

Identification of an Unstable ARMA Equation*

YULIA R. GEL^{a, b, †} and VLADIMIR N. FOMIN^b

^aMalardalen University, PO Box 883, S-721 23 Vasteras, Sweden; ^bSt. Petersburg State University, Bibliotechnaya pl., 2 198904 St. Petersburg, Russia

(Received 23 May 2000)

Usually the coefficients in a stochastic time series model are partially or entirely unknown when the realization of the time series is observed. Sometimes the unknown coefficients can be estimated from the realization with the required accuracy. That will eventually allow optimizing the data handling of the stochastic time series.

Here it is shown that the recurrent least-squares (LS) procedure provides strongly consistent estimates for a linear autoregressive (AR) equation of infinite order obtained from a minimal phase regressive (ARMA) equation. The LS identification algorithm is accomplished by the Padé approximation used for the estimation of the unknown ARMA parameters.

Keywords: Identification; Unstable ARMA models; Least squares method; Linear estimation

1. INTRODUCTION

Usually in applications parameters of the mathematical model of a stochastic time series are partially or entirely unknown when the realization is observed. Sometimes unknown coefficients can be estimated from the realization with the required accuracy, which eventually allows optimizing the data handling of stochastic time

* This project is partially supported by Russian Fund of Basic Research N98-01-00581.

† Corresponding author. Tel.: 46-21-101661, Fax: 46-21-101330, e-mail: yulia.gel@mdh.se, yulia.gel@lpx.ff.phys.spbu.ru

series. Problems arising from the identification of an autoregressive (AR) equation form a standard topic in mathematical statistics and are covered by many research works. We dwell on some well known results.

In the standard literature, most attention is devoted to stable AR equation (when the polynomial $a(\cdot)$ has no roots inside the closed unit disk $\mathbf{D}_1 = \{\lambda: |\lambda| \leq 1\}$). In this case a time series is assumed to be stationary, and for the reconstruction of the AR coefficients spectral methods are widely used. However, the problem of the spectral density reconstruction is comparable in difficulty to the problem of the identification of the AR coefficients. Therefore, methods that are not connected with such a reconstruction have received wide acceptance. Among them are, for example, the Yule–Walker method of identification of the stable AR model of a stationary time series with the rational spectral density, or the least-squares (LS) recurrent procedure, and the stochastic approximation method. For modeling a weakly stationary time series the AR equation is widely used, where the polynomial $a(\cdot)$ has roots on the unit circle (see [1]). Convergence proofs for the LS identification of weakly stationary processes have been published by several researchers. In [12] convergence in probability is proved for the system with an independent and identically distributed input. A proof in the almost sure sense given in [9] contains some incorrect computations (see [2]). The LS foundation for an unstable AR equation (when the polynomial $a(\cdot)$ may have roots inside but not on the unit \mathbf{D}_1) is given in [2]. However, there is no theoretical proof of the consistency of the LS estimates derived regardless of stability of the process. The reason, probably, is that realizations of an unstable time series increase indefinitely.

For the regressive (ARMA) equation the LS estimates turn out to be shifted. Then for the identification of the AR coefficients, variations of the instrumental variable method are employed. They are based on the finite correlativity of the righthand side of the ARMA model. In [3, 4, 10, 11] convergence of the extended LS method for a stable ARMA equation is proved when the positive realness of the transfer function is assumed. It ensures strictly consistent estimates of the unknown coefficients. In [5, 6] this method is extended to the case of a weakly stable ARMA equation with the above-mentioned positive-real property. In [7, 8] is proposed and theoretically proved an

identification method related to the transformation of a stable ARMA equation into an AR equation of infinite order. To this end the method of truncation of the infinite Yule–Walker system and the Padé method are used.

In this paper it is shown that the recurrent LS procedure provides consistent estimates for an AR equation of general form (when roots of the polynomial $a(\cdot)$ are located arbitrarily in the complex plane). Furthermore, the study directs the way to the transformation of the identification problem of a regressive equation into the identification problem of an AR equation of infinite order. To this end an algorithm based on the recurrent LS procedure is developed. From these considerations it follows that if coefficients of the original ARMA equation should be reconstructed then it is sufficient to know some finite number of coefficients of the above-mentioned infinite AR equation. To this end the recurrent LS procedure is taken, followed by the Padé approximation. Such an identification procedure is free from assumptions of stability or weak stability, except that the polynomial $b(\cdot)$ must have no roots inside \mathbf{D}_1 .

2. THE EMPIRICAL FUNCTIONAL METHOD

The ARMA model under study is

$$a(\nabla)y_t = \sigma^2 b(\nabla)v_t, \quad t \in \mathbf{T}, \quad (1)$$

where $a(\cdot)$ is a polynomial of the backward shift operator ∇ ($\nabla y_t = y_{t-1}, \nabla v_t = v_{t-1}$), $a(\lambda) = 1 + \lambda a_1 + \dots + \lambda^p a_p$, $b(\lambda) = 1 + \lambda b_1 + \dots + \lambda^q b_q$, $b(\lambda) \neq 0$, $|\lambda| \leq 1, \sigma^2$ is a positive constant, $v = \{v_t, t \in \mathbf{T}\}$ is a standard discrete white noise process ($E v_t \equiv 0, E v_t v_{t'} = \delta_{tt'}$).

We rewrite (1) in the form of the AR equation

$$\tilde{a}(\nabla)y_t = \sigma^2 v_t, \quad t \in \mathbf{N}, \quad (2)$$

where $\tilde{a}(\cdot)$ is the Taylor expansion of the rational function

$$\tilde{a}(\lambda) = \frac{a(\lambda)}{b(\lambda)} = 1 + \sum_{t=1}^{\infty} \lambda^t \tilde{a}_t. \quad (3)$$

If (3) does not degenerate to a polynomial, then (2) is of infinite order. Recall that $y = \text{col}(y_t, y_{t-1}, \dots) \in \mathbf{l}_2(\mathbf{N})$ is finite ($y_{t'} = 0$ for $t' < -p$). The quantities $y_0, y_{-1}, \dots, y_{-p}$ play the role of initial conditions for (2).

Loosely speaking, the correlation operator of y is not stationary. Hence, the identification method presented in [7, 8], which exploits the stationarity of the correlation operator, needs to be revised. The empirical functional method can be used for approximate but as accurate as desired reconstruction of the vector τ_* of unknown coefficients of such equations. It is based on a predicting model of y . The sequence $\tau[t] = \text{col}(\tau^{(1)}[t], \tau^{(2)}[t], \dots) \in \mathbf{l}_2(\mathbf{N})$ of estimates of the coefficients

$$\tau_* = \text{col}(\tau_*^{(1)}, \tau_*^{(2)}, \dots) = \text{col}(\tilde{a}_1, \tilde{a}_2, \dots) \in \mathbf{l}_2(\mathbf{N}) \quad (4)$$

is determined from the minimum condition for the functional

$$\begin{aligned} J_T(\tau) &= \frac{1}{T} \sum_{t=1}^T (y_t + \Phi_{t-1}^* \tau)^2 = \frac{1}{T} \left(\tau^* R_T \tau + 2\tau^* r_T + \sum_{t=1}^T y_t^2 \right) \\ &= \frac{1}{T} \left((\tau_* - \tau)^* R_T (\tau_* - \tau) - 2(\tau_* - \tau)^* \tilde{r}_T + \sigma^4 \sum_{t=1}^T |v_t|^2 \right), \end{aligned} \quad (5)$$

where

$$\begin{aligned} R_T &= \sum_{t=1}^T \Phi_{t-1} \Phi_{t-1}^* = \sum_{t=0}^{T-1} \Phi_t \Phi_t^*, \\ r_T &= \sum_{t=1}^T \Phi_{t-1} y_t, \quad \tilde{r}_T = \sum_{t=1}^T \Phi_{t-1} v_t, \\ \tau &= \text{col}(\tau^{(1)}, \tau^{(2)}, \dots), \quad \Phi_{t-1}^* = (y_{t-1}, y_{t-2}, \dots). \end{aligned} \quad (6)$$

In (5) it was taken into account that (2) can be rewritten in terms of (6) as

$$y_t + \Phi_{t-1}^* \tau_* = \sigma^2 v_t. \quad (7)$$

If (3) does not degenerate to a polynomial, (*i.e.* (2) is not an AR equation of finite order), then the operator R_T in (6) is considered as a linear operator in the Hilbert space $\mathbf{l}_2(\mathbf{N})$. Here Φ_t , $t \leq T-1$, are

elements (vectors) of this space:

$$\Phi_{t-1} = \text{col}(y_{t-1}, y_{t-2}, \dots, y_0, y_{-1}, \dots, y_{-p}, 0, 0, \dots).$$

Here the range of the operator R_T coincides with a linear span of the elements Φ_{t-1} , $t \leq T$. It implies, in particular, that for each $T < \infty$ the spectrum of the operator R_T consists of one zero eigenvalue of infinite multiplicity and just a finite number of nonzero eigenvalues. The number of nonzero eigenvalues of the nonnegative operator R_T (with regard to multiplicity) coincides with the number of linearly independent vectors in Φ_t , $t \leq T-1$. In $\mathbf{l}_2(\mathbf{N})$, owing to the finiteness of the vectors Φ_{t-1} , $t \leq T$, the matrix of infinite order corresponding to the operator R_T has only a finite number of nonzero elements for every T .

3. THE LEAST SQUARES ESTIMATES

The minimum in τ of the functional (5) is attained on the stochastic vector

$$\hat{\tau}[T] = \underset{\tau}{\text{argmin}} J_T(\tau) = -R_T^+ r_T = \tau_* - R_T^+ \tilde{r}_T. \quad (8)$$

This value is equal to

$$\begin{aligned} \inf_{\tau} J_T(\tau) &= \frac{1}{T} \left(-[r_T]^* R_T^+ r_T + \sum_{i=1}^T y_i^2 \right) \\ &= \frac{1}{T} \left(-[\tilde{r}_T]^* R_T^+ \tilde{r}_T + \sigma^4 \sum_{i=1}^T v_i^2 \right). \end{aligned} \quad (9)$$

Here R_T^+ denotes pseudoinversion of the operator $R_T: \mathbf{l}_2(\mathbf{N}) \rightarrow \mathbf{l}_2(\mathbf{N})$: $R_T^+ = \text{Pr}_T R_T^{-1} \text{Pr}_T + (I_2 - \text{Pr}_T)$, where Pr_T is the orthoprojector onto the subspace of the range of the operator R_T and $\text{Pr}_T R_T^{-1} \text{Pr}_T$ is the inversion of the operator $\text{Pr}_T R_T \text{Pr}_T$ in the invariant subspace $\text{Pr}_T \mathbf{l}_2(\mathbf{N})$. (It is assumed that $\text{Pr}_T R_T \text{Pr}_T > \varepsilon \text{Pr}_T$ for some $\varepsilon > 0$.) Note that the operator R_T^+ is unambiguously defined by the relation $R_T^+ R_T = R_T R_T^+ = \text{Pr}_T$. If the operator R_T is invertible ($\text{Pr}_T = I_2$), then obviously $R_T^+ = R_T^{-1}$.

The vector $\hat{\tau}[T]$ is called the estimate of the least squares method (the LS estimate) of the vector of coefficients of (7).

If τ_* has a finite number of nonzero elements, *i.e.*, $\tilde{a}(\cdot)$ is a polynomial, then it can be shown that the sequence $R_T^+ \tilde{r}_T$ converges to zero with probability one as $T \rightarrow \infty$. Thus, (9) implies that the vectors (8) converge to the vector τ_* with probability one as $T \rightarrow \infty$. In other words, the LS estimates are strongly consistent. If τ_* has an infinite number of nonzero elements, *i.e.*, $b(\lambda)$ has nonzero degree, then it is not hard to see that (9) takes the form $\inf_{\tau} J_T(\tau) = 0$ for arbitrary T . Thus, LS estimates cannot be consistent. However, LS estimates that are regularized in a special way are found to be strongly consistent.

4. REGULARIZED LS ESTIMATES

To avoid the pseudo-inversion of the operators R_T , instead of (8) we consider the estimates

$$\hat{\tau}_\varepsilon[T] = -(R_T + \varepsilon R)^{-1} r_T, \quad (10)$$

where ε is a fixed positive number and R is a positive definite operator in the Hilbert space \mathbf{l}_2 under the inner product $\langle \cdot, \cdot \rangle$.

The operator R is called *the regularizator*. If the number of coefficients in (2) is infinite then the Hilbert space is infinite-dimensional, $\mathbf{l}_2 = \mathbf{l}_2(\mathbf{N})$. Then we additionally assume that $R^{-\delta}$ is an operator of trace class for arbitrary small $\delta > 0$. (For example, for R we can take any operator to which a diagonal matrix with eigenvalues, $\mu_k = \exp\{\alpha k\}$, $\alpha > 0$ corresponds in $\mathbf{l}_2(\mathbf{N})$.)

The estimates (10) are said to be ε -regularized LS estimates.

5. THE RECURRENT FORM OF THE LS ESTIMATES

The LS estimates satisfy the recurrent relations representing the special variant of the Kalman–Bucy filter. The latter are convenient for practical implementation.

LEMMA 1 *The estimates (10) satisfy*

$$\begin{aligned}\hat{\tau}_\varepsilon[t+1] &= \hat{\tau}_\varepsilon[t] - \frac{y_{t+1} + \Phi_t^* \hat{\tau}_\varepsilon[t]}{1 + \Phi_t^* \Gamma_\varepsilon[t] \Phi_t} \Gamma_\varepsilon[t] \Phi_t, \\ \Gamma_\varepsilon[t+1] &= \Gamma_\varepsilon[t] - \frac{\Gamma_\varepsilon[t] \Phi_t \Phi_t^* \Gamma_\varepsilon[t]}{1 + \Phi_t^* \Gamma_\varepsilon[t] \Phi_t}, \quad t \in N,\end{aligned}\tag{11}$$

and are defined by these relations under the initial conditions

$$\hat{\tau}_\varepsilon[1] = 0, \quad \Gamma_\varepsilon[1] = (R_1 + \varepsilon R)^{-1}.\tag{12}$$

Proof of Lemma 1 Using the familiar matrix identity

$$\begin{aligned}(R_t + \varepsilon R + \Phi_t \Phi_t^*)^{-1} &= (R_t + \varepsilon R)^{-1} \\ &- (R_t + \varepsilon R)^{-1} \Phi_t (1 + \Phi_t^* (R_t + \varepsilon R)^{-1} \Phi_t)^{-1} (R_t + \varepsilon R)^{-1},\end{aligned}\tag{13}$$

we obtain the second formula of (11) for $\Gamma_\varepsilon[t] = (R_t + \varepsilon R)^{-1}$. With formulas (10), (13) and using some elementary operations we arrive at the first formula of (11). ■

Note that the estimates $\hat{\tau}_{\varepsilon,\tau}[t]$ obtained by (11) under the initial conditions $\hat{\tau}_\varepsilon[1] = \tau$, $\Gamma_\varepsilon[1] = (R_1 + \varepsilon R)^{-1}$ (τ is an arbitrary vector in \mathbf{I}_2) are related to the estimates (11), (12) by the formula

$$\hat{\tau}_{\varepsilon,\tau}[t] = \Gamma_\varepsilon[t] (\Gamma_\varepsilon[1])^{-1} \tau + (I_2 - \varepsilon \Gamma_\varepsilon[t] R) \tau_* + \hat{\tau}_\varepsilon[t].$$

Usually $\Gamma_\varepsilon[t] \rightarrow \mathbf{0}$ as $t \rightarrow \infty$ (with probability one). Therefore, the consistency of the estimates $\hat{\tau}_\varepsilon[t]$ implies the consistency of $\hat{\tau}_{\varepsilon,\tau}[t]$ (and vice versa).

6. CONVERGENCE OF THE LS ESTIMATES

Let us begin justification of the consistency of the regularized LS estimates by establishing the following assertion. This assertion is valid for the ARMA equation of general form (1) (regardless of stability of the process).

THEOREM 1 Consider the ARMA equation (1) with the standard white noise $v = \{v_t, t \in T\}$ as a disturbance. Then for any $\delta > 0$,

$$\lim_{t \rightarrow \infty} \langle \tau_* - \hat{\tau}_\varepsilon[t], (R_T + \varepsilon R)^{1-\delta} (\tau_* - \hat{\tau}_\varepsilon[t]) \rangle = 0 \quad (14)$$

with probability one and in the mean square sense. (Here the matrix R_T is defined in (6).)

The proof of Theorem 1 is essentially based on the following assertion, which is of interest in itself.

LEMMA 2 For the operators $R_T : \mathcal{L}_2(N) \rightarrow \mathcal{L}_2(N)$ and elements $\Phi_{t-1} \in \mathcal{L}_2(N)$, $t \in N$ related by

$$R_T = \sum_{t=1}^T \Phi_{t-1} \Phi_{t-1}^* \quad (15)$$

the following inequality is satisfied for any $\delta > 0$:

$$\sum_{t=1}^{\infty} \langle \Phi_t, (R_{t+1} + \varepsilon R)^{-1-\delta} \Phi_t \rangle \leq \frac{Sp\{R^{-\delta}\}}{\varepsilon^\delta \delta}. \quad (16)$$

The proof of Lemma 2 is given in Appendix A.

Proof of Theorem 1 Let δ be an arbitrary positive number from $(0, 1)$. Consider a stochastic variable

$$\begin{aligned} V_T &= (\tau_* - \hat{\tau}_\varepsilon[T])(\Gamma_\varepsilon[T])^{-1+\delta} (\tau_* - \hat{\tau}_\varepsilon[T]) \\ &= (\tau_* - \hat{\tau}_\varepsilon[T])(R_T + \varepsilon R)^{1-\delta} (\tau_* - \hat{\tau}_\varepsilon[T]). \end{aligned}$$

The first relation of (11) can be rewritten as

$$\begin{aligned} \hat{\tau}_\varepsilon[t+1] - \tau_* &= \hat{\tau}_\varepsilon[t] - \tau_* + ((\Phi_t), \tau_* - \hat{\tau}_\varepsilon[t]) - \sigma^2 v_{t+1} \Gamma_\varepsilon[t+1] \Phi_t \\ &= (I_2 - \Gamma_\varepsilon[t+1] \Phi_t \Phi_t^*) (\hat{\tau}_\varepsilon[t] - \tau_*) \\ &\quad - \sigma^2 v_{t+1} \Gamma_\varepsilon[t+1] \Phi_t. \end{aligned} \quad (17)$$

Here we used the equality

$$\frac{\Gamma_\varepsilon[t] \Phi_t}{1 + \Phi_t^* \Gamma_\varepsilon[t] \Phi_t} = \Gamma_\varepsilon[t+1] \Phi_t,$$

which is satisfied by virtue of the second relation of (11).

Considering (17) and the measurability of $\{\Phi_t, 1 \leq t \leq T\}$ with respect to the σ -algebra generated by $v^T = \{v_t, 1 \leq t \leq T\}$, we obtain

the conditional mathematical expectation

$$\begin{aligned}
\mathbf{M}(V_{T+1}|v^T) &= (\hat{\tau}_\varepsilon[T] - \tau_*)^* (\mathbf{I}_2 - \Gamma_\varepsilon[T+1] \Phi_T \Phi_T^*)^* \\
&\quad \times (\Gamma_\varepsilon[T+1])^{-1+\delta} \\
&\quad \times (\mathbf{I}_2 - \Gamma_\varepsilon[T+1] \Phi_T \Phi_T^*) (\hat{\tau}_\varepsilon[T] - \tau_*) \\
&\quad + \sigma^4 \langle \Phi_T, (\Gamma_\varepsilon[T+1])^{1+\delta} \Phi_T \rangle \\
&= V_T + (\hat{\tau}_\varepsilon[T] - \tau_*)^* \\
&\quad \times ((\mathbf{I}_2 - \Gamma_\varepsilon[T+1] \Phi_T \Phi_T^*)^* (\Gamma_\varepsilon[T+1])^{-1+\delta} \\
&\quad \times (\mathbf{I}_2 - \Gamma_\varepsilon[T+1] \Phi_T \Phi_T^*) - (\Gamma_\varepsilon[T])^{-1+\delta}) \\
&\quad \times (\hat{\tau}_\varepsilon[T] - \tau_*) + \sigma^4 \langle \Phi_T, (\Gamma_\varepsilon[T+1])^{1+\delta} \Phi_T \rangle \\
&= V_T + (\hat{\tau}_\varepsilon[T] - \tau_*)^* \\
&\quad \times ((R_{T+1} + \varepsilon R - \Phi_T \Phi_T^*) (R_{T+1} + \varepsilon R)^{-1-\delta} \\
&\quad \times (R_{T+1} + \varepsilon R - \Phi_T \Phi_T^*) - (R_T + \varepsilon R)^{1-\delta}) \\
&\quad \times (\hat{\tau}_\varepsilon[T] - \tau_*) + \sigma^4 \langle \Phi_T, (R_{T+1} + \varepsilon R)^{-1-\delta} \Phi_T \rangle \\
&= V_T + (\hat{\tau}_\varepsilon[T] - \tau_*)^* ((R_T + \varepsilon R) (R_{T+1} + \varepsilon R)^{-1-\delta} \\
&\quad \times (R_T + \varepsilon R) - (R_T + \varepsilon R)^{1-\delta}) \\
&\quad \times (\hat{\tau}_\varepsilon[T] - \tau_*) + \sigma^4 \langle \Phi_T, (R_{T+1} + \varepsilon R)^{-1-\delta} \Phi_T \rangle \\
&= V_T + (\hat{\tau}_\varepsilon[T] - \tau_*)^* \\
&\quad \times (R_T + \varepsilon R) ((R_{T+1} + \varepsilon R)^{-1-\delta} - (R_T + \varepsilon R)^{-1-\delta}) \\
&\quad \times (R_T + \varepsilon R) (\hat{\tau}_\varepsilon[T] - \tau_*) \\
&\quad + \sigma^4 \langle \Phi_T, (R_{T+1} + \varepsilon R)^{-1-\delta} \Phi_T \rangle. \tag{18}
\end{aligned}$$

Because $R_{T+1} \geq R_T$, from (18) follows

$$\mathbf{M}(V_{T+1}|v^T) \leq V_T + \sigma^4 \langle \Phi_T, (R_{T+1} + \varepsilon R)^{-1-\delta} \Phi_T \rangle.$$

By virtue of (16) the corollary to the Doob theorem can be used, by which the finite limit

$$\lim_{T \rightarrow \infty} V_T = \lim_{T \rightarrow \infty} (\tau_* - \hat{\tau}_\varepsilon[T]) (\Gamma_\varepsilon[T])^{-1+\delta} (\tau_* - \hat{\tau}_\varepsilon[T]) = V_*$$

exists with probability one and in the mean square sense. Here V_* is some non-negative stochastic process, $\mathbf{M}V_* < \infty$. Therefore, the stochastic variables

$$(\tau_* - \hat{\tau}_\varepsilon[T]) (\Gamma_\varepsilon[T])^{-1+\delta} (\tau_* - \hat{\tau}_\varepsilon[T])$$

converge for all $\delta > 0$ with probability one and in the mean square sense. This implies that their limit is equal to zero:

$$\lim_{T \rightarrow \infty} V_T = \lim_{T \rightarrow \infty} (\tau_* - \hat{\tau}_\varepsilon[T])((R_T + \varepsilon R)^{1-\delta}(\tau_* - \hat{\tau}_\varepsilon[T]) = 0. \quad (19)$$

■

COROLLARY 1 *With probability one and in the mean square sense,*

$$\lim_{t \rightarrow \infty} |\tau_* - \tau_\varepsilon[t]|^2 = 0. \quad (20)$$

Indeed, from the positive definiteness of the operator $R_T + \varepsilon R$, the limit relation (20) obviously follows from (19).

7. SIMULATION EXAMPLE

Example 1 Consider the stable ARMA model

$$y_t + 0.7y_{t-1} + 0.1y_{t-2} = v_t - v_{t-1} + 0.21v_{t-2}.$$

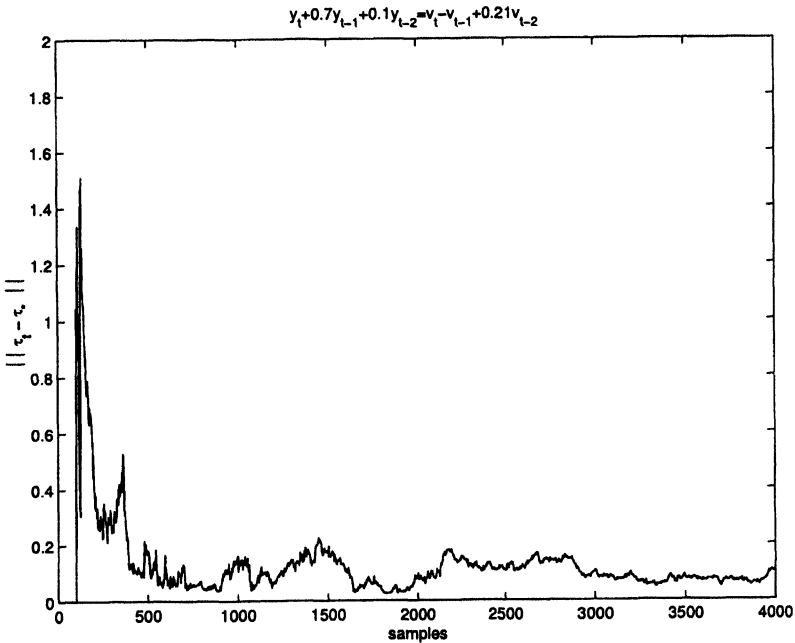


FIGURE 1 Stable ARMA model.

The initial conditions for the LS procedure are $\hat{\tau}_{[1]} = 0$, $\Gamma_0 = 1000000$. The Padé approximation is chosen of order (2, 2). In Figure 1 the estimation error for the ARMA parameters is plotted with respect to the number of samples. Here $\tau_* = \text{col}(1 \ 0.7 \ 0.1 \ 1 \ -1 \ 0.21)$; τ_t is the estimate derived at step t by the following procedure: obtaining the estimate $\hat{\tau}_{[t]}$ of the parameters of the infinite AR model is combined with the Padé approximation for the identification of the original ARMA parameters.

Notice that the model does not satisfy the above-mentioned condition of positive realness.

Example 2 Consider now the weakly stable ARMA model

$$y_t + 1.5y_{t-1} + 0.5y_{t-2} = v_t - 0.3v_{t-1} + 0.03v_{t-2} - 0.001v_{t-3}.$$

By analogy with the first example, $\hat{\tau}_{[1]} = 0$, $\Gamma_0 = 1000000$ are the initial conditions; $\tau_* = \text{col}(1 \ 1.5 \ 0.5 \ 1 \ -0.3 \ 0.03 \ -0.001)$. The Padé approximation is sought of the order (2, 3). Figure 2 shows a graph of convergence of the estimated ARMA parameters.

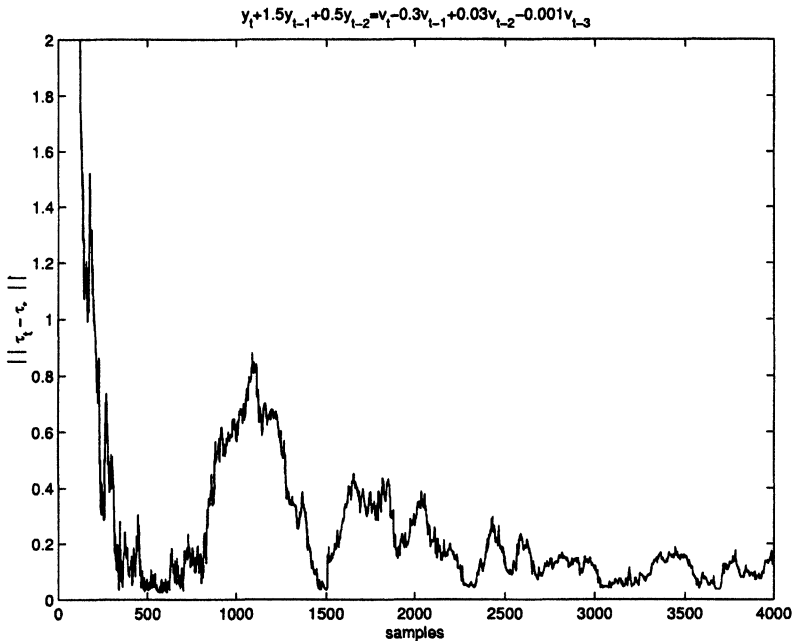


FIGURE 2 Weakly stable model.

The obtained estimates are of relatively high accuracy. This allows us to say that the identification method, which is based on the LS procedure followed by the Padé approximation, is effective for the estimation of the ARMA parameters.

8. CONCLUSION

In Section 6 we proved that the LS estimates of the infinite AR model converges almost surely to the true parameters, regardless of the location of the roots of the autoregressive equation, *i.e.*, regardless of stability of the process. The identification problem of the ARMA-model is transformed to the identification problem of the infinite AR-model. It allows us to propose a new identification algorithm based on the LS method by the Padé approximation. This identification method is free from the assumption of positive realness of the transfer function of the linear formative filter.

References

- [1] Box, G. E. P. and Jenkins, G. M., *Time Series Analysis. Prediction and Control*. Holden-Day, San-Francisco, 1970.
- [2] Barabanov, A. E. (1983). On strong convergence of the method of least squares. *Automatica and Telemekhanika*, **10**, 119–127.
- [3] Agafanov, S., Barabanov, A. and Fomin, V. (1982). Adaptive filtering of stochastic processes. In book: *Problems of Cybernetics*. Actual Adaptive Control Problems, Cybernetics Counsel of USSR, Acad. Sc., N. 4–30.
- [4] Fomin, V., *Recurrent Estimation and Adaptive Filtering*. Nauka, Moscow, 1985.
- [5] Fomin, V. and Shepelyavaya, N., Extended least squares method in the problem of the identification of a weakly stable dynamic plant. *VINITI*, 13.10.88, N7386–B88, 19 p.
- [6] Fomin, V., Regression models of nonstationary time series. In book: *Modern Problems of Computer Data Analysis and Modelling* (Edited by Kharin, Y.), **2**, 269–274. Minsk, 1995.
- [7] Gel, Y. and Fomin, V. (1998). *Identification of a Linear Model of a Stationary Process by its Realization*. Vestnik St.Petersburg University: Mathematics, **31**(2), 21–28; translation from Vestnik St.-Peterburg. Univ. Mat. Mekh. Astronom., **1**(2), 24–32, 1998.
- [8] Gel, Y. and Fomin, V. (1998). Reconstruction of a Regressive Model of a Stationary Time Series. Vestnik St.-Petersburg State University. *Mat. Mekh. Astronom.*, **4**(22), 11–15.
- [9] Graupe, D. (1980). On convergence of least-squares identifiers of autoregressive models having stable and unstable roots. *IEEE Trans. Automat. Control*, **AC-25**(5), 761–771.

- [10] Ljung, L. and Söderström, T., *Theory and Practice of Recursive Identification*. The MIT press, London, 1983.
- [11] Ljung, L. (1977). On positive real transfer functions and the convergence of some recursive schemes. *IEEE Trans. Automat. Control*, **AC-22**(4), 539–551.
- [12] Mann, H. and Wald, A. (1943). On statistical treatment of linear stochastic difference equations. *Econometrica*, **11**.

APPENDIX PROOF OF LEMMA 2

Lemma 2 was formulated and proved in [3] for finite-dimensional vectors Φ_t , $t = 1, 2, \dots, T$. Below the proof is modified for the more general case.

Let Φ_{k-1} , $k \in \mathbf{N}$, be an arbitrary sequence of elements (vectors) in the Hilbert space $\mathbf{l}_2(\mathbf{N})$ under the inner product $\langle \cdot, \cdot \rangle$. Given the operator function $R(t)$, $t \in [0, \infty)$,

$$\begin{aligned} R(t) &= R_l^{(\varepsilon)} + (t-l)\Phi_{l-1}\Phi_{l-1}^* \\ &= R_l^{(\varepsilon)} - (l-t)\Phi_{l-1}\Phi_{l-1}^*, \quad l-1 \leq t \leq l, \quad l \in \mathbf{N}, \end{aligned} \quad (21)$$

where $R(0) = \varepsilon R: \mathbf{l}_2(\mathbf{N}) \rightarrow \mathbf{l}_2(\mathbf{N})$ is a regularizator (see (10)) and Φ_l^* is a linear functional generated by element $\Phi_l: \Phi_l^* = \langle \Phi_l, \cdot \rangle$ in $\mathbf{l}_2(\mathbf{N})$. Obviously, in terms of (6), $R_l^{(\varepsilon)} = R_l + \varepsilon R$.

First we prove some lemmas.

LEMMA 3 *Let $r(\cdot)$ be a differentiable function defined on the semiaxis $[0, \infty)$, the derivative $r'(\cdot)$ of which is bounded on the spectrum of the operator $R(t)$.*

Then

$$\begin{aligned} Sp \left\{ \frac{d}{dt} r(R(t)) \right\} &= Sp \{ (r'(R(t))\Phi_{l-1}\Phi_{l-1}^* \\ &\quad + \Phi_{l-1}\Phi_{l-1}^* r'(R(t))) \}, \\ l-1 < t \leq l, \quad r'(R) &= \left. \frac{d}{dt} r(t) \right|_{t=R}. \end{aligned} \quad (22)$$

Proof of Lemma 3 By virtue of (21) for $l \leq t < l+1$ and small enough Δt we have $R(t + \Delta t) - R(t) = \Delta t \Phi_{l-1} \Phi_{l-1}^*$. Hence,

$$r(R(t + \Delta t)) = r(R(t) + \Delta t \Phi_{l-1} \Phi_{l-1}^*). \quad (23)$$

Denoting $A = R(t)$, $\varepsilon B = \Delta t \Phi_{l-1} \Phi_{l-1}^*$, we arrive at the problem of computing

$$\frac{d}{d\varepsilon} r(A + \varepsilon B) \Big|_{\varepsilon=0}.$$

Assume that C is an arbitrary smooth contour enclosing the spectrum of the operator A and such that for a small enough ε the spectrum of the operator $A + \varepsilon B$ is enclosed by it as well.

The operator function $r(A + \varepsilon B)$ can be represented in terms of contour integrals:

$$r(A + \varepsilon B) = \frac{1}{2\pi i} \oint_C r(\mu) (\mu I_2 - (A + \varepsilon B))^{-1} d\mu. \quad (24)$$

For a small enough ε the formula (24) may be rewritten as

$$\begin{aligned} r(A + \varepsilon B) &= \frac{1}{2\pi i} \oint_C r(\mu) (\mu I_2 - A)^{-1} (I_2 - \varepsilon B (\mu I_2 - A)^{-1})^{-1} d\mu \\ &= \frac{1}{2\pi i} \oint_C r(\mu) (\mu I_2 - A)^{-1} d\mu \\ &\quad + \varepsilon \frac{1}{2\pi i} \oint_C r(\mu) (\mu I_2 - A)^{-1} B (\mu I_2 - A)^{-1} d\mu + o(\varepsilon^2), \end{aligned}$$

which implies

$$\frac{dr(A + \varepsilon B)}{d\varepsilon} \Big|_{\varepsilon=0} = \frac{1}{2\pi i} \oint_C r(\mu) (\mu I_2 - A)^{-1} B (\mu I_2 - A)^{-1} d\mu. \quad (25)$$

Computing the trace of both sides of (25) and using the relations

$$\text{Sp}\{AB\} = \text{Sp}\{BA\}, \quad \frac{1}{2\pi i} \oint_C r(\mu) (\mu I_2 - A)^{-2} d\mu = r'(A),$$

we arrive at (22). ■

LEMMA 4 *Assume $\rho(\cdot)$ is defined on the positive semiaxis and is a measurable scalar real function. Assume the function*

$$r(t) = \int_0^t \rho(t') dt'$$

is defined and bounded on the spectrum of the operator R . Then

$$\begin{aligned} Sp\{r(R_k^{(\varepsilon)})\} - Sp\{r(\varepsilon R)\} \\ = \sum_{l=1}^{k-1} \langle \Phi_{l-1}, \int_{l-1}^l \rho(R_l^{(\varepsilon)} - (l-t)\Phi_{l-1}\Phi_{l-1}^*) dt \Phi_{l-1} \rangle. \end{aligned} \tag{26}$$

Proof of Lemma 4 From 22, we have

$$\begin{aligned} \int_0^k Sp\left\{ \frac{d}{dt} r(R(t)) dt \right\} &= \sum_{l=1}^k Sp\left\{ \left(\int_{l-1}^l r'(R(t)) dt \Phi_{l-1} \Phi_{l-1}^* \right) \right\} \\ &= \sum_{l=1}^k \int_{l-1}^l \langle \Phi_{l-1}, r'(R(t)) \rangle dt \Phi_{l-1}, \end{aligned}$$

which clearly concludes the proof of Lemma 4. ■

LEMMA 5 Assume that R is a linear positive definite operator in the Hilbert space H and $\rho(\cdot)$ is a monotonically non-increasing scalar function bounded on the spectrum of the operator R . Then for an arbitrary element $\Phi \in H$ the following inequality is satisfied:

$$\Phi^* \rho(R - \Phi \Phi^*) \Phi \geq \Phi^* \rho(R) \Phi,$$

where Φ^* is a linear functional generated by Φ .

Proof of Lemma 5 By virtue of the obvious inequality $R - \Phi \Phi^* \leq R$ and the monotonic non-increase of the function $\rho(\cdot)$ we have $\rho(R - \Phi \Phi^*) \geq \rho(R)$, which immediately proves Lemma 5. ■

LEMMA 6 Assume that under the conditions of Lemma 4 the function $\rho(\cdot)$ is non-negative, monotonically non-increasing, and bounded on the spectrum of the operator R . Then

$$\sum_{l=0}^{k-1} \langle \Phi_l, \rho(R_{l+1}^{(\varepsilon)}) \Phi_l \rangle \leq Sp\{r(R_k^{(\varepsilon)}) - r(\varepsilon R)\}.$$

Proof of Lemma 6 Owing to Lemma 5, the following inequality is satisfied:

$$\rho(R_l^{(\varepsilon)} - (l-t)\Phi_{l-1}\Phi_{l-1}^*) \geq \rho(R_l^{(\varepsilon)}).$$

With (26) this proves Lemma 6. ■

We now turn to the proof of Lemma 2. We choose the function $\rho(t) = \mathbf{1}(t - t_0)t^{-1-\delta}$ as $\rho(\cdot)$ in Lemma 4. Here t_0 is an arbitrary positive number satisfying $t_0 I_1 \leq R^{(\varepsilon)}$ (see (10)) and $\mathbf{1}(t)$ is the Heaviside function. In accordance with Lemma 6 the corresponding function $r(\cdot)$ is expressed as

$$r(t) = \int_0^t (t')^{-1-\delta} dt' = \frac{1}{\delta} (t_0^{-\delta} - t^{-\delta}).$$

This function is monotonically non-increasing and bounded on the spectrum of the operator R . The inequality (16) has the form

$$\begin{aligned} \sum_{l=0}^{k-1} \langle \Phi_{l-1}, (R_l^{(\varepsilon)})^{-1-\delta} \Phi_{l-1} \rangle &\leq \frac{1}{\delta} Sp\{(\varepsilon R)^{-\delta} - (R_k^{(\varepsilon)})^{-\delta}\} \\ &\leq \frac{1}{\delta} Sp\{(\varepsilon R)^{-\delta}\}, \end{aligned}$$

which proves Lemma 2.