ADAPTIVE TIME SERIES FILTERS OBTAINED BY MINIMISATION OF THE KULLBACK-LEIBLER DIVERGENCE CRITERION

Elena L'vovna Pervukhina[1], Jean-François Emmenegger[2]
[1]Sevastopol National Technical University, UKRAINE
e-mail: elena@pervuh.sebastopol.ua
[2]Department of Quantitative Economics, University of Fribourg, SWITZERLAND
e-mail: jean-francois.emmenegger@unifr.ch

AMS Subject Classification: 93E11, 37M10, 60G35
Key Words: state space time series models, Kalman filter, Kullback-Leibler divergence information,

Abstract

This study presents an extension of the *Kalman filter* techniques used in *state space* time series filtering and proposes adaptive filters, based on the minimisation of the Kullback-Leibler divergence criterion, measuring the difference between two distributions of random variables. The paper discusses algorithms working under conditions of deficient information concerning the knowledge of the distribution of the error terms of the system described in state space form.

1. Introduction

Linear filtering of state space time series comprises a framework of basic, fundamental and popular methods of system analysis. The largely known approach of R. E. Kalman is based on a modelisation of signals or data by stochastic differential or difference equations. The present theory aims to obtain optimal or approximately optimal filters, using all available *a priori information*[1] on the observations and their stochastic characteristics and also on the analysed

[1]An *a priori information* concerns the knowledge of the probability density function of the distribution of the noise or error term of a model. *A priori information* on a set of parameters of the identified system provides the initial values of the estimating algorithms.

system [2], [3], [7], [21]. But usually *a priori information* is deficient or incorrect.

Modern *adaptive filter* methods, able to *learn* from the data, open new possibilities in filtering theory. The main example treated in this paper is a *linear Gaussian state space model*, see Shilman [21]. In the present paper an *adaptive filter algorithm* for a non trivial filter is proposed to solve a state space time series problem that needs as *a priori information* the theoretical variance of the theoretical state and the running empirical variances, computed on the basis of the history of the observations.

2. State-space representation of a dynamic system

Let $\mathbf{y}_t$ denote an $(n \times 1)$ vector of variables $y_{1t}, y_{2t}, \ldots, y_{nt}$ observed at discrete dates t=1, 2, ...T and a $(r \times 1)$ vector $\mathbf{x}_t$ of generally unobserved variables $x_{1t}, x_{2t}, \ldots, x_{nt}$, known as *state vector*. Then, a *simplified version* of a *time-homogenous state-space representation* of the dynamics of the vector $\mathbf{y}_t$ is defined by a first order Markov process, described by a so-called *state equation*

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{v}_t \tag{1}$$

and secondly by an *observation equation*

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{w}_t, \tag{2}$$

where $\mathbf{A}$ and $\mathbf{H}$ are $(r \times r)$, respectively $(n \times r)$ matrices of parameters, see Harvey ([6], p.101). The $(n \times 1)$ disturbance vector $\mathbf{v}_t$ and the $(r \times 1)$ $\mathbf{w}_t$ disturbance vector are themselves serially uncorrelated white noise vectors with $E(\mathbf{w}_t) = E(\mathbf{v}_t) = \mathbf{0}$ and $Var(\mathbf{w}_t) = \mathbf{Q}_t$, $Var(\mathbf{v}_t) = \mathbf{R}_t$. Furthermore, the disturbance vectors $\mathbf{v}_t$ and $\mathbf{w}_t$ are mutually uncorrelated: $E(\mathbf{v}_t\mathbf{w}'_\tau) = \mathbf{0}$ for all t and τ, see Kalman, ([8], p. 100). It is assumed that the initial state vector $\mathbf{x}_1$ is uncorrelated with any realisations of $\mathbf{v}_t$ or $\mathbf{w}_t$. This implies that $\mathbf{v}_t$ is uncorrelated with lagged values of $\mathbf{x}$ or $\mathbf{y}$: $E(\mathbf{v}_t\mathbf{x}'_\tau) = E(\mathbf{v}_t\mathbf{y}'_\tau) = E(\mathbf{w}_t\mathbf{y}'_\tau) = \mathbf{0}$, $\tau = t-1, t-2, ..., 1$ and similarly $E(\mathbf{w}_t\mathbf{x}'_\tau) = 0$, $\tau = 1, ..., T$, see Hamilton ([5], p. 372). Note that a state space representation of a dynamic system is not unique. The concept of state-space representation is now illustrated by some examples.

1. Example: The one dimensional *local level* model, see Harvey ([6], p. 102) and Durbin-Koopman ([12], p. 9). The settings are: $r = n = 1$, with $\mathbf{x}_t = [x_t]$, $\mathbf{y}_t = [y_t]$, $\mathbf{A} = [1]$, $\mathbf{w}_t = [\nu_t]$, then the *observation equation* is

$$y_t = x_t + \nu_t \quad ; \quad Var(\nu_t) = \sigma_\nu^2 \tag{3}$$

and with $\mathbf{v}_t = [\eta_t]$, $\mathbf{H} = [1]$, the *state equation*

$$x_t = x_{t-1} + \eta_t \quad ; \quad Var(\eta_t) = \sigma_\eta^2 \tag{4}$$

that is a *time-homogenous* state space model with the state x_t and with stationary first differences Δy_t of the observations.

2. Example: *The univariate $ARMA(p,q)$ process*, $n = 1$, see Hamilton ([5], p. 375), using the backward shift operator B, $By_t = y_{t-1}$. The autoregression polynomial Φ of order p with coefficients $\phi_1 \dots \phi_p$,

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p, \tag{5}$$

and the moving average polynomial Θ of order q with coefficients $\theta_0, \ldots, \theta_q$

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q. \tag{6}$$

are defined. An univariate $ARMA(p,q)$ process

$$\Phi(B) y_t = \Theta(B) \varepsilon_t \tag{7}$$

has more generally a state-space representation form by defining the number of rows and columns $r = max\{p, q+1\}$ of matrix $\mathbf{A}$, where one defines $\phi_j = 0$ for $j = p+1, \dots r$ and $\theta_i = 0$ for $i = q+1, \dots r$. In this case the *state vector* is $\mathbf{x}_t = [x_t, x_{t-1}, \dots x_{t-r+1}]'$. Then the $(r \times 1)$ error-vector $\boldsymbol{\varepsilon}_t = [\varepsilon_t, 0, \ldots, 0]'$, the $(r \times r)$ transition matrix $\mathbf{A}$

$$\mathbf{A} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \ldots & \phi_{r-1} & \phi_r \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \ldots & & & \ldots & & \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \tag{8}$$

and the $(n \times r) - matrix$ $\mathbf{H} = \boldsymbol{\Theta} = [1, \theta_1, \theta_2, \ldots, \theta_{r-1}]$ are defined. A state-space representation is defined with the *state equation* (1)

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{9}$$

and, for a single observation, $n = 1$, the *observation equation* (2) of the single component vector $\mathbf{y}_t = [y_t]$ becomes, written without error term,

$$y_t = \boldsymbol{\Theta}\mathbf{x}_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_{r-1} B^{r-1}) x_t. \tag{10}$$

The first row of equation (8) yields

$$x_t - \phi_1 x_{t-1} - \ldots - \phi_r x_{t-r} = (1 - \phi_1 B - \ldots - \phi_r B^r) x_t = \varepsilon_t. \tag{11}$$

Then, equation (10) is multiplied by the polynomial Φ, equation (5), resulting in the univariate $ARMA(p,q)$ model of the form (7)[2]

$$(1 - \phi_1 B - \ldots - \phi_r B^r) y_t = (1 + \theta_1 B + \ldots + \theta_{r-1} B^{r-1}) \varepsilon_t. \tag{12}$$

[2]The necessary requirement for stationarity of the model (7) is that the polynomial Φ in (5) of degree p must have all its roots outside the unit circle within the Gaussian plane.

Once a model is put in a state space form, a number of important algorithms are applied. At the center is the Kalman filter algorithm that is a recursive procedure for computing the optimal estimator of the *state vector* at time t, based on the information available at time t. This information consists of the observations up to and including $\mathbf{y}_t$. All the system matrices together with the initial expectation vector $E(\mathbf{x_0})$ and covariance matrix of the estimation error $\mathbf{P_0} = Var(\mathbf{x_0})$ are assumed to be known in all time periods, see Harvey ([6], p. 101).

3. The Kalman filter

The derivation of the Kalman filter rests on the assumption that the disturbances and *initial state vector* $\mathbf{x}_0$ are normally distributed. It enables the likelihood function to be calculated via what is known as the prediction error decomposition[3].

Kalman used the covariance functions. But it is also possible to describe the dynamics of the series $\mathbf{y}_t$, $\mathbf{x}_t$ by linear stochastic equations. Rewriting a "process in such a form has the aim to obtain a convenient summary of the system's dynamics, and this is the reason to be interested in state-space representation of any system", see Hamilton ([5], p. 374).

Consider the *state equation* (1) and the *observation equation* (2). Consider the linear orthogonal projection of the state vector $\mathbf{x}_t$ on the subspace $\mathcal{Y}_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ..., \mathbf{y}_1\}$, generated by the observations previous to $\mathbf{y}_t$, denoted by $\hat{\mathbf{x}}_{t|t-1} := \hat{E}(\mathbf{x}_t|\mathcal{Y}_{t-1})$[4], see Hamilton ([5], p. 377). The state estimator $\hat{\mathbf{x}}_{t|t-1}$ is "defined as the best (minimum variance, unbiased) estimate of the state $\mathbf{x}_t$, given the set of observations $\mathcal{Y}_{t-1}$", see Ohap and Stubberud ([13], p. 589). Analogously, the estimate $\hat{\mathbf{x}}_t := \hat{E}(\mathbf{x}_t|\mathcal{Y}_t)$ is the best estimate of $\mathbf{x}_t$ given $\mathcal{Y}_t$. Let $\mathbf{P}_{t-1}$ denote the $(r \times r)$ estimation error covariance matrix of $\mathbf{x}_t$,

$$\mathbf{P}_{t-1} = E[(\mathbf{x}_{t-1} - \hat{\mathbf{x}}_{t-1})(\mathbf{x}_{t-1} - \hat{\mathbf{x}}_{t-1})'] = Var(\mathbf{x}_{t-1}). \tag{13}$$

Given the optimal estimators $\hat{\mathbf{x}}_{t-1}$ and $\mathbf{P}_{t-1}$, their estimates are

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{A}\hat{\mathbf{x}}_{t-1}, \tag{14}$$

[3]A standard result on the multivariate normal distribution is then used to calculate recursively the distribution of $\mathbf{x}_t$, conditional on the information set at time t, for all $t = 1, \ldots, T$. These conditional distributions are themselves normal and hence are completely specified by their means and covariance matrices. The Kalman filter algorithm computes these quantities. It is known that the mean of the conditional distribution of the state vector $\mathbf{x}_t$ is an optimal estimator of $\mathbf{x}_t$ in the sense that is minimises the mean square errors (MSE), see Harvey ([6], p. 105).

[4]Suppose we were to find a matrix $\boldsymbol{\alpha}$ such that the forecast error $(\mathbf{x}_t - \boldsymbol{\alpha}\mathcal{Y}_{t-1})$ is uncorrelated with $\mathcal{Y}_{t-1}$, $E[(\mathbf{x}_t - \boldsymbol{\alpha}\mathcal{Y}_{t-1})\mathcal{Y}_{t-1}] = \mathbf{0}$. For this case, the *linear projection* notation $\hat{P}(\mathbf{x}_t|\mathcal{Y}) = \boldsymbol{\alpha}\mathcal{Y}_{t-1}$ is used. The operator $\hat{E}$ defines the linear projection $\hat{P}$ of $\mathbf{x}_t$ on the set $\mathcal{Y}_{t-1}$, along with a constant term, $\hat{E}(\mathbf{x}_t|\mathcal{Y}_{t-1}) = \hat{P}(\mathbf{x}_t|1, \mathcal{Y}_{t-1})$, see Hamilton ([5], p. 74).

while the associated approximate covariance matrix is

$$\mathbf{P}_{t|t-1} = E[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})'] = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}' + \mathbf{R}_t. \tag{15}$$

The equations (14), (15) are called *prediction equations*. When the new observation $\mathbf{y}_t$ is available, the estimator $\hat{\mathbf{x}}_t$ of the state vector $\mathbf{x}_t$ is computed on the basis of the estimator $\hat{\mathbf{x}}_{t|t-1}$ with the subsequent *updating equations*

$$\begin{aligned} \mathbf{z}_t &= \mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1} \\ \hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}'\mathbf{F}^{-1}\mathbf{z}_t, \end{aligned} \tag{16}$$

where $\mathbf{z}_t$ is a *white noise process*, see Ohap ([13], p. 589), called innovations and represent the *one-step forecast error* of $\mathbf{y}_t$, see Durbin and Koopman ([12], p. 66). The computation of the estimate $\hat{\mathbf{x}}_t$ (16) is based on a *lemma in multivariate regression theory*, see Anderson ([1], Theorem 2.5.1). The covariance matrix of the innovations

$$\mathbf{F}_t = Var(\mathbf{z}_t) = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}' + \mathbf{Q}_t \tag{17}$$

and the estimation error covariance matrix $\mathbf{P}_t$ are presented[5],

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}'\mathbf{F}_t^{-1}\mathbf{H}\mathbf{P}_{t|t-1}. \tag{18}$$

Taken together (16) and (18) make up the *Kalman filter*. The recursions can be computed directly from $\hat{\mathbf{x}}_{t|t-1}$ to $\hat{\mathbf{x}}_{t+1|t}$ with (14) and (16):

$$\hat{\mathbf{x}}_{t+1|t} = (\mathbf{A} - \mathbf{K}_t\mathbf{H})\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\mathbf{y}_t = \mathbf{A}\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\mathbf{z}_t, \tag{19}$$

where the gain matrix $\mathbf{K}_t$ is given by

$$\mathbf{K}_t = \mathbf{A}\mathbf{P}_{t|t-1}\mathbf{H}'\mathbf{F}_t^{-1} \quad ; \quad t = 1, \ldots T, \tag{20}$$

and the recursion for the error covariance matrix is known as the *Riccati equation*, resulting from (15) and (18)

$$\mathbf{P}_{t+1|t} = \mathbf{A}(\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}'\mathbf{F}_t^{-1}\mathbf{H}\mathbf{P}_{t|t-1})\mathbf{A}' + \mathbf{R}_t. \tag{21}$$

The procedure of filtering has the aim to obtain an estimated signal $\hat{\mathbf{x}}_t$ at each moment t, corresponding to each element of the set of observations $(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots)$.

3. Example: Consider again *the local level model* with observations (3) and scalar states (4). Then the estimation of the scalar state x_t is given by the "best (minimum variance, unbiased) estimate" $\hat{x}_{t|t-1}$ of the state x_t and the

[5]In this paper it is assumed that the inverse of $\mathbf{F}_t$ exists.

associated approximate variance $P_{t|t-1} := E[(x_t - \hat{x}_{t|t-1})^2]$. The innovations z_t and the variance of innovations f_t are computed from equations (16) and (17).

$$\begin{aligned} z_t &= y_t - \hat{x}_{t|t-1} \\ f_t &= P_{t|t-1} + \sigma_\nu^2 \end{aligned} \tag{22}$$

Then, the Kalman filter k_t and the recursion for the best estimate $\hat{x}_{t+1|t}$ can be computed from equations (20), (19), getting

$$\begin{aligned} k_t &= P_{t|t-1} f_t^{-1} \\ \hat{x}_{t+1|t} &= \hat{x}_{t|t-1} + k_t z_t. \end{aligned} \tag{23}$$

Finally, one computes the Riccati equation (21)

$$P_{t+1|t} = P_{t|t-1}(1 - P_{t|t-1} f_t^{-1}) + \sigma_\eta^2. \tag{24}$$

Subsequently, a computer simulation of $N = 2000$ values of each series x_t and y_t is realised, where the variances $\sigma_\eta^2 = 16$ and $\sigma_\nu^2 = 4$ have been set[6]. It is known that "non-trivial *local level* models have a steady state solution" ([12], p. 33). The variance of this steady state solution is computed by setting $P = P_{t+1|t} = P_{t|t-1}$ into equation (24)[7]. Then, empirical variances are computed from the simulated series to get estimations of $Var(\eta_t) = Var(x_t - x_{t-1})$ and $Var(\nu_t)$ at each time t and the innovation variance (22), the Kalman filter (23) and the associated approximate variance (24) are computed from the simulated series, confirming the theoretical results, see Fig. 1.

4. Minimisation of the Kullback-Leibler information divergence as criterion of estimation

The realisation of the Kalman filter under complete information implies that the signal values $\hat{\mathbf{x}}_t$ are calculated at each moment t. In real applications, there are different deviations from the above described text book descriptions.

Adaptive filtering. The Kalman filter is just one of many *adaptive filtering* or estimation algorithms. There are many alternative of the Kalman filter. In the present case it is assumed that the covariance functions are partly or completely unknown.

Deficient *a priori* **information**. When the description of the signals by the system equations (1), (2), the matrices $\mathbf{A}, \mathbf{H}, \mathbf{Q}_t, \mathbf{R}_t$ or the initial conditions $E(\mathbf{x}_0)$, $Var(\mathbf{x}_0)$ considered to be partially or completely unknown, one is

[6]The computer simulation has been realised with a computer program written by the authors in RATS, *Regression Analysis of Time Series*, T. A. Doan, Estima, Evanston IL 60201 USA.

[7]Through simple algebra the steady state associated approximate variance $P = 19.314$ of the estimated state and consequently also the steady state Kalman filter $k = 0.8284$ and the steady state innovation variance $f = 23.314$ are obtained.

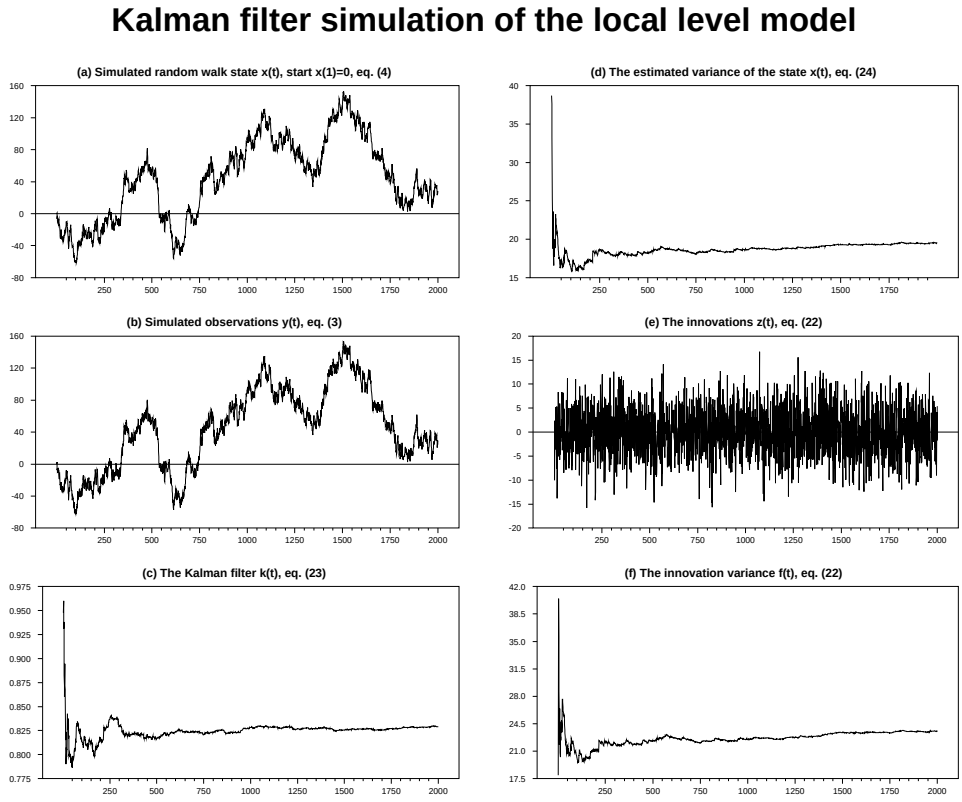

Figure 1: Simulation of the local level model, computing the running *empirical variances* of the *state* (3) and the *observations* (4)

in presence of the particular situation of deficient *a priori* information. In this case, only the observations $\mathbf{y}_t$ and realisations connected with the state vector $\mathbf{x}_t$ are available. These empirical data can be utilised in two directions. With indirect methods the data are used to restore the description of observations and the optimal filter is synthesised by standard methods. With direct methods, the filters are formed on the basis of observations without restoration of the description. This approach is easier from the computational point of view. Moreover, it often offers the possibility to solve problems in real time regime, see Shilman [21].

As it is known, see Harvey ([6], p. 105), under the conditions of deficient *a priori* information on noises and signals, it is not possible to obtain optimal estimations of the state vector of the system (1), (2). The available filtering algorithms are suboptimal from the computational point of view and do not quickly converge. Moreover, when the covariance matrices of noise are unknown, the analytical solution of the problem is at present not known.

For the present development of adaptive filters, it is assumed that the

covariance matrices $\mathbf{Q}_t$, $\mathbf{R}_t$ of the noise $\mathbf{v}_t$, $\mathbf{w}_t$ of the state and observation equations (1), (2) are unknown.

Kullback-Leibler divergence information. It is assumed that the true probability density of the random vector $\mathbf{y}_t$, $t = 1, ..., T$, here written without time index, $\mathbf{y}$, is $g(\mathbf{y})$ and the sample probability density of $\mathbf{y}$ is $f(\mathbf{y})$. The idea is to introduce the *Kullback-Leibler divergence criterion*, defined for the probability densities g and f as, see Kullback-Leibler ([11], p. 6),

$$J_{g,f}(\mathbf{y}) = J(g, f) = \int [g(\mathbf{y}) - f(\mathbf{y})] \ln \frac{g(\mathbf{y})}{f(\mathbf{y})} d\mathbf{y}. \tag{25}$$

In other words, the information divergence (25) reflects the difference between the unknown true density of the random variable $\mathbf{y}$ and its estimation. The Kullback-Leibler information divergence has following properties: (i) $J(g, f) > 0$ if $g(\mathbf{y}) \neq f(\mathbf{y})$, (ii) $J(g, f) = 0$ if $g(\mathbf{y}) = f(\mathbf{y})$, (iii) $J(g, f) = J(f, g)$. Then, it is additive for independent observations $\mathbf{y}_1, \ldots, \mathbf{y}_m$,

$$(iv) \qquad J_{g,f}(\mathbf{y}_1, \ldots, \mathbf{y}_m) = \sum_{k=1}^{n} J_{g,f}(\mathbf{y}_k). \tag{26}$$

Furthermore, an additional measurement never reduces the $Kullback-Leibler$ information divergence[8]

$$(v) \qquad J_{g;f}(\mathbf{y}_1, \ldots, \mathbf{y}_m) \leq J_{g;f}(\mathbf{y}_1, \ldots, \mathbf{y}_m, \mathbf{y}_{m+1}). \tag{27}$$

For the parametric case, there is a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ of parameters, specifying a hypothetical distribution h. One sets $g(\mathbf{y}) = h(\mathbf{y}, \boldsymbol{\theta})$ and $f(\mathbf{y}) = h(\mathbf{y}, \boldsymbol{\theta} + \Delta\boldsymbol{\theta})$ for neighbouring points $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + \Delta\boldsymbol{\theta}$ in a parameter space. The information divergence is represented in this case by the elements $g_{\alpha\beta} = \int f(\mathbf{y}, \theta) \frac{1}{f(\mathbf{y},\theta)} \frac{\partial f(\mathbf{y},\theta)}{\partial \theta_\alpha} \frac{1}{f(\mathbf{y},\theta)} \frac{\partial f(\mathbf{y},\theta)}{\partial \theta_\beta} d\mathbf{y}$ of *Fisher's information matrix* $\mathbf{G} = \{g_{\alpha\beta}\}$, see Kullback and Leibler ([11], p. 28), [9],

$$J(\theta, \theta + \Delta\theta) \approx \sum_{\alpha=1}^{k} \sum_{\beta=1}^{k} g_{\alpha\beta} \Delta\theta_\alpha \Delta\theta_\beta. \tag{28}$$

Estimation. For the construction of the *estimation algorithm*, the *Kullback-Leibler* information divergence $J(g, f)$ (25) is considered as a matrix functional of matrices $\mathbf{K}_t \in \mathcal{M}_{n\times m}$. The necessary conditions to determine the argument that provide the minimum to the matrix functional, written in form of matrix differentiation [7], [17] with respect to the gain matrix $\mathbf{K_t}$ are calculated,

$$\frac{\partial J(g, f)}{\partial \mathbf{K}} = \mathbf{0}. \tag{29}$$

[8]The mentioned properties of the Kullback-Leibler information divergence explain its application to discover, recognise and construct signals.

The second order necessary conditions for existence of minima have also to be verified, but are not explicitly mentioned here. The adaptive filter $\mathbf{K_t}$ is the solution of an equation

$$\mathbf{K_t} = \arg\min_{\mathcal{M}_{n\times m}}\{J(g,f)\}, \tag{30}$$

see Pervukhina [15], [17], [20].

Estimations are stochastic values. For this reason, one considers an estimation as optimal, when the distribution parameters are closest to the distribution parameters of the true *state vector*. A measure of the mean degree of correspondence of the probability distribution to the true distribution will be the criterion of quality for every step of estimation.

Then, the estimation, optimal from the point of view of the *Kullback-Leibler* information criterion, based on the observations till time t will give the expression [17]

$$\hat{\mathbf{x}}_t = \xi(\mathbf{K}_t, \hat{\mathbf{x}}_{t|t-1}, \mathbf{y}_t). \tag{31}$$

Here ξ is an estimating function, applying the space $R^n \times R^m$ into R^n, whereas $\mathbf{K}_t$ are $n \times m$ matrices of gain coefficients of the space $\mathcal{M}_{n\times n}$ that are the solution of the equation (30).

5. Adaptive filtering for multivariate Gaussian populations

Here a direct analysis is made by the method of *adaptive filtering* over current observations $\mathbf{y}_t$. Generally no other empirical data are used.

The case is considered, where the $(r\times 1)$ *state vector* $\mathbf{x}_t$ has a *r-dimensional normal distribution* with $E(\mathbf{x}_t) = \mathbf{q}$ and a $(r \times r)$ covariance matrix $\mathbf{\Sigma}_t = E[(\mathbf{x}_t - \mathbf{q})(\mathbf{x}_t - \mathbf{q})']$. The estimations $\hat{\mathbf{x}}_t$ are found in a class of unbiased estimations which are calculated by equation (31). Furthermore, the estimations of the state vector $\hat{\mathbf{x}}_t$ are linear combinations of observations under the normality assumption. Thereby, $\hat{\mathbf{x}}_t$ are also normally distributed with mathematical expectation $E(\hat{\mathbf{x}}_t) = \hat{\mathbf{q}}$ and covariance matrix $\hat{\mathbf{\Sigma}} = E[(\hat{\mathbf{x}}_t - \hat{\mathbf{q}})(\hat{\mathbf{x}}_t - \hat{\mathbf{q}})']$. The estimation requires to calculate a matrix $\mathbf{K}_t$ that is provided by the minimum of the *Kullback-Leibler* divergence information criterion.

When the vectors $\mathbf{x}_t$ and $\hat{\mathbf{x}}_t$ are normally distributed, its probability densities are

$$g(\mathbf{x}) = \frac{1}{|2\pi\mathbf{\Sigma}|^{1/2}}\exp(-\frac{1}{2}(\mathbf{x}-\mathbf{q})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{q})), \tag{32}$$

and

$$f(\mathbf{x}) = \frac{1}{|2\pi\hat{\mathbf{\Sigma}}|^{1/2}}\exp(-\frac{1}{2}(\hat{\mathbf{x}}-\hat{\mathbf{q}})'\hat{\mathbf{\Sigma}}^{-1}(\hat{\mathbf{x}}-\hat{\mathbf{q}})). \tag{33}$$

With (32), (33) the function $ln\frac{g(\mathbf{x})}{f(\mathbf{x})}$ in (25) is computed, see Hawks [9]. Using the trace operator tr for square matrices, the integral (25) is computed and

gives

$$J(g,f) = \frac{1}{2}tr[(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}})(\hat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1})] + \frac{1}{2}tr[(\boldsymbol{\Sigma}^{-1} + \hat{\boldsymbol{\Sigma}}^{-1})(\mathbf{q} - \hat{\mathbf{q}})(\mathbf{q} - \hat{\mathbf{q}})']. \quad (34)$$

Under the assumption of equal population means, $\mathbf{q}$=$\hat{\mathbf{q}}$, equation (34) is written, where n is the dimension of $\mathbf{x}_t$,

$$J(g,f) = \frac{1}{2}tr[(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}})(\hat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1})] = \frac{1}{2}tr[\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}] - n. \quad (35)$$

On the other hand, under the assumption of equal population covariances, $\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}$, equation (34) is written,

$$J(g,f) = tr[\boldsymbol{\Sigma}^{-1}(\mathbf{q} - \hat{\mathbf{q}})(\mathbf{q} - \hat{\mathbf{q}})'] = (\mathbf{q} - \hat{\mathbf{q}})'\boldsymbol{\Sigma}^{-1}(\mathbf{q} - \hat{\mathbf{q}}), \quad (36)$$

giving the generalised distance of Mahalanobis, see Kullback-Leibler ([11], p. 190).

6. A Kullback-Leibler adaptive filter algorithm

The proposed algorithm (37)-(39) computes an *adaptive filter* as an alternative to the well known *Kalman filter* algorithm (14) - (21). The following version of this alternative so called *Kullback-Leibler adaptive filter algorithm* has been described, see Pervukhina [15]-[19] and is a result of the solution of the optimisation problem (29). As it will be seen in this paper there is a system when the subsequent algorithm does not result in all the cases in non trivial filters[9].

First, the estimate $\hat{\mathbf{x}}_{t|t-1}$ (14) and the innovations $\mathbf{z}_t$ (16) are identical

$$\begin{aligned} \hat{\mathbf{x}}_{t|t-1} &= \mathbf{A}\hat{\mathbf{x}}_{t-1} \\ \mathbf{z}_t &= \mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}. \end{aligned} \quad (37)$$

Second, the covariance matrix of the observations $\mathbf{T}_t$, the covariance matrix $\boldsymbol{\Sigma}_t$ of the state vector $\mathbf{x}_t$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}_{t-1}$ of its estimations $\hat{\mathbf{x}}_{t-1}$ are defined to compute the proposed *adaptive filter*,

$$\mathbf{K}_t = \mathbf{A}\hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{A}'\mathbf{H}'[\mathbf{H}\mathbf{A}\hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{A}'\mathbf{H}' + \mathbf{T}_t]^{-1}, \quad (38)$$

as an alternative of the Kalman filter (20). Then the covariance matrix of the estimations $\hat{\mathbf{x}}_{t-1}$ of the state is updated through equation

$$\hat{\boldsymbol{\Sigma}}_t = [\mathbf{I} - \mathbf{K}_t\mathbf{H}]\mathbf{A}\hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{A}'[\mathbf{I} - \mathbf{K}_t\mathbf{H}]' + \mathbf{K}_t\mathbf{T}_t\mathbf{K}_t'. \quad (39)$$

Finally, a recursion equation, analogue to (19), is set up to compute the estimated state vectors

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{A}\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\mathbf{z}_t. \quad (40)$$

The main advantage of this algorithm resides in the possibility to estimate the system state vector in real time regime, when the covariance matrices $\mathbf{Q}_t$ and $\mathbf{R}_t$ are unknown, see the remark above.

[9] A trivial filter is a filter that is a null-matrix, $\mathbf{K}_t = \mathbf{0}$, for all t.

7. An example of a Linear Gaussian State Space Model

The Kalman filter algorithm (14) - (21) and the Kullback-Leibler adaptive filter algorithm (37)-(40) are applied to a *one-dimensional linear Gaussian state space model* (1), (2) that has been treated by Shilman [21].

4. Example: The settings of the presented *linear Gaussian state space model* are; $n = r = 1$, $\mathbf{x}_t = [x_t]$, $\mathbf{y}_t = [y_t]$, $\mathbf{H} = [1]$, $\mathbf{w}_t = [w_t]$ $\mathbf{A} = [a]$ and $\mathbf{v}_t = [bv_t]$. The *observations* are described by

$$y_t = x_t + w_t \quad ; \quad w_t \sim WN(0, \sigma_w^2), \tag{41}$$

where w_t is a Gaussian white noise with $\sigma_w = 1$. The *scalar state*, defined as a stationary $AR(1)$ process x_t is

$$x_t = ax_{t-1} + bv_t \quad ; \quad v_t \sim WN(0, \sigma_v^2), \tag{42}$$

where v_t is independent from w_t, $E[v_t w_t] = 0$, with $\sigma_v = 1$, parameters $a = e^{-1}$, $b = \sqrt{15(1 - e^{-2})}$ and the covariance function $\gamma_m = 15e^{-|m|}$. As (42) is stationary, its variance is time independent, calculated as $\sigma^2 = Var(x_t) = \frac{b^2}{1-a^2} = 15$. This implies $T = Var(y_t) = 16$ for the variance of the observations.

a) **The Kalman filter**. The Kalman filter algorithm (13)-(21) is applied to model (41), (42). The innovation z_t, the innovation variance f_t, the Kalman filter k_t, the estimated state $\hat{x}_{t+1|t}$ and the associated approximate variance $P_{t+1|t}$ are set up,

$$\begin{aligned} z_t &= y_t - \hat{x}_{t|t-1} \\ f_t &= P_{t|t-1} + \sigma_w^2 \\ k_t &= aP_{t|t-1}f_t^{-1} \\ \hat{x}_{t+1|t} &= a\hat{x}_{t|t-1} + k_t z_t \\ P_{t+1|t} &= a^2 P_{t|t-1}(1 - P_{t|t-1}f_t^{-1}) + b^2\sigma_v^2. \end{aligned} \tag{43}$$

Then, the steady state solutions ([12], p. 33) are investigated, setting for the Kalman filter $k = k_t$, for the innovation variance $f = f_t$, for the associated approximate variance $P = P_{t+1|t} = P_{t|t-1}$ and solving the system (43)[10].

b) **Simulation of the Kalman filter**. A computer simulation of $N = 2000$ values of the $AR(1)$ model x_t (42) and the series y_t (41) is realised with a RATS-program. The theoretical variances $\sigma_v^2 = 1$ and $\sigma_w^2 = 1$ are replaced by estimations from the simulated series[11]. The theoretical results under a) are confirmed by this simulation, see Fig. 2.

[10] Through simple algebra the steady state associated approximate variance $P = 13.0957$ of the estimated state, the steady state Kalman filter $k = 0.34178$ and the steady state innovation variance $f = 14.0957$ are obtained.

[11] Then empirical variances are computed at each time $t; t = 2, ..., N$ for both simulated series $\{w_\tau\}$ and $\{x_\tau - ax_{\tau-1}\}$, $\tau = 1, ..., t$. These estimations replace the theoretical variances in equations of system (43) for the simulation.

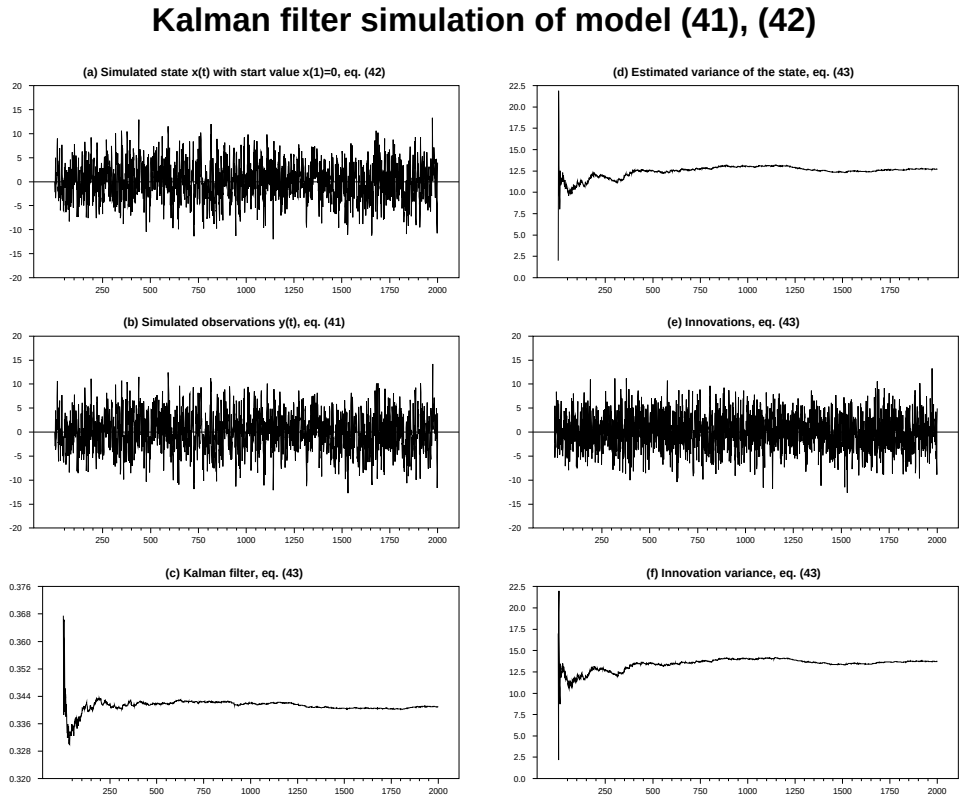

Figure 2: Computation of the Kalman filter following Harvey [6], Ohap and Stubberud [13] and Durbin and Koopman [12], see equation (43)

c) **The Kullback-Leibler adaptive filter**. The innovation $z_t = y_t - \hat{x}_t$ is defined. Then, the optimal estimator $\hat{x}_{t+1}$ is set up in application of the updating equation (40),

$$\hat{x}_{t+1} = a\hat{x}_t + k_t z_t = a\hat{x}_t + k_t[y_t - \hat{x}_t] = a(1-k_t)\hat{x}_t + k_t y_t. \tag{44}$$

As the innovations represent a *white noise* process, one is in presence of a stationary $AR(1)$-process (44), as is the process (42). Then, the expectations $\mu = E[\hat{x}_{t+1}] = E[x_t] = 0$ vanish and therefore the estimated variance becomes $\hat{\sigma}^2_{t+1} = E[\hat{x}^2_{t+1}]$. It is developed as follows

$$\begin{aligned} \hat{\sigma}^2_{t+1} &= E[(a(1-k_t)\hat{x}_t + k_t y_t)^2] \\ &= a^2(1-k_t)^2\hat{\sigma}^2_t + k_t^2 E[y_t^2] + 2a(1-k_t)k_t E[\hat{x}_t y_t]. \end{aligned} \tag{45}$$

The last term in (45) becomes with the conditions mentioned on page 3,

$$E[\hat{x}_t w_t] = E[\hat{x}_t v_t] = E[x_{t-1} w_{t-1}] = 0.$$

$$\begin{aligned}
E[\hat{x}_t y_t] &= E[\hat{x}_t (x_t + w_t)] = E[\hat{x}_t x_t] + \underbrace{E[\hat{x}_t w_t]}_{=0} \\
&= E[\hat{x}_t (a x_{t-1} + b v_t)] = a E[\hat{x}_t x_{t-1}] + b \underbrace{E[\hat{x}_t v_t]}_{=0} \\
&= a E[(a(1-k_{t-1})\hat{x}_{t-1} + k_{t-1} y_{t-1}) x_{t-1}] \\
&= a^2 \underbrace{E(\hat{x}_{t-1} x_{t-1}]}_{=0} (1-k_{t-1}) + a k_{t-1} E[y_{t-1} x_{t-1}] \\
&= a k_{t-1} E[(x_{t-1} + w_{t-1}) x_{t-1}] = a k_{t-1} E[x_{t-1}^2] = a k_{t-1} \sigma_{t-1}^2.
\end{aligned} \tag{46}$$

The result of (46) is now introduced in equation (45) with $T_t = E[y_t^2]$, giving

$$\hat{\sigma}_{t+1}^2 = a^2 (1-k_t)^2 \hat{\sigma}_t^2 + T_t k_t^2 + 2\sigma_{t-1}^2 a^2 (1-k_t) k_t k_{t-1}. \tag{47}$$

It is known that the norm of the third term in (47) is small and can be skipped, see Pervukhina [15]-[19], also in the multi-dimensional case. This leads to a second degree equation for the adaptive filter k_t[12],

$$\hat{\sigma}_{t+1}^2 = a^2 (1-k_t)^2 \hat{\sigma}_t^2 + T_t k_t^2. \tag{48}$$

d) **Steady state solutions**. The steady state variance, $\hat{\sigma}^2 = \hat{\sigma}_t^2 = \hat{\sigma}_{t+1}^2$, the steady state adaptive filter $k = k_t = k_{t-1}$ and the steady state observation variance $T = T_t$ are introduced in (47) and (48). In this case, by skipping the third term of equation (47) that has nearly no influence on the searched filter of this steady state investigation, a second degree function g and a third degree function f

$$\begin{aligned}
f(k) &= a^2 (1-k)^2 \hat{\sigma}^2 + T k^2 + 2\sigma^2 a^2 (1-k) k^2 - \hat{\sigma}^2 \\
g(k) &= a^2 (1-k)^2 \hat{\sigma}^2 + T k^2 - \hat{\sigma}^2.
\end{aligned} \tag{49}$$

are deduced. Both functions f and g are presented in Fig 3. The zeros of function f are $\{-0646904, 0.963736, 5.12400\}$, the zeros of function g are $\{-0.742996, 0.968179\}$. Clearly, there is a *zero* nearly around $\{0.965\}$ common to f and g. This common zero is advocated as the value of the steady state adaptive filter of (48) and (47). In the subsequent development only the second degree equation (48) is considered, much simpler to treat.

e) **Minimisation of the Kullback-Leibler information divergence**. The Kullback-Leibler criterion is formulated for this linear Gaussian state space model. The condition is that there is a theoretical state x_t and an

[12] All the subsequent calculations have also been realised by the authors with the corresponding third degree equation (47).

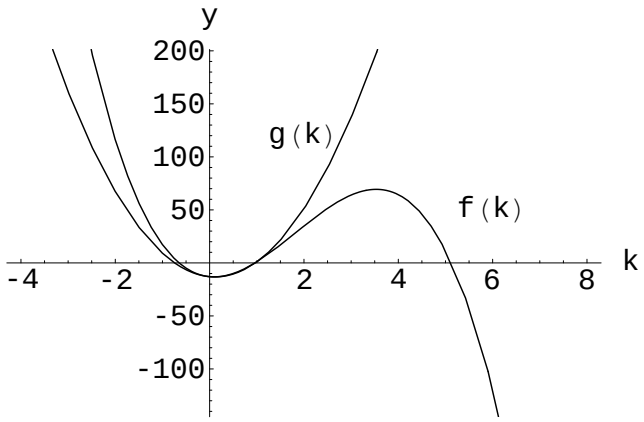

Figure 3: Functions g and f of the adaptive filter k

expected state $\hat{x}_t$ with identical expectations $\mu = E[x_t] = E[\hat{x}_t] = 0$ and variance σ^2 of x_t, respectively, variance $\hat{\sigma}^2$ of the expected state $\hat{x}_t$ being generally different, $\hat{\sigma}^2 \neq \sigma^2$. Both single-variate distributions are supposed to be normal. For the purpose of this analysis it is supposed that σ is *known* and $\hat{\sigma}$ is estimated from the data. The divergence criterion (35) becomes in this case

$$J(\sigma^2, \hat{\sigma}^2) = \frac{1}{2}(\frac{\sigma^2}{\hat{\sigma}^2} + \frac{\hat{\sigma}^2}{\sigma^2}) - 1 = \frac{(\sigma^2 - \hat{\sigma}^2)^2}{2\sigma^2\hat{\sigma}^2} \tag{50}$$

and measures the divergence between both normal distributions. Clearly, for $\sigma = \hat{\sigma}$ this *divergence information criterion* becomes zero, $J(\sigma^2, \sigma^2) = 0$. The estimation criterion (29) is applied to minimise the present *divergence information criterion* $J(\sigma^2, \hat{\sigma}^2)$ with respect to the steady state adaptive filter k, the derivation being set to zero,

$$\frac{\partial J(\sigma^2, \hat{\sigma}^2)}{\partial k} = \frac{\partial J(\sigma^2, \hat{\sigma}^2)}{\partial \hat{\sigma}^2} \cdot \frac{\partial \hat{\sigma}^2}{\partial k} = \frac{1}{2}(\frac{-\sigma^2}{\hat{\sigma}_2^4} + \frac{1}{\sigma^2})\frac{\partial \hat{\sigma}^2}{\partial k} = 0. \tag{51}$$

Consequently, one of both factors in (51) is zero. This means either, (α) $\frac{\partial \hat{\sigma}^2}{\partial k} = 0$, the derivation of $\hat{\sigma}^2$ is zero, or, (β) $(\frac{-\sigma^2}{\hat{\sigma}_2^4} + \frac{1}{\sigma^2}) = 0$, the parenthesis is zero, leading to the equality $\hat{\sigma} = \sigma$.

(α) The operational criterion is $\frac{\partial \hat{\sigma}^2}{\partial k} = 0$ together with the equation (48) for the variance $\hat{\sigma}^2$ gives the algorithm for the steady state

$$\begin{aligned} \hat{\sigma}^2 &= a^2(1-k)^2\hat{\sigma}^2 + Tk^2 \\ k &= \frac{a^2\hat{\sigma}^2}{a^2\hat{\sigma}^2 + T}. \end{aligned} \tag{52}$$

This is an interesting system, because the variance σ^2 of the state x_t has disappeared and the system (52) only depends on the steady state variance

T of the observations y_t. Unfortunately, this one-dimensional case only has as solutions either the trivial filter $k = 0$ or the negative filter $k = \frac{a^2-1}{a^2} < 0$ coupled with a meaningless negative estimated variance $\hat{\sigma}^2 = T\frac{a^2-1}{a^2} < 0$ of the estimated state $\hat{x}_t$.

(β) The operational criterion is the equality $\hat{\sigma} = \sigma$, meaning that the variance of the estimated state $\hat{x}_t$ is optimal when it is forced to be the variance of the theoretical state x_t.

f) **The Kullback-Leibler adaptive filter algorithm** is now calculated for the model (41), (42), initially for time dependent parameters. The quadratic equation (48) is *expanded* into powers of k_t, giving the quadratic equation

$$0 = (a^2\hat{\sigma}_t^2 + T_t)k_t^2 - 2a^2\hat{\sigma}_t^2 k_t + \hat{\sigma}_t^2 a^2 - \hat{\sigma}_{t+1}^2. \tag{53}$$

Its single positive solution, approximately a solution of (47), is chosen as the subsequently aimed adaptive filter k_t. The *Kullback-Leibler adaptive filter algorithm* is set up, based on the coefficient a, the values T_t, y_t, $\hat{x}_t$, comprising the three coefficients of equation (53) in the variable k_t, the adaptive filter k_t, the innovation z_t and the estimated state $\hat{x}_{t+1}$. The single remaining operational criterion (β) is applied, $\sigma = \hat{\sigma} = \hat{\sigma}_t = \hat{\sigma}_{t+1} = 15$. At the start, the initial value $\hat{x}_1$ of the state must be chosen, here $\hat{x}_1 = 0$, the computation is performed in a loop for $t = 2, ..., N$ and all the observations y_t must be available at each step. The algorithm comprises the equations

$$\begin{aligned} &a_t = a^2\sigma^2 + T_t;\ \ b = -2a^2\sigma^2;\ \ c = \sigma^2(a^2-1);\ \ z_t = y_t - \hat{x}_t;\ \ \overline{\hat{x}}_t = \frac{1}{t}\sum_{\tau=1}^{t}\hat{x}_\tau \\ &k_t = \frac{-b + \sqrt{b^2 - 4a_t c}}{2a_t};\ \ \hat{x}_{t+1} = a\hat{x}_t + k_t z_t;\ \ s_{\hat{x}_t}^2 = \frac{1}{t-1}\sum_{\tau=1}^{t}(\hat{x}_\tau - \overline{\hat{x}}_t)^2. \end{aligned} \tag{54}$$

It is easily seen that neither estimations of the variances σ_w^2 nor σ_w^2 are necessary for the computation of the algorithm (54). Moreover the variances T_t for y_t have to be computed at each step t, similarly to those of $\hat{x}_t$.

g) **The steady state solution**. The steady state adaptive filter $k = k_t$, see Durbin and Koopman ([12], p. 33) is now investigated. The additional setting is: $T = T_t = 16$. Then the steady state corresponding to the coefficients a_t, b and c of the system (54) are computed and the steady state solution $k = 0.968$ for the adaptive filter is found.

h) **Simulation of the Kullback-Leibler adaptive filter**. A computer simulation of the adaptive filter algorithm (54) is realised with a RATS program[13]. The theoretical results under g) are confirmed by this simulation, see

[13]First, $N = 2000$ observations $y_t; t = 1, ..., N$ are simulated. Second, the algorithm (54) is applied. The variances T_t of the *simulated* observations y_t are computed at each time t on the basis of the history of the values $y_\tau; \tau = 1, ..., t$, $t = 2, ..., N$.

Fig. 4. The six slides (a)-(f) are self explaining and following convergences are observed, $T_t \to 16, s^2_{\hat{x}_t} \to 15, k_t \to 0.968$.

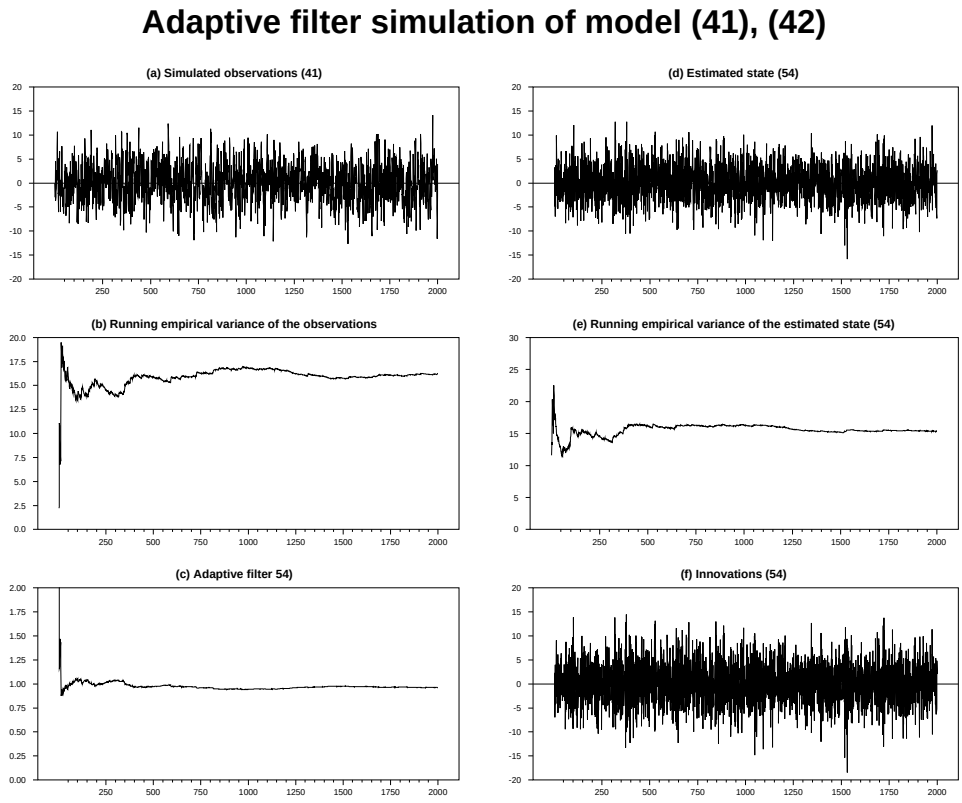

Figure 4: Simulation of the Kullback-Leibler adaptive filter (54)

8. Conclusion

The *Kalman filter algorithm* has been applied to a one-dimensional non stationary *local level model* and to a one-dimensional stationary *linear Gaussian model*. In the sense of an extension, an alternative principle, the minimisation of the *Kullback-Leibler information criterion* is proposed, in order to obtain algorithms to compute so called *Kullback-Leibler adaptive filters*. This procedure leads to different operational criterions, depending on the chosen problem and the corresponding Kullback-Leibler criterion. In this paper, essentially, a one-dimensional stationary *linear Gaussian model* has been treated. The derivation of the Kullback-Leibler information criterion gives two operational criterions: (a) Either, the requirement is that the estimated variance of the estimated state is minimal with respect to the searched filter, and this leads in this one-dimensional case to an algorithm that only has the trivial

adaptive filter, $k = 0$. Non-trivial filters are described for this criterion in higher-dimensional systems, see Pervukhina [15]- [20], not treated in this paper. (b) Or, the requirement is that the estimated variance of the estimated state is equal to the *a priori known* variance of the *theoretical* state. This second case, where a steady state adaptive filter $k = 0.968$ is identified, is of special interest, because it is a case of realisation of an *adaptive filter*, working under condition of *deficient a priori information*. The *variances* of the *state noise* and of the *observation noise* need not to be known or computed, contrary to the case of the Kalman filter.

The present results are illustrated by simulations of the *Kalman filter* and the *Kullback-Leibler adaptive filter*, applied to the discussed simulated non-stationary *local level series* and to the simulated stationary *linear Gaussian time series*.

Acknowledgements

The authors express their deep gratitude to the Department of Quantitative Economy of the University of Fribourg for the encouragement and support of this research.

References

[1] T. W. Anderson, *An Introducton to Multivariate Statistical Analysis*, 2nd. Edition. New York, John Wiley & Sons, (1984).

[2] G. E. P. Box, and G. M. Jenkins and G. C. Reinsel, *Time Series Analysis, forecasting and control*, Prentice Hall, Englewood Cliffs, New Jersey 07632 (1994).

[3] T. A. Doan, *RATS Software Package, User's Manual, Version 4.0.* Estima, Evanston, Illinois (1992).

[4] J. A. Doornik, and D. F. Hendry, *Empirical Econometric Using PcGive for Windows*, International Thomson Business Press, London (1996).

[5] J. D. Hamilton, *Time Series Analysis.* Princeton University Press, Princeton, New Jersey (1994).

[6] A. C. Harvey, *Forecasting, structural time series models and Kalman filter*, Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sidney, (1996).

[7] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York, (1997).

[8] R. E. Kalman, *A new approach to linear filtering and prediction problems* Journal of Basic Engineering, **82**, 35-45, (1960).

[9] R. Hawks, J. T. Moor, *The scopes of quality indexes at the adaptive estimation.* IEEE Trans. Autom. Contr, Vol. 64, No: 8, (1976), pp. 28-36.

[10] B. Kovacevic, Z. Durovic, S. Glavaski, *On robust Kalman filtering*, International Journal of Control, Vol.56 - No: 3., (1992), pp. 547-562.

[11] S. Kullback, *Information Theory and Statistics*, Dover Publications, Inc., New York, (1967).

[12] J. Durbin, S. J. Koopman, *Time Series Analysis by State Space methods*, Oxford Statistical Science Series, Nr. 24, Oxford University Press, (2001).

[13] R. F. Ohap and A. R. Stubberud, *Adaptive Minimum Variance Estimation in Discrete-Time Linear Systems* in "Control and Dynamic Systems", edited by C. T. Leondes, Academic Press, New York, vol. 12, (1976).

[14] J. R. Magnus, H. Neudecker *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, New York, (1988).

[15] E. Pervukhina *Identification of technical system on trials for adaptation to specific functioning model*, Zeitschrift fuer Angewandte Mathematik und Mechanik, Vol. 76, suppl.3, (1996) pp. 533-534.

[16] E. Pervukhina *Method for Identification of Stochastic Dynamic Objects*, Abstracts of 11. International Colloquium of Numerical Analysis and Computer Science with Applications. Plovdiv, Bulgaria, August 12-17, (2002), p. 51.

[17] E. Pervukhina *Minimum Informative Divergence Estimation in Time Discrete Linear Systems with Applications to Adaptive Filtering* Abstracts of International Open Workshop " Computational Mechanics and Mathematics", Moscow, Russia, October 16-20, (1995), p. 23.

[18] E. Pervukhina *The Forming of Structure for Matrix Function Derivatives with Respect to Matrix Argument*, Bulletin of Sevastopol State Technical University "Modeling and Experiment in Engineering" , Vol.1, (1995), pp. 22-25.

[19] E. Pervukhina, J.-F. Emmenegger *Time Series Models and the Kalman Filter for Analysis of Irregularly Observed data,* Proceedings in Computational Statistics, 14^{th} Symposium at Utrecht, Netherlands, August (2000), p. 79-80.

[20] E. Pervukhina *Use of Information Measure in Procedures for Estimation of Discrete Stochastic Systems at Unknown Noises Covariance*, Transactions of Russian Academy of Natural Sciences "Mathematics. Mathematical Modeling. Informatics and Control, Vol.3, No: 3, (1999), pp. 100-106.

[21] S. Shilman *Adaptive Filtering of Time Series.* Nizhny Novgorod, (1995).

[22] T. Tou, R. Gonzalez, *Pattern Recognation Principles*, Addison-Wesley Publishing Company, London, 1974.