15th Triennial World Congress, Barcelona, Spain

CONVERGENCE ANALYSIS OF THE LEAST-SQUARES ESTIMATES FOR INFINTE AR MODELS

Yulia Gel *, **Andrey Barabanov** **

* *Saint-Petersburg State University, Russia and University of Washington, Box 354320, Seattle, WA 98195, USA*
** *Saint-Petersburg State University, Bibliotechnaya sqr. 2, 198904 Saint-Petersburg, Russia*

Abstract: In this paper time series identification problem amounts to estimating the unknown parameters of an ARMA model, which is transformed to an infinite AR model and the least-squares method is proposed for its identification. The convergence analysis of the LS estimates almost surely is carried out for an infinite case. Moreover, it is established the result on the estimate of the degree of convergence of the LS estimates for infinite AR model. Such an approach has been studied before for the "long" AR models but an overall convergence analysis has been lacking. In addition, a complimentary result on the convergence of semi-martingales is presented here, which is a corner-stone for proof of all theorems here, but is of interest by itself.

Keywords: ARMA models, autoregressive models, parameter estimation, identification, least-squares method

1. INTRODUCTION

One of the most used linear models of stochastic time series is a regressive equation, i.e. an ARMA model. The problem of estimating unknown parameters of a regressive equation is a corner-stone of mathematical statistics, and an extensive literature is devoted to this question. A special modification of the least-squares method (LSM), known as the extended least-squares method (ELSM), was introduced and justified under the positive realness condition on the transfer function of a filter. The estimates provided by ELSM are unbiased and strongly consistent, e.g. converge almost surely to the unknown parameters of an ARMA model. However, the positive realness condition is rather a severe restriction on a class of the considered time series. It is not surprising that the development of new algorithms without the positive-realness condition is of constant interest to both theoreticians and engineers, as reflected in the significant number of publications. Let us discuss briefly the results achieved.

An alternative approach to this identification problem is to transform the ARMA model to an AR model of infinite order. A historical overview of this approach is given by Mari et al. (2000). According to Mari et al. (2000) the idea of approximating an ARMA process by AR processes of high order go back to Wold (1938), Durbin (1959) and Whittle (1953). To that list it may also be added the book of Marple-Jr. (1987). Following such an approach, Mari and coauthors suggested an identification algorithm for ARMA models based on a three-step procedure: 1) empirical estimation of a partial covariance sequence; 2) covariance extension by the maximum-entropy method, leading to a high order AR model with the transfer function $\hat{W}_\upsilon(z) = z^\upsilon/\hat{\phi}_\upsilon(z)$, where $\hat{\phi}_\upsilon(z)$ is the normalized Szego polynomial of degree υ, which is com-

[1] Supported by the Russian Foundation for Basic Research under grant N 01-01-00306, by the grant for Russian scientific schools N 00-15-96-028 and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the US Office of Naval Research under Grant N00014-01-10745.

puted from the estimated covariance data; 3) determination of a reduced-degree approximation $\hat{W}_v(z)$ of $\hat{W}(z)$ via stochastically balanced truncation. The proposed algorithm shares certain features with the subspace method of identification. In particular, both of them are based on partial stochastic realization theory. However, in contrast to subspace methods the method presented by Mari et al. (2000) guarantees the minimal phase property. It is stipulated by using stochastically balanced truncation by Mari et al. (2000). The authors developed a simple computational procedure and provided theoretical analysis and a simulation example. The idea of using model truncation for systems identification appears in the works of Wahlberg (1989), Green (1995).

However, it should be emphasized that all above-mentioned papers considered the approximation of the ARMA model by ”long” AR models of finite order. For the first time the problem of transforming an ARMA model to an AR model of infinite order was introduced by V.N. Fomin (Gel and Fomin (1998)). In the paper of Gel and Fomin (1998) an identification method for the stationary stochastic time series was proposed. The method is based on the analysis of an infinite AR equation, coefficients of which are estimated by the Yule-Walker method, and on the subsequent reconstruction of the parameters of the ARMA model by the Padé approximation. It should be mentioned that the Padé approximation needs to estimate only a finite number of coefficients of the infinite AR model to reconstruct coefficients of the ARMA model. The proposed identification algorithm is easily implemented as it involves only linear algebra operations and no nonconvex optimization computations are required. The detailed investigation of connection between the approximation by ”long” autoregressive processes and the AIC and BIC criterions is carried out in the book of Hannan and Diestler (1988).

Here the approach suggested by Gel and Fomin (1998, 2001) is developed further, and an AR model of infinite order is considered. The least-squares method is used to estimate unknown parameters of the infinite AR model. In this paper presented the analysis of consistency and the degree of convergence of the LS estimates for an infinite case, which is by no means a new result for systems identification.

2. THE PROBLEM STATEMENT

Time series identification problem studied here amounts to estimating the unknown system parameters for the ARMA model

$$a(\nabla)y_t = \sigma^2 b(\nabla)v_t \tag{1}$$

from the data $\{y_t\}_{t=1}^{\infty}$, where $a(\lambda)$ and $b(\lambda)$ are polynomials ($a(\lambda) = 1 + \lambda a_1 + \cdots + \lambda^p a_p, b(\lambda) = 1 + \lambda b_1 + \cdots + \lambda^p b_q$); v_t is the martingale difference ($\mathbf{E}(v_t|\mathcal{F}_{t-1}) \equiv 0$, $\mathbf{E}(v_t^2|\mathcal{F}_{t-1}) = 1$ a.s.; here $\mathcal{F}_{t-1}$ is the σ-algebra generated by the stochastic variables $v_1, v_2, \ldots, v_{t-1}$), and $\sup_t \mathbf{E}v_t^4 < \infty$; ∇ is the shift-back operator ($\nabla y_t = y_{t-1}$).

If system (1) is minimal phase ($b(\lambda) \neq 0, |\lambda| \leq 1$), it may be transformed to

$$\tilde{a}(\nabla)y_t = \sigma^2 v_t, \qquad \tilde{a}(\lambda) = \frac{a(\lambda)}{b(\lambda)} = \sum_{k=0}^{\infty} \lambda^k \tilde{a}_k. \tag{2}$$

This linear system is called an AR model of infinite order. Write it in the form of a linear observation scheme

$$y_t = \mathbf{\Phi}_{t-1}^* \boldsymbol{\tau}_* + \sigma^2 v_t, \tag{3}$$

where $\mathbf{\Phi}_{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_1, 0, \ldots)$, $\boldsymbol{\tau}_* = -(\tilde{a}_1, \tilde{a}_2, \ldots)$.

The vector of unknown parameters $\boldsymbol{\tau}_*$ is estimated using the recursive LS method

$$\boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{\gamma}_t^{\varepsilon} \mathbf{\Phi}_t (y_{t+1} - \mathbf{\Phi}_t^* \boldsymbol{\tau}_t) \tag{4}$$

$$\boldsymbol{\gamma}_{t+1}^{\varepsilon} = \boldsymbol{\gamma}_t^{\varepsilon} - \boldsymbol{\gamma}_t^{\varepsilon} \mathbf{\Phi}_{t-1} (1 + \mathbf{\Phi}_t^* \boldsymbol{\gamma}_t^{\varepsilon} \mathbf{\Phi}_t)^{-1} \mathbf{\Phi}_t^* \boldsymbol{\gamma}_t^{\varepsilon}. \tag{5}$$

The matrix $\boldsymbol{\gamma}_t^{\varepsilon}$ is inverse to the information matrix, $\boldsymbol{\gamma}_t^{\varepsilon} = (\mathbf{R}_t^{\varepsilon})^{-1}$, where $\mathbf{R}_t^{\varepsilon} = \sum_{k=1}^{t} \mathbf{\Phi}_k \mathbf{\Phi}_k^* + \varepsilon \mathbf{R}$, a $\mathbf{R} = diag\{e^{\mu k}\}_{k=1}^{\infty}$ is a regularizer. The idea of such a regularizer for the LSM of infinite order and its theoretical basis belong to V.N. Fomin (see Gel and Fomin (2001)). The estimates (4) are called the regularized estimates of the least-squares method.

The main results from the analysis of consistency of the LS estimates $\boldsymbol{\tau}_t$ for the infinite case are the following:

1. The estimates $\boldsymbol{\tau}_t$ converge with probability 1 to the vector of unknown parameters $\boldsymbol{\tau}_*$, i.e. are strongly consistent. In this connection, it is shown that the regularized information matrix $\mathbf{R}_t^{\varepsilon} = (\sum_{k=1}^{T} \mathbf{\Phi}_k \mathbf{\Phi}_k^* + \varepsilon \mathbf{R})$ is strictly positive definite as $t \to \infty$.

2. It is established the degree of convergence with probability 1 of the estimates $\boldsymbol{\tau}_t$ to the vector $\boldsymbol{\tau}_*$.

3. THE DEGREE OF CONVERGENCE OF SEMI-MARTINGALES

The result formulated in this section forms the foundation for all subsequent theorems in this paper on the convergence analysis of the LS estimates for the infinite AR model. The idea is suggested by the result on convergence of semi-martingales (see Fomin (1999)) based on the Doob inequality. According to V.N. Fomin formulate it as follows.

Theorem 1. Assume that the sequence of nonnegative stochastic variables $(\xi_n)_{n=0}^{\infty}$ satisfy

$$\mathbf{E}(\xi_{n+1}|\xi_1,\ldots,\xi_n) \le (1+\alpha_n)\xi_n + \zeta_n \qquad (6)$$

where $\alpha_n \ge 0$, $\zeta_n = \zeta_n(\xi_1,\ldots,\xi_n) \ge 0$ and $\sum_{k=1}^{\infty}\alpha_n < \infty$, $\sum_{n=1}^{\infty}\mathbf{E}\zeta_n < \infty$. Then $\xi_n \to \xi$ almost surely and $\mathbf{E}\xi < \infty$.

In the lemma stated below the degree of convergence of the stochastic variables $\{\xi_k\}$ is estimated, thereby extending the previous results obtained for the limiting case.

Lemma 1. Assume the stochastic variables $\xi_t \ge 0$ and ζ_t satisfy

1) $\xi_0 = 0, \quad \forall t \ge 0 \ \ \mathbf{E}(\xi_{t+1}|\xi_t,\ldots,\xi_1) \le \xi_t + \zeta_t;$
2) $\sum_{t=0}^{\infty}\mathbf{E}|\zeta_t| = C < \infty.$

Then

$$\forall X > 0 \quad P\{\forall T \ge 0,\ \ \xi_T \le X\} \ge 1 - \frac{C}{X}. \quad (7)$$

Proof of Lemma. Let $X > 0$. Define the random stopping time τ by

$$\tau = \min\{t \ge 0 \mid \xi_t > X\}$$

with $\tau = \infty$ if the full trajectory is below the level X. For any $t \ge 0$ define the random characteristic function $\chi_\tau(t)$ which is equal to 1 if $\tau > t$ and else equals to 0. Minimum of τ and t will be denoted by $\tau \wedge t$. Then for any $t \ge 0$ it holds

$$\xi_{\tau\wedge t} = \sum_{k=0}^{t-1}\chi_\tau(k)(\xi_{k+1} - \xi_k).$$

Denote the flow of σ–algebras associated with $(\xi_t)_{t=0}^{\infty}$ by $(\mathcal{F}_t)_{t=0}^{\infty}$. Then for any $t > 0$ the random variable $\chi_\tau(t)$ is measurable with respect to $\mathcal{F}_t$ and therefore

$$\begin{aligned}\mathbf{E}\xi_{\tau\wedge t} &= \mathbf{E}\sum_{k=0}^{t-1}\chi_{\tau>k}(\mathbf{E}\{\xi_{k+1} \mid \mathcal{F}_k\} - \xi_k) \quad (8)\\ &\le \sum_{k=0}^{t-1}\mathbf{E}|\zeta_k| \le \sum_{k=0}^{\infty}\mathbf{E}|\zeta_k| = C.\end{aligned}$$

Since $\xi_t \ge 0$ and $\xi_\tau(\omega) > X$ when $\tau(\omega) < \infty$ it holds

$$XP\{\exists t > 0 \ \mid\ \xi_t > X\} \le \liminf_{t\to\infty}\mathbf{E}\xi_{\tau\wedge t} \le C$$

and assertion of Lemma 1 follows. □

Corollary 1. Let $\psi_T = \mu_T\sum_{t=1}^{T}\nu_t\eta_t$, in which μ_T is a decreasing positive function, the random process (η_t) satisfies the relation $\mathbf{E}\{\eta_t|\eta_1,\eta_2,\ldots,\eta_{t-1}\} = 0$; for any $t > 0$ the random variable ν_t is measurable with respect to $(\eta_{t-1},\eta_{t-2},\ldots,\eta_1)$; and

$$\sum_{t=1}^{\infty}\mu_t^2\ \mathbf{E}\nu_t^2\ \mathbf{E}\eta_t^2 \le C < \infty.$$

Then

$$\forall X > 0 \quad P\{\forall T > 0,\ \ |\psi_T|^2 \le X\} \ge 1 - \frac{C}{X}.$$

The assertion directly follows from Lemma 1 with $\xi_t = \psi_t^2$.

Let $\delta_1 > 0$. Corollary 1 allows to describe a degree of convergence to zero of the random variable $\kappa_t = t^{-\delta_1}|\psi_t|^2$ almost surely:

$$\begin{aligned}\forall Y > 0,\ \ \forall T > 0, \quad & P\{\forall t \ge T,\ \ \kappa_t \le t^{\delta_2-\delta_1}Y\}\\ & \ge 1 - \frac{C}{YT^{\delta_2}} \qquad (9)\end{aligned}$$

with $0 < \delta_2 < \delta_1$.

4. THE DEGREE OF CONVERGENCE OF THE LS ESTIMATES FOR AN AR MODEL OF INFINITE ORDER

The question of consistency of the LS estimates for the AR equation of finite order is very well worked out (for an overview see Fomin (1998), Ljung (2000)). However, not so many papers cover the analysis of the degree of convergence even for the finite case (Lai and Wei (1982), Barabanov (1983)). Below the result on the estimate of the degree of convergence of the LS estimates for the AR equation of infinite order is stated.

Assertion. For any positive δ_1 and δ_2 there is a constant $C > 0$ such that

$$\begin{aligned}\forall T_0 > 0 \quad & P\left\{\forall T \ge T_0,\ \ |\boldsymbol{\tau}_T - \boldsymbol{\tau}_*|^2 \le \frac{1}{T^{1-\delta_1-\delta_2}}\right\}\\ & \ge 1 - \frac{C}{T_0^{\delta_2}}. \qquad (10)\end{aligned}$$

The proof of the main assertion follows readily from the following two theorems.

Theorem 2. For all $\delta > 0$,

$$\begin{aligned}\lim_{T\to\infty}\ & (\boldsymbol{\tau}_{T+1} - \boldsymbol{\tau}_*)^*\frac{1}{T^\delta}(\sum_{t=1}^{T}\mathbf{\Phi}_t\mathbf{\Phi}_t^* + \varepsilon\mathbf{R})\\ & \times(\boldsymbol{\tau}_{T+1} - \boldsymbol{\tau}_*) = 0 \qquad (11)\end{aligned}$$

holds with probability 1. Moreover, there is a positive constant C_δ such that

$$\forall X > 0,$$

$$P\Big\{\forall T \geq 0, (\boldsymbol{\tau}_{T+1} - \boldsymbol{\tau}_*)^* \frac{1}{T^\delta}(\sum_{t=1}^{T} \boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^* + \varepsilon \mathbf{R}) \times (\boldsymbol{\tau}_{T+1} - \boldsymbol{\tau}_*) \leq X\Big\} \geq 1 - \frac{C_\delta}{X}. \quad (12)$$

Let us briefly discuss the main idea of the proof of Theorem 2. The justification of assertion (11) is based on the convergence properties of the stochastic variable

$$V_T = (\boldsymbol{\tau}_T - \tau_*)^* \frac{1}{(T-1)^\delta}(\sum_{t=1}^{T-1} \boldsymbol{\Phi}_t {\boldsymbol{\Phi}_t}^* + \varepsilon \mathbf{R}) \times (\tau_T - \tau_*). \quad (13)$$

It is shown that

$$\mathbf{E}(V_{T+1}|V_1, \ldots, V_T) \leq V_T + \frac{\sigma^2}{T^\delta} \boldsymbol{\Phi}_T{}^* \boldsymbol{\gamma}_T \boldsymbol{\Phi}_T,$$

and that

$$\sigma^2 \sum_{t=1}^{T} \frac{1}{t^\delta} \mathbf{E} \boldsymbol{\Phi}_t{}^* \boldsymbol{\gamma}_t \boldsymbol{\Phi}_t < \infty. \quad (14)$$

In view of the inequality in Lemma 1 and arbitrariness of $\delta > 0$ it follows that the stochastic variables V_T converge to 0 with probability 1 as $T \to \infty$, i.e. the assertion of Theorem 2.

The main assertion (10) will follow directly from Theorem 2 if to show that the information matrix $T^{-\delta}(\sum_{t=1}^{T} \boldsymbol{\Phi}_t \boldsymbol{\Phi}^* + \varepsilon \mathbf{R})$ is bounded away from 0 for $T > 0$. Below derived and justified the estimate of the probability that the information matrix is uniformly bounded away from 0. The relation between the consistency of the LS estimates and the behavior of the information matrix in the finite case is studied in the works of Lai and Wei (1982), Barabanov (1983). Extension to the infinite case requires the special bounding of infinite-dimensional matrices, in which the regularizer $\mathbf{R}$ plays a significant role.

Theorem 3. For any $\alpha \in (0, 1)$ there are positive constants C_0, β and T_0 such that for all $T_1 > T_0$

$$P\Big\{\forall T \geq T_1, \ \ \frac{1}{T}(\sum_{k=1}^{T} \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^* + \varepsilon \mathbf{R}) \geq \beta I\Big\} \geq 1 - \frac{C_0}{T_1^\alpha}$$

where $\boldsymbol{\Phi}_k = (y_k, y_{k-1}, \ldots, y_1, 0, 0, \ldots)^*$.

Just as for the previous theorem let us discuss briefly the main idea of the proof.

Choose an arbitrary positive integer N, $N < T$. Afterwards $N = N(T)$ will be chosen as a determistic function of T. The vector $\boldsymbol{\Phi}_k$ has the form

$$\boldsymbol{\Phi}_k = \mathbf{A}^N \Phi_{k-N} + \sum_{j=0}^{N-1} \mathbf{A}^j \mathbf{B} v_{k-j} \quad (15)$$

when $k > N$. The information matrix $\mathbf{R}_T = T^{-1} \sum_{k=1}^{T} \boldsymbol{\Phi}_k \Phi_k^*$ may be divided into three sums (below all present quantities with negative indexes are 0):

$$\frac{1}{T} \sum_{k=1}^{T} \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^* = \mathbf{Q}_{1,T,N} + \mathbf{Q}_{2,T,N} + \mathbf{Q}_{3,T,N}, \quad (16)$$

where

$$\mathbf{Q}_{1,T,N} = \frac{1}{T} \sum_{k=1}^{T} \mathbf{A}^N \boldsymbol{\Phi}_{k-N} \boldsymbol{\Phi}_{k-N}^* \mathbf{A}^{*N},$$
$$\mathbf{Q}_{2,T,N} = 2 \sum_{j=0}^{N-1} \mathbf{A}^N (\frac{1}{T} \sum_{k=1}^{T} \boldsymbol{\Phi}_{k-N} v_{k-j}) \mathbf{B}^* \mathbf{A}^{*j},$$
$$\mathbf{Q}_{3,T,N} = \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} \mathbf{A}^j \mathbf{B} (\frac{1}{T} \sum_{k=1}^{T} v_{k-j} v_{k-i}) \mathbf{B}^* \mathbf{A}^{*i}.$$

Clearly, the matrix $\mathbf{Q}_{1,T,N}$ is nonnegative. Let us derive a lower bound for the matrix $\mathbf{Q}_{2,T,N} + \varepsilon_1 \mathbf{R}/T$, where $\varepsilon = \varepsilon_1 + \varepsilon_2, \ \ \varepsilon_1, \varepsilon_2 > 0$.

Lemma 2. There exist $C_1, C_2, C_3 > 0$ such that for any $X_0 > 0$ and $\varepsilon_3 > 0$ it holds

$$P\Big\{\forall T > 0, \ \ \forall N \in [\frac{\log T}{\mu}, \frac{C_3 T^{1/2 - \varepsilon_3}}{X_0}], \ \mathrm{Re}\, \mathbf{Q}_{2,T,N} + \frac{\varepsilon_1}{T} \mathbf{R} \geq -\frac{C_2 N^2 X_0}{T^{1/2 - \varepsilon_3}} \ I\Big\} \geq 1 - \frac{C_1}{\varepsilon_3 X_0^2}$$

where I is the identity operator.

Turn now to a bound on the matrix $\mathbf{Q}_{3,T,N}$. Divide it into $\mathbf{Q}_{3,T,N} = \sigma^2 \mathbf{U}_N + \mathbf{W}_{T,N}$ where

$$\mathbf{U}_N = \sum_{i=0}^{N-1} \mathbf{A}^i \mathbf{B} \mathbf{B}^* \mathbf{A}^{*i},$$
$$\mathbf{W}_{T,N} = \sum_{i=0}^{N-1} \mathbf{A}^i \mathbf{B} \frac{1}{T} \sum_{k=1}^{T} (v_{k-i}^2 - \sigma^2) \mathbf{B}^* \mathbf{A}^{*i} + 2\mathrm{Re} \sum_{i=0}^{N-1} \sum_{j=i+1}^{N-1} \mathbf{A}^i \mathbf{B} (\frac{1}{T} \sum_{k=1}^{T} v_{k-i} v_{k-j}) \mathbf{B}^* \mathbf{A}^{*j}$$

and $\mathrm{Re}\mathbf{X} = (\mathbf{X}^T + \mathbf{X})/2$ for any square matrix $\mathbf{X}$.

The sums $\mathbf{U}_N$ and $\mathbf{W}_{T,N}$ are bounded in the following assertion.

Lemma 3. 1. There exist $C_4 > 0$, $C_5 > 0$ such that for any $\varepsilon_4 \in (0, 1/2)$

$\forall Y_0 > 0,$
$$P\left\{\forall T > 0,\ \ \forall N > 0,\ \ \|\mathbf{W}_{T,N}\| \le \frac{C_4 N^2 Y_0}{T^{1/2-\epsilon_4}}\right\}$$
$$\ge 1 - \frac{C_5}{\epsilon_4 Y_0^2}.$$

2. There exist $K > 0$, $\alpha > 0$ such that for any $N \ge K$
$$P_{N-K} \sum_{i=0}^{N-1} \mathbf{A}^i \mathbf{B}\mathbf{B}^* \mathbf{A}^{*i} P_{N-K} \ge \alpha P_{N-K}.$$

Here P_m is a standard projector
$$P_m \mathbf{c} = (c_0, c_1, \ldots, c_{m-1}, 0, 0, \ldots),$$
when $\mathbf{c} = (c_0, c_1, \ldots)$.

Finally, collecting the bounds for the matrixes $\mathbf{Q}_{2,T}$ and $\mathbf{Q}_{3,T}$ yields the required result of Theorem (3).

From (10) follows directly the power degree of convergence with probability 1 of the LS estimates for the infinite-dimensional AR equation.

Corollary 2.
$$\lim_{T\to 0} T^{1-\delta} |\boldsymbol{\tau}_T - \boldsymbol{\tau}_*|^2 = 0 \qquad (17)$$
with probability 1.

CONCLUSION

In the standard estimation algorithms the number of parameters is fixed and their convergence with probability 1 is proved as the number of observations tends to infinity. If the number of observations is fixed then an attempt to estimate too many parameters leads to an unreliable result. It is recommended to choose a model with the number of parameters proportional to $\log T$, where T is the number of observations (the AIC criterion). If T increases then the model can be made richer, but the number of parameters increases much slower than T and proportional to $\log T$. The LS algorithm in the infinite dimensional regression model does not satisfy this logarithmic relation.

The number of parameters that are obtained after T observations in the infinite linear regression model is equal to T for the RLS algorithm. Therefore it is impossible to extract a small set of parameters that are estimated precisely and to expand this set slowly with time T. Let a function $F(p)$ be the accuracy of the estimates of the first p parameters. In the standard approaches (Mari et al. (2000)) this function is small for the first $\log T$ parameters and arbitrary for others that were not estimated after T observations. This function can be smoothed in the standard LS algorithm if the initial conditions are chosen properly.

The initial covariance of parameter estimates in the RLS algorithm is equal to the inverse of the initial information matrix, namely the regularizer R. Small covariances imply small correction gains. The rate of correction is inversely proportional to the values of R. It is proposed in in this paper to define R as a diagonal matrix with exponential entries $e^{\mu k}$ on the diagonal. A sum of correction gains at time T for parameter number k achieves some fixed value when $k \sim \log T$. The number of parameters that can be precisely estimated is proportional to $\log T$, and estimates of other parameters cannot move far from their initial values. For this reason the initial information matrix R determines a smoothed number of parameters with reliable estimates at time T.

The degree of convergence with probability 1 is obtained from the lemma on convergence of semi-martingales. This lemma presented in Section 4 gives a simple expression for probability of the exceptional set if a semi-martingale converges. It is used in the proofs of different assertions in this paper.

Proofs of the main theorems about the degree of convergence of the RLS estimates for the AR(∞) model are based on the standard Lyapunov function associated with the LS approach (Barabanov (1983). A special mathematical technique was developed to find appropriate lower and upper bounds on the information matrix and to analyze their asymptotic behavior almost surely.

BARABANOV, A. Ye. (1983). Strong convergence of the method of least-squares. *Automat. Remote Control* **10**, part 2, 1338–1346.

DURBIN, J. (1959). The fitting of time-series models. *Rev. Inst. Int. Stat.*, 223–243.

FOMIN, V.N. (1999). *Optimal filtering. Vol.1. Filtering of stochastic processes.* Kluwer Academic Publishers, Dordrecht.

GEL, Yu.R. and FOMIN,, V.N. (1998). Linear model approximation of stochastic stationary time series. *Vestnik of St.-Petersburg University* **2**, 24–31.

GEL, Yu.R. and FOMIN, V.N. (2001). Identification of an unstable ARMA equation. *Mathematical Problems in Engineering.* **7**, 97–112.

GREEN, M. (1995).*Linear robust control.* Prentice Hall, New Jersey.

HANNAN, E.J. and DIESTLER, M. (1988). *The statistical theory of linear systems.* John Wiley, New York.

LAI, T.L. and WEI, C.Z. (1982). Asymptotic properties of projections with application to stochastic regression problems. *J. Multivariate Analysis* **12**, 346–370.

LJUNG, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control* **22**, 551–575.

LJUNG, L. (1999). *System Identification - Theory For the User.* Prentice Hall, New Jersey.

MARI, J., DAHLEN, A., and A. LINDQUIST (2000). A covariance extension approach to identification of time series. *Automatica* **36**, no. 3, 379–398.

MARPLE, S. L., Jr. (1987).*Digital spectral analysis with applications.* Prentice Hall, New Jersey.

WAHLBERG, B. (1989). Estimation of autoregressive moving-average models via high-order autoregressive approximations. *Journal of Time Series Analysis* **10**, 283–299.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat. Astr. Fys.* **2**, 423–434.

WOLD, H. (1938). A study in the analysis of Stationary Time Series. *Almqvist and Wiksell*.