15th Triennial World Congress, Barcelona, Spain

A TWO - STEP MAXIMUM - LIKELIHOOD IDENTIFICATION OF NON - GAUSSIAN SYSTEMS

Gerencsér, László [*] **Michaletzky, György** [**] **Reppa, Zoltán** [***]

[*] *Computer and Automation Institute of the Hungarian Academy of Sciences.*
[**] *Eötvös Loránd University, Dept. of Probability Theory and Statistics.*
[***] *Eötvös Loránd University, Dept. of Probability Theory and Statistics, Computer and Automation Institute of the Hungarian Academy of Sciences, and National Bank of Hungary.*

Abstract: ARMA modeling of many economic time series leads to processes with heavy - tailed marginal distribution. We present methods of estimating the parameters of such processes. Asymptotic properties of the full information maximum likelihood and partially adaptive estimates are discussed. We give an asymptotic description of the estimation error process in both cases. The results are generalizations of (Philips, 1994) and (Gerencsér, 1990). *Copyright ©2002 IFAC.*

Keywords: Linear systems, non - Gaussian processes, identification, ARMA parameter estimation.

1. INTRODUCTION

It has long been known that least squares (LS) estimates of ARMA parameters is not efficient if the distribution of the error term is not normal, and even if it is normal, LS is very sensible to the presence of outliers. Several methods of robust estimation have been proposed as alternatives to least squares which are asymptotically more efficient than LS. Most of these estimation procedures are special cases of M - estimation, differing only in the choice of the particular score function. For example, (Tiku *et al.*, 2000) consider a transformation of Student's t distribution, (Goldfeld and Quandt, 1981) uses a generalization of the Laplace distribution, and (McDonald, 1989) introduces the generalized t distribution, which includes many other distributions as special or limiting cases.

In these papers the authors do not assume any particular knowledge of the marginal distribution of the process or the distribution of the innovation process. On the other hand, it is quite a frequent situation that one of these distributions is known to belong to a parameterized family of distributions. In (Li and McLeod, 1988) the authors examine ARMA processes with gamma and log - normal innovations, and mention that these distributions may arise in hydrology in connection with daily precipitation, while (Abraham and Balakrishna, 1999) describes statistical inference on autoregressive processes with inverse Gaussian marginal distributions in their study that had been motivated by problems in lifetime models. We should also mention the recent results of (Barndorff-Nielsen, 1977), (Eberlein and Keller, 1995), (Eberlein *et al.*, 1998) and (Eberlein, 1999) concerning the distribution of daily stock returns. The results of a survey carried out along the same line, examining the marginal distribution of shares traded on the Budapest Stock Exchange, are shown in Figures 1, 2 and 3.

If the distribution of the innovation process is known, then maximum likelihood (ML) estimation of the ARMA parameters can be carried out, and this estimate is asymptotically more efficient than LS. In most cases however, the true value of the parameter describing the distribution of the innovation is unknown and

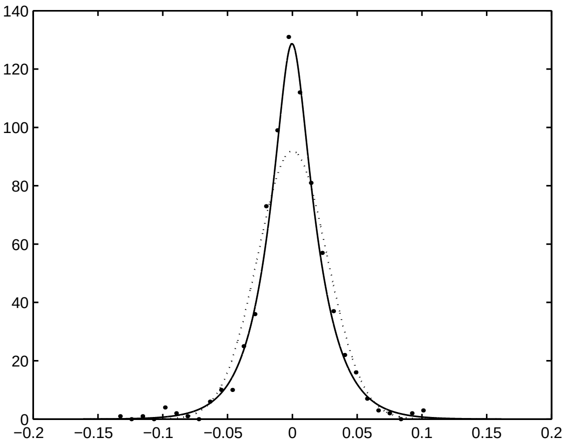

Figure 1. Daily returns of MOL, and fitted normal and hyperbolic densities.

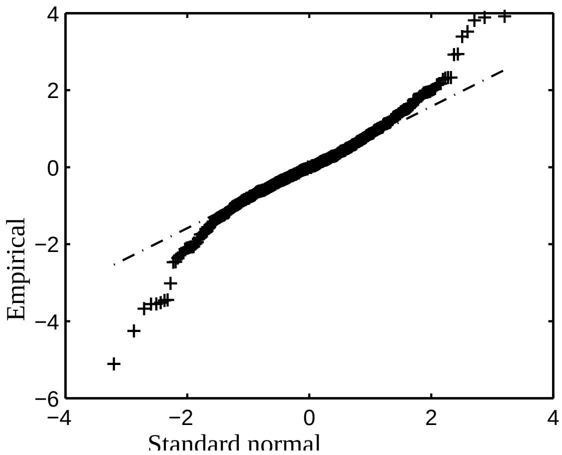

Figure 2. Standard normal qq - plot.

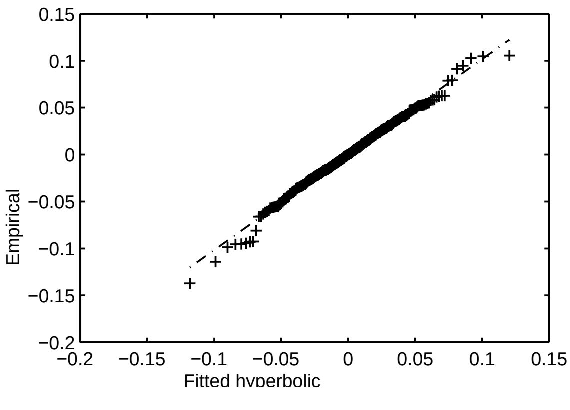

Figure 3. Fitted hyperbolic qq - plot.

has to be estimated too. In (Li and McLeod, 1988)) the authors describe the properties of the simultaneous or *full information maximum likelihood* (FIML) estimation (cf. also (McDonald and Xu, 1994)), while (Philips, 1994) proposes an iterative or *partially adaptive* estimation method, in which the ML problem is decoupled into two separate problems: the estimation of the system parameters and the estimation of the unknown parameters in the distribution of the innovation. This method goes back to (Beran, 1976). However, this latter approach is carried out as a version of M - estimation, with a score function that is a mixture of zero mean normals. Both of these estimates are shown to be $\sqrt{N}$ consistent and asymptotically normal.

The aim of the present paper is to generalize a result of (Gerencsér, 1990), concerning the rate of convergence of the LS or *prediction error* estimate of ARMA parameters, to the case of ML estimates. The prediction error process is shown to be of order $O_M\left(N^{-1}\right)$. This result holds for both FIML and partially adaptive estimates. We also prove that the asymptotic covariance matrix of the partially adaptive estimate is equal to that of the ML estimate with known distribution parameters. The proposed method, with slight modification, is applicable also for misspecified models, in which the distributions that we consider does not contain the true distribution of the innovation. The analysis is carried out using the techniques of (Gerencsér, 1989).

2. NOTATION AND ASSUMPTIONS

Let $y = (y_n), n = 0, \pm 1, \pm 2, \ldots$ be a second order stationary ARMA (p, q) process defined by the following difference equation:

$$y_n + a_1^* y_{n-1} + \cdots + a_p^* y_{n-p} \\ = e_n + c_1^* e_{n-1} + \cdots + c_q^* e_{n-q}. \qquad (1)$$

Let q^{-1} denote the backward shift operator acting as $(q^{-1}y)_n = y_{n-1}$ and let us define

$$A^*(q^{-1}) = \sum_{i=0}^{p} a_i^* q^{-i},$$
$$C^*(q^{-1}) = \sum_{i=0}^{q} c_i^* q^{-i},$$

with $a_0^* = c_0^* = 1$. Then (1) can be written as $A^* y = C^* e$. Replacing q by the complex variable z we get the polynomials $A^*(z^{-1})$ and $C^*(z^{-1})$.

The following condition is standard in the system identification literature (cf. (Caines, 1988), (Hannan and Deistler, 1988), (Ljung, 1987) and (Söderström and Stoica, 1989)).

Condition 1. The two polynomials $A^*(z^{-1})$ and $C^*(z^{-1})$ are relative prime, and all of their roots have absolute value strictly less than one, i.e. $A^*(z^{-1})$ and $C^*(z^{-1})$ are stable.

Condition 2. The process $e = (e_n)$ consists of a sequence of independent, identically distributed (i.i.d.) random variables with zero mean, finite variance and probability distribution function $f(x, \eta^*)$, where the parameter vector η^* is in an open domain $F \subseteq \mathbb{R}^d$. We assume furthermore that

i. $\log f$ is three times differentiable in both x and η, and its third derivative is locally Lipschitz -

continuous with a Lipschitz constant $K(x, \eta)$ that is polynomially increasing in both of its variables,

ii. f satisfies certain standard regularity conditions (conditions RR in (Borovkov, 1998), Chapter 24.).

Among others, Condition 1 ensures that a wide sense stationary solution of (1) exists, while Condition 2 is introduced to enable us to use the results of (Gerencsér, 1989).

Remark 1. It is easy to verify, using a Taylor - expansion, that all the derivatives of $\log f$ up to the order three satisfy the conditions that we imposed on the third derivative, and these derivatives are also polynomially increasing. This consequence of Condition 2 will be necessary in the proof of Theorem 7.

3. ESTIMATION OF THE ARMA PARAMETERS

Let us introduce the notation $\theta^* = (a_1^*, \ldots, a_p^*, c_1^* \ldots, c_q^*)^T$. The maximum likelihood (or ML) estimation of θ^* – assuming that η^* is known – can be carried out using ideas which are well - known in the engineering literature (cf. (Caines, 1988) and (Ljung, 1987)). This can roughly be described as follows. Suppose there exist a known compact domain D_0 in $\mathbb{R}^{p+q}$ such that $\theta^* \in \text{int} D_0 \subseteq D$, where D is the (open) set of those θ - s in $\mathbb{R}^{p+q}$ for which Condition 1 is satisfied. For a given $\theta \in D_0$, let $\varepsilon_n(\theta)$ denote the estimated prediction error process defined as

$$\varepsilon = (A/C)\, y, \quad n \geq 0, \tag{2}$$

using zero initial conditions. Then find the value $\hat{\theta}_N$ of θ such that

$$V_N(\theta) = -\sum_{n=1}^{N} \log f\left(\varepsilon_n(\theta), \eta^*\right) \tag{3}$$

is minimized in D_0. The details of this minimization procedure will be described below. Define

$$W(\theta) = \lim_{n\to\infty} E\left(-\log f(\varepsilon_n(\theta), \eta^*)\right).$$

It is easy to see that the equation

$$\frac{\partial}{\partial\theta} W(\theta) = W_\theta(\theta) = 0 \tag{4}$$

is solved by $\theta = \theta^*$.

Condition 3. Equation (4) has a unique solution in D_0, and the Hessian - matrix $W_{\theta\theta}(\theta^*)$ is nonsingular.

This condition is in general difficult to verify, but for Gaussian processes it has been verified in (Åström and Söderström, 1974).

We define the maximum likelihood estimate $\hat{\theta}_N$ of θ^* as the solution of the equation

$$\sum_{n=1}^{N} \frac{\partial}{\partial\theta} \log f(\varepsilon_n(\theta), \eta^*) = 0 \tag{5}$$

in D_0 if such a solution exists, and an arbitrary point in D_0, ensuring only that $\hat{\theta}_N$ is measurable, if such a solution does not exists or there are more than one solutions. The following result describes the asymptotic behaviour of this estimate. Here, Σ_{ML} denotes the asymptotic covariance matrix of $\hat{\theta}_N$, i.e.

$$\Sigma_{ML} = \lim_{N\to\infty} NE\left(\left(\hat{\theta}_N - \theta^*\right)\left(\hat{\theta}_N - \theta^*\right)^T\right).$$

Theorem 2. Assume that Conditions 1, 2 and 3 are satisfied. Then

$$\begin{aligned} &\hat{\theta}_N - \theta^* \\ &= -W_{\theta\theta}^{-1}(\theta^*) \frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial\theta} \log f(\varepsilon_n(\theta), \eta^*)\Big|_{\theta=\theta^*} \\ &\quad + O_M(N^{-1}). \end{aligned} \tag{6}$$

For the asymptotic covariance matrix of $\hat{\theta}_N$, we have

$$\Sigma_{ML} = W_{\theta\theta}^{-1}(\theta^*) P^* (W_{\theta\theta}^T(\theta^*))^{-1}, \tag{7}$$

where P^* is given by

$$\begin{aligned} P^* = \lim_{m\to\infty} E\Bigg(&\left(\frac{\partial}{\partial\theta} \log f(\varepsilon_m(\theta), \eta^*)\right)^T \\ &\times \left(\frac{\partial}{\partial\theta} \log f(\varepsilon_m(\theta), \eta^*)\right)\Bigg|_{\theta=\theta^*}\Bigg). \end{aligned} \tag{8}$$

PROOF. Equation (6) is a special case of Theorem A3 of the Appendix (see also Remark A4). Equation (7) is a direct consequence of (6).

Remark 3. The analogous results for the prediction error (PE) estimate $\tilde{\theta}$ of θ^*, defined as the value that minimizes

$$V_N(\theta) = \sum_{n=1}^{N} \varepsilon_n^2(\theta)$$

in D_0 were proved in (Gerencsér, 1990) (cf. also (Caines, 1988) for more on the prediction error method).

Let Σ_{PE} denote the asymptotic covariance matrix of the prediction error estimation. It is well known that that $\Sigma_{PE} = (R^*)^{-1}(\sigma^*)^2$ with

$$R^* = \lim_{n\to\infty} E\left[\varepsilon_{\theta n}^T(\theta^*) \varepsilon_{\theta n}(\theta^*)\right].$$

The following Lemma gives us a similar formula for Σ_{ML}.

Lemma 4. Suppose Condition 1, 2 and 3 are satisfied, then the asymptotic covariance matrix of the ML estimation is

$$\Sigma_{ML} = (R^*)^{-1} \mu^{-1}, \tag{9}$$

with the notation

$$\mu = \lim_{n\to\infty} E\left(\frac{f'(e_n)}{f(e_n)}\right)^2,$$

with the simplified notation $f(x) = f(x, \eta^*)$.

Remark 5. The quantity μ introduced above is the Fisher-information of the density f with respect to the location parameter.

PROOF. Define $h = f'/f$. For the matrix P^* defined under (8), we have

$$P^* = \lim_{m\to\infty} E\left[\varepsilon_{\theta,m}^T(\theta^*)\varepsilon_{\theta,m}(\theta^*)\right] Eh^2(e_m) = R^*\mu. \quad (10)$$

Here we used the fact that $\varepsilon_n(\theta^*) = e_n + O_M(\alpha^n)$ with some $0 < \alpha < 1$ and that e_m is independent from $\varepsilon_{\theta,m}$.

Similarly, $W_{\theta\theta}(\theta^*)$ equals

$$-\lim_{n\to\infty} E\left(\frac{f''(\varepsilon_n(\theta^*))}{f(\varepsilon_n(\theta^*))}\varepsilon_{\theta,n}^T(\theta^*)\varepsilon_{\theta,n}(\theta^*) - \frac{f'(\varepsilon_n(\theta^*))}{f^2(\varepsilon_n(\theta^*))}f'(\varepsilon_n(\theta^*))\varepsilon_{\theta,n}^T(\theta^*)\varepsilon_{\theta,n}(\theta^*) + \frac{f'(\varepsilon_n(\theta^*))}{f(\varepsilon_n(\theta^*))}\varepsilon_{\theta\theta,n}(\theta^*)\right).$$

Since e_n is independent from $\varepsilon_{\theta\theta,n}(\theta^*)$, the last term in this equation is equal to zero, for we have due to the regularity conditions imposed on f

$$Eh(e_n) = \int_{-\infty}^{\infty} \frac{f'(x)}{f(x)} f(x)dx = \int_{-\infty}^{\infty} f'(x)dx = 0,$$

while for the first term we get

$$\lim_{n\to\infty} E\left(\frac{f''(\varepsilon_n(\theta^*))}{f(\varepsilon_n(\theta^*))}\varepsilon_{\theta,n}^T(\theta^*)\varepsilon_{\theta,n}(\theta^*)\right) = R^* \int_{-\infty}^{\infty} \frac{f''(x)}{f(x)} f(x)dx = 0,$$

using again the regularity of f. Thus we get, using the same arguments

$$W_{\theta\theta}(\theta^*) = Eh^2(e_n)R^* = R^*\mu, \quad (11)$$

and (9) follows from (7).

4. ESTIMATION WITH UNKNOWN DISTRIBUTION PARAMETERS

A more realistic situation is when the parameter vector η describing the distribution of the e_n's and the system parameter θ are both unknown and to be estimated from data. This estimation can be carried out in two ways, by using the *„full information maximum likelihood"* (FIML) method, or by using an iterative, or *„partially adaptive"* method (see (Beran, 1976), (McDonald and Xu, 1994) and (Philips, 1994)). In this section we concentrate on the first one, while the iterative estimation is discussed in the next section.

Let us introduce the notation $\nu^T = (\theta^T, \eta^T)$ and $\nu^{*T} = (\theta^{*T}, \eta^{*T})$. The FIML method gives us an estimate $\hat{\nu}_N$ of ν^* in essentially the same way as the estimate $\hat{\theta}_N$ was given in the preceding section. Define

$$V_N(\nu) = -\sum_{n=1}^{N} \log f(\varepsilon_n(\theta), \eta),$$
$$W(\nu) = -\lim_{n\to\infty} E\left(\log f(\varepsilon_n(\theta), \eta)\right). \quad (12)$$

Let F denote the interior of the set of those η-s in $\mathbb{R}^d$ for which $f(x, \eta)$ is defined and suppose that there exist a known compact domain F_0 in $\mathbb{R}^d$ such that $\eta^* \in \mathrm{int}F_0 \subseteq F$.

Condition 4. ν^* is the unique solution of the equation $W_\nu(\nu) = 0$ in $D_0 \times F_0$, and $W_{\nu\nu}(\nu^*)$ is a positive definite matrix.

The FIML estimate $\hat{\nu}_N = (\hat{\theta}_N, \hat{\eta}_N)$ of ν^* is the solution of the equation

$$V_{\nu,N} = -\sum_{n=1}^{N} \frac{\partial}{\partial\nu} \log f(\varepsilon_n(\theta), \eta) = 0, \quad (13)$$

in $D_0 \times F_0$, if such a solution exists, and an arbitrary point in $D_0 \times F_0$, ensuring only that $\hat{\nu}_N$ is measurable, if such a solution does not exists or there are more than one solutions. The following lemma describes a very useful property of this estimate, which is proved by calculating the asymptotic covariance matrix of $\hat{\nu}_N$.

Lemma 6. Suppose Condition 1, 2 and 4 are satisfied. Then the estimates $\hat{\theta}_N$ and $\hat{\eta}_N$ are asymptotically uncorrelated.

PROOF. The statement is a straightforward consequence of Theorem A3, but the calculations are quite cumbersome, therefore we present only the final form of the asymptotic covariance matrix of $\hat{\nu}_N$, which we denote by Σ. With Δ defined as

$$\Delta = -\lim_{n\to\infty} E\left(\frac{f_{\eta\eta}(e_n, \eta^*)}{f(e_n, \eta^*)} - \frac{f_\eta^T(e_n, \eta^*)f_\eta(e_n, \eta^*)}{f^2(e_n, \eta^*)}\right) = \lim_{n\to\infty} E\left(\frac{f_\eta^T(e_n, \eta^*)f_\eta(e_n, \eta^*)}{f^2(e_n, \eta^*)}\right),$$

we have

$$\Sigma = \begin{pmatrix} (R^*)^{-1}\mu^{-1} & 0 \\ 0 & \Delta \end{pmatrix}. \quad (14)$$

The fact that the FIML estimates of the system parameter θ^* and the distribution parameter η^* are uncorrelated suggests (following Philips, (Philips, 1994)), that if we are initially given a "sufficiently good" estimate of the distribution parameters, and we use this value instead of the true parameter vector η^* to estimate the system parameters as in Section 2, then the final $\hat{\theta}$ will also have "good" properties. This is described in the next section.

5. PARTIALLY ADAPTIVE ESTIMATION OF ARMA PARAMETERS

Let us be given an initial estimate $\hat{\eta}_N$ of η^* which satisfies

$$\hat{\eta}_N - \eta^* = O_M\left(N^{-1/2}\right). \qquad (15)$$

We assume also that $\hat{\eta}_N \in F_0$, and therefore we can define the partially adaptive estimate $\overline{\theta}_N(\hat{\eta}_N)$ of θ^* as the solution of the equation

$$V_N(\theta) = -\sum_{n=1}^{N} \frac{\partial}{\partial\theta} \log f(\varepsilon_n(\theta), \hat{\eta}_N) = 0, \qquad (16)$$

if such a solution exists, and an arbitrary point in D_0 if such a solution does not exist. The next theorem describes the asymptotic behaviour of this estimate.

Theorem 7. Under Condition 1, 2 and 4 we have

$$\begin{aligned} &\overline{\theta}_N(\hat{\eta}_N) - \theta^* \\ &= -(R^*)^{-1}\mu^{-1}\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\theta}\log f(\varepsilon_n(\theta),\eta^*)\Big|_{\theta=\theta^*} \\ &\quad + O_M(N^{-1}). \end{aligned} \qquad (17)$$

PROOF. For a fixed $\eta \in F$ we may write, due to Theorem A3 and Remark A4

$$\begin{aligned} &\overline{\theta}(\eta) - \theta^* \\ &= -W_{\theta\theta}^{-1}(\theta,\eta)\Big|_{\theta=\theta^*}\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\theta}\log\left(f(\varepsilon_n(\theta),\eta\right)\Big|_{\theta=\theta^*} \\ &\quad + O_M(N^{-1}), \end{aligned} \qquad (18)$$

with $W(\theta,\eta)$ defied as in (12). Let $F' \subseteq F_0$ be a compact domain such that $\eta^* \in F'$ and for any $\eta \in F'$ the matrix $W_{\theta\theta}(\theta^*,\eta)$ is bounded away from singularity, or more precisely

$$\sup_{\eta\in F'} \|W_{\theta\theta}^{-1}(\theta^*,\eta)\| \le K < \infty. \qquad (19)$$

The existence of such an F' is justified by Condition 4 and the continuity of the function $W_{\theta\theta}(\theta^*,\eta)$ in η. We modify our initial estimate $\hat{\eta}_N$ so that $\hat{\eta}_N \in F'$. Let us introduce the notation

$$g_n(\theta,\eta) = \frac{\partial}{\partial\theta}\log(f(\varepsilon_n(\theta),\eta)).$$

Since the process $(g_n(\theta^*,\eta))$ is L-mixing (uniformly in $\eta \in F'$) by Theorem A1, it follows from Theorem A2 that

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\theta}\log\left(f(\varepsilon_n(\theta),\eta)\right)\Big|_{\theta=\theta^*} = O_M\left(N^{-1/2}\right), \qquad (20)$$

uniformly in $\eta \in E'$. Furthermore, it follows from (15) and (19) that

$$\|W_{\theta\theta}^{-1}(\theta^*,\hat{\eta}_N) - W_{\theta\theta}^{-1}(\theta^*,\eta^*)\| = O_M\left(N^{-1/2}\right), \qquad (21)$$

except for an event of probability $O_M\left(N^{-1/2}\right)$. If we truncate $W_{\theta\theta}^{-1}(\theta^*,\hat{\eta}_N)$ on this event to be the identity matrix of appropriate dimension, then we see from (20) and (21) that the asymptotic behaviour of the estimation error $\overline{\theta}_N(\hat{\eta}_N) - \theta^*$ is – up to an error of order $O_M(N^{-1})$ – determined by the sum in (18), that is

$$\begin{aligned} &\overline{\theta}_N(\hat{\eta}_N) - \theta^* \\ &= -(R^*)^{-1}\mu^{-1}\frac{1}{N}\sum_{n=1}^{N} g_n(\theta^*,\hat{\eta}_N) \\ &\quad + O_M(N^{-1}). \end{aligned} \qquad (22)$$

To handle this expression we use the Taylor-expansion of $g_n(\theta^*,\eta)$ around η^*, which is

$$\begin{aligned} &g_n(\theta^*,\hat{\eta}_N) \\ &= g_n(\theta^*,\eta^*) + \frac{\partial g_n}{\partial\eta}(\theta^*,\eta^*)\,(\hat{\eta}_N - \eta) + O_M(N^{-1}). \end{aligned}$$

The statement of the theorem now follows from (15) and from Theorem A1 and A2 applied to the process $u_n = \frac{\partial g_n}{\partial\eta}(\theta^*,\eta^*)$.

Remark 8. Comparing (17) and (24) we see that the asymptotic covariance matrix of the partially adaptive estimator of θ^* is equal to the upper-left diagonal block of the FIML estimator of ν^*.

The following lemma describes one of the many possible methods that one can obtain an initial estimation $\hat{\eta}$ satisfying (15).

Lemma 9. Let $\tilde{\theta}_N$ denote the PE estimator of θ, and let $\hat{\eta}_N$ be the solution of

$$V_N(\eta) = \frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\eta}\log f(\varepsilon_n(\tilde{\theta}_N),\eta) = 0. \qquad (23)$$

Then $\hat{\eta}_N$ satisfies (15).

PROOF. The proof is similar to that of Theorem A2.

6. CONCLUSION

A strong approximation result for the maximum-likelihood estimation of non-Gaussian ARMA processes has been presented, which is a direct generalization of the results of (Gerencsér, 1990). Further more we have developed a computationally efficient method to get a statistically efficient estimator of the system parameters. The application of these results for quatifying the error of adaptive predictors, following (Gerencsér, 1994) is under progress.

7. ACKNOWLEDGEMENT

The first author acknowledges the support of the National Research Foundation of Hungary (OTKA) under Grant no. T 032932.

8. REFERENCES

Abraham, B. and N. Balakrishna (1999). Inverse Gaussian autoregressive models. *Journal of Time Series Analysis* **20**(6), 605–618.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society London A* **353**, 401–419.

Beran, R. (1976). Adaptive estimates for autoregressive processes. *Annals of the Institute of Statistical Mathematics* **28**, 77–89.

Borovkov, A. A. (1998). *Mathematical Statistics*. ELTE. In Hungarian.

Caines, P. (1988). *Linear Stochastic Systems*. Wiley.

Eberlein, E. (1999). Application of generalized hyperbolic Lévy motion to finance.

Eberlein, E. and U. Keller (1995). Hyperbolic distributions in finance. *Bernoulli* **1**, 281–299.

Eberlein, E., U. Keller and K. Prause (1998). New insights into smile, mispricing and value at risk: The hyperbolic model. *Journal of Business* **71**, 371–406.

Gerencsér, L. (1989). On a class of mixing processes. *Stochastics* **26**, 165–191.

Gerencsér, L. (1990). On the martingale approximation of the estimation error of ARMA parameters. *System & Control Letters* **15**, 417–423.

Gerencsér, L. (1994). On Rissanen's predictive stochastic complexity for stationary ARMA processes. *Statistical Planning and Inference* **41**, 303–325.

Goldfeld, S. M. and R. E. Quandt (1981). Econometric modelling with non-normal disturbances. *Journal of Econometrics* **17**, 141–155.

Hannan, E.J. and M. Deistler (1988). *The statistical theory of linear systems*. Wiley.

Li, W. K. and A. I. McLeod (1988). ARMA modelling with non-Gaussian innovations. *Journal of Time Series Analysis* **9**(2), 155–168.

Ljung, L. (1987). *System identification: theory for the user*. Prentice Hall. Englewood Cliffs.

McDonald, J. B. (1989). Partially adaptive estimation of ARMA time series models. *International Journal of Forecasting* **5**, 217–230.

McDonald, J. B. and Yexiao Xu (1994). Some forecasting applications of partially adaptive estimators of ARIMA models. *Economics Letters* **45**, 155–160.

Philips, R. F. (1994). Partially adaptive estimation via a normal mixture. *Journal of Econometrics* **64**, 123–144.

Åström, K.J. and T. Söderström (1974). Uniqueness of the maximum-likelihood estimates of the parameters of an ARMA modell. *IEEE Trans. Automat. Contr.* **AC-19**, 769–773.

Söderström, T. and P. Stoica (1989). *System identification*. Prentice Hall.

Tiku, M. L., Wing-Keung Wong, D. C. Vaughan and Gouroi Bian (2000). Time series models in non-normal situations: symmetric innovations. *Journal of Time Series Analysis* **21**(5), 571–596.

9. APPENDIX

For the definition of M - boundedness, L - mixing and some related properties the reader is asked to consult (Gerencsér, 1989) and (Gerencsér, 1990).

Theorem A1. Let (x_n) be an L - mixing process with respect to $(\mathfrak{F}_n, \mathfrak{F}_n{}^+)$, and let $f(x)$ be a Lipschitz - continuous function. Then the process $(y_n) = (f(x_n))$ is also L - mixing with respect to $(\mathfrak{F}_n, \mathfrak{F}_n{}^+)$.

PROOF. Let us define $x^+_{n,n-\tau} = E\left(x_n | \mathfrak{F}^+_{n-\tau}\right)$ and $y^{++}_{n,n-\tau} = f(x^+_{n,n-\tau})$. Let K be the Lipschitz constant of f. Then we have

$$\gamma_q(\tau, y) \leq 2E^{1/q}\left|y_n - y^{++}_{n,n-\tau}\right|^q$$
$$= 2E^{1/q}\left|f(x_n) - f(x^+_{n,n-\tau})\right|^q \leq 2K\gamma_q(x,\tau),$$

by the previous lemma, and the claim follows by definition.

Define

$$\Delta x/\Delta^\alpha\theta = \left|x_n(\theta + h) - x_n(\theta)\right|/|h|^\alpha$$

for $n \geq 0$, $\theta \neq \theta + h \in D$ with $0 < \alpha \leq 1$.

Theorem A2. Suppose that the processes $(x_n(\theta))$ and $\Delta x/\Delta\theta$ are both L - mixing uniformly in $\theta, \theta + h \in D$, and $Ex_n(\theta) = 0$ for all $n \geq 0$ and $\theta \in D$. Then, for a compact domain $D_0 \subseteq \text{int}D$

$$\sup_{\theta \in D_0}\left|\frac{1}{N}\sum_{n=1}^{N} x_n(\theta)\right| = O_M\left(N^{-1/2}\right).$$

Theorem A3. Let the process (y_n) be given by (1), and assume that Conditions 1, 2 and 4 are satisfied. Then we have

$$\hat{\nu}_N - \nu^* = -W^{-1}_{\nu\nu}(\nu^*)\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\nu}\log f(\varepsilon_n(\theta), \eta)\bigg|_{(\theta^*,\eta^*)} + O_M(N^{-1}). \qquad (24)$$

PROOF. The proof of this result is analogous to the proof of Theorem 2.1 of (Gerencsér, 1990) and is omitted for the sake of brevity.

Remark A4. It is easy to see that the statements of equation (6) and (18) are consequences of this result. Indeed, since by Lemma 6 the two components of $\hat{\nu}_N = (\hat{\theta}_N, \hat{\nu}_N)$ are asymptotically uncorrelated, equation (24) holds "componentwise", and this is equivalent to saying that (18) is true for any $\eta \in F_0$, and (6) is (18) with $\eta = \eta^*$.