if } z\in D_1 \\ \text{a point in } D_2 & \text{if } z\notin D_1. \end{cases}$$

Let $\overline{D}$ be a compact part of D_R such that the trajectories of the following ODE $\partial x(t)/\partial t = f(x(t))$ starting from a point in $\overline{D}$, stay in a closed part $\overline{D}_R$ of D_R. Suppose that the ODE possesses an invariant set D_c with its domain of attraction D_A such that $\overline{D} \subset D_A$. Let $\widetilde{D} = D_1 \backslash D_2$ and suppose there exists a twice differentiable function $U(x) \geqslant 0$ defined over a neighbourhood of $\widetilde{D}$ and such that:

$$\begin{aligned}
&\sup_{x \in \widetilde{D}} U'(x) f(x) < 0, \\
&U(x) \geqslant c_1 \qquad \textit{for } x \notin D_1, \\
&U(x) \leqslant c_2 < c_1 \quad \textit{for } x \in D_2.
\end{aligned}$$

Then $\widehat{x}_t \to D_c$ almost surely when $t \to \infty$.

Appendix 2

Lemma 2.1. *Consider $K_t = \overline{\sigma}_t \widetilde{\beta}_t$ defined in Section 3.2. Then*

$$\begin{aligned}
K_t = &\sum_{k=1}^{t} A_{1,k} + \sum_{k=1}^{t} A_{2,k} + \sum_{k=2}^{t} F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k + T_t \widetilde{\beta}_{t-1} - \sum_{k=1}^{t-1} T_k \left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) \\
&- F^{-1}(\beta^*) S_t \widetilde{\beta}_{t-1} + F^{-1}(\beta^*) \sum_{k=1}^{t-1} S_k \left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right), \qquad (56)
\end{aligned}$$

where

$$\begin{aligned}
A_{1,t} = &\left(\varepsilon_{t-1}^2(\widehat{\beta}_{t-1}) - e_{t-1}^2\right) \widetilde{\beta}_{t-1} - F^{-1}(\widehat{\beta}_{t-1}) \left(\psi_t(\widehat{\beta}_{t-1}) - \psi_t(\beta^*)\right) \psi_t^T(\beta^*) \widetilde{\beta}_{t-1} \\
&- \left(F^{-1}(\widehat{\beta}_{t-1}) - F^{-1}(\beta^*)\right) \psi_t(\beta^*) \psi_t^T(\beta^*) \widetilde{\beta}_{t-1} \\
&- \frac{1}{2} F^{-1}(\widehat{\beta}_{t-1}) \psi_t \widetilde{\beta}_{t-1}^T \left(\frac{\partial\, \psi_t(\beta)}{\partial \beta^T}\right)_{\beta = \varkappa_t} \widetilde{\beta}_{t-1}, \qquad (57)
\end{aligned}$$

$$\begin{aligned}
A_{2,t} = &\left(\varepsilon_{t-1}^2 - \varepsilon_{t-1}^2(\widehat{\beta}_{t-1})\right) \widetilde{\beta}_{t-1} + F^{-1}(\widehat{\beta}_{t-1}) \psi_t \left(\varepsilon_t - \varepsilon_t(\widehat{\beta}_{t-1})\right) \\
&- F^{-1}(\widehat{\beta}_{t-1}) \left(\psi_t - \psi_t^T(\widehat{\beta}_{t-1})\right) \psi_t^T(\beta^*) \widetilde{\beta}_{t-1}, \qquad (58)
\end{aligned}$$

$$S_k = \sum_{j=1}^{k} \left[\psi_j(\beta^*) \psi_j^T(\beta^*) - \sigma_e^2 F(\beta^*)\right] \textit{ with } S_0 = 0, \qquad (59)$$

$$T_k = \sum_{j=1}^{k} \left(e_{j-1}^2 - \sigma_e^2\right) \textit{ with } T_0 = 0. \qquad (60)$$

Proof of Lemma 2.1.

By (41), the definition of $\overline{\sigma}_t = t\widehat{\sigma}_t^2$ and (37),

$$K_t = \overline{\sigma}_t \widetilde{\beta}_t = \overline{\sigma}_{t-1} \widetilde{\beta}_{t-1} + \varepsilon_{t-1}^2 \widetilde{\beta}_{t-1} + F^{-1}(\widehat{\beta}_{t-1})\psi_t \varepsilon_t,$$

and we have

$$\begin{aligned} K_t = K_{t-1} &+ \left(\varepsilon_{t-1}^2 - \varepsilon_{t-1}^2(\widehat{\beta}_{t-1})\right)\widetilde{\beta}_{t-1} + \left(\varepsilon_{t-1}^2(\widehat{\beta}_{t-1}) - e_{t-1}^2\right)\widetilde{\beta}_{t-1} + e_{t-1}^2\widetilde{\beta}_{t-1} \\ &+ F^{-1}(\widehat{\beta}_{t-1})\psi_t \left(\varepsilon_t - \varepsilon_t(\widehat{\beta}_{t-1})\right) + F^{-1}(\widehat{\beta}_{t-1})\psi_t \left(\varepsilon_t(\widehat{\beta}_{t-1}) - e_t\right) + F^{-1}(\widehat{\beta}_{t-1})\psi_t e_t. \end{aligned}$$

Note that $K_0 = 0$. We can use (42) so that

$$\begin{aligned} K_t = K_{t-1} &+ \left(\varepsilon_{t-1}^2 - \varepsilon_{t-1}^2(\widehat{\beta}_{t-1})\right)\widetilde{\beta}_{t-1} + \left(\varepsilon_{t-1}^2(\widehat{\beta}_{t-1}) - e_{t-1}^2\right)\widetilde{\beta}_{t-1} + e_{t-1}^2\widetilde{\beta}_{t-1} \\ &+ F^{-1}(\widehat{\beta}_{t-1})\psi_t \left(\varepsilon_t - \varepsilon_t(\widehat{\beta}_{t-1})\right) - F^{-1}(\widehat{\beta}_{t-1}) \left(\psi_t - \psi_t(\widehat{\beta}_{t-1})\right)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} \\ &- F^{-1}(\widehat{\beta}_{t-1}) \left(\psi_t(\widehat{\beta}_{t-1}) - \psi_t(\beta^*)\right)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} \\ &- \left(F^{-1}(\widehat{\beta}_{t-1}) - F^{-1}(\beta^*)\right)\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} - F^{-1}(\beta^*)\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1} \\ &- \frac{1}{2}F^{-1}(\widehat{\beta}_{t-1})\psi_t\widetilde{\beta}_{t-1}^T \left(\frac{\partial\,\psi_t(\beta)}{\partial\beta^T}\right)_{\beta=\varkappa_t}\widetilde{\beta}_{t-1} + F^{-1}(\widehat{\beta}_{t-1})\psi_t e_t. \end{aligned}$$

Moving some terms leads to

$$\begin{aligned} K_t &= K_{t-1} + A_{1,t} + A_{2,t} + F^{-1}(\widehat{\beta}_{t-1})\psi_t e_t + \left(e_{t-1}^2 - \sigma_e^2\right)\widetilde{\beta}_{t-1} \\ &\quad - F^{-1}(\beta^*)\left[\psi_t(\beta^*)\psi_t^T(\beta^*) - \sigma_e^2 F(\beta^*)\right]\widetilde{\beta}_{t-1} \\ &= K_0 + \sum_{k=1}^t A_{1,k} + \sum_{k=1}^t A_{2,k} + \sum_{k=2}^t F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k + \sum_{k=1}^t \left(e_{k-1}^2 - \sigma_e^2\right)\widetilde{\beta}_{k-1} \\ &\quad - F^{-1}(\beta^*)\sum_{k=1}^t \left[\psi_k(\beta^*)\psi_k^T(\beta^*) - \sigma_e^2 F(\beta^*)\right]\widetilde{\beta}_{k-1}. \end{aligned}$$

Introducing S_t, defined by (59), yields after some algebra

$$\begin{aligned} \sum_{k=1}^t \left[\psi_k(\beta^*)\psi_k^T(\beta^*) - \sigma^2 F(\beta^*)\right]\widetilde{\beta}_{k-1} &= \sum_{k=1}^t \left(S_k - S_{k-1}\right)\widetilde{\beta}_{k-1} \\ &= S_t\widetilde{\beta}_{t-1} - \sum_{k=1}^{t-1} S_k \left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right). \end{aligned}$$

Similarly using T_k defined by (60),

$$\begin{aligned} \sum_{k=1}^t \left(e_{k-1}^2 - \sigma^2\right)\widetilde{\beta}_{k-1} &= \sum_{k=1}^t \left(T_k - T_{k-1}\right)\widetilde{\beta}_{k-1} \\ &= T_t\widetilde{\beta}_{t-1} - \sum_{k=1}^{t-1} T_k \left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right). \end{aligned}$$

■

We will use a lemma due to Hannan (1976) for which the following notations are needed. Let X_t, $t = 1, ..., N$, be a stationary multivariate random process, with components $X_t(a)$, $a = 1, ..., v$. Consider the autocovariances of a realization of length N

$$C_k(a,b) = \frac{1}{N}\sum_{t=1}^{N-k}\left\{X_t(a) - \overline{X}(a)\right\}\left\{X_{t+k}(b) - \overline{X}(b)\right\},$$

where $\overline{X}(a) = (1/N)\sum_{t=1}^{N} X_t(a)$. Suppose that $X_t = \sum_{j=0}^{\infty} A_j e_{t-j}$, $E\left\{e_t e_s^T\right\} = \delta_{ts}G$, $\sum_0^\infty \|A_j\|^2 < \infty$ and $E\{e_t\} = 0$, with Dirac δ_{ts}, G an invertible $v \times v$ matrix and random vectors e_t. Let $\mathcal{F}_t$ the $\sigma-$algebra spanned by $X_s(a)$, $s \leq t$, $a = 1, ..., v$. Suppose that for all a, b, c, d , $E\{e_t(a)/\mathcal{F}_{t-1}\}$, $E\{e_t(a)e_t(b)/\mathcal{F}_{t-1}\}$, $E\{e_t(a)e_t(b)e_t(c)/\mathcal{F}_{t-1}\}$, $E\{e_t(a)e_t(b)e_t(c)e_t(d)/\mathcal{F}_{t-1}\}$ are constants. Denote the latter by κ_{abcd}. Let

$$h(\omega) = \sum_0^\infty A_j e^{ij\omega},\ f(\omega) = (2\pi)^{-1}h(\omega)Gh^*(\omega), \tag{61}$$

where $h^*(\omega) = \overline{h^T(\omega)}$ it the conjugated matrix of $h^T(\omega)$. Let

$$\gamma_k(a,b) = E\{X_t(a)X_{t+k}(b)\} \text{ and } Z_k(a,b) = N^{1/2}\{C_k(a,b) - \gamma_k(a,b)\}. \tag{62}$$

We use the asymptotic covariance between $Z_k(a,b)$ and $Z_k(c,d)$ defined by:

$$\begin{aligned} &2\pi\int_{-\pi}^{\pi}\left\{f_{ac}(\omega)\overline{f_{bd}(\omega)}e^{-i(k-t)\omega} + f_{ad}(\omega)\overline{f_{bc}(\omega)}e^{i(k+t)\omega}\right\}d\omega \\ &+\sum_p^v\sum_q^v\sum_r^v\sum_s^v \kappa_{pqrs}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{h_{ap}(\omega)\overline{h_{bq}(\omega)}e^{it\omega} + h_{cr}(\omega)\overline{h_{ds}(\omega)}e^{-ik\omega}\right\}d\omega, \end{aligned} \tag{63}$$

where $h_{ap}(\omega)$ is the element (a,p) of matrix $h(\omega)$.

Lemma 2.2. *Under those conditions, a necessary and sufficient for asymptotic normality of any vector composed of $Z_k(a,b)$ with variance-covariance matrix whose components are given by (63), is that the square of $f_{aa}(\omega)$, $a = 1, ..., v$, defined in (61), be integrable.*

We will use Lemma 2.2 for $k = 0$, so with $Z_0(a,b) = N^{1/2}\{C_0(a,b) - \gamma_0(a,b)\}$. Note that for an ARMA process defined by (1), given Assumption 1, the spectral density $f(\omega) = (1/2\pi)\sigma^2\left|\Theta(e^{i\omega})\right|^2 / \left|\Phi(e^{i\omega})\right|^2$ is square integrable. Note also that if the vectors $\{e_t\}$ are independent, then $E\{e_t(a)/\mathcal{F}_{t-1}\} = E\{e_t(a)\}$, $E\{e_t(a)e_t(b)/\mathcal{F}_{t-1}\} = E\{e_t(a)e_t(b)\}$, $E\{e_t(a)e_t(b)e_t(c)/\mathcal{F}_{t-1}\} = E\{e_t(a)e_t(b)e_t(c)\}$ and $E\{e_t(a)e_t(b)e_t(c)e_t(d)/\mathcal{F}_{t-1}\} = E\{e_t(a)e_t(b)e_t(c)e_t(d)\}$.

Lemma 2.3. *Let S_t and T_t defined by (59) and (60), respectively. Then $(1/\sqrt{t})S_t$ and $(1/\sqrt{t})T_t$ converge in law to a normal distribution when $t \to \infty$.*

Consider $R_t(\beta^*)$ defined by (39), then $\sqrt{t}\left(R_t(\beta^*) - \sigma_e^2\, F(\beta^*)\right)$ converges in law to a normal distribution when $t \to \infty$ as well as $\sqrt{t}\left(\sigma_t^2(\beta^*) - \sigma_e^2\right)$, where $\sigma_t^2(\beta^*)$ is defined by (37).

Proof of Lemma 2.3. Let $\psi_t(\beta^*) = (\psi_{1,t}(\beta^*), \cdots, , \psi_{p+q,t}(\beta^*))^T$. From (38) we have $\Theta^*(B)\psi_t(\beta^*) = \varphi_{t-1}^1$, where $\varphi_{t-1}^1 = (y_{t-1}, \cdots, y_{t-p}, -e_{t-1} \cdots, -e_{t-q})^T$, hence

$$\begin{aligned}\Phi^*(B)\psi_{i,t}(\beta^*) &= \Phi^*(B)\Theta^*(B)^{-1}y_{t-i},\ i = 1, ..., p,\\ \Theta^*(B)\psi_{p+i,t}(\beta^*) &= -e_{t-i},\ i = 1, ..., q,\end{aligned}$$

but we know that $\Phi^*(B)\Theta^*(B)^{-1}y_{t-1} = e_{t-1}$, hence

$$\begin{aligned}\Phi^*(B)\psi_{i,t}(\beta^*) &= e_{t-i},\ i = 1, ..., p,\\ \Theta^*(B)\psi_{p+i,t}(\beta^*) &= -e_{t-i},\ i = 1, ..., q.\end{aligned}$$

Consequently the $\{\psi_{i,t}(\beta^*),\ i = 1, 2, ..., p\}$ are autoregressive processes with $\Phi^*(B)$ as autoregressive polynomial, and the $\{\psi_{i,t}(\beta^*),\ i = p+1, ..., p+q\}$ are autoregressive processes with $\Theta^*(B)$ as autoregressive polynomial. Since the roots of $\Phi^*(B)$ are outside of the unit circle, there exist constants a_i, $i \geqslant 1$, such that $\sum_{i=1}^{\infty} |a_i| < \infty$ and $\psi_{j,t}(\beta^*) = e_{t-j} + \sum_{i=1}^{\infty} a_i e_{t-j-i}$, $j = 1, ..., p$. Remark that $\psi_{2,t}(\beta^*) = \psi_{1,t-1}(\beta^*)$, ..., $\psi_{p,t}(\beta^*) = \psi_{1,t-p+1}(\beta^*)$.

Similarly, since the roots of $\Theta^*(B)$ are outside of the unit circle, there exist constants b_i, $i \geqslant 1$, such that $\sum_{i=1}^{\infty} |b_i| < \infty$ and $\psi_{p+j,t}(\beta^*) = e_{t-j} + \sum_{i=1}^{\infty} b_i e_{t-j-i}$, $j = 1, ..., q$. Hence $\psi_t(\beta^*)$ can be written under the form

$$\begin{pmatrix} \psi_{1,t}(\beta^*) \\ ... \\ \psi_{p,t}(\beta^*) \\ \psi_{p+1,t}(\beta^*) \\ ... \\ \psi_{p+q,t}(\beta^*) \end{pmatrix} = \begin{pmatrix} 1 & a_1 & ... & a_{p-1} & a_p & a_{p+1} & ... & a_{p+q-1} \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & a_1 & a_2 & ... & a_q \\ 1 & b_1 & ... & b_{p-1} & b_p & b_{p+1} & ... & b_{p+q-1} \\ ... & ... & ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... & ... & ... \end{pmatrix} \begin{pmatrix} e_{t-1} \\ ... \\ e_{t-p} \\ e_{t-p-1} \\ ... \\ e_{t-p-q)} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{p+q} & a_{p+q+1} & \cdots & a_{2p+q-1} & a_{2p+q} & a_{2p+q+1} & \cdots & a_{2p+2q-1} \\ ... & ... & ... & ... & ... & ... & ... & ... \\ a_{q+1} & a_{q+2} & ... & a_{q+p} & a_{q+p+1} & a_{q+p+2} & \cdots & a_{q+p+q} \\ b_{p+q} & b_{p+q+1} & \cdots & b_{2p+q-1} & b_{2p+q} & b_{2p+q+1} & \cdots & b_{2p+2q-1} \\ ... & ... & ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... & ... & ... \end{pmatrix} \begin{pmatrix} e_{t-p-q-1} \\ ... \\ e_{t-2p-q} \\ e_{t-2p-q-1} \\ ... \\ e_{t-2p-2q)} \end{pmatrix}$$

$$+ ...$$

Let $Z_{p+q,t}^j = (e_{t-j(p+q)-1}, e_{t-j(p+q)-2}, ..., e_{t-(j+1)(p+q)})^T$, and note that the $\left\{Z_{p+q,t}^j, j \geqslant 0\right\}$ are i. i. d. random vectors with mean 0 and variance-covariance matrix $\sigma_e^2 I_{p+q}$ which is invertible. We may write $\psi_t(\beta^*)$ under the form $\psi_t(\beta^*) = \sum_{j=1}^{\infty} A_j Z_{p+q,t}^j$, where the components A_j, $j \geqslant 1$, are the a_i, $i \geqslant 1$ and b_i, $i \geqslant 1$.

Let us use the following norm for a matrix $M = (m_{ij})_{i,j} \in \mathbb{R}^{n\times n}$: $\|M\| = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{n} |m_{ij}|$. Since $\sum_{i=1}^{\infty} |a_i| < \infty$ and $\sum_{i=1}^{\infty} |b_i| < \infty$, then $\sum_{i=1}^{\infty} \|A_i\| < \infty$ and hence $\sum_{i=1}^{\infty} \|A_i\|^2 \leqslant (\sum_{i=1}^{\infty} \|A_i\|)^2 < \infty$. Note that $\psi_t(\beta^*)$ satisfies the conditions of Lemma 2.2. Let $\overline{\psi}_t(\beta^*) = (1/t)\sum_{k=1}^{t}\psi_k(\beta^*)$. Note that $E(e_t)$, $E(e_t^2)$, $E(e_t^3)$, $E(e_t^4)$ do exist because the random variable e_t is bounded, so

$$\sqrt{t}\left[\frac{1}{t}\sum_{k=1}^{t}\psi_k(\beta^*)\psi_k^T(\beta^*) - \overline{\psi}_t(\beta^*)\overline{\psi}_t^T(\beta^*) - E\left\{\psi_k(\beta^*)\psi_k^T(\beta^*)\right\}\right]$$

converges in law to a normal distribution, and since (24) and (39), we have that

$$\sqrt{t}\left(R_t(\beta^*) - \sigma_e^2\, F(\beta^*)\right) - \sqrt{t}\left(\overline{\psi}_t(\beta^*)\overline{\psi}_t^T(\beta^*)\right)$$

converges also in law to a normal distribution. For each $i = 1, ..., p+q$, $\psi_{i,t}(\beta^*)$ is a stationary autoregressive process, so $(1/t)\sum_{k=1}^{t}\psi_{i,k}(\beta^*) \to 0$ a.s. when $t \to \infty$, and $(1/\sqrt{t})\sum_{k=1}^{t}\psi_{i,k}(\beta^*)$ converges in law to the normal distribution when $t \to \infty$, i.e. $\overline{\psi}_t(\beta^*) \to 0$ a.s. when $t \to \infty$ and $\sqrt{t}\left[\overline{\psi}_t(\beta^*)\right]$ converges in law to the normal distribution when $t \to \infty$. Consequently $\sqrt{t}(\overline{\psi}_t(\beta^*)\overline{\psi}_t^T(\beta^*)) \to 0$ a.s. when $t \to \infty$ and $(1/\sqrt{t})S_t = \sqrt{t}\left(R_t(\beta^*) - \sigma_e^2\, F(\beta^*)\right)$ converges in law to the normal distribution when $t \to \infty$. We have

$$\widehat{\sigma}_t^2(\beta^*) = \frac{1}{t}\sum_{k=1}^{t}\varepsilon_{k-1}^2(\beta^*) = \frac{1}{t}\sum_{k=1}^{t}e_{k-1}^2, \text{ and } E(e_k^2) = \sigma_e^2.$$

The e_k^2 are independent and $E(e_k^4)$ is finite, so by Lindeberg-Feller central limit theorem, $(1/\sqrt{t})T_t = \sqrt{t}\left(\sigma_t^2(\beta^*) - \sigma_e^2\right)$ converges in law to the normal distribution. ∎

The following lemma is taken from Ljung and Söderström (1983, pp. 441-444) but the proof is more detailed here.

Lemma 2.4. *$\forall\beta \in D_R$, $\psi_t(\beta)$, $\partial\psi_t(\beta)/\partial\beta$ and the sequence h_t in algorithm (30) are bounded, and there exists a positive constant M such that $\left\|h_t - h_t(\widehat{\beta}_t)\right\| \leqslant M/t$* where $\forall\beta \in D_R$, $h_t(\beta) = A(\beta)h_{t-1}(\beta) + B(\beta)z_t$. *Moreover $h_t - h_t(\beta^*) \to 0$ a.s, when $n \to \infty$.*

Proof of Lemma 2.4. $\forall\beta \in D_R$ we may write $h_t(\beta)$ under the form

$$h_t(\beta) = B\,(\beta)z_t + \sum_{i=1}^{t-1}A(\beta)^i B\,(\beta)z_{t-i}(\beta) + A(\beta)^t h_0(\beta),$$

and as z_t is bounded and the eigenvalues of $A(\beta)$ are in the unit circle, $h_t(\beta)$ is bounded, implying that $\psi_t(\beta)$ is bounded. We have also

$$\frac{\partial\, h_t(\beta)}{\partial\beta^T} = A(\beta)\frac{\partial\, h_{t-1}(\beta)}{\partial\beta^T} + \left(h_{t-1}(\beta)\otimes I_v\right)\frac{\partial \mathrm{vec}A(\beta)}{\partial\beta^T} + \left(z_t\otimes I_v\right)\frac{\partial \mathrm{vec}B(\beta)}{\partial\beta^T}.$$

where $v = q(p+q+1)$. Let

$$G_t(\beta) = (h_{t-1}(\beta) \otimes I_v) \frac{\partial \text{vec} A(\beta)}{\partial \beta^T} + (z_t \otimes I_v) \frac{\partial \text{vec} B(\beta)}{\partial \beta^T}.$$

We can write $\partial\, h_t(\beta)/\partial \beta^T$ under the form

$$\frac{\partial\, h_t(\beta)}{\partial \beta^T} = G_t(\beta) + \sum_{i=1}^{\infty} A(\beta)^i G_{t-i}(\beta).$$

$A(\beta)$ and $B(\beta)$ are bounded since they are continuous functions of β. Since $h_{t-1}(\beta)$ and z_t are bounded then $G_t(\beta)$ is bounded and, because the eigenvalues of $A(\beta)$ are in the unit circle, then $\partial h_t(\beta)/\partial \beta^T$ is bounded, thus $\partial \psi_t(\beta)/\partial \beta^T$ is bounded.

Let us now show that h_t is bounded. We know that $\beta^* \in D_R$, so $\|A(\beta^*)^t\| \leqslant C\lambda^t$ for some $\lambda < 1$. Furthermore, for β_k belonging to a neighbourhood of β^* small enough, we have also $\left\|\prod_{k=1}^t A(\beta_k)\right\| \leqslant C\lambda_1^t$ for some $\lambda_1 < 1$ since $\prod_{k=1}^t A(\beta)$ is a continuous function of β. In Section 3.1, we have proved that $\widehat{\beta}_t \to \beta^*$ a.s. when $t \to \infty$, then for a large enough t, $\exists T > 0$, such as $\forall s > T$, $\left\|A(\widehat{\beta}_s)\right\| < \lambda_1$, so $\forall t > T$,

$$\left\|\prod_{k=1}^{t} A(\widehat{\beta}_k)\right\| \leqslant \left\|\prod_{k=1}^{T-1} A(\widehat{\beta}_k)\right\| \left\|\prod_{k=T}^{t} A(\widehat{\beta}_k)\right\| \leqslant C_0 C \lambda_1^{t-T} = C_2 \lambda_1^t, \tag{64}$$

where C_0 can be taken as $(C_1)^T$, for example, where $C_1 = \sup_{\beta \in D_R} \|A(\beta)\|$. According to (30) we have

$$h_t = A(\widehat{\beta}_{t-1})\, h_{t-1} + B(\widehat{\beta}_{t-1})\, z_t.$$

Since h_t contains ε_t and $\psi_t, \psi_{t-1}, ..., \psi_{t-q}$, we can suppose that $h_0 = 0$ hence

$$\begin{aligned} h_t &= A(\widehat{\beta}_{t-1}) A(\widehat{\beta}_{t-2})\, h_{t-2} + A(\widehat{\beta}_{t-1}) B(\widehat{\beta}_{t-2})\, z_{t-1} + B(\widehat{\beta}_{t-1})\, z_t = ... \\ &= \sum_{k=1}^{t} \left[\prod_{j=k}^{t-1} A(\widehat{\beta}_j)\right] B(\widehat{\beta}_{k-1})\, z_k + \left[\prod_{j=k}^{t} A(\widehat{\beta}_j)\right] h_0, \end{aligned} \tag{65}$$

with the convention $\prod_{j=t}^{t-1} A(\widehat{\beta}_j) = I_{p+q}$. Since $\forall \beta \in D_R$, there exists a positive real C such that $\|B(\beta)\| < C$, and z_t is bounded so that (64) and (65) imply

$$\|h_t\| \leqslant C \sum_{k=1}^{t} \lambda_1^{t-k} \|z_k\| \leqslant C_3. \tag{66}$$

Let us now show that there exists a constant M such that $\left\|h_t - h_t(\widehat{\beta}_t)\right\| \leqslant M/t$. Let $\widetilde{h}_k(\beta) = h_k - h_k(\beta)$, $\widetilde{A}(\widehat{\beta}_k, \beta) = A(\widehat{\beta}_k) - A(\beta)$, $\widetilde{B}(\widehat{\beta}_k, \beta) = B(\widehat{\beta}_k) - B(\beta)$.

For $k \leqslant t$, we have similarly to (65)

$$\widetilde{h}_t(\beta) = \sum_{k=1}^{t} \left(\prod_{j=k}^{t-1} A(\widehat{\beta}_j) \right) \left[\widetilde{A}(\widehat{\beta}_{k-1}, \beta) h_{k-1}(\beta) + \widetilde{B}(\widehat{\beta}_{k-1}, \beta) z_k \right]$$

since $\widetilde{h}_0(\beta) = 0$. Hence for $\beta = \widehat{\beta}_{t-1}$, we have

$$\widetilde{h}_t(\widehat{\beta}_{t-1}) = \sum_{k=1}^{t} \left(\prod_{j=k}^{t-1} A(\widehat{\beta}_j) \right) \left[\widetilde{A}(\widehat{\beta}_{k-1}, \widehat{\beta}_{t-1}) h_{k-1}(\widehat{\beta}_{t-1}) + \widetilde{B}(\widehat{\beta}_{k-1}, \widehat{\beta}_{t-1}) z_k \right]. \tag{67}$$

According to (66), h_t is bounded, and since ε_t and ψ_t are components of h_t, they are also bounded, and $F^{-1}(\widehat{\beta}_{t-1})$ is bounded because $\widehat{\beta}_{t-1} \in D_R$, and we know that $\widehat{\sigma}_t^2$ is bounded. By the recurrence formula of $\widehat{\beta}_t$ (28), we obtain $\left\| \widehat{\beta}_t - \widehat{\beta}_{t-1} \right\| < C_0/t$, which implies

$$\left\| \widehat{\beta}_{k-1} - \widehat{\beta}_{t-1} \right\| \leqslant C_4 \log(\frac{t-1}{k-1}) \text{ for } t > k > 0,$$

and for $k = 0$, $\left\| \widehat{\beta}_0 - \widehat{\beta}_{t-1} \right\| \leqslant C_4 \log(t-1)$. Since $A(\beta)$ and $B(\beta)$ are Lipschitz continuous, there exists a constant C_{AB} such that

$$\left\| \widetilde{A}(\widehat{\beta}_{k-1}, \widehat{\beta}_{t-1}) \right\| \leqslant C_{AB} \left\| \widehat{\beta}_{k-1} - \widehat{\beta}_{t-1} \right\| \ , \ \left\| \widetilde{B}(\widehat{\beta}_{k-1}, \widehat{\beta}_{t-1}) \right\| \leqslant C_{AB} \left\| \widehat{\beta}_{k-1} - \widehat{\beta}_{t-1} \right\|$$

hence,

$$\left\| \widetilde{A}(\widehat{\beta}_{k-1}, \widehat{\beta}_{t-1}) \right\| + \left\| \widetilde{B}(\widehat{\beta}_{k-1}, \widehat{\beta}_{t-1}) \right\| \leqslant C \log(\frac{t-1}{k-1}) \text{ pour } t \geqslant k-1.$$

Using (66) and the fact that $h_t(\widehat{\beta}_t)$ and z_t are bounded by (67), we have

$$\begin{aligned} \widetilde{h}_t &\leqslant M \sum_{k=2}^{t} \lambda_1^{t-k} \log(\frac{t-1}{k-1}) + M \lambda_1^{t-1} \log(t) \\ &\leqslant M \sum_{k=1}^{t} \lambda_1^{t-k} \log(\frac{t-1}{k}) \leqslant M/t. \end{aligned}$$

Now let $\widetilde{h}_k(\beta^*) = h_k - h_k(\beta^*)$, $\widetilde{A}(\widehat{\beta}_k, \beta^*) = A(\widehat{\beta}_k) - A(\beta^*)$, $\widetilde{B}(\widehat{\beta}_k, \beta^*) = B(\widehat{\beta}_k) - B(\beta^*)$. We have

$$\widetilde{h}_t(\beta^*) = \sum_{k=1}^{t} \left(\prod_{j=k}^{t-1} A(\widehat{\beta}_j) \right) \left[\widetilde{A}(\widehat{\beta}_{k-1}, \beta^*) h_k(\beta^*) + \widetilde{B}(\widehat{\beta}_{k-1}, \beta^*) z_k \right].$$

Because $h_k(\beta^*)$ and z_t are bounded and $\widetilde{A}(\widehat{\beta}_{k-1},\beta^*)$, $\widetilde{B}(\widehat{\beta}_{k-1},\beta^*)$ converge to 0, hence $\widetilde{h}_k \longrightarrow 0$ a.s, when $n \longrightarrow \infty$. ■

Lemma 2.5. *The exists a positive real C such that $\forall t \geqslant 1$, $\|A_{1,t}\| < C \left\|\widetilde{\beta}_{t-1}\right\|^2$, where $A_{1,t}$ is defined by (57).*
Proof of Lemma 2.5. A sketch of the proof is given by Ljung and Söderström (1983, p. 444), Lemma 4.B.4. A more detailed proof is as follows. We know that $\widehat{\beta}_t \to \beta^*$ a.s., and $F^{-1}(\beta)$ being continuous over D_R, $F^{-1}(\widehat{\beta}_{t-1}) \to F^{-1}(\beta^*)$ a.s. By Lemma 2.4, $\partial\,\psi_t(\beta)/\partial\beta^T$ and ψ_t are bounded, so there exists a positive constant C_1 such that

$$\left\|\frac{1}{2}F^{-1}(\widehat{\beta}_{t-1})\psi_t\widetilde{\beta}_{t-1}^T\left(\frac{\partial\,\psi_t(\beta)}{\partial\beta^T}\right)_{\beta=\varkappa_t}\widetilde{\beta}_{t-1}\right\| < C_1\left\|\widetilde{\beta}_{t-1}\right\|^2.$$

Similarly, since $\psi_t(\beta)$, $\partial\,\psi_t(\beta)/\partial\beta^T$ and ψ_t are bounded, there exist positive constants C_2 and C_3 such that

$$\left\|F^{-1}(\widehat{\beta}_{t-1})\psi_t\left(\psi_t^T(\widehat{\beta}_{t-1}) - \psi_t^T(\beta^*)\right)\widetilde{\beta}_{t-1}\right\| < C_2\left\|\widetilde{\beta}_{t-1}\right\|^2,$$
$$\left\|\left(\varepsilon_{t-1}^2(\widehat{\beta}_{t-1}) - e_{t-1}^2\right)\widetilde{\beta}_{t-1}\right\| = \left\|\left(\varepsilon_{t-1}^2(\widehat{\beta}_{t-1}) - \varepsilon_{t-1}^2(\beta^*)\right)\widetilde{\beta}_{t-1}\right\| < C_3\left\|\widetilde{\beta}_{t-1}\right\|^2.$$

Also, $\partial F(\beta)/\partial\beta = \partial E\left(\psi_t(\beta)\psi_t^T(\beta)\right)/\partial\beta$ is bounded and $F^{-1}(\beta)$ is bounded over D_R so that by Lemma 1.2, $\partial F^{-1}(\beta)/\partial\beta$ is bounded and there exists a positive constant C_4 such that

$$\left\|\left(F^{-1}(\widehat{\beta}_{t-1}) - F^{-1}(\beta^*)\right)\psi_t(\beta^*)\psi_t^T(\beta^*)\widetilde{\beta}_{t-1}\right\| < C_4\left\|\widetilde{\beta}_{t-1}\right\|^2.$$

We conclude that there exists a positive constant C such that $\|A_{1,t}\| < C\left\|\widetilde{\beta}_t\right\|^2$.■

From Chung (1968, p. 117) we have the following lemma.
Lemma 2.6 (Kronecker). *Let x_k be a sequence of real numbers, a_k a sequence of positive numbers which converges to ∞. Then*

$$\sum_n \frac{x_n}{a_n} < \infty \Rightarrow \frac{1}{a_n}\sum_{i=1}^n x_i \to 0.$$

Lemma 2.7. *Consider $A_{2,t}$ defined by (58). For all $\delta > 0$, $t^{-1/2-\delta}\sum_{k=1}^t A_{2,k}$ converges to 0 a.s. when $t \to \infty$.*
Proof of Lemma 2.7. A sketch of the proof is given by Ljung and Söderström (1983, p. 444), Lemma 4.B.4. A more detailed proof is as follows. Let us

consider the series $Z_t = \sum\limits_{k=1}^{t} k^{-1/2-\delta} A_{2,k}$ obtained by

$$
\begin{aligned}
Z_t = & \sum_{k=1}^{t} k^{-1/2-\delta} \left(\varepsilon_{k-1} - \varepsilon_{k-1}(\widehat{\beta}_{k-1}) \right) \left(\varepsilon_{k-1} + \varepsilon_{k-1}(\widehat{\beta}_{k-1}) \right) \widetilde{\beta}_{k-1} \\
& + \sum_{k=1}^{t} k^{-1/2-\delta} F^{-1}(\widehat{\beta}_{k-1}) \psi_k \left(\varepsilon_{k-1} - \varepsilon_{k-1}(\widehat{\beta}_{k-1}) \right) \\
& - \sum_{k=1}^{t} k^{-1/2-\delta} F^{-1}(\widehat{\beta}_{k-1}) \left(\psi_k - \psi_k(\widehat{\beta}_{k-1}) \right) \psi_k^T(\beta^*) \widetilde{\beta}_{k-1},
\end{aligned}
$$

By Lemma 2.4, since $(\psi_t - \psi_t(\hat{\beta}_{t-1}))$ and $(\varepsilon_t - \varepsilon_t(\hat{\beta}_{t-1}))$ are components of $(h_t - h_t(\hat{\beta}_{t-1}))$, there exists a constant M such that $\forall \beta \in D_R$, $\left\| \psi_t - \psi_t(\hat{\beta}_{t-1}) \right\| < M/t$ and $\left| \varepsilon_t - \varepsilon_t(\hat{\beta}_{t-1}) \right| < M/t$. We know that $F^{-1}(\beta)$, $\psi_t^T(\beta)$, ε_t and $\varepsilon_t(\beta)$ are bounded over D_R, thus Z_t is finite, hence by Lemma 2.6, $t^{-1/2-\delta} \sum_{k=1}^{t} A_{2,k} \to 0$ a.s. when $t \to \infty$. ■

Lemma 2.8. *For all* $\delta > 0$, $t^{-1/2-\delta} \sum_{k=1}^{t} F^{-1}(\widehat{\beta}_{k-1}) \psi_k e_k \to 0$ *a.s. when* $t \to \infty$.

Proof of Lemma 2.8. A sketch of the proof is given by Ljung and Söderström (1983, p. 442), Lemma 4.B.3. A more detailed proof is as follows. Consider the series

$$
s_t = \sum_{k=1}^{t} k^{-1/2-\delta} F^{-1}(\widehat{\beta}_{k-1}) \psi_k e_k.
$$

That random vector is a martingale with respect to the σ-algebra $\mathcal{F}_{t-1}$, spanned by the e_i, $i \leqslant t-1$. Indeed

$$
\begin{aligned}
E(s_t | \mathcal{F}_{t-1}) &= s_{t-1} + E\left(t^{-1/2-\delta} F^{-1}(\widehat{\beta}_{k-1}) \psi_k e_k \middle| \mathcal{F}_{t-1} \right) \\
&= s_{t-1} + t^{-1/2-\delta} F^{-1}(\widehat{\beta}_{k-1}) \psi_k E\left(e_t | \mathcal{F}_{t-1} \right) = s_{t-1},
\end{aligned}
$$

since $F^{-1}(\hat{\beta}_{t-1})$ and ψ_t do not depend of e_i, $i \leqslant t-1$. Moreover

$$
E \left\| s_t \right\|^2 \leqslant \sum_{k=1}^{t} k^{-1-2\delta} E \left\| F^{-1}(\widehat{\beta}_{k-1}) \psi_k \right\|^2 E \left| e_k \right|^2 \leqslant C \sum_{k=1}^{t} k^{-1-2\delta} < \infty,
$$

where C denotes a constant. Hence s_t is a martingale with a bounded variance, and according to Chung, (1968, p. 310), s_t converges a.s. to a finite limit s_∞.

Hence by Lemma 2.6, $t^{-1/2-\delta}\sum_{k=1}^{t} F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k \to 0$ a.s. when $t \to \infty$. ■

Lemma 2.9. *Let S_t defined by* (59) *and T_k defined by* (60)*. For all $\delta > 0$,*

$$G_t = t^{-1/2-\delta}\left\{T_t\widetilde{\beta}_{t-1} - \sum_{k=1}^{t-1} T_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right)\right.$$
$$\left. -F^{-1}(\beta^*)S_t\widetilde{\beta}_{t-1} + F^{-1}(\beta^*)\sum_{k=1}^{t-1} S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right)\right\}$$

converges to 0 a.s. when $t \to \infty$.

Proof of Lemma 2.9. By Lemma 2.3, $(1/\sqrt{t})S_t$ and $(1/\sqrt{t})T_t$ converge in law to the normal distribution when $t \to \infty$, hence for all $\delta > 0$, $t^{-1/2-\delta}S_t$ and $t^{-1/2-\delta}T_t$ converge to 0 a.s. when $t \to \infty$, which implies that $t^{-1/2-\delta}T_t\widetilde{\beta}_{t-1}$ and $t^{-1/2-\delta}F^{-1}(\beta^*)S_t\widetilde{\beta}_{t-1}$ converge to 0 a.s. when $t \to \infty$. Let us now prove that

$$t^{-1/2-\delta}F^{-1}(\beta^*)\sum_{k=1}^{t-1} S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) \to 0, \text{ a.s. when } t \to \infty.$$

Consider the series

$$F^{-1}(\beta^*)\sum_{k=1}^{t-1} k^{-1/2-\delta} S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right).$$

From (41), we know that $\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) = O_P(1/k)$ and $\forall\alpha > 0$, $k^{-1/2-\alpha}S_k = o_p(1)$. Let $0 < \alpha < \delta$, so

$$k^{1/2-\delta}S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) = k^{1/2-\alpha}S_k k^{\alpha-\delta}\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) = o_p(k^{-1-(\delta-\alpha)}).$$

since $-1-(\delta-\alpha) < -1$, hence $F^{-1}(\beta^*)\sum_{k=1}^{t} k^{-1/2-\delta}S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right)$ is finite, and by Lemma 2.7, when $t \to \infty$, $t^{-1/2-\delta}F^{-1}(\beta^*)\sum_{k=1}^{t} S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) \to 0$ a.s. . Similarly, we show that $t^{-1/2-\delta}\sum_{k=1}^{t} T_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) \to 0$, a.s. . ■

Lemma 2.10. *$\widehat{\sigma}_t^2$ converges to σ_e^2 almost surely when $t \to \infty$.*

Proof of Lemma 2.10. Indeed we have by (37)

$$\widehat{\sigma}_t^2 = \frac{1}{t}\sum_{k=1}^{t}\varepsilon_{k-1}^2 = \frac{1}{t}\sum_{k=1}^{t} e_{k-1}^2 + \frac{1}{t}\sum_{k=1}^{t}\left(\varepsilon_{k-1} - e_{k-1}\right)\left(\varepsilon_{k-1} + e_{k-1}\right)$$

and, according to Lemma 2.4, $\varepsilon_{t-1} - e_{t-1} \to 0$ a.s. when $t \to \infty$, and, since ε_k and e_k are bounded, the second term converges to 0 a.s.when $t \longrightarrow \infty$. According to Lemma 2.3, $(1/t)\sum_{k=1}^{t} e_{k-1}^2 \to \sigma_e^2$ a.s. when $t \to \infty$, hence $\widehat{\sigma}_t^2 \to \sigma_e^2$ a.s. when $t \longrightarrow \infty$. ■

Lemma 2.11 (Ljung and Söderström, 1983, p. 445). *Let b_t be a real sequence such that $b_t > 0$, $b_t \to 0$, when $t \to \infty$ and, for some $C > 0$ and $0 < \alpha < 1$,*

$$tb_t < C\left(\sum_{k=1}^{t-1} b_k^2 + t^{\alpha}\right).$$

Then

$$\sum_{k=1}^{t-1} b_k^2 < t^{\alpha'},$$

where $\alpha' = \max(0, 2\alpha - 1)$ for $\alpha \neq 1/2$, $\alpha' = \varepsilon > 0$ (arbitrary) for $\alpha = 1/2$.

Lemma 2.12. *Using notations in the proof of Theorem 3, we have*

$$\forall \delta > 0, t^{1/2-\delta}\left\|\widetilde{\beta}_t\right\| \to 0 \text{ a.s. } \textit{when } t \to \infty. \tag{68}$$

Proof of Lemma 2.12.
From (56) and Lemma 2.1, we have

$$\begin{aligned} t\widehat{\sigma}_t^2\widetilde{\beta}_t = &\sum_{k=1}^{t} A_{1,k} + \sum_{k=1}^{t} A_{2,k} + \sum_{k=2}^{t} F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k + T_t\widetilde{\beta}_{t-1} - \sum_{k=1}^{t-1} T_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) \\ &- F^{-1}(\beta^*)S_t\widetilde{\beta}_{t-1} + F^{-1}(\beta^*)\sum_{k=1}^{t-1} S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) \end{aligned} \tag{69}$$

hence

$$\begin{aligned} t\widetilde{\beta}_t = &\widehat{\sigma}_t^{-2}\sum_{k=1}^{t} A_{1,k} + \widehat{\sigma}_t^{-2}\sum_{k=1}^{t} A_{2,k} + \widehat{\sigma}_t^{-2}\sum_{k=2}^{t} F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k + \widehat{\sigma}_t^{-2} T_t\widetilde{\beta}_{t-1} \\ &- \widehat{\sigma}_t^{-2}\sum_{k=1}^{t-1} T_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) - \widehat{\sigma}_t^{-2}F^{-1}(\beta^*)S_t\widetilde{\beta}_{t-1} + \widehat{\sigma}_t^{-2}F^{-1}(\beta^*)\sum_{k=1}^{t-1} S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right), \end{aligned}$$

which implies

$$\begin{aligned} \left\|t\widetilde{\beta}_t\right\| \leqslant &\left|\widehat{\sigma}_t^{-2}\right| \sum_{k=1}^{t} \|A_{1,k}\| + \left|\widehat{\sigma}_t^{-2}\right| \left\|\sum_{k=1}^{t} A_{2,k}\right\| + \left|\widehat{\sigma}_t^{-2}\right| \left\|\sum_{k=1}^{t} F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k\right\| \\ &+ \left|\widehat{\sigma}_t^{-2}\right| \Bigg\| T_t\widetilde{\beta}_{t-1} + \sum_{k=1}^{t-1} T_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) + F^{-1}(\beta^*)S_t\widetilde{\beta}_{t-1} \\ &+ F^{-1}(\beta^*)\sum_{k=1}^{t-1} S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right)\Bigg\|. \end{aligned} \tag{70}$$

By Lemma 2.5, there exists a constant C such that $\forall k \geqslant 1$, $\|A_{1,k}\| < C \left\|\widetilde{\beta}_k\right\|^2$, hence $\sum_{k=1}^t \|A_{1,k}\| < C\sum_{k=1}^t \left\|\widetilde{\beta}_k\right\|^2$, and by Lemma 2.10, we know that $\widehat{\sigma}_t^2 \to \sigma_e^2$ a.s., and $\widehat{\sigma}_t^{-2} \to \sigma_e^{-2}$ a.s., so there exists a constant $C_1 > 0$ such that $\forall k > 1$,

$$\left|\widehat{\sigma}_t^{-2}\right| \sum_{k=1}^t \|A_{1,k}\| < C_1 \sum_{k=1}^t \left\|\widetilde{\beta}_k\right\|^2.$$

By Lemma 2.8, we know that for all $\delta > 0$, $t^{-1/2-\delta}\sum_{k=1}^t F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k \to 0$, a.s when $t \to \infty$, so there exists a constant $C_2 > 0$ such that $\forall k \geqslant 1$,

$$\left|\widehat{\sigma}_t^{-2}\right| \left\|\sum_{k=1}^t F^{-1}(\widehat{\beta}_{k-1})\psi_k e_k\right\| < C_2 t^{1/2+\delta}.$$

By Lemma 2.9, we have that

$$G_t = t^{-1/2-\delta}\left\{T_t\widetilde{\beta}_{t-1} - \sum_{k=1}^t T_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right) - F^{-1}(\beta^*)S_t\widetilde{\beta}_{t-1} + F^{-1}(\beta^*)\sum_{k=1}^t S_k\left(\widetilde{\beta}_k - \widetilde{\beta}_{k-1}\right)\right\}$$

converges to 0 a.s when $t \to \infty$ and, by Lemma 2.7, $t^{-1/2-\delta}\sum_{k=1}^t A_{2,k} \to 0$ a.s. when $t \to \infty$, so that there exists a constant $C_3 > 0$ such that $\forall k \geqslant 1$,

$$\left|\widehat{\sigma}_t^{-2}\right| \left\|\sum_{k=1}^t A_{2,k}\right\| + \left|\widehat{\sigma}_t^{-2}\right| \left\| T_t\widetilde{\beta}_{t-1} + \sum_{k=1}^t T_k\left(\widetilde{\beta}_{k-1} - \widetilde{\beta}_{k-2}\right) + F^{-1}(\beta^*)S_t\widetilde{\beta}_{t-1} \right.$$

$$\left. + F^{-1}(\beta^*)\sum_{k=1}^t S_k\left(\widetilde{\beta}_{k-1} - \widetilde{\beta}_{k-2}\right)\right\| < C_3 t^{1/2+\delta}.$$

From (70), we can conclude that for each $\delta > 0$, there exists a constant $C > 0$, such that

$$t\left\|\widetilde{\beta}_t\right\| \leqslant C\sum_{k=1}^t \left\|\widetilde{\beta}_{k-1}\right\|^2 + Ct^{1/2+\delta}. \tag{71}$$

Applying Lemma 2.11 on (71) with $b_n = \left\|\widetilde{\beta}_n\right\|$ et $\alpha = 1/2 + \delta$, we obtain

$$\sum_{k=1}^t \left\|\widetilde{\beta}_{k-1}\right\|^2 < Ct^{2\delta}, \tag{72}$$

hence (72) inserted in (71) gives, $\forall\delta > 0, t\left\|\widetilde{\beta}_t\right\| \leqslant Ct^{2\delta} + Ct^{1/2+\delta}$, so that $\forall\delta', \delta > 0$

$$t^{1/2-\delta'}\left\|\widetilde{\beta}_t\right\| \leqslant Ct^{-1/2-\delta'+2\delta} + Ct^{\delta-\delta'}, \tag{73}$$

and thus $\forall\delta' > 0$, by taking $\delta < \delta'$, $t^{1/2-\delta'} \left\|\widetilde{\beta}_t\right\|$ converges to 0 almost surely when $t \to \infty$.

Lemma 2.13. *For all* $\delta > 0$, $t^{1/2-\delta}\left(\widehat{\sigma}_t^2 - \sigma_e^2\right) \to 0$ *a.s. when* $t \to \infty$.

Proof of Lemma 2.13. From (37) we have

$$\begin{aligned} t^{1/2-\delta}\left(\widehat{\sigma}_t^2 - \sigma_e^2\right) &= t^{1/2-\delta}\left[\frac{1}{t}\sum_{k=1}^{t}\{\varepsilon_{k-1}^2 - \varepsilon_{k-1}^2(\widehat{\beta}_{k-2})\}\right. \\ &\left.+\frac{1}{t}\sum_{k=1}^{t}\{\varepsilon_{k-1}^2(\widehat{\beta}_{k-2}) - \varepsilon_{k-1}^2(\beta^*)\} + \frac{1}{t}\sum_{k=1}^{t}\{\varepsilon_{k-1}^2(\beta^*) - \sigma_e^2\}\right], \\ &= \frac{1}{t^{1/2+\delta}}\sum_{k=1}^{t}\Big(\varepsilon_{k-1} - \varepsilon_{k-1}(\widehat{\beta}_{k-2})\Big)\Big(\varepsilon_{k-1} + \varepsilon_{k-1}(\widehat{\beta}_{k-2})\Big) \quad (74) \\ &+ \frac{1}{t^{1/2+\delta}}\sum_{k=1}^{t}\Big(\varepsilon_{k-1}(\widehat{\beta}_{k-2}) - \varepsilon_{k-1}(\beta^*)\Big)\Big(\varepsilon_{k-1}(\widehat{\beta}_{k-2}) + \varepsilon_{k-1}(\beta^*)\Big) \\ & \qquad\qquad (75) \\ &+ \frac{1}{t^{1/2+\delta}}\sum_{k=1}^{t}\left(\varepsilon_{k-1}^2(\beta^*) - \sigma_e^2\right). \qquad (76) \end{aligned}$$

By Lemma 2.5, we know that $\left|\varepsilon_t - \varepsilon_t(\hat{\beta}_{t-1})\right| = O_p(1/t)$, hence

$$\sum_{k=1}^{t} k^{-1/2-\delta}\Big(\varepsilon_{k-1} - \varepsilon_{k-1}(\widehat{\beta}_{k-2})\Big)\Big(\varepsilon_{k-1} + \varepsilon_{k-1}(\widehat{\beta}_{k-2})\Big)$$

converges to a finite limit and, by Lemma 2.6, (74) $\to 0$ a.s. There exists a constant C such that

$$\begin{aligned} &\sum_{k=1}^{t} k^{-1/2-\delta}\left|\Big(\varepsilon_{k-1}(\widehat{\beta}_{k-2}) - \varepsilon_{k-1}(\beta^*)\Big)\Big(\varepsilon_{k-1}(\widehat{\beta}_{k-2}) + \varepsilon_{k-1}(\beta^*)\Big)\right| \\ &\qquad \leqslant C\sum_{k=1}^{t} k^{-1/2-\delta}\left|\varepsilon_{k-1}(\widehat{\beta}_{k-2}) + \varepsilon_{k-1}(\beta^*)\right|\left\|\widetilde{\beta}_{k-1}\right\|. \end{aligned}$$

By (68) and Lemma 2.12, we have, $\forall\epsilon > 0$, $\left\|t^{1/2-\epsilon}\widetilde{\beta}_t\right\| \to 0$ a.s., hence for $\epsilon = \delta - \delta' > 0$, where $\delta' > 0$, we have

$$k^{1/2-(\delta-\delta')}\left|\varepsilon_{k-1}(\widehat{\beta}_{k-2}) + \varepsilon_{k-1}(\beta^*)\right|\left\|\widetilde{\beta}_{k-1}\right\| = o_p(1),$$

which implies

$$k^{-1/2-\delta}\left|\varepsilon_{k-1}(\widehat{\beta}_{k-2}) + \varepsilon_{k-1}(\beta^*)\right|\left\|\widetilde{\beta}_{k-1}\right\| = o_p(1/k^{1+\delta'}),$$

from which $\sum_{k=1}^{t} k^{-1/2-\delta}\left(\varepsilon_{k-1}(\widehat{\beta}_{k-2})-\varepsilon_{k-1}(\beta^*)\right)\left(\varepsilon_{k-1}(\widehat{\beta}_{k-2})+\varepsilon_{k-1}(\beta^*)\right)$ converges to a finite limit, and, by Lemma 2.6, (75) $\to 0$ a.s. By Lemma 2.3, (76) $\to 0$ a.s. when $t\to\infty$. ■

Lemma 2.14. (Theorem 1, Brown, 1971)
Let $\{S_t,\mathcal{F}_t,t=1,...\}$ be a martingale. Let

$$X_t=S_t-S_{t-1}, V_t^2=\sum_{k=1}^{t} E\left(X_k^2\middle|\,\mathcal{F}_{k-1}\right), s_t^2=EV_t^2=ES_t^2.$$

Suppose that $V_t^2 s_t^{-2}$ converges in probability to 1 when $t\to\infty$, and that the following Lindeberg condition is satisfied: $\forall\delta>0$, $s_t^{-2}\sum_{j=1}^{t} EX_j^2 I\left(X_j\geqslant\delta s_t\right)$ converges in probability to 1 when $t\to\infty$. Then $S_t/\ s_t$ converges in law to the normal distribution with mean 0 and variance 1.

Lemma 2.15.

$$(1/\sqrt{t})\sum_{k=1}^{t} R_k(\beta^*)\widehat{\sigma}_k^{-2}\,F^{-1}(\beta^*)\psi_k(\beta^*)e_k \tag{77}$$

converges in law to the normal distribution $N(0,\sigma_e^4 F(\beta^*))$.

Proof of Lemma 2.15. A sketch of the proof is given by Ljung and Söderström (1983, p. 448), Lemma 4.B.7. A more detailed proof is as follows.
We use Lemma 2.14 with the Cramér-Wold device. Let

$$Y_t^2=\sum_{k=1}^{t}E\left[R_k(\beta^*)\widehat{\sigma}_k^{-2}\,F^{-1}(\beta^*)\psi_k(\beta^*)e_k\left(R_k(\beta^*)\widehat{\sigma}_k^{-2}\,F^{-1}(\beta^*)\psi_k(\beta^*)e_k\right)^T\middle|\,\mathcal{F}_{k-1}\right]$$

and $M_t^2=EY_t^2$. We have

$$\begin{aligned}Y_t^2=&\sum_{k=1}^{t}\{R_k(\beta^*)-\sigma_e^2F(\beta^*)\}\widehat{\sigma}_k^{-4}\,F^{-1}(\beta^*)\psi_k(\beta^*)\psi_k^T(\beta^*)\,F^{-1}(\beta^*)R_k(\beta^*)\sigma_e^2\\&+\sum_{k=1}^{t}\sigma_e^4\widehat{\sigma}_k^{-4}\,\psi_k(\beta^*)\psi_k^T(\beta^*)\,F^{-1}(\beta^*)\{R_k(\beta^*)-\sigma_e^2F(\beta^*)\}\\&+\sum_{k=1}^{t}\sigma_e^6\left(\sigma_e^2-\widehat{\sigma}_k^2\right)\left(\sigma_e^2+\widehat{\sigma}_k^2\right)\widehat{\sigma}_k^{-4}\,\sigma_e^{-4}\psi_k(\beta^*)\psi_k^T(\beta^*)+\sum_{k=1}^{t}\sigma_e^2\psi_k(\beta^*)\psi_k^T(\beta^*)\end{aligned} \tag{78}$$

Let x_k be the first component of $R_k(\beta^*)\widehat{\sigma}_k^{-2}\,F^{-1}(\beta^*)\psi_k(\beta^*)$. To prove that $S_t=\sum_{k=1}^{t} x_k e_k$ converges in law to the normal distribution, we check the two conditions of Lemma 2.14, with $V_t^2=\sum_{k=1}^{t} E\left(|x_k e_k|^2\middle|\,\mathcal{F}_{k-1}\right)$ and $s_t^2=E(V_t^2)$.

First, by Lemma 2.13, $\left(\sigma_e^2-\widehat{\sigma}_k^2\right)=o_p(\ t^{1/2-\delta})$ and by Lemma 2.3,

$$\left(R_t(\beta^*)-\sigma_e^2F(\beta^*)\right)=o_p(t^{1/2-\delta}),$$

so by applying Lemma 2.6 to (78), the first three terms of Y_t^2/t converge to 0 a.s. when $t \to \infty$, hence $Y_t^2/t \to \sigma_e^4 F(\beta^*)$. It is obvious that $M_t^2/t \to \sigma_e^4 F(\beta^*)$ so $V_t^2/s_t^2 \to 1$ in probability when $t \to \infty$, where $s_t/\sqrt{t}$ is the square root of the element (1,1) of M_t^2/t.

Secondly, $x_k e_k = R_k(\beta^*)\widehat{\sigma}_k^{-2} F^{-1}(\beta^*)\psi_k(\beta^*)e_k$ is bounded, so there exists $t_0 > 0$ such that $\forall t \geqslant t_0$, $E\left|x_k e_k\right|^2 I\left(|x_k e_k| > \delta s_t\right) = 0$ and, since $s_t \to \infty$, then $(1/s_t^2)\sum_{k=1}^t E\left|x_k e_k\right|^2 I\left(|x_k e_k| > \delta s_t\right) \to 0$ when $t \to \infty$. Hence Lemma 2.14 implies convergence in law of $(S_t/\sqrt{t})/(s_t/\sqrt{t})$. In a similar way, we can show that any linear combination of the components of (77) converges in law to a normal distribution. ■

Lemma 2.16. *The series H_t and L_t defined by (47) and (48) converge to 0 a.s. when $t \to \infty$.*

Proof of Lemma 2.16. Let us show that L_t converges to 0 a.s. when $t \to \infty$. Consider the series

$$L_t^1 = \sum_{k=1}^t \frac{1}{\sqrt{k}} R_k(\beta^*)\widehat{\sigma}_k^{-2} \left(F^{-1}(\widehat{\beta}_{k-1}) - F^{-1}(\beta^*)\right)\psi_k(\hat{\beta}_{k-1})e_k$$
$$+ \sum_{k=1}^t \frac{1}{\sqrt{k}} R_k(\beta^*)\widehat{\sigma}_k^{-2}\, F^{-1}(\beta^*)\left(\psi_k(\hat{\beta}_{k-1}) - \psi_k(\beta^*)\right) e_k.$$

L_t^1 is a martingale and

$$E\left\|L_t^1\right\|^2 \leqslant \sum_{k=1}^t \frac{1}{k}\left\|R_k(\beta^*)\widehat{\sigma}_k^{-2} \left(F^{-1}(\widehat{\beta}_{k-1}) - F^{-1}(\beta^*)\right)\psi_k(\hat{\beta}_{k-1})e_k\right\|^2$$
$$+ \sum_{k=1}^t \frac{1}{k}\left\|R_k(\beta^*)\widehat{\sigma}_k^{-2}\, F^{-1}(\beta^*)\left(\psi_k(\hat{\beta}_{k-1}) - \psi_k(\beta^*)\right) e_k\right\|^2.$$

We know that $F^{-1}(\beta)$ and $\partial F(\beta)/\partial\beta$ are bounded, then by Lemma 1.2, we have that $\partial F^{-1}(\beta)/\partial\beta$ is bounded, and Lemma 2.4 implies that $\psi_k(\beta)$ et $\partial\psi_t(\beta)/\partial\beta^T$ are bounded, then there exists a constant C such that

$$E\left\|L_t^1\right\|^2 \leqslant C\sum_{k=1}^t \frac{1}{k} E\left\|\widetilde{\beta}_{k-1}\right\|^2 E\left|e_k\right|^2.$$

Using (68) and Lemma 2.12, for all ϵ positive, $\left\|t^{1/2-\epsilon}\widetilde{\beta}_t\right\|^2 = t^{1-2\epsilon}\left\|\widetilde{\beta}_t\right\|^2$ converges to 0 a.s. hence $E\left\|L_t^1\right\|^2$ is bounded. Then L_t^1 is a martingale with a bounded variance, and using Chung, (1968, p. 310) L_t^1 converges a.s. to a finite limit when $t \to \infty$, hence $R_t^{-1}(\beta^*)L_t^1$ converges, and by Lemma 2.6, L_t converges a.s. to 0 when $t \to \infty$.

Let us now show that $H_t \to 0$ a.s. when $t \to \infty$. Let

$$H_t^1 = R_t^{-1}(\beta^*) \sum_{k=1}^{t} \frac{1}{\sqrt{k}} \left(\sigma_e^2 \, F(\beta^*) - R_k(\beta^*)\right) \widehat{\sigma}_k^{-2} F^{-1}(\widehat{\beta}_{k-1}) \psi_k(\beta^*) \psi_k(\beta^*) \widetilde{\beta}_k$$

$$+ R_t^{-1}(\beta^*) \sum_{k=1}^{t} \frac{1}{\sqrt{k}} \left(\widehat{\sigma}_k^2 - \sigma_e^2\right) F(\beta^*) \widehat{\sigma}_k^{-2} \; F^{-1}(\widehat{\beta}_{k-1}) \psi_k(\beta^*) \psi_k(\beta^*) \widetilde{\beta}_{k-1}$$

$$+ R_t^{-1}(\beta^*) \sum_{k=1}^{t} \frac{1}{\sqrt{k}} B_{1,k} + R_t^{-1}(\beta^*) \sum_{k=1}^{t} \frac{1}{\sqrt{k}} B_{2,k},$$

By Lemma 2.3, we know that $\sqrt{t}\left(R_t(\beta^*) - \sigma_e^2 \, F(\beta^*)\right)$ converges in law, hence $\forall \delta > 0$,

$$t^{1/2-\delta}\left(R_t(\beta^*) - \sigma_e^2 \, F(\beta^*)\right) = o_p(1),$$

and since for every positive ϵ, $t^{1/2-\epsilon}\widetilde{\beta}_k \to 0$, then

$$t^{1-\delta-\epsilon}\left\{\left(\sigma_e^2 \, F(\beta^*) - R_k(\beta^*)\right) \widehat{\sigma}_k^{-2} \, F^{-1}(\widehat{\beta}_{k-1}) \psi_k(\beta^*) \psi_k(\beta^*) \widetilde{\beta}_k\right\} = o_p(1)$$

so that for all positive ϵ and δ

$$\frac{1}{\sqrt{k}} \left(\sigma_e^2 \, F(\beta^*) - R_k(\beta^*)\right) \widehat{\sigma}_k^{-2} \, F^{-1}(\widehat{\beta}_{k-1}) \psi_k(\beta^*) \psi_k(\beta^*) \widetilde{\beta}_k = o_p(t^{-1-1/2+\delta+\epsilon}),$$

i.e.

$$R_t^{-1}(\beta^*) \sum_{k=1}^{t} \frac{1}{\sqrt{k}} \left(\sigma_e^2 \, F(\beta^*) - R_k(\beta^*)\right) \widehat{\sigma}_k^{-2} F^{-1}(\widehat{\beta}_{k-1}) \psi_k(\beta^*) \psi_k(\beta^*) \widetilde{\beta}_k$$

is convergent. In the same manner, we have

$$\frac{1}{\sqrt{k}} \left(\widehat{\sigma}_k^2 - \sigma_e^2\right) F(\beta^*) \widehat{\sigma}_k^{-2} \; F^{-1}(\widehat{\beta}_{k-1}) \psi_k(\beta^*) \psi_k(\beta^*) \widetilde{\beta}_{k-1} = o_p(t^{-1-1/2+\delta+\epsilon}).$$

hence

$$R_t^{-1}(\beta^*) \sum_{k=1}^{t} \frac{1}{\sqrt{k}} \left(\widehat{\sigma}_k^2 - \sigma_e^2\right) F(\beta^*) \widehat{\sigma}_k^{-2} \; F^{-1}(\widehat{\beta}_{k-1}) \psi_k(\beta^*) \psi_k(\beta^*) \widetilde{\beta}_{k-1}$$

is convergent.

In the same way as in Lemma 2.5, we can show from (43) that $\|B_{1,t}\| < C \left\|\widetilde{\beta}_t\right\|^2$, where C is a positive constant, and since $\forall \epsilon > 0$, $t^{1-2\epsilon} \left\|\widetilde{\beta}_t\right\|^2$ converges to 0, then $R_t^{-1}(\beta^*) \sum_{k=1}^{t} B_{1,k}/\sqrt{k}$ converges a.s. Let us now consider $B_{2,t}$ defined by (44). We know by Lemma 2.4, that $\left\|\psi_t - \psi_t(\hat{\beta}_{t-1})\right\| = O_p(1/t)$ and $\left|\varepsilon_t - \varepsilon_t(\hat{\beta}_{t-1})\right| = O_p(1/t)$, and since $\psi_t(\beta)$ and $\varepsilon_t(\beta)$ and $\widehat{\sigma}_t^{-2}$ are bounded, then $R_t^{-1}(\beta^*) \sum_{k=1}^{t} B_{2,k}/\sqrt{k}$ converges a.s. Finally H_t^1 is finite. According to Lemma 2.6, $H_t \to 0$ a.s. when $t \to \infty$. ∎

References

[1] ASTRÖM, K. J. (1980). Maximum likelihood and prediction error methods, *Automatica,* **16,** 457-476.

[2] ASTRÖM, K. J. and SÖDERSTRÖM, T. (1974). Uniqueness of the Maximum Likelihood Estimates of the parameters of an ARMA Model, *IEEE Transactions on Automatic Control, AC* **19**, 769-773.

[3] BOX, G. E. P., JENKINS, G. M., and REINSEL G. C.(1994). *Time Series Analysis, Forecasting and Control*, Prentice Hall Press, Englewood Cliffs, N.J. (3rd edition).

[4] BROWN, B. M. (1971). Martingale central limit theorems*, Annals of Statistics,* **42**, 59-66.

[5] BROWN, R. L., DURBIN, J., and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time, *J. Roy. Statist. Soc. Ser. B,* **37** , 149-163 (with discussion: 163-192).

[6] CARTAN, M. H. (1967). Calcul différentiel, Hermann, Paris.

[7] CHUNG, K. L. (1968). *A Course In Probability Theory.* Harcourt, Brace and World, New York.

[8] DAHLHAUS, R., and SUBBA RAO, S. (2004). ???, working paper.

[9] DUFLO, M. (1997). Random Iterative Models*,* Springer-Verlag, Berlin.

[10] FABIAN, V. (1968). On asymptotic normality in stochastic approximation, *Annals of Statistics,* 39, 1327-1332.

[11] GUO, L. (1994). Stability of recursive stochastic tracking algorithms, *SIAM J. Control and Optimization,* 32, 1195-1225.

[12] HANNAN, E. J. (1976). The asymtotic distribution of serial covariance, *Annals of Statistics*, **8**, 762-777.

[13] HARVILLE, D. A. (1997). *Matrix algebra from statistician's perspective,* New York : Springer.

[14] KALMAN, R. E. (1960). A new approach to general filtering and prediction problems. *Trans. ASME Ser. D. Basic Eng.,* **83,** 95-108.

[15] KLEIN, A. and MÉLARD, G. (1989). On algorithms for computing the covariance matrix of estimates in autoregressive-moving average models. *Computational Statistics Quarterly,* **5,** 1-9.

[16] KLEIN, A. and SPREIJ, P. J. C. (1993). On Fisher's information matrix of an ARMA process and Sylvester's resultant matrices, Research Memorandum 1993-33, Dep. of Econometrics, Vrije Univ, Amsterdam.

[17] KUSHNER, H. J. and HUANG, H. (1979). Rates of convergence for stochastic approximation type algorithms, *SIAM J. Control and Optimization,* **17,** 607-617.

[18] KUSHNER, H. J. and YIN, G. (1997). *Stochastic Approximation Algorithms and Applications.* Springer-Verlag, Berlin.

[19] LJUNG, L. PFLUG, G. and WALK, H. (1992). *Stochastic Approximation and Optimisation of Random Systems.* DMV Seminar Band 17, Birkhäuser, Boston.

[20] LJUNG, L. and SÖDERSTRÖM, T. (1983). *Theory and Practice of recursive identification,* MIT Press, Cambridge, Massachussets.

[21] LJUNG, L. (1977). Analysis of recursive stochastic algorithms, *IEEE Transactions on Automatic Control, AC* **22**, 551-575.

[22] LJUNG, L. (1978). Convergence analysis of parametric identification methods, *IEEE Transactions on Automatic Control, AC* **23** , 770-783.

[23] LJUNG, L. (1985). Estimation of parameters in dynamical systems in *Handbook of Statistics*, Vol 5: *Time series in the time domain* (Hannan, E. J. Krishnaiah, P. R. and Rao, M. M.), Elsevier Science Publishers, 189-211.

[24] LJUNG, L. (2000). *System Identification Toolbox User's Guide : For Use with MATLAB*, The MathWorks Inc. 4th edition for Version 5.0, Natick, MA.

[25] MÉLARD, G. (1984). Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models, *J. Roy. Statist. Soc. Ser. C Appl. Statist.,* **33,** 104-114.

[26] MÉLARD, G. (1989). Méthode d'identification récurrente pour la modélisation des données chronologiques économiques, *Actes du Salon Int. de l'Inf. de la Tél., séance III: Intelligence artificielle et génie logiciel,* Association Marocaine des utilisateurs de l'informatique, Direction de la statistique, Rabat.

[27] MOULINES, E., PRIOURET, P., and ROUEFF, F. (2004). On recursive estimation for locally stationary time varying autoregressive processes, submitted.

[28] OUAKASSE, A. (2004). Estimation par récurrence dans le contexte d'un modèle dynamique. Propriétés des estimateurs et mise en oeuvre de la méthode, Thèse de doctorat, Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles.

[29] OUAKASSE, A. and MÉLARD, G. (2005). Recursive estimation of SISO models, in preparation.

[30] OUAKASSE, A., MÉLARD, G.,and ZAHAF, T. (2005). Seasonal models and recursive estimation, in preparation.

[31] PLACKETT, R. L. (1950). Some theorems in least squares, *Biometrika,* **37** , 149-157.

[32] ROUCHE, N. and MAWHIN, J. (1980). *Ordinary Differential Equations*, Pitman Advanced Publishing Program.

[33] SÖDERSTRÖM, T. (1973). An on-line algorithm for approximate maximum likelihood identification of linear dynamic systems. Report 7308, Departement of Automatic Control, Lund Institute of Technology, Lund, Sweden.

[34] SPSS Inc. (2004). SPSS 12.0 for Windows, Headquarters, 233 S. Wacker Drive, 11th floor Chicago, Illinois 60606.

[35] YOUNG, P. (1984). *Recursive estimation and time series analysis*, Springer-Verlag, Berlin.

[36] YOUNG, P. (1985). Recursive identification, estimation and control in *Handbook of Statistics*, Vol 5: *Time series in the time domain* (Hannan, E. J. Krishnaiah, P. R. and Rao, M. M.), Elsevier Science Publishers, 213-255.

[37] ZAHAF, T. (1999). Contributions à l'estimation des paramètres de modèles de séries chronologique, Thèse de doctorat, Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles.