Tomáš Cipra

Asymmetric recursive methods for time series

Applications of Mathematics, Vol. 39 (1994), No. 3, 203–214

Persistent URL: http://dml.cz/dmlcz/134253

ASYMMETRIC RECURSIVE METHODS FOR TIME SERIES

TOMÁŠ CIPRA, Praha

(Received January 22, 1993)

Summary. The problem of asymmetry appears in various aspects of time series modelling. Typical examples are asymmetric time series, asymmetric error distributions and asymmetric loss functions in estimating and predicting. The paper deals with asymmetric modifications of some recursive time series methods including Kalman filtering, exponential smoothing and recursive treatment of Box-Jenkins models.

Keywords: asymmetric recursive methods, time series, Kalman filter, exponential smoothing, asymmetric time series, autoregressive model, split-normal distribution

AMS classification: 62M10 (62M20, 60G35, 93E11)

1. INTRODUCTION

By considering various forms of asymmetry of time series data we can improve results of the corresponding analysis.

A typical example is the case of asymmetric time series which respond to innovations with one of two different rules according to whether the innovation is positive or negative (see e.g. [14]). Sometimes an asymmetric loss function may be suitable for the construction of predictions due to a practical motivation (see e.g. [7]). If prediction errors are analyzed as asymmetric (e.g. in investory control) the resulting confidence intervals may significantly reduce costs (see [9]). The asymmetric curve analysis is presented in [13].

This paper suggests asymmetric modifications of some recursive time series methods since the recursive procedures are popular in practical time series analysis. A simple asymmetric modification of the Kalman filter based on the asymmetric least squares is described in Section 3. Asymmetric modifications of exponential smoothing procedures motivated by the relations of exponential smoothing to Box-Jenkins models are given in Section 4. Asymmetric recursive estimation in autoregressive models is considered in Section 5 including a convergence result for asymmetric trimming. Section 2 contains some preliminaries necessary for further text.

2. Preliminaries

2.1. Kalman filter. In applications to univariate time series it is sufficient to consider the Kalman filter of the form

$$x_t = F_t x_{t-1} + w_t, \tag{2.1}$$

$$y_t = h_t x_t + v_t, \tag{2.2}$$

where $E(w_t) = 0$, $E(v_t) = 0$, $E(w_s w_t') = \delta_{st} Q_t$, $E(v_s v_t) = \delta_{st} r_t$, $E(w_s v_t) = 0$ and some initial conditions are fulfilled. The state equation (2.1) describes the behavior of an n-dimensional state vector x_t in time while the observation equation (2.2) describes the relation of x_t to the scalar observations y_t.

The Kalman filter gives recursive formulas for construction of the linear minimum variance estimator $\hat{x}_t^t$ of the state x_t and for its error covariance matrix $P_t^t = E(x_t - \hat{x}_t^t)(x_t - \hat{x}_t^t)'$ in a current time period t using the previous information $\{y_0, y_1, \dots, y_t\}$:

$$\hat{x}_t^t = \hat{x}_t^{t-1} + \frac{P_t^{t-1} h_t'}{h_t P_t^{t-1} h_t' + r_t}(y_t - h_t \hat{x}_t^{t-1}), \tag{2.3}$$

$$P_t^t = P_t^{t-1} - \frac{P_t^{t-1} h_t' h_t P_t^{t-1}}{h_t P_t^{t-1} h_t' + r_t}, \tag{2.4}$$

where

$$\hat{x}_t^{t-1} = F_t \hat{x}_{t-1}^{t-1}, \quad \hat{y}_t^{t-1} = h_t \hat{x}_t^{t-1}, \tag{2.5}$$

$$P_t^{t-1} = F_t P_{t-1}^{t-1} F_t' + Q_t \tag{2.6}$$

are the predictive values constructed for time t at time $t-1$. The state value $\hat{x}_t^t$ can be obtained for the given predictive values $\hat{x}_t^{t-1}$ and P_t^{t-1} by the least squares minimization

$$\hat{x}_t^t = \operatorname*{argmin}_{x_t \in \mathsf{R}^n} \left\{ (\hat{x}_t^{t-1} - x_t)'(P_t^{t-1})^{-1}(\hat{x}_t^{t-1} - x_t) + r_t^{-1}(y_t - h_t x_t)^2 \right\}. \tag{2.7}$$

2.2. Exponential smoothing. The exponential smoothing procedures are popular in practice for their numerical simplicity (see e.g. [2]). For instance, the simple exponential smoothing suitable for time series $\{y_t\}$ with a locally constant trend has the form

$$\hat{y}_t^t = \hat{y}_t^{t-1} + \alpha(y_t - \hat{y}_t^{t-1}), \tag{2.8}$$

$$\hat{y}_{t+k}^t = \hat{y}_t^t, \qquad k \geqslant 1, \tag{2.9}$$

where α is a suitable smoothing constant ($0 < \alpha < 1$) and $\hat{y}_{t+k}^t$ denotes the prediction of y_{t+k} at time t (in particular, $\hat{y}_t^t$ is the smoothed value at time t). It can be shown that the predictions provided by (2.8), (2.9) are the same as the recursive ones by the model ARIMA(0,1,1)

$$(1 - B)y_t = \varepsilon_t - (1 - \alpha)\varepsilon_{t-1}, \tag{2.10}$$

where B denotes the backward-shift operator fulfilling $By_t = y_{t-1}$.

2.3. Asymmetric time series. The asymmetric time series which respond to innovations in two different ways according to whether the innovation is positive or negative form an important class of nonlinear time series. For instance, the asymmetric process MA(1) has the form

$$y_t = \varepsilon_t + \vartheta_{11}\varepsilon_{t-1}^- + \vartheta_{21}\varepsilon_{t-1}^+, \tag{2.11}$$

where

$$\varepsilon_t^- = \min(\varepsilon_t, 0), \quad \varepsilon_t^+ = \max(\varepsilon_t, 0) \tag{2.12}$$

are negative and positive innovations (white noise), respectively, and ϑ_{11}, ϑ_{21} are parameters of the model. If $\vartheta_{11} = \vartheta_{21}$ then (2.11) is the classical process MA(1).

In practice, the predictions in the models of the type (2.11) are constructed recursively as

$$\hat{y}_{t+k}^t = \begin{cases} \vartheta_{11}(y_t - \hat{y}_t^{t-1})^- + \vartheta_{21}(y_t - \hat{y}_t^{t-1})^+, & k = 1, \\ 0, & k > 1, \end{cases} \tag{2.13}$$

although the invertibility of the model should be verified theoretically (a model is invertible if the innovations $\{\varepsilon_t\}$ can be estimated from the data $\{y_t\}$). For instance, the condition of invertibility of the model (2.11) has the form $\max(|\vartheta_{11}|, |\vartheta_{21}|) < 1$ (see [14]).

2.4. Split-normal distribution. The split-normal distribution is an asymmetric distribution that is suitable just for the purpose of asymmetric prediction errors since its parameters can be estimated in a simple recursive way (see [9]). It is sufficient to confine oneself to the split-normal distribution $N(0; \sigma_1^2, \sigma_2^2)$ with zero mean value that has the probability density of the form

$$f(x) = \begin{cases} \dfrac{2\sigma_2}{(\sigma_1 + \sigma_2)\sigma_1}\varphi\Big(\dfrac{x}{\sigma_1}\Big), & x < 0, \\ \dfrac{2\sigma_1}{(\sigma_1 + \sigma_2)\sigma_2}\varphi\Big(\dfrac{x}{\sigma_2}\Big), & x \geqslant 0, \end{cases} \tag{2.14}$$

where $\varphi(\cdot)$ is the probability density of the standard normal distribution $N(0, 1)$.

One can easily verify that $X \sim N(0; \sigma_1^2, \sigma_2^2)$ satisfies

$$E(X) = 0, \tag{2.15}$$

$$\operatorname{var}(X) = \sigma_1\sigma_2, \tag{2.16}$$

$$E\big[(X^-)^2 \mid X \leqslant 0\big] = \sigma_1^2, \quad E\big[(X^+)^2 \mid X \geqslant 0\big] = \sigma_2^2 \tag{2.17}$$

(the symbols X^- and X^+ have the same meaning as in (2.12)).

If we denote the one-step-ahead prediction error by the symbol

$$e_t = y_t - \hat{y}_t^{t-1} \tag{2.18}$$

(this symbol will be used troughout the following text) then updated estimates $\hat{\sigma}_{1t}^2$, $\hat{\sigma}_{2t}^2$ of σ_1^2, σ_2^2 can be obtained as

$$\begin{aligned} \hat{\sigma}_{1t}^2 &= \hat{\sigma}_{1,t-1}^2 + \delta z_{t-1}(e_{t-1}^2 - \hat{\sigma}_{1,t-1}^2), \\ \hat{\sigma}_{2t}^2 &= \hat{\sigma}_{2,t-1}^2 + \delta z_{t-1}(e_{t-1}^2 - \hat{\sigma}_{2,t-1}^2), \end{aligned} \tag{2.19}$$

where z_t equals 1 if $e_t < 0$, otherwise it is 0, and δ $(0 < \delta < 1)$ is a damping constant (see [9]).

3. Asymmetric Kalman filter

A simple approach to asymmetry consists in replacing the least squares (LS) estimation by the asymmetric least squares (ALS) estimation (see e.g. [10] for linear regression models).

In the case of the Kalman filter we can use this approach in the framework of the LS minimization (2.7). Let us consider the simplest situation when (2.7) is replaced by the ALS minimization of the form

$$\begin{aligned} \hat{x}_t^t = \operatorname*{argmin}_{x_t \in \mathrm{R}^n} \Big\{ &(\hat{x}_t^{t-1} - x_t)'(P_t^{t-1})^{-1}(\hat{x}_t^{t-1} - x_t) \\ &+ r_{1t}^{-1}\big[(y_t - h_t x_t)^-\big]^2 + r_{2t}^{-1}\big[(y_t - h_t x_t)^+\big]^2 \Big\} \end{aligned} \tag{3.1}$$

for suitable positive values r_{1t}, r_{2t}. If $v_t \sim N(0; \sigma_{1t}^2, \sigma_{2t}^2)$ in (2.2) then the natural choice of r_{1t}, r_{2t} obviously is σ_{1t}^2, σ_{2t}^2.

The result of the minimization (3.1) has the explicit form

$$\hat{x}_t^t = \hat{x}_t^{t-1} + \frac{P_t^{t-1} h_t'}{h_t P_t^{t-1} h_t' + r_{1t}} e_t^- + \frac{P_t^{t-1} h_t'}{h_t P_t^{t-1} h_t' + r_{2t}} e_t^+, \tag{3.2}$$

where $e_t = y_t - \hat{y}_t^{t-1} = y_t - h_t \hat{x}_t^{t-1}$.

This can be shown in the following way. Let x_t be such that $y_t - h_t x_t \geqslant 0$. Then it is not difficult to derive that the unique value of x_t for which the derivative (according to x_t) of ALS in (3.1) is zero has the form

$$x_t^* = \hat{x}_t^{t-1} + \frac{P_t^{t-1} h_t'}{h_t P_t^{t-1} h_t' + r_{2t}} e_t.$$

Since

$$y_t - h_t x_t^* = \frac{r_{2t}}{h_t P_t^{t-1} h_t' + r_{2t}} e_t,$$

we have the equivalence $y_t - h_t x_t^* \geqslant 0$ iff $e_t \geqslant 0$. In the case $y_t - h_t x_t < 0$ the results are analogous with r_{2t} replaced by r_{1t}. This proves the explicit formula (3.2).

If $r_{1t} = r_{2t} = r_t$ then (3.2) becomes the symmetric formula (2.3). One of possible applications of the asymmetric Kalman filter described above will be shown in Section 5.

4. Asymmetric exponential smoothing

One can use a similar approach as to the robustification of exponential smoothing when a suitable robustifying function is applied to the prediction errors e_t in the corresponding recursive formulas of exponential smoothing (see [3], [6]), or as to the exponential smoothing in the L_1-norm when the absolute value is used (see [5]). In the case of asymmetry, it is natural to take the prediction errors e_t in the recursive formulas of exponential smoothing in the same asymmetric way as in the models of asymmetric time series (see Section 2.3). Moreover, if this approach is used then for each case of asymmetric exponential smoothing it is possible to find an asymmetric time series analogue of the type (2.11) that provides the same recursive predictions.

Let us start with the simple exponential smoothing (2.8), (2.9). According to the above discussion its asymmetric modification is

$$\hat{y}_t^t = \hat{y}_t^{t-1} + \alpha_1 e_t^- + \alpha_2 e_t^+, \tag{4.1}$$

$$\hat{y}_{t+k}^t = \hat{y}_t^t, \qquad k \geqslant 1, \tag{4.2}$$

where $\alpha_1, \alpha_2 \in (0,1)$ are the smoothing constants. The same recursive predictions can be obtained by means of the asymmetric model ARIMA(0,1,1) of the form

$$(1-B)y_t = \varepsilon_t - (1-\alpha_1)\varepsilon_{t-1}^- - (1-\alpha_2)\varepsilon_{t-1}^+. \tag{4.3}$$

Namely, in (4.3) we have (compare with (2.13))

$$\begin{aligned} \hat{y}_{t+k}^t &= y_t - (1-\alpha_1)e_t^- - (1-\alpha_2)e_t^+ \\ &= y_t - e_t + \alpha_1 e_t^- + \alpha_2 e_t^+ \\ &= \hat{y}_t^{t-1} + \alpha_1 e_t^- + \alpha_2 e_t^+, \qquad k \geqslant 1. \end{aligned}$$

These predictions are equal to those in (4.1), (4.2).

Further, let us consider the Holt model of exponential smoothing that is suitable for time series $\{y_t\}$ with a locally constant linear trend. Its asymmetric modification can be written as

$$S_t = S_{t-1} + T_{t-1} + \alpha_1 e_t^- + \alpha_2 e_t^+, \tag{4.4}$$

$$T_t = T_{t-1} + \alpha_1\gamma_1 e_t^- + \alpha_2\gamma_2 e_t^+, \tag{4.5}$$

$$\hat{y}_{t+k}^t = S_t + kT_t, \qquad k \geqslant 0, \tag{4.6}$$

where S_t and T_t denote the level and trend of $\{y_t\}$ at time t, respectively, and α_1, α_2, γ_1, $\gamma_2 \in (0,1)$ are smoothing constants. The same recursive predictions can be obtained by means of the asymmetric model ARIMA(0,2,2) of the form

$$\begin{aligned} (1-B)^2 y_t = \varepsilon_t &+ (\alpha_1 + \alpha_1\gamma_1 - 2)\varepsilon_{t-1}^- + (1-\alpha_1)\varepsilon_{t-2}^- \\ &+ (\alpha_2 + \alpha_2\gamma_2 - 2)\varepsilon_{t-1}^+ + (1-\alpha_2)\varepsilon_{t-2}^+. \end{aligned} \tag{4.7}$$

The proof is similar to the case of the asymmetric simple exponential smoothing.

Finally, the third example of the exponential smoothing important from the practical point of view is the Holt-Winters model that is suitable for seasonal time series $\{y_t\}$ with a locally constant seasonality of length p. The asymmetric modification of the additive Holt-Winters model can be written as

$$S_t = S_{t-1} + T_{t-1} + \alpha_1 e_t^- + \alpha_2 e_t^+, \tag{4.8}$$

$$T_t = T_{t-1} + \alpha_1\gamma_1 e_t^- + \alpha_2\gamma_2 e_t^+, \tag{4.9}$$

$$I_t = I_{t-p} + \delta_1(1-\alpha_1)e_t^- + \delta_2(1-\alpha_2)e_t^+, \tag{4.10}$$

$$\hat{y}_{t+k}^t = S_t + kT_t + I_{t+k-p}, \qquad k = 1, \dots, p, \tag{4.11}$$

where S_t, T_t and I_t denote the level, trend and seasonal index of $\{y_t\}$ at time t, respectively, and α_1, α_2, γ_1, γ_2, δ_1, $\delta_2 \in (0,1)$ are smoothing constants. The case of the multiplicative Holt-Winters model is similar. The same recursive predictions can be obtained by means of the asymmetric seasonal ARIMA model of the form

$$(4.12)\qquad (1-B)(1-B^p)y_t = \varepsilon_t + \vartheta_{11}\varepsilon_{t-1}^{-} + \ldots + \vartheta_{1,p+1}\varepsilon_{t-p-1}^{-} + \vartheta_{21}\varepsilon_{t-1}^{+} + \ldots + \vartheta_{2,p+1}\varepsilon_{t-p-1}^{+},$$

where $\vartheta_{i1} = -1 + \alpha_i + \alpha_i\gamma_i$, $\vartheta_{ip} = -1 + \alpha_i\gamma_i + \delta_i(1-\alpha_i)$, $\vartheta_{i,p+1} = (1-\alpha_i)(1-\delta_i)$, $\vartheta_{ij} = \alpha_i\gamma_i$ $(i = 1,2;\ j = 2,\ldots,p-1)$.

5. Asymmetric recursive procedures in autoregressive models

Let us deal with a problem of asymmetric recursive estimation in an AR(p) process $\{y_t\}$ which can be written for this purpose in the Kalman filter form

$$(5.1)\qquad x_t = x_{t-1},$$

$$(5.2)\qquad y_t = h_t x_t + v_t,$$

where $h_t = (y_{t-1},\ldots,y_{t-p})$ and $\{v_t\}$ is a white noise.

Moreover, let the innovations $\{v_t\}$ of the process $\{y_t\}$ be distributed asymmetrically, the assumption $v_t \sim \operatorname{iid} N(0;\sigma_1^2,\sigma_2^2)$ being acceptable. Then the result of Section 3 can be used producing the recursive formulas

$$(5.3)\qquad \hat{x}_t^t = \hat{x}_{t-1}^{t-1} + \frac{P_{t-1}^{t-1}h_t'}{h_t P_{t-1}^{t-1}h_t' + \sigma_1^2}e_t^{-} + \frac{P_{t-1}^{t-1}h_t'}{h_t P_{t-1}^{t-1}h_t' + \sigma_2^2}e_t^{+},$$

$$(5.4)\qquad P_t^t = P_{t-1}^{t-1} - \frac{P_{t-1}^{t-1}h_t' h_t P_{t-1}^{t-1}}{h_t P_{t-1}^{t-1}h_t' + \sigma_1\sigma_2}$$

(obviously, $\hat{x}_t^{t-1} = \hat{x}_{t-1}^{t-1}$, $P_t^{t-1} = P_{t-1}^{t-1}$, $e_t = y_t - h_t\hat{x}_{t-1}^{t-1}$). The formula (5.4) is only an approximative one (it is taken from the symmetric procedure (2.3), (2.4)). In practice, the parameters σ_1^2, σ_2^2 can be also estimated recursively using the procedure (2.19) parallelly with (5.3), (5.4).

Various forms of trimming of prediction errors are typical for recursive estimation in autoregressive processes (see e.g. [1], [4], [12]). It makes it possible not only to face outliers in data but it may be important for the proof of convergence of the corresponding recursive formulas. The next theorem is an example of such convergence results for asymmetric trimming.

Theorem. *In the model* AR(1)

$$(5.5)\qquad y_t = \varphi y_{t-1} + v_t, \qquad t = \ldots,-1,0,1,\ldots$$

with

$$v_t \sim \text{iid}\, N(0; \sigma_1^2, \sigma_2^2) \tag{5.6}$$

let an estimate of the parameter φ be given by means of the recursive formulas

$$\hat{x}_t^t = \hat{x}_{t-1}^{t-1} + \frac{P_t^t y_{t-1}}{P_t^t y_{t-1}^2 + \sigma_1\sigma_2}\left[\sigma_1\psi\Big(\frac{e_t^-}{\sigma_1}\Big) + \sigma_2\psi\Big(\frac{e_t^+}{\sigma_2}\Big)\right], \qquad t = 1, 2, \dots, \tag{5.7}$$

$$P_t^t = \frac{P_{t-1}^{t-1}\sigma_1\sigma_2}{P_{t-1}^{t-1} y_{t-1}^2 + \sigma_1\sigma_2}, \qquad t = 1, 2, \dots \tag{5.8}$$

with initial (random) values $\hat{x}_0^0$ and P_0^0, where $e_t = y_t - y_{t-1}\hat{x}_{t-1}^{t-1}$ and

$$\psi(z) = \begin{cases} z, & |z| \leqslant c, \\ c\,\text{sgn}(z), & |z| > c \end{cases} \tag{5.9}$$

($c > 0$ is a constant). Let the following assumptions be fulfilled:

$$|\varphi| < 1, \tag{5.10}$$

$$E(\hat{x}_0^0)^2 < \infty, \quad P_0^0 > 0 \text{ a.s.}, \quad \hat{x}_0^0,\ P_0^0,\ v_t \text{ are independent.} \tag{5.11}$$

Then

$$\hat{x}_t^t \to \varphi \text{ a.s.} \tag{5.12}$$

Proof. See Appendix. □

If we denote

$$\tilde{\psi}(z) = \begin{cases} -c\sigma_1, & z < -c\sigma_1, \\ z, & -c\sigma_1 \leqslant z \leqslant c\sigma_2, \\ c\sigma_2, & z > c\sigma_2 \end{cases} \tag{5.13}$$

then (5.7) can be rewritten to the form

$$\hat{x}_t^t = \hat{x}_{t-1}^{t-1} + \frac{P_t^t y_{t-1}}{P_t^t y_{t-1}^2 + \sigma_1\sigma_2}\tilde{\psi}(e_t) \tag{5.14}$$

so that one can really speak about asymmetric trimming. If $\sigma_1 = \sigma_2 = \sigma$ then (5.7) becomes the typical recursive estimation formula for the autoregressive processes (see e.g. [12]).

Appendix: Proof of Theorem

For simplicity we will omit the upper indices in the subsequent text (e.g. we shall write $\hat{x}_t$ instead of $\hat{x}_t^t$).

Lemma 1. *Let $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \ldots \subset \mathscr{F}$ be a sequence of σ-algebras in a probability space $(\Omega, \mathscr{F}, P)$. Let z_t, β_t, ξ_t, η_t $(t = 0, 1, \ldots)$ be non-negative $\mathscr{F}$-measurable random variables such that*

$$E(z_t \mid \mathscr{F}_{t-1}) \leqslant (1 + \beta_{t-1}) z_{t-1} + \xi_{t-1} - \eta_{t-1}, \qquad t = 1, 2, \ldots, \tag{A.1}$$

$$\sum_{t=0}^{\infty} \beta_t < \infty \text{ a.s.}, \quad \sum_{t=0}^{\infty} \xi_t < \infty \text{ a.s.} \tag{A.2}$$

Then the sequence z_t converges a.s.

Proof. See [11]. □

Lemma 2. *In the model from Theorem let an estimate $\hat{x}_t$ of the parameter φ be given by means of the recursive formulas*

$$\hat{x}_t = \hat{x}_{t-1} + a_{t-1} y_{t-1} \tilde{\psi}(y_t - y_{t-1} \hat{x}_{t-1}), \qquad t = 1, 2, \ldots \tag{A.3}$$

with an initial (random) value $\hat{x}_0$. Here a_t $(t = 0, 1, \ldots)$ are $\mathscr{F}_t$-measurable random variables for $\mathscr{F}_t = \sigma\{\hat{x}_0, v_t, v_{t-1}, \ldots\}$ fulfilling

$$0 \leqslant a_t^{(1)} \leqslant a_t \leqslant a_t^{(2)}, \qquad \sum_{t=0}^{\infty} a_t^{(1)} = \infty, \ \sum_{t=0}^{\infty} (a_t^{(2)})^2 < \infty \tag{A.4}$$

for deterministic sequences $a_t^{(1)}$, $a_t^{(2)}$. Then

$$\hat{x}_t \to \varphi \text{ a.s.} \tag{A.5}$$

Proof. Put $\tilde{x}_t = \hat{x}_t - \varphi$. Then (A.3) can be rewritten as

$$\tilde{x}_t = \tilde{x}_{t-1} + a_{t-1} y_{t-1} \tilde{\psi}(-y_{t-1} \tilde{x}_{t-1} + v_t).$$

Hence one obtains

$$\tilde{x}_t^2 \leqslant \tilde{x}_{t-1}^2 + 2 a_{t-1} y_{t-1} \tilde{x}_{t-1} \tilde{\psi}(-y_{t-1} \tilde{x}_{t-1} + v_t) + [a_{t-1}^{(2)} C y_{t-1}]^2 \tag{A.6}$$

and for conditional expectations

$$E(\tilde{x}_t^2 \mid \mathscr{F}_{t-1}) \leqslant \tilde{x}_{t-1}^2 + 2a_{t-1}y_{t-1}\tilde{x}_{t-1}E\{\tilde{\psi}(-y_{t-1}\tilde{x}_{t-1} + v_t) \mid \mathscr{F}_{t-1}\}$$
$$+ [a_{t-1}^{(2)} C y_{t-1}]^2,$$

where $C = c\max(\sigma_1, \sigma_2)$.

Let us apply Lemma 1 with $z_t = \tilde{x}_t^2$, $\beta_{t-1} = 0$, $\xi_{t-1} = [a_{t-1}^{(2)} C y_{t-1}]^2$, $\eta_{t-1} = -2a_{t-1}y_{t-1}\tilde{x}_{t-1}E\{\tilde{\psi}(-y_{t-1}\tilde{x}_{t-1} + v_t) \mid \mathscr{F}_{t-1}\}$. The only problem may be to verify that $\eta_t \geqslant 0$ a.s.: Let us denote

$$\varrho(b) = E_v\tilde{\psi}(b + v), \qquad -\infty < b < \infty. \tag{A.7}$$

Since $v_t \sim N(0; \sigma_1^2, \sigma_2^2)$ one can easily show that

$$b\varrho(b) > 0, \qquad b \neq 0, \tag{A.8}$$

which guarantees $\eta_t \geqslant 0$.

According to Lemma 1 there exists a (finite) random variable $\tilde{x}$ such that

$$\tilde{x}_t \to \tilde{x} \text{ a.s.} \tag{A.9}$$

For an arbitrary n it follows from (A.6) that

$$\tilde{x}_n^2 \leqslant \tilde{x}_0^2 + 2\sum_{t=1}^{n} a_{t-1}y_{t-1}\tilde{x}_{t-1}\tilde{\psi}(-y_{t-1}\tilde{x}_{t-1} + v_t) + C^2\sum_{t=1}^{n}(a_{t-1}^{(2)}y_{t-1})^2$$

and hence

$$-2\sum_{t=1}^{\infty} E\{a_{t-1}y_{t-1}\tilde{x}_{t-1}\tilde{\psi}(-y_{t-1}\tilde{x}_{t-1} + v_t)\} \leqslant E(\tilde{x}_0^2) + (C\sigma_y)^2\sum_{t=1}^{\infty}(a_{t-1}^{(2)})^2,$$

where $\sigma_y^2 = \mathrm{var}(y_t) = E(y_t^2)$. Therefore according to (A.4) one has

$$-\sum_{t=1} a_{t-1}^{(1)} E\{y_{t-1}\tilde{x}_{t-1}\tilde{\psi}(-y_{t-1}\tilde{x}_{t-1} + v_t)\} < \infty.$$

Since $\sum a_t^{(1)} = \infty$ there exists a subsequence such that

$$\sum_{j=1}^{\infty} E\{-y_{t_j-1}\tilde{x}_{t_j-1}\tilde{\psi}(-y_{t_j-1}\tilde{x}_{t_j-1} + v_{t_j})\} < \infty.$$

Hence

$$-y_{t_j-1}\tilde{x}_{t_j-1}E\{\tilde{\psi}(-y_{t_j-1}\tilde{x}_{t_j-1}+v_{t_j}) \mid \mathscr{F}_{t_j-1}\} \to 0 \text{ a.s.}$$

or equivalently

$$-y_{t_j-1}\tilde{x}_{t_j-1}\varrho(-y_{t_j-1}\tilde{x}_{t_j-1}) \to 0 \text{ a.s.}$$

Due to (A.8) this implies

$$y_{t_j-1}\tilde{x}_{t_j-1} \to 0 \text{ a.s.} \tag{A.10}$$

Further, one can write

$$v_{t_j}\tilde{x}_{t_j-1} = y_{t_j}(\tilde{x}_{t_j-1}-\tilde{x}_{t_j}) + y_{t_j}\tilde{x}_{t_j} - \varphi y_{t_j-1}\tilde{x}_{t_j-1}. \tag{A.11}$$

Since y_{t_j} are identically distributed and the limit relations (A.9) and (A.10) hold all three summands on the right-hand side of (A.11) converge in probability to zero, i.e.

$$v_{t_j}\tilde{x}_{t_j-1} \to 0 \text{ in probability.} \tag{A.12}$$

Due to independence of $\tilde{x}_{t_j-1}$ and v_{t_j}, where v_{t_j} are identically distributed, and due to (A.9) we conclude

$$\tilde{x}_t \to 0 \text{ a.s.}$$

□

Proof of Theorem. We have

$$P_t = \left[P_{t-1}^{-1} + y_{t-1}^2/(\sigma_1\sigma_2)\right]^{-1} = \left[P_0^{-1} + (y_0^2+\ldots+y_{t-1}^2)/(\sigma_1\sigma_2)\right]^{-1}.$$

Hence

$$(2\sigma_1\sigma_2)^{-1} \leqslant (P_t y_{t-1}^2 + \sigma_1\sigma_2)^{-1} \leqslant (\sigma_1\sigma_2)^{-1}$$

and further, due to the properties of the process y_t (see [8, p. 210, Theorem 6]),

$$tP_t \to \sigma_1\sigma_2/\sigma_y^2 \text{ a.s.} \tag{A.13}$$

Let us choose an arbitrary $\varepsilon > 0$ and $0 < \delta < \sigma_1\sigma_2/\sigma_y^2$. By virtue of (A.13) there exists t_0 such that

$$P\Big(\bigcap_{t\geqslant t_0}\left[|tP_t - \sigma_1\sigma_2/\sigma_y^2| < \delta\right]\Big) > 1-\varepsilon.$$

Put

$$\bar{x}_t = \begin{cases} \hat{x}_t, & t = 0,1,\ldots,t_0-1,\\ \bar{x}_{t-1} + \frac{P_t y_{t-1}}{P_t y_{t-1}^2+\sigma_1\sigma_2}\tilde{\psi}(y_t - y_{t-1}\bar{x}_{t-1}), & t\geqslant t_0,\ |tP_t-\sigma_1\sigma_2/\sigma_y^2| < \delta,\\ \bar{x}_{t-1} + \frac{1}{t}\frac{\sigma_1\sigma_2/\sigma_y^2}{P_t y_{t-1}^2+\sigma_1\sigma_2}y_{t-1}\tilde{\psi}(y_t - y_{t-1}\bar{x}_{t-1}), & t\geqslant t_0,\ |tP_t-\sigma_1\sigma_2/\sigma_y^2| \geqslant \delta. \end{cases}$$

Then Lemma 2 with $a_t^{(1)} = \frac{1}{t}(\sigma_1\sigma_2/\sigma_y^2 - \delta)(2\sigma_1\sigma_2)^{-1}$ and $a_t^{(2)} = \frac{1}{t}(\sigma_1\sigma_2/\sigma_y^2 + \delta)(\sigma_1\sigma_2)^{-1}$ yields

$$\bar{x}_t \to \varphi \text{ a.s.}$$

Finally, one can write

$$P(\hat{x}_t \to \varphi) \geqslant P\Big(\Big(\bigcap_{t \geqslant t_0} [\bar{x}_t = \hat{x}_t] \cap [\bar{x}_t \to \varphi]\Big) = P\Big(\bigcap_{t \geqslant t_0} [\bar{x}_t = \hat{x}_t]\Big) \geqslant$$
$$\geqslant P\Big(\bigcap_{t=t_0} \big[|tP_t - \sigma_1\sigma_2/\sigma_y^2| < \delta\big]\Big) > 1 - \varepsilon.$$

Since $\varepsilon > 0$ can be arbitrary we have $\hat{x}_t \to \varphi$ a.s. □

References

[1] *K. Campbell*: Recursive computation of M-estimates for the parameters of a finite autoregressive process. Annals of Statistics *10* (1982), 442–453.

[2] *T. Cipra*: Some problems of exponential smoothing. Aplikace matematiky *34* (1989), 161–169.

[3] *T. Cipra*: Robust exponential smoothing. Journal of Forecasting *11* (1992), 57–69.

[4] *T. Cipra and R. Romera*: Robust Kalman filter and its application in time series analysis. Kybernetika *27* (1991), 481–494.

[5] *T. Cipra and R. Romera*: Recursive time series methods in L_1-norm. L_1-Statistical Analysis and Related Methods (Y. Dodge, ed.). North Holland, Amsterdam, 1992, pp. 233–243.

[6] *T. Cipra, A. Rubio and L. Canal*: Robustified smoothing and forecasting procedures. Czechoslovak Journal of Operations Research *1* (1992), 41–56.

[7] *C. W. J. Granger*: Prediction with a generalized cost of error function. Operational Research Quarterly *20* (1969), 199–207.

[8] *E. J. Hannan*: Multiple Time Series. Wiley, New York, 1970.

[9] *P. Lefrançois*: Allowing for asymmetry in forecast errors: Results from a Monte-Carlo study. International Journal of Forecasting *5* (1989), 99–110.

[10] *W. K. Newey and J. L. Powell*: Asymmetric least squares estimation and testing. Econometrica *55* (1987), 819–847.

[11] *H. Robbins and D. Siegmund*: A convergence theorem for non negative almost supermartingales and some applications. Optimizing Methods in Statistics (J. S. Rustagi, ed.). Academic Press, New York, 1971, pp. 233–257.

[12] *K. Sejling, H. Madsen, J. Holst, U. Holst and J.-E. Englund*: A method for recursive robust estimation of AR-parameters. Preprint. Technical University of Lyngby, Denmark and University of Lund, Sweden, 1990.

[13] *M. J. Silvapulla*: On M-method in growth curve analysis with asymmetric errors. Journal of Statistical Planning and Inference *32* (1992), 303–309.

[14] *W. E. Wecker*: Asymmetric time series. Journal of the American Statistical Association *76* (1981), 16–21.

Author's address: Tomáš Cipra, Matematicko-fyzikální fakulta UK (Faculty of Mathematics and Physics, Charles University), Sokolovská 83, 186 00 Praha 8, Czech Republic.