KYBERNETIKA — VOLUME 44 (2008), NUMBER 3, PAGES 385 – 399

EXPONENTIAL SMOOTHING FOR IRREGULAR TIME SERIES

Tomáš Cipra and Tomáš Hanzák

The paper deals with extensions of exponential smoothing type methods for univariate time series with irregular observations. An alternative method to Wright's modification of simple exponential smoothing based on the corresponding ARIMA process is suggested. Exponential smoothing of order m for irregular data is derived. A similar method using a DLS (*discounted least squares*) estimation of polynomial trend of order m is derived as well. Maximum likelihood parameters estimation for forecasting methods in irregular time series is suggested. The suggested methods are compared with the existing ones in a simulation numerical study.

Keywords: ARIMA model, exponential smoothing of order m, discounted least squares, irregular observations, maximum likelihood, simple exponential smoothing, time series

AMS Subject Classification: 62M10, 62M20, 90A20, 60G35

1. INTRODUCTION

Methods denoted generally as exponential smoothing are very popular in practical time series smoothing and forecasting. They are all recursive methods which makes them easy to implement and highly computationally efficient. Some extensions of these methods to the case of irregular time series have been presented in past. It is simple exponential smoothing (see [7]), Holt method (see [7]), Holt–Winters method (see [5]) and double exponential smoothing (see [6]).

In this paper we suggest further methods of this type. In Section 2 we derive a method alternative to Wright's simple exponential smoothing, based on the assumption that the series is an irregularly observed ARIMA$(0, 1, 1)$ process. Prediction intervals for this method are a natural outcome of this assumption. In Section 3 we derive an exponential smoothing of order m for irregular time series. It is a generalization of simple and double exponential smoothing for irregular data presented before. In Section 4 a similar but not equivalent method is derived using a DLS (*discounted least squares*) estimate of polynomial trend of order m.

In Section 5 maximum likelihood parameters estimation is suggested for forecasting methods in irregular time series when the variances of individual one-step-ahead forecasting errors are not equal (i. e. the case of heteroscedasticity). If normal distribution is assumed, this is a generalization of classical MSE (*mean square error*) minimization.

All methods and techniques presented in this paper are applicable to time series with general time irregularity in their observations. Time series with missing observations are just a special case of them. All these methods have the software form by the authors (application DMITS) and have been applied both to real and simulated data. Various modifications are possible (e. g. for non-negative time series, see [3, 8]).

In Section 6 a simulation numerical study is presented comparing the predictive performance of the suggested methods and the existing ones. Some practical conclusions are then made based on the results.

2. IRREGULARLY OBSERVED ARIMA(0, 1, 1)

It is well known that the simple exponential smoothing procedure is optimal for certain ARIMA(0, 1, 1) model of regular time series, see [4] p. 90. On the other hand if we assume one step ahead forecasting errors produced by this procedure to form a white noise, our time series is necessarily driven by this concrete ARIMA(0, 1, 1) model.

Wright [7] has suggested an extension of classical simple exponential smoothing to the case of irregular observations in a very intuitive way, just following the basic idea of *exponential smoothing*. In this section alternative method to the one by Wright will be derived based on the relation to ARIMA(0, 1, 1) model.

Let $\{y_t, t \in \mathbb{Z}\}$ be a time series driven by ARIMA(0, 1, 1) model. The series of its first differences $\{\Delta y_t, t \in \mathbb{Z}\}$ is then driven by MA(1) model which can be expressed e. g. as

$$\Delta y_t = y_t - y_{t-1} = e_t + (\alpha - 1)e_{t-1}, \tag{1}$$

where $\{e_t, t \in \mathbb{Z}\}$ is a white noise with finite variance $\sigma^2 > 0$. Let us suppose that $\alpha \in (0, 1)$ (i. e. the MA(1) process is invertible). If we denote for $t \in \mathbb{Z}$

$$S_t = y_t - (1 - \alpha)e_t\,, \tag{2}$$

we can express the model of $\{y_t, t \in \mathbb{Z}\}$ using equations

$$y_{t+1} = S_t + e_{t+1}\,, \tag{3}$$

$$S_{t+1} = (1 - \alpha)S_t + \alpha\, y_{t+1}\,. \tag{4}$$

From this notation the relation to the simple exponential smoothing with smoothing constant α is obvious.

Let us consider increasing sequence $\{t_j, j \in \mathbb{Z}\}$ representing the time grid on which we observe values of series $\{y_t, t \in \mathbb{Z}\}$. So we can observe only values $y_{t_j}, j \in \mathbb{Z}$, while values y_t for $t \notin \{t_j, j \in \mathbb{Z}\}$ are unobservable for us. We are interested in resulting time series $\{y_{t_j}, j \in \mathbb{Z}\}$ with missing observations. In particular we will look for a forecasting method optimal in the sense of minimal one step ahead forecasting error variance (i. e. from time t_n to time t_{n+1}).

Let us suppose we have already observed values y_{t_j} for $j \le n$ and based on them we have got a random variable $\widetilde{S}_{t_n}$ representing a forecast of unknown value S_{t_n}. It is realistic to suppose that the random variable $y_{t_n} - \widetilde{S}_{t_n}$ has a finite variance and is uncorrelated with random variables e_t for $t > t_n$. From (2) it follows that the same properties will have the random variable $S_{t_n} - \widetilde{S}_{t_n}$ as well. Let us denote

$$v_{t_n} = \frac{\operatorname{var}(S_{t_n} - \widetilde{S}_{t_n})}{\sigma^2} < \infty\,. \tag{5}$$

Let us look for a forecast of $\widetilde{S}_{t_{n+1}}$ in the form of

$$\widetilde{S}_{t_{n+1}} = (1-a)\widetilde{S}_{t_n} + a\, y_{t_{n+1}}, \tag{6}$$

where $y_{t_{n+1}}$ is a newly observed value of series y. The parameter $a \in \mathbb{R}$ will be chosen to minimize the variance

$$\operatorname{var}(S_{t_{n+1}} - \widetilde{S}_{t_{n+1}}) = \sigma^2 v_{t_{n+1}}\,. \tag{7}$$

From (1) and (2) we can easily derive the formulas

$$S_{t_{n+1}} = S_{t_n} + \alpha(e_{t_n+1} + e_{t_n+2} + \cdots + e_{t_{n+1}-1} + e_{t_{n+1}})\,, \tag{8}$$

$$y_{t_{n+1}} = S_{t_n} + \alpha(e_{t_n+1} + e_{t_n+2} + \cdots + e_{t_{n+1}-1}) + e_{t_{n+1}}\,. \tag{9}$$

Substituting (9) into (6) we obtain

$$\widetilde{S}_{t_{n+1}} = (1-a)\widetilde{S}_{t_n} + a\left[S_{t_n} + \alpha\left(e_{t_n+1} + e_{t_n+2} + \cdots + e_{t_{n+1}-1}\right) + e_{t_{n+1}}\right]. \tag{10}$$

Subtracting equations (8) and (10) we get

$$\begin{aligned} S_{t_{n+1}} - \widetilde{S}_{t_{n+1}} &= (1-a)(S_{t_n} - \widetilde{S}_{t_n}) \\ &\quad + \alpha(1-a)(e_{t_n+1} + e_{t_n+2} + \cdots + e_{t_{n+1}-1}) + (\alpha - a)\, e_{t_{n+1}}\,. \end{aligned} \tag{11}$$

Since random variables $S_{t_n} - \widetilde{S}_{t_n}$ and e_t for $t > t_n$ are uncorrelated, it is

$$\operatorname{var}(S_{t_{n+1}} - \widetilde{S}_{t_{n+1}}) = \sigma^2\left[(1-a)^2 v_{t_n} + \alpha^2(1-a)^2(t_{n+1} - t_n - 1) + (\alpha - a)^2\right]. \tag{12}$$

So we are solving

$$\min_{a\in\mathbb{R}}\left[(1-a)^2 v_{t_n} + \alpha^2(1-a)^2(t_{n+1} - t_n - 1) + (\alpha - a)^2\right]. \tag{13}$$

It is a minimization of the convex quadratic function of variable a so we find the minimizing $\hat{a}$ very easily as

$$\hat{a} = \frac{v_{t_n} + \alpha^2(t_{n+1} - t_n - 1) + \alpha}{v_{t_n} + \alpha^2(t_{n+1} - t_n - 1) + 1}\,. \tag{14}$$

This formula is a generalization of that from [2] where the special case of a gap after regularly observed data is concerned. The achieved minimal variance value is

$$v_{t_{n+1}} = \frac{\operatorname{var}(S_{t_{n+1}} - \widetilde{S}_{t_{n+1}})}{\sigma^2} = (1-\hat{a})^2[v_{t_n} + \alpha^2(t_{n+1} - t_n - 1)] + (\alpha - \hat{a})^2\,. \tag{15}$$

Let us notice that $\hat{a} \in (0,1)$ and so the formula

$$\widetilde{S}_{t_{n+1}} = (1-\hat{a})\widetilde{S}_{t_n} + \hat{a}\, y_{t_{n+1}} \tag{16}$$

computes the forecast $\widetilde{S}_{t_{n+1}}$ as a convex linear combination of the current forecast $\widetilde{S}_{t_n}$ and the newly observed value $y_{t_{n+1}}$.

From (14) it can be seen that $\hat{a}$ is an increasing function of arguments v_{t_n}, α and $t_{n+1} - t_n$ which is consistent with our intuitive view. A higher value of v_{t_n} means that $\widetilde{S}_{t_n}$ is not a good forecast of real value S_{t_n} and so more weight in formula (16) is given to the newly observed value $y_{t_{n+1}}$. Similarly a longer time step $t_{n+1} - t_n$ means that the value S_{t_n}, of which $\widetilde{S}_{t_n}$ is a forecast, is further in past from the new observation $y_{t_{n+1}}$. Parameter α represents the smoothing constant when using a simple exponential smoothing to the series $\{y_t, t \in \mathbb{Z}\}$ so its relation to $\hat{a}$ is not a surprise. For $\alpha \to 1$, $t_{n+1} - t_n \to \infty$ or $v_{t_n} \to \infty$ have have $\hat{a} \to 1$. If $v_{t_n} = 0$ and $t_{n+1} - t_n = 1$, which corresponds to regular time series $\{y_{t_j}, j \in \mathbb{Z}\}$ with $t_j = j$, then $\hat{a} = \alpha$.

The derived method consists of the formula (14) for computation of the optimal smoothing coefficient $\hat{a}$ in the current step, the recursive formula (16) for the smoothed value $\widetilde{S}$ update and the recursive formula (15) for the variance factor v update. The forecast of future unknown value $y_{t_n+\tau}$, $\tau > 0$, from time t_n is the smoothed value $\hat{y}_{t_n} = \widetilde{S}_{t_n}$ as it is in the case of simple exponential smoothing. For variance of the error $e_{t_n+\tau}(t_n) = y_{t_n+\tau} - \widetilde{S}_{t_n}$ of this forecast we obtain

$$\operatorname{var}\left[e_{t_n+\tau}(t_n)\right] = \sigma^2\left[v_{t_n} + \alpha^2(\tau-1) + 1\right] \tag{17}$$

since random variables $S_{t_n} - \widetilde{S}_{t_n}$ and e_t for $t > t_n$ are uncorrelated. It can be seen that this variance is minimal if and only if v_{t_n} is minimal. So the specified smoothing coefficient $\hat{a}$ is optimal also when minimal forecasting error variance is concerned.

If we assume $e_t \sim \mathrm{N}(0,\sigma^2)$ and $S_{t_n} - \widetilde{S}_{t_n} \sim \mathrm{N}(0, \sigma^2 v_{t_n})$ then

$$e_{t_n+\tau}(t_n) \sim \mathrm{N}\left(0, \sigma^2\left[v_{t_n} + \alpha^2(\tau-1) + 1\right]\right) \tag{18}$$

and the corresponding prediction interval with confidence $1-\theta$ has borders

$$\widetilde{S}_{t_n} \pm \mu_{1-\theta/2}\, \sigma \sqrt{v_{t_n} + \alpha^2(\tau-1) + 1}\,, \tag{19}$$

where $\mu_{1-\theta/2}$ is the $1-\theta/2$ percent quantile of the standard normal distribution.

Let us notice that once the random variable $S_{t_n} - \widetilde{S}_{t_n}$ is uncorrelated with e_t for $t > t_n$ then because of (11) and because $\{e_t, j \in \mathbb{Z}\}$ are uncorrelated, the same holds true automatically for all $m \geq n$. Similarly if $e_t \sim \mathrm{N}(0,\sigma^2)$ then from normality of $S_{t_n} - \widetilde{S}_{t_n}$ the normality of $S_{t_m} - \widetilde{S}_{t_m}$ for all $m \geq n$ already follows.

If we have at our disposal observations of y starting with time t_1 and want to start the recursive computation using formulas (14), (15) and (16), we have to determine the initial values t_0, $\widetilde{S}_{t_0}$ and v_{t_0} first. Let us denote q the average time spacing of our time series $\{y_{t_j}, j \in \mathbb{Z}\}$ and set $t_0 = t_1 - q$. The initial smoothed value $\widetilde{S}_{t_0}$ will be computed as a weighted average of several observations from the beginning of the series y. These weights can decrease into future with the discount factor $\beta = 1 - \alpha$. The value v_{t_0} can be determined as a fixed point of formula (15) with $t_{n+1} - t_n \equiv q$, i. e. as the component v of a solution of the following system of equations

$$a = \frac{v + \alpha^2(q-1) + \alpha}{v + \alpha^2(q-1) + 1}, \tag{20}$$

$$v = (1-a)^2[v + \alpha^2(q-1)] + (\alpha - a)^2 \tag{21}$$

with unknowns v and a. After some algebraic operations we obtain

$$v_{t_0} = \frac{(1-\widetilde{a})^2\alpha^2(q-1) + (\widetilde{a} - \alpha)^2}{\widetilde{a}(2-\widetilde{a})}, \tag{22}$$

where the value $\widetilde{a} \in (0,1)$ is computed as

$$\widetilde{a} = \frac{\alpha^2 q - \sqrt{\alpha^4 q^2 + 4(1-\alpha)\alpha^2 q}}{2(\alpha - 1)}. \tag{23}$$

The described method has a similar character as Wright's simple exponential smoothing. Also here the smoothed value of the series is recomputed using a recursive formula of typical form, namely (16). The smoothing coefficient $\hat{a}$ changes in particular steps and therefore the variance factor v has to be recomputed as well.

Although this method has been explicitly derived only for time series with missing observations, it can be used in practice for general irregular time series. The fact that the time step $t_{n+1} - t_n$ is not generally an integer value does not prevent us in using formulas of this method in a reasonable way.

3. EXPONENTIAL SMOOTHING OF ORDER m

Exponential smoothing of order m for regular time series is an adaptive recursive method with one parameter – *smoothing constant* α. It estimates a local polynomial trend of order m using the *discounted least squares* (DLS) method with the discount factor $\beta = 1 - \alpha$. The estimates of polynomial coefficients are expressed as linear combinations of the first $m+1$ *smoothing statistics* $S_t^{[p]}$, $p = 1, 2, \ldots, m+1$. These smoothing statistics are computed in a recursive way using very simple formulas.

It is possible to extend exponential smoothing of order m to the case of irregular time series in two different ways. In this section we will derive a method working with smoothing statistics $S_t^{[p]}$, $p = 1, 2, \ldots, m+1$. The second extension uses explicitly the DLS estimation method and will be shown in Section 4.

Let us consider an irregular time series $y_{t_1}, y_{t_2}, \ldots, y_{t_n}, y_{t_{n+1}}, \ldots$ observed at times $t_1 < t_2 < \cdots < t_n < t_{n+1} < \ldots$. Let $m \in \mathbb{N}_0$ be the order of exponential smoothing, i. e. the order of considered local polynomial trend. Let us suppose $n \geq m+1$ and consider the regression model

$$y_{t_j} = b_0 + b_1(t_n - t_j) + b_2(t_n - t_j)^2 + b_m(t_n - t_j)^m + \varepsilon_{t_j}, \quad j = 1, 2, \ldots, \tag{24}$$

where $b_0, b_1, \ldots, b_m \in \mathbb{R}$ are unknown parameters and residuals ε_{t_j} have zero expected values (there are no assumptions on their covariance structure). Let us consider a smoothing constant $\alpha \in (0,1)$ and the corresponding discount factor $\beta = 1-\alpha$.

We will construct unbiased estimates $\hat{b}_0(t_n), \hat{b}_1(t_n), \ldots, \hat{b}_m(t_n)$ of the parameters $b_0, b_1, \ldots, b_m$ based on the first n observations of time series y. For this purpose we define smoothing statistics $S_{t_j}^{[p]}$, $p = 1, 2, \ldots, m+1$, in the following way:

$$S_{t_j}^{[1]} = \alpha_{t_j} \sum_{i=1}^{j} y_{t_i} \beta^{t_j - t_i}\,, \tag{25}$$

$$S_{t_j}^{[p+1]} = \alpha_{t_j} \sum_{i=1}^{j} S_{t_i}^{[p]} \beta^{t_j - t_i}\,, \quad p = 1, 2, \ldots, m\,, \tag{26}$$

where $\alpha_{t_j} = \left(\sum_{i=1}^{j} \beta^{t_j - t_i} \right)^{-1}$. For $k = 0, 1, \ldots, m$ and $j = 1, 2, \ldots$ let us denote

$${}^{k}T_{t_j}^{[1]}(t_n) = \alpha_{t_j} \sum_{i=1}^{j} (t_n - t_i)^k \beta^{t_j - t_i}\,, \tag{27}$$

$${}^{k}T_{t_j}^{[p+1]}(t_n) = \alpha_{t_j} \sum_{i=1}^{j} {}^{k}T_{t_i}^{[p]}(t_n) \beta^{t_j - t_i}\,, \quad p = 1, 2, \ldots, m\,, \tag{28}$$

Obviously ${}^{0}T_{t_j}^{[p]}(t_n) \equiv 1$. To simplify the notation let us denote ${}^{k}T_{t_n}^{[p]} = {}^{k}T_{t_n}^{[p]}(t_n)$. Now let us look at our model (24). It is

$$\mathrm{E}(y_{t_j}) = b_0 + b_1(t_n - t_j) + b_2(t_n - t_j)^2 + \cdots + b_m(t_n - t_j)^m\,. \tag{29}$$

Applying linear smoothing operator of order p to (29) we can express in our notation

$$\mathrm{E}\left(S_{t_n}^{[p]}\right) = b_0 + b_1\ {}^{1}T_{t_n}^{[p]} + b_2\ {}^{2}T_{t_n}^{[p]} + \cdots + b_m\ {}^{m}T_{t_n}^{[p]}\,, \quad p = 1, 2, \ldots, m+1\,. \tag{30}$$

This is a system of $m+1$ linear equations for $m+1$ unknown parameters $b_0, b_1, \ldots, b_m$. These are (as the solution of the system) linear functions of the left hand sides in (30). Replacing the expected values $\mathrm{E}(S_{t_n}^{[p]})$ directly by the values $S_{t_n}^{[p]}$, we obtain unbiased estimates $\hat{b}_0(t_n), \hat{b}_1(t_n), \ldots, \hat{b}_m(t_n)$ of parameters $b_0, b_1, \ldots, b_m$. Their unbiasedness follows from linearity of expected value. So we get our estimates of the polynomial trend at time t_n as a solution of

$$b_0 + b_1\ {}^{1}T_{t_n}^{[p]} + b_2\ {}^{2}T_{t_n}^{[p]} + \cdots + b_m\ {}^{m}T_{t_n}^{[p]} = S_{t_n}^{[p]}\,, \quad p = 1, 2, \ldots, m+1\,. \tag{31}$$

The smoothed value at time t_n and the forecast for $\tau > 0$ time units ahead can be obtained simply as

$$\hat{y}_{t_n} = \hat{b}_0(t_n)\,, \tag{32}$$

$$\hat{y}_{t_n+\tau}(t_n) = \hat{b}_0(t_n) + \hat{b}_1(t_n)(-\tau) + \cdots + \hat{b}_m(t_n)(-\tau)^m\,. \tag{33}$$

After receiving a new observation $y_{t_{n+1}}$ we move from time t_n to time t_{n+1} and estimate parameters $b_0, b_1, \ldots, b_m$ in the updated model

$$y_{t_j} = b_0 + b_1(t_{n+1} - t_j) + \cdots + b_m(t_{n+1} - t_j)^m + \varepsilon_{t_j}\,, \quad j = 1, 2, \ldots\,. \tag{34}$$

These estimates $\hat{b}_0(t_{n+1}), \hat{b}_1(t_{n+1}), \dots, \hat{b}_m(t_{n+1})$ are solutions of the updated system (31):

$$b_0 + b_1\,{}^1T^{[p]}_{t_{n+1}} + b_2\,{}^2T^{[p]}_{t_{n+1}} + \dots + b_m\,{}^mT^{[p]}_{t_{n+1}} = S^{[p]}_{t_{n+1}}\,, \quad p = 1, 2, \dots, m+1. \tag{35}$$

We will derive recursive formulas which allow us to compute coefficients of this updated system (35) using coefficients of the original system (31). It is obviously

$$\alpha_{t_{n+1}} = \frac{\alpha_{t_n}}{\alpha_{t_n} + \beta^{t_{n+1}-t_n}}\,, \tag{36}$$

$$S^{[1]}_{t_{n+1}} = (1 - \alpha_{t_{n+1}}) S^{[1]}_{t_n} + \alpha_{t_{n+1}} y_{t_{n+1}}\,, \tag{37}$$

$$S^{[p+1]}_{t_{n+1}} = (1 - \alpha_{t_{n+1}}) S^{[p+1]}_{t_n} + \alpha_{t_{n+1}} S^{[p]}_{t_{n+1}}\,, \quad p = 1, 2, \dots, m\,. \tag{38}$$

Further from binomial theorem and linearity of the smoothing operator we can derive for $k = 1, 2, \dots, m$ and $p = 1, 2, \dots, m+1$ the formula

$${}^kT^{[p]}_{t_n}(t_{n+1}) = \sum_{i=0}^{k} \left[\binom{k}{i} (t_{n+1} - t_n)^{k-i}\,{}^iT^{[p]}_{t_n} \right]. \tag{39}$$

And finally formulas analogous to (37) and (38) are ($k = 1, 2, \dots, m$)

$${}^kT^{[1]}_{t_{n+1}} = (1 - \alpha_{t_{n+1}})\,{}^kT^{[1]}_{t_n}(t_{n+1})\,, \tag{40}$$

$${}^kT^{[p+1]}_{t_{n+1}} = (1 - \alpha_{t_{n+1}})\,{}^kT^{[p+1]}_{t_n}(t_{n+1}) + \alpha_{t_{n+1}}\,{}^kT^{[p]}_{t_{n+1}}\,, \quad p = 1, 2, \dots, m\,. \tag{41}$$

The main difference against the same method for regular time series is that now, besides smoothing statistics $S^{[p]}_{t_n}$, we must recalculate at each time step also the variable smoothing coefficient α_{t_n} and the left hand side coefficients ${}^kT^{[p]}_{t_n}$ of the system (31). Their variability also forces us to solve a new system of $m+1$ linear equations at each time step. The computational demand of this method is naturally rapidly growing with higher orders m. However for $m = 0, 1, 2$ the formulas are still quite simple and easy to implement. The case $m = 0$ corresponds to the Wright's simple exponential smoothing for time series with local constant trend, see [7]. The case $m = 1$ is equivalent to the double exponential smoothing for time series with local linear trend presented in [6]. The case $m = 2$ which is the last one with practical importance is a triple exponential smoothing for time series with local quadratic trend.

The method has been derived explicitly for time series with finite history (but the recursive formulas would be exactly the same if we assumed infinite history). Namely, there is no argument which would prevent us to compute α_{t_1}, $S^{[p]}_{t_1}$ and ${}^kT^{[p]}_{t_1}$ using formulas (25), (26), (27) and (28). It is

$$\alpha_{t_1} = 1\,, \quad S^{[p]}_{t_1} = y_{t_1}\,, \quad {}^0T^{[p]}_{t_1} = 0\,, \quad {}^kT^{[p]}_{t_1} = 1 \tag{42}$$

for $p = 1, 2, \dots, m+1$ and $k = 1, 2, \dots, m$. Further we can continue with recursive update from time t_1 to time t_2 etc. Having first n observations of time series y at our disposal, where $n \geq m+1$, we can successively compute statistics up to time t_n.

Here we can already generate smoothed value and forecasts in time series y. The condition $n \geq m+1$ is necessary for the system (31) to have a unique solution.

Although there is no problem with initialization of this recursive method, see (42), we will show a possible way how to compute initial values α_{t_0}, $S_{t_0}^{[p]}$ and ${}^kT_{t_0}^{[p]}$ which allows us to construct the smoothed value and forecasts already from time t_0. We will proceed in the same way as Cipra [6] did in the case of double exponential smoothing. Let us denote again q the average time spacing of our time series y and set

$$t_0 = t_1 - q \quad \text{and} \quad \alpha_{t_0} = 1 - (1-\alpha)^q \,. \tag{43}$$

Naturally we take ${}^0T_{t_0}^{[p]} = 1$ for $p = 1, 2, \ldots, m+1$. The values ${}^kT_{t_0}^{[p]}$ for $p = 1, 2, \ldots, m+1$ and $k = 1, 2, \ldots, m$ can be easily obtained as a fixed point of the corresponding formulas (40) and (41) where we use $t_{n+1} - t_n \equiv q$ and $\alpha_{t_n} \equiv \alpha_{t_0}$. Finally the values $S_{t_0}^{[p]}$, $p = 1, 2, \ldots, m+1$, can be taken as

$$S_{t_0}^{[p]} = \hat{b}_0(t_0) + \hat{b}_1(t_0)\, {}^1T_{t_0}^{[p]} + \hat{b}_2(t_0)\, {}^2T_{t_0}^{[p]} + \cdots + \hat{b}_m(t_0)\, {}^mT_{t_0}^{[p]} \,, \tag{44}$$

where $\hat{b}_0(t_0), \hat{b}_1(t_0), \ldots, \hat{b}_m(t_0)$ are estimates of parameters $b_0, b_1, \ldots, b_m$ in the regression model

$$y_{t_j} = b_0 + b_1(t_0 - t_j) + b_2(t_0 - t_j)^2 + \cdots + b_m(t_0 - t_j)^m + \varepsilon_{t_j} \tag{45}$$

based on several starting observations (at least $m+1$) of time series y. Specially, the DLS method with weights decreasing exponentially *into future* with the discount factor β can be used to obtain these regression estimates.

4. METHOD BASED ON DLS ESTIMATION

In this section we will show the second possibility how to extend the exponential smoothing of order m to the case of irregular time series. This method will be explicitly based on using a DLS estimation method to get polynomial trend parameters.

Let us consider again an irregular time series $y_{t_1}, y_{t_2}, \ldots, y_{t_n}, y_{t_{n+1}}, \ldots$ observed at times $t_1 < t_2 < \cdots < t_n < t_{n+1} < \ldots$. Let $m \in \mathbb{N}_0$ be the degree of considered local polynomial trend. Let us suppose $n \geq m+1$ and consider the regression model

$$y_{t_j} = b_0 + b_1(t_n - t_j) + b_2(t_n - t_j)^2 + b_m(t_n - t_j)^m + \varepsilon_{t_j} \,, \quad j = 1, 2, \ldots \,, \tag{46}$$

where $b_0, b_1, \ldots, b_m \in \mathbb{R}$ are unknown parameters and random error components ε_{t_j} have zero expected values (there are no assumptions on their covariance structure). Let us consider smoothing constant $\alpha \in (0,1)$ and the corresponding discount factor $\beta = 1 - \alpha$.

We will estimate the unknown parameters $b_0, b_1, \ldots, b_m$ of model (46) based on first n observations $y_1, y_2, \ldots, y_n$ of time series y using a DLS (*discounted least squares*) method with discount factor β (see [1], chapters 2.13 and 3.5, for an overview of this estimation method). The corresponding system of normal equations for these estimates $\hat{b}_0(t_n), \hat{b}_1(t_n), \ldots, \hat{b}_m(t_n)$ has the form

$$b_0 + b_1 T_{t_n}^{(k)} + b_2 T_{t_n}^{(k+1)} + \cdots + b_m T_{t_n}^{(k+m)} = Y_{t_n}^{(k)} \,, \quad k = 0, 1, \ldots, m \,, \tag{47}$$

where we have denoted

$$T_{t_n}^{(k)} = \sum_{i=1}^{n} (t_n - t_i)^k \beta^{t_n - t_i}\,, \qquad k = 0, 1, \ldots, 2m\,, \tag{48}$$

$$Y_{t_n}^{(k)} = \sum_{i=1}^{n} y_{t_i} (t_n - t_i)^k \beta^{t_n - t_i}\,, \quad k = 0, 1, \ldots, m\,. \tag{49}$$

If $n \geq m + 1$ and $t_1 < t_2 < \cdots < t_n$ then the system (47) has a unique solution. The smoothed value of time series y at time t_n and the forecast for $\tau > 0$ time units ahead from time t_n are obtained again as

$$\hat{y}_{t_n} = \hat{b}_0(t_n), \tag{50}$$

$$\hat{y}_{t_n+\tau}(t_n) = \hat{b}_0(t_n) + \hat{b}_1(t_n)(-\tau) + \cdots + \hat{b}_m(t_n)(-\tau)^m\,. \tag{51}$$

Since DLS estimates $\hat{b}_0(t_n), \hat{b}_0(t_2), \ldots, \hat{b}_0(t_m)$ are unbiased, the above smoothed value and forecast are unbiased as well, i. e.

$$\mathrm{E}\,(\hat{y}_{t_n}) = \mathrm{E}\,(y_{t_n}) \quad \text{and} \quad \mathrm{E}\,[\hat{y}_{t_n+\tau}(t_n)] = \mathrm{E}\,[y_{t_n+\tau}]\,. \tag{52}$$

When we get a new observation $y_{t_{n+1}}$, we move from time t_n to time t_{n+1} and we will estimate the parameters $b_0, b_1, \ldots, b_m$ in the updated model

$$y_{t_j} = b_0 + b_1(t_{n+1} - t_j) + b_2(t_{n+1} - t_j)^2 + \cdots + b_m(t_{n+1} - t_j)^m + \varepsilon_{t_j}\,,\; j = 1, 2, \ldots \tag{53}$$

using the same DLS method. These estimates $\hat{b}_0(t_{n+1}), \hat{b}_1(t_{n+1}), \ldots, \hat{b}_m(t_{n+1})$ will be obtained by solving the system (47) shifted to time t_{n+1}, i. e. the system

$$b_0 + b_1 T_{t_{n+1}}^{(k)} + b_2 T_{t_{n+1}}^{(k+1)} + \cdots + b_m T_{t_{n+1}}^{(k+m)} = Y_{t_{n+1}}^{(k)}\,, \quad k = 0, 1, \ldots, m\,. \tag{54}$$

We need recursive formulas which enable us to get coefficients of the new system (54) using coefficients of the original system (47). It can be shown easily that

$$T_{t_{n+1}}^{(k)} = \beta^{t_{n+1} - t_n} \sum_{i=0}^{k} \left[\binom{k}{i} (t_{n+1} - t_n)^{k-i}\, T_{t_n}^{(i)} \right], \quad k = 1, 2, \ldots, 2m, \tag{55}$$

$$T_{t_{n+1}}^{(0)} = 1 + \beta^{t_{n+1} - t_n}\, T_{t_n}^{(0)} \tag{56}$$

and analogously

$$Y_{t_{n+1}}^{(k)} = \beta^{t_{n+1} - t_n} \sum_{i=0}^{k} \left[\binom{k}{i} (t_{n+1} - t_n)^{k-i}\, Y_{t_n}^{(i)} \right], \quad k = 1, 2, \ldots, m, \tag{57}$$

$$Y_{t_{n+1}}^{(0)} = y_{t_{n+1}} + \beta^{t_{n+1} - t_n}\, Y_{t_n}^{(0)}\,. \tag{58}$$

So the application of this method is exactly the same as of the exponential smoothing from Section 3. Of course, the computational demand is rapidly growing with higher orders m. But for $m = 0, 1, 2$ the formulas are still quite simple. For larger m this method is computationally less demanding than the previous one.

As in the Section 3, the method has been derived for time series with finite history again (the recursive formulas would stay unchanged again if we assumed infinite history). Therefore we can again compute $T_{t_1}^{(k)}$ and $Y_{t_1}^{(k)}$ using formulas (48) and (49). We get

$$T_{t_1}^{(0)} = 1\,, \quad T_{t_1}^{(k)} = 0\,, \quad Y_{t_1}^{(0)} = y_{t_1}\,, \quad Y_{t_1}^{(k)} = 0 \tag{59}$$

for $k \geq 1$. Having first n observations of time series y, $n \geq m+1$, we can proceed with recurrent computation up to time t_n and generate the smoothed value and forecasts here. The condition $n \geq m+1$ is again necessary for the system (47) to have a unique solution.

Similarly as in Section 3 we will show a possible way of selection initial values t_0, $T_{t_0}^{(0)}$ and $Y_{t_0}^{(0)}$ which enables us to construct the smoothed value and forecast already from time t_0. Again let us set $t_0 = t_1 - q$ where q is the average time spacing of the time series y. The values $T_{t_0}^{(k)}$ will be constructed using the assumption that the fictive infinite history of y starting in t_0 and going into past has been observed regularly with time intervals of length q. So we will take

$$T_{t_0}^{(k)} = \sum_{j=0}^{\infty} (jq)^k \beta^{jq}\,, \quad k = 0, 1, \ldots, 2m\,. \tag{60}$$

If we denote $T_k(x) = \sum_{j=0}^{\infty} j^k x^j$ for $|x| < 1$ and $k \geq 0$, we can write

$$T_{t_0}^{(k)} = q^k\, T_k(\beta^q)\,. \tag{61}$$

Values $T_k(x)$ for a fixed x can be computed recursively. For $k \geq 0$ it is

$$T_{k+1}(x) = \frac{x}{1-x} \sum_{i=0}^{k} \binom{k+1}{i} T_i(x) \tag{62}$$

and $T_0(x) = \frac{1}{1-x}$. Initial values $Y_{t_0}^{(k)}$ for $k = 0, 1, \ldots, m$ will be taken as

$$Y_{t_0}^{(k)} = \hat{b}_0(t_0)\, T_{t_0}^{(k)} + \hat{b}_1(t_0)\, T_{t_0}^{(k+1)} + \cdots + \hat{b}_m(t_0)\, T_{t_0}^{(k+m)}\,, \tag{63}$$

where $\hat{b}_0(t_0), \hat{b}_1(t_0), \ldots, \hat{b}_m(t_0)$ are estimates of parameters $b_0, b_1, \ldots, b_m$ in model

$$y_{t_j} = b_0 + b_1(t_0 - t_j) + b_2(t_0 - t_j)^2 + \cdots + b_m(t_0 - t_j)^m + \varepsilon_{t_j}\,, \quad j = 1, 2, \ldots\,. \tag{64}$$

These estimates will be based on several (at least $m+1$) initial observations of time series y using the DLS method with weights decreasing exponentially *into future* with discount factor β.

It is trivial that the case $m = 0$ of this method is nothing else but Wright's simple exponential smoothing. The case $m = 1$ (i. e. the method for local linear trend) is similar but not equivalent to Cipra's double exponential smoothing from [6]. And the both methods are not a special case of Wright's modification of the Holt method as it is in the case of regular time series. So we have three different recursive methods for irregular time series with local linear trend. In a concrete numerical example, any of these methods can provide the best results, see Section 6.

5. MAXIMUM LIKELIHOOD PARAMETERS ESTIMATION

All of exponential smoothing type methods have one or more numerical parameters whose values must be selected somehow for the best performance of the method. Widely used approach in context of regular time series is to choose the value which minimizes a certain criterion like MSE (*mean square error*). When we know the formula for forecasting error variance $\mathrm{var}[e_{t_n+\tau}(t_n)]$ and we assume its distribution type then we can estimate parameters of the method using the maximum likelihood method. This will be a generalization of the approach mentioned above which will take time irregularity of observations into account.

Let us consider a forecasting method with k-dimensional parameter $\alpha \in A$ where $A \subseteq \mathbb{R}^k$. We will construct the estimate $\hat{\alpha}$ of this parameter based on observed forecasting errors $e_{t_1}, e_{t_2}, \ldots, e_{t_n}$ at times $t_1, t_2, \ldots, t_n$. Let us denote more rigorously $e_{t_1}(\alpha), e_{t_2}(\alpha), \ldots, e_{t_n}(\alpha)$ these forecasting errors occurred when using the method with parameter value $\alpha \in A$. Let for the true value of α be

$$\mathrm{E}[e_{t_j}(\alpha)] = 0 \quad \text{and} \quad \mathrm{var}[e_{t_j}(\alpha)] = \sigma^2 v_{t_j}(\alpha)\,, \quad j = 1, 2, \ldots, n\,, \tag{65}$$

where $v_{t_j}(\alpha) > 0$ are known positive functions and $\sigma^2 > 0$ is another unknown parameter. To achieve a particular form of the likelihood function we must assume a specific distribution of errors $e_{t_j}(\alpha)$. Let us suppose for example that this distribution is normal, i. e.

$$e_{t_j}(\alpha) \sim \mathrm{N}\left(0, \sigma^2 v_{t_j}(\alpha)\right). \tag{66}$$

Of course, it is possible to consider a different distribution type here as well. Further let us assume that for the true value of α, the forecasting errors $e_{t_j}(\alpha)$ are uncorrelated. Now we can already write down the likelihood function

$$L\left(\alpha, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-n/2} \left[\prod_{j=1}^{n} v_{t_j}(\alpha)\right]^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{e_{t_j}^2(\alpha)}{v_{t_j}(\alpha)}\right\} \tag{67}$$

and the log-likelihood function

$$l\left(\alpha, \sigma^2\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{j=1}^{n} \ln\left[v_{t_j}(\alpha)\right] - \frac{1}{2\sigma^2}\sum_{j=1}^{n} \frac{e_{t_j}^2(\alpha)}{v_{t_j}(\alpha)}\,. \tag{68}$$

So we solve a minimization problem

$$\min_{\alpha\in A,\, \sigma^2>0} \left\{ n\ln\sigma^2 + \sum_{j=1}^{n} \ln\left[v_{t_j}(\alpha)\right] + \frac{1}{\sigma^2}\sum_{j=1}^{n} \frac{e_{t_j}^2(\alpha)}{v_{t_j}(\alpha)} \right\}. \tag{69}$$

Maximum likelihood estimates can be then expressed as

$$\hat{\alpha} \;=\; \arg\min_{\alpha\in A} \left\{ \ln \sum_{j=1}^{n} \frac{e_{t_j}^2(\alpha)}{v_{t_j}(\alpha)} + \frac{1}{n}\sum_{j=1}^{n} \ln\left[v_{t_j}(\alpha)\right] \right\}, \tag{70}$$

$$\hat{\sigma}^2 \;=\; \frac{1}{n}\sum_{j=1}^{n} \frac{e_{t_j}^2(\hat{\alpha})}{v_{t_j}(\hat{\alpha})}\,. \tag{71}$$

As one can see from (70) the weights of square errors $e_{t_j}^2(\alpha)$ in minimized expression are inversely proportional to the corresponding variance factors $v_{t_j}(\alpha)$. The suggested minimization must be done numerically. In (71) the parameter σ^2 is estimated as a mean square of values $\widetilde{e}_{t_j}(\alpha) = \frac{e_{t_j}(\alpha)}{\sqrt{v_{t_j}(\alpha)}}$ for which

$$\widetilde{e}_{t_j}(\alpha) \sim \mathrm{N}(0, \sigma^2) \tag{72}$$

if α is the true value of the parameter. These *normalized forecasting errors* form a white noise and are useful in testing adequacy of applying the method for a particular time series.

In the case of regular time series we can without loss of generality suppose that $v_{t_j}(\alpha) \equiv 1$ and the formulas (70) and (71) simplify to the form

$$\hat{\alpha} = \arg\min_{\alpha \in A} \sum_{j=1}^{n} e_{t_j}^2(\alpha) \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} e_{t_j}^2(\hat{\alpha}), \tag{73}$$

i. e. to the classical MSE minimization.

6. NUMERICAL STUDY

We have done a simulation numerical study to compare the predictive performance of the suggested methods with the existing ones from [7] and [6]. We have compared (1) Wright's simple exponential smoothing with the irregularly observed ARIMA$(0,1,1)$ method from Section 2 and (2) Wright's version of Holt method with the double exponential smoothing from Section 3 and DLS linear trend method from Section 4.

The methodology was the same in both cases. We generated a regular time series using certain ARIMA process (with the white noise variance equal to 1) and then we sampled it to create an irregular time series. The individual sampling steps were taken randomly from $\{1, 2, \dots N\}$. The resulted irregular time series had 3000 observations. Then we used the concerned methods with smoothing constants minimizing mean square one-step-ahead forecasting error (MSE) through the whole data. Initial values didn't play a considerable role because of the length of the series.

For the first comparison the ARIMA$(0,1,1)$ process was used with $\alpha = 0.1, 0.2, 0.4$ (we use the parametrization of simple exponential smoothing). We took $N = 2, 3, 5, 10$ which led to $3 \times 4 = 12$ different simulated irregular time series. The results are presented in Table 1. Although we generated the series by the model for which our suggested method is optimal, the differences in achieved RMSE are inconsiderable. While for the suggested method the optimal values of α are always close to the value used for generating and don't depend on N, for the simple exponential smoothing they decrease significantly with increasing value of N. The reason is that the smoothing coefficient in Wright's method tends to 1 exponentially fast as the time step tends to infinity (see [7]) while the optimal convergence speed is that of the linear-rational function in (14). So this is compensated by the lower α value when the average time step is higher. This is annoying since we get different optimal α values for the same series depending just on the observation frequency.

Our suggested method doesn't suffer from this phenomenon. This together with the model based prediction intervals formula (19) are arguments for using the suggested method rather than the Wright's one.

Table 1. Wright's simple exponential smoothing and irregularly observed ARIMA$(0,1,1)$ method: optimal α and achieved RMSE for 12 simulated time series.

		Simple exp. smoothing		ARIMA$(0,1,1)$	
α	N	optimal α	RMSE	optimal α	RMSE
0.1	2	0.0896	1.0138	0.1093	1.0138
0.1	3	0.0709	1.0192	0.0997	1.0192
0.1	5	0.0663	1.0371	0.1129	1.0368
0.1	10	0.0453	1.0780	0.1040	1.0780
0.2	2	0.1687	1.0153	0.2033	1.0153
0.2	3	0.1495	1.0327	0.2063	1.0330
0.2	5	0.1142	1.0906	0.1926	1.0905
0.2	10	0.0917	1.1596	0.2068	1.1587
0.4	2	0.3426	1.0520	0.4068	1.0525
0.4	3	0.2989	1.0899	0.4020	1.0885
0.4	5	0.2441	1.1779	0.3955	1.1777
0.4	10	0.1867	1.4019	0.4042	1.3959

For the second comparison the ARIMA$(0,2,2)$ process was used with nine different combinations of α_H and γ_H (we use the parametrization of Holt method). It is well known that in the context of regular time series double exponential smoothing with smoothing constant α is equivalent to Holt method with smoothing constants

$$\alpha_H = \alpha\,(2-\alpha) \quad \text{and} \quad \gamma_H = \frac{\alpha}{2-\alpha}\,. \tag{74}$$

The first three generating combinations correspond to $(\alpha_H(\alpha), \gamma_H(\alpha))$ for $\alpha = 0.1, 0.2, 0.4$. The next three combinations have lower α_H and higher γ_H value when compared to the first three ones and the last three combinations just in the opposite way. Shift by ± 1 in the argument of the logistic curve $1/[1+\exp(-x)]$ is used to make the values *lower* or *higher*. All the values of α_H and γ_H were rounded to three decimal digits.

Together with taking $N = 2, 3, 5$ this led to $3 \times 3 \times 3 = 27$ different simulated irregular time series. The results are presented in Table 2. The optimal α values and achieved RMSE are almost the same for double exponential smoothing and DLS linear trend method. Holt method has slightly worse performance for the first nine series while for the rest of the series it does usually better, gaining from the flexibility of two independent smoothing constants.

The autocorrelation of normalized forecasting errors (see Section 5) from the methods with one parameter was not significantly different from 0 for the first nine rows of Table 2, was negative for the next nine rows and positive for the last nine rows of the table. This empirical fact is consistent with our intuition and can be interpreted as the following practical recommendation: when forecasting errors from one-parameter method are not correlated then Holt method will probably not offer better results. Especially for short time series it is then maybe more reasonable to use the one-parameter method which can prevent us from over-fitting and can provide better out-of-sample results. This also prevents us from the more complicated 2-dimensional smoothing constants optimization.

Table 2. Holt method, double exponential smoothing and DLS linear trend method for irregular time series: optimal α (and γ) and achieved RMSE for 27 simulated time series.

			Holt method			Double exp. smooth.		DLS linear trend	
α	γ	N	opt. α	opt. γ	RMSE	opt. α	RMSE	opt. α	RMSE
0.190	0.053	2	0.1592	0.0472	1.0398	0.0877	1.0388	0.0876	1.0391
0.190	0.053	3	0.1522	0.0360	1.0694	0.0800	1.0646	0.0797	1.0649
0.190	0.053	5	0.1427	0.0341	1.1103	0.0767	1.1000	0.0766	1.1003
0.360	0.111	2	0.3358	0.1142	1.0641	0.1927	1.0599	0.1928	1.0589
0.360	0.111	3	0.2997	0.0826	1.1403	0.1630	1.1328	0.1628	1.1326
0.360	0.111	5	0.2930	0.0719	1.3063	0.1500	1.2793	0.1507	1.2777
0.640	0.250	2	0.5813	0.2240	1.1486	0.3529	1.1426	0.3546	1.1428
0.640	0.250	3	0.5606	0.2303	1.3652	0.3425	1.3513	0.3507	1.3467
0.640	0.250	5	0.4829	0.2554	1.7656	0.3169	1.7595	0.3327	1.7506
0.079	0.131	2	0.0958	0.0836	1.0185	0.0797	1.0271	0.0795	1.0269
0.079	0.131	3	0.0949	0.0593	1.0691	0.0687	1.0720	0.0687	1.0718
0.079	0.131	5	0.1004	0.0364	1.1275	0.0631	1.1212	0.0632	1.1213
0.171	0.254	2	0.1864	0.1902	1.0646	0.1641	1.0863	0.1642	1.0864
0.171	0.254	3	0.1866	0.1459	1.1486	0.1472	1.1620	0.1470	1.1617
0.171	0.254	5	0.1809	0.1302	1.2343	0.1410	1.2287	0.1416	1.2244
0.395	0.475	2	0.3980	0.4174	1.1760	0.3534	1.1989	0.3568	1.1951
0.395	0.475	3	0.3755	0.4326	1.3057	0.3406	1.3354	0.3488	1.3326
0.395	0.475	5	0.3810	0.4152	1.7325	0.3314	1.7506	0.3556	1.7189
0.389	0.020	2	0.3364	0.0129	1.0496	0.1463	1.0671	0.1460	1.0660
0.389	0.020	3	0.2756	0.0183	1.1120	0.1151	1.1271	0.1152	1.1279
0.389	0.020	5	0.2558	0.0186	1.1886	0.1034	1.1986	0.1035	1.1979
0.605	0.044	2	0.5250	0.0423	1.1158	0.2237	1.1405	0.2240	1.1396
0.605	0.044	3	0.4459	0.0533	1.1931	0.1945	1.2078	0.1935	1.2107
0.605	0.044	5	0.4230	0.0334	1.3991	0.1587	1.4129	0.1593	1.4123
0.829	0.109	2	0.7949	0.1147	1.2151	0.3667	1.2348	0.3664	1.2364
0.829	0.109	3	0.6443	0.1176	1.3676	0.2946	1.3755	0.2973	1.3770
0.829	0.109	5	0.5510	0.1210	1.7847	0.2452	1.7828	0.2482	1.7858

ACKNOWLEDGEMENT

The paper is a part of research project MSM 0021620839 of the Ministry of Education, Youth and Sports of the Czech Republic.

(Received September 25, 2007.)

REFERENCES

[1] B. Abraham and J. Ledolter: Statistical Methods for Forecasting. Wiley, New York 1983.
[2] M. Aldrin and E. Damsleth: Forecasting non-seasonal time series with missing observations. J. Forecasting *8* (1989), 97–116.
[3] J. Anděl and J. Zichová: A method for estimating parameter in nonnegative MA(1) models. Comm. Statist. Theory Methods *31* (2002), 2101–2111.
[4] C. Chatfield: Time-Series Forecasting. Chapman & Hall/CRC, 2002.
[5] T. Cipra, J. Trujillo, and A. Rubio: Holt–Winters method with missing observations. Manag. Sci. *41* (1995), 174–8.
[6] T. Cipra: Exponential smoothing for irregular data. Appl. Math. *51* (2006), 597–604.
[7] D. J. Wright: Forecasting data published at irregular time intervals using extension of Holt's method. Manag. Sci. *32* (1986), 499–510.
[8] J. Zichová: On a method of estimating parameters in non-negative ARMA models. Kybernetika *32* (1996), 409–424.

Tomáš Cipra and Tomáš Hanzák, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics — Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mails: cipra@karlin.mff.cuni.cz, tomas.hanzak@post.cz