Propagation of initial errors on the parameters for linear and Gaussian state space models

Salima El Kolei

Abstract

For linear and Gaussian state space models parametrized by $\theta_0 \in \Theta \subset \mathbb{R}^r, r \geq 1$ corresponding to the vector of parameters of the model, the Kalman filter gives exactly the solution for the optimal filtering under weak assumptions. This result supposes that θ_0 is perfectly known. In most real applications, this assumption is not realistic since θ_0 is unknown and has to be estimated. In this paper, we analysis the Kalman filter for a biased estimator θ of θ_0. We show the propagation of this bias on the estimation of the hidden state. We give an expression of this propagation for linear and Gaussian state space models and we extend this result for almost linear models estimated by the Extended Kalman filter. An illustration is given for the autoregressive process with measurement noises widely studied in econometrics to model economic and financial data.

Keywords: Kalman filter, Extended Kalman filter, State space models, Autoregressive process

1   Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P}_{\theta_0})$ be a probability space parametrized by $\theta_0 \in \Theta \subset \mathbb{R}^r, r \geq 1$ corresponding to the vector of parameters of the model. We define two real vectors $\{x_t, t \in \mathbb{N}\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P}_{\theta_0})$ with value in $\mathcal{X}$ and $\{y_t, t \in \mathbb{N}^*\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P}_{\theta_0})$ with value in $\mathcal{Y}$. The process $\{x_t, t \in \mathbb{N}\}$ (respectively $\{y_t, t \in \mathbb{N}^*\}$) is called the unobserved signal process (resp. the observation process).

The Kalman filter (KF) and the Extended Kalman filter (EKF) commonly used in some engineering applications have been successfully employed in various areas. These filters may be easily understood by reading the first publication of Kalman in 1960 [Kal60] or the Bayesian interpretation of Harrison and Stevens in 1971 [HS71].

---

[1]Affiliation: S. El Kolei, Laboratoire de Mathématiques J.A. Dieudonné
UMR n 7351 CNRS UNSA
Université de Nice - Sophia Antipolis
06108 Nice Cedex 02 France
Tel.: +33-04-92-07-62-56
salima@unice.fr

1.1 The (Extended) Kalman filter: motivation

Let $y_1, \cdots, y_T$ be the data (which may be either a scalar or a vector) at time $1, \cdots, T$. We assume that y_t depends on the unobservable variable x_t. The aim of the (Extended) Kalman filter is to make inference about the hidden state x_t (which may also be a scalar or a vector) conditionally to the data $y_1, \cdots, y_t$. The relationship between the observed variable y_t and the hidden state x_t is linear and described by a function h depending on the unknown vector of parameters θ_0. This relation is specified by the following observation equation:

$$y_t = h(\theta_0, x_t) + \sigma_{\theta_0}^{\varepsilon} \varepsilon_t, \qquad t \geq 1$$

where ε_t is the vector of noises assumed to be normally distributed with mean zero and unit variance, denoted as: $\varepsilon_t \sim \mathcal{N}(0, I_{n_y \times n_y})$ where n_y is the dimension of the observation space $\mathcal{Y}$.

The hidden state x_t is assumed to be varying with time and its dynamic feature is given by the following state equation:

$$x_t = b(\theta_0, x_{t-1}) + \sigma_{\theta_0}^{\eta} \eta_t$$

where b is a known function and η_t is the state error assumed to be normally distributed with mean zero and unit variance, i.e $\eta_t \sim \mathcal{N}(0, I_{n_x \times n_x})$ where n_x is the dimension of the state space $\mathcal{X}$.

In addition to the usual Kalman filter assumptions (see [Kal60]), we also assume that the noises ε_t and η_t are independent.

Hence, this paper is concerned with the following discrete time state space model with additives noises:

$$\begin{cases} y_t = h(\theta_0, x_t) + \sigma_{\theta_0}^{\varepsilon} \varepsilon_t, & t \geq 1 \\ x_t = b(\theta_0, x_{t-1}) + \sigma_{\theta_0}^{\eta} \eta_t \end{cases} \qquad (1)$$

Under the usual Kalman assumptions, the model (1) can be rewritten as follows:

$$\begin{cases} y_t = d_t(\theta_0) + C_{\theta_0} x_t + \sigma_{\theta_0}^{\varepsilon} \varepsilon_t & t \geq 1, \\ x_t = u_t(\theta_0) + A_{\theta_0} x_{t-1} + \sigma_{\theta_0}^{\eta} \eta_t, \end{cases} \qquad (2)$$

If the vector of parameters θ_0 is perfectly known, the optimal filtering $p_{\theta_0}(x_t|y_{1:t})$ is Gaussian and the Kalman filter gives exactly the two first conditional moments: $\hat{x}_t = \mathbb{E}_{\theta_0}[x_t|y_{1:t}]$ and $P_t = \mathbb{E}_{\theta_0}[(x_t - \hat{x}_t)(x_t - \hat{x}_t)'|y_{1:t}]$ where $'$ stands for the transpose. In particular, the Kalman filter estimator is the *BLUE* (Best Linear and Unbiased Estimator) among linear estimators. Nevertheless, in most applications the linearity assumption of the functions h and b is not always satisifed. A linearization by a one order Taylor series expansion can be performed and the Extended Kalman filter consists in applying the Kalman filter on this linearized model.

For the EKF, the matrix C_{θ_0} is the differential of the function h with respect to (w.r.t.) x computed at the point $(\theta_0, \hat{x}_t^-)$ where $\hat{x}_t^-$ corresponds to the conditional expectation $\mathbb{E}_{\theta_0}[x_t|y_{1:t-1}]$. Additionally, the matrix A_{θ_0} is the differential of the function b w.r.t. x computed at the point $(\theta_0, \hat{x}_{t-1})$. Furthermore, the functions $u_t(\theta_0)$ and $d_t(\theta_0)$ are defined as:

$$\begin{cases} u_t(\theta_0) = b(\theta_0, \hat{x}_{t-1}) - A_{\theta_0}\hat{x}_{t-1} \\ d_t(\theta_0) = h(\theta_0, \hat{x}_t^-) - C_{\theta_0}\hat{x}_t^- \end{cases}$$

In this paper, we assume that the vector of parameters θ_0 is not perfectly known such that the inference of the hidden state x_t conditionnally to $y_{1:t}$ is made with errors of specification. This typical case is frequent in practice since in general the vector of parameters is unknown and need to be estimated by an ordinary method. The resulting estimator can be biased and consequently this bias is propagated on the estimation of the hidden state. More precisely, if we denote by $\hat{\theta}$ a biased estimator of θ_0 such that $\mathbb{E}_{\theta_0}[\hat{\theta}] = \theta = \theta_0 + \epsilon$ where ϵ is a fixed and unknown error corresponding to the bias, we want to evaluate the propagation of the error a posteriori and of the residues a posteriori given by:

$$e_t = x_t - \mathbb{E}_\theta[x_t|y_{1:t}] \tag{3}$$

$$\xi_t = y_t - \mathbb{E}_\theta[y_t|y_{1:t}]. \tag{4}$$

Many papers concerned the propagation of the initial error on the state $(x_0 - \hat{x}_0)$ through the filter, and, to the best of our knowledge, there don't exist in the literature, an analysis of the propagation of the initial errors on the vector of parameters. In this paper, we derive an expression of these propagations for the Kalman and the Extended Kalman filters. Our main result shows that a correlation between the error a posteriori e_t and the unobserved state x_t appeared at each time t of the filter. The Kalman filter is now a biased estimator and a new Lyapunov dynamic equation for the variance matrix P_t is induced.

Applications of this result include epidemiology, meteorology, neuroscience, ecology (see [IBAK11]) and finance (see [JPS09]). For example, our result can be applied to the five ecological state space models described in [PHH10]. Although the scope of our method is general, we have chosen to focus on the so-called autoregressive process AR(1) with measurement noise which has been widely studied and on which our main result can be easily applied and understood. A full illustration of this result is given for a more complex model as the Heston model which is very used in finance for pricing options and hedging portfolios (see [ElK12]).

The paper is organized as follows. Section 2 presents the model assumptions and states all of the theoretical results. The application is given and discussed in Section 3. Some concluding remarks are provided in the last section. The proofs are gathered in Appendix 4.

2 Main result

2.1 General setting and assumptions

In this section, we introduce some preliminary main notations and provide the assumptions of model (2).

2.1.1 Notations

Subsequently, we denote by E_t the pair of the unobservable states vectors given by $\begin{pmatrix} e_t \\ x_t \end{pmatrix}$ and by $\mathfrak{E}_t$ the pair of the observations vectors $\begin{pmatrix} \xi_t \\ y_t \end{pmatrix}$ where e_t and ξ_t are defined in (3) respectively. Their variances matrix are denoted by Σ_t^x and Σ_t^y respectively.

Regarding the partial derivatives, for any function h, $[\partial h/\partial \theta]$ is the vector of the partial derivatives of h w.r.t θ.

Finally, R_{θ_0} denotes $\sigma_{\theta_0}^\varepsilon \sigma_{\theta_0}^{'\varepsilon}$ and Q_{θ_0} denotes $\sigma_{\theta_0}^\eta \sigma_{\theta_0}^{'\eta}$ and are the covariances matrix of ε_t and of η_t respectively.

2.1.2 Assumptions

We consider the state space models (2), the following assumption ensures some smoothness for the functions h and b.

(A1) The functions b and h are differentiable with respect to θ_0 and x.

2.1.3 Main result

Before running into the main theorem of this paper, let us explain some existing results. It is well known that if the vector of parameters is exactly known, the error a posteriori e_t is given by the following formula:

$$e_t = (I_{n_x \times n_x} - K_t C_{\theta_0})A_{\theta_0}e_{t-1} - K_t(\sigma_{\theta_0}^\varepsilon \varepsilon_t + C_{\theta_0}\sigma_{\theta_0}^{\eta_0}\eta_t) + \sigma_{\theta_0}^\eta \eta_t$$

where K_t is called the Kalman matrix that minimizes the variance matrix P_t. Under some assumptions on the model (2), a CLT is obtained for e_t as t tends to infinity (see [dNCdL94]). The following Theorem gives the propagation of the error a posteriori e_t and of ξ_t for the Kalman filter and the Extended Kalman filter when θ_0 is not exactly known. In this respect, we further assume that assumption **(A1)** holds true and the usual Kalman assumptions are satisfied.

Theorem 2.1. *Consider the model (2). If $\epsilon << 1$, then:*

$$\begin{aligned}
e_t &= (I_{n_x \times n_x} - K_t C_\theta)A_\theta e_{t-1} - K_t(\sigma_\theta^\varepsilon \varepsilon_t + C_\theta \sigma_\theta^\eta \eta_t) + \sigma_\theta^\eta \eta_t \\
&\quad + \mathcal{E}_x^\epsilon(\theta,t) + \mathcal{F}_x^\epsilon(\theta,t)x_{t-1} + \mathcal{W}_x^\epsilon(\theta,t) + o(\epsilon)
\end{aligned} \tag{5}$$

with:

$$\mathcal{E}_x^\epsilon(\theta, t) = -\epsilon\left((I_{n_x \times n_x} - K_t C_\theta)\frac{\partial u_t}{\partial \theta}(\theta) - K_t\frac{\partial d_t}{\partial \theta}(\theta) - K_t\frac{\partial C_\theta}{\partial \theta}u_t(\theta)\right) \quad (6)$$

$$\mathcal{F}_x^\epsilon(\theta, t) = -\epsilon\left((I_{n_x \times n_x} - K_t C_\theta)\frac{\partial A_\theta}{\partial \theta} - K_t\frac{\partial C_\theta}{\partial \theta}A_\theta\right) \quad (7)$$

$$\mathcal{W}_x^\epsilon(\theta, t) = -\epsilon\left(\frac{\partial \sigma_\theta^\eta}{\partial \theta}\eta_t - K_t C_\theta\frac{\partial \sigma_\theta^\eta}{\partial \theta}\eta_t - K_t\sigma_\theta^\eta\frac{\partial C_\theta}{\partial \theta}\eta_t - K_t\frac{\partial \sigma_\theta^\varepsilon}{\partial \theta}\varepsilon_t\right) \quad (8)$$

Additionally, the propagation of ξ_t is equal to:

$$\xi_t = C_\theta e_t + \sigma_\theta^\varepsilon\varepsilon_t + \mathcal{E}_y^\epsilon(\theta, t) + \mathcal{F}_y^\epsilon(\theta, t)x_t + \mathcal{W}_y^\epsilon(\theta, t) + o(\epsilon) \quad (9)$$

with:

$$\mathcal{E}_y^\epsilon(\theta, t) = -\epsilon\frac{\partial d_t}{\partial \theta}(\theta), \quad \mathcal{F}_y^\epsilon(\theta, t) = -\epsilon\frac{\partial C_\theta}{\partial \theta}, \quad \mathcal{W}_y^\epsilon(\theta, t) = -\epsilon\frac{\partial \sigma_\theta^\varepsilon}{\partial \theta}\varepsilon_t \quad (10)$$

Moreover, when the linearity assumption of the funtions b and h is not satisfied, the formulas above remain true with the notations of the EKF defined in Section 1.

Proof. See Appendix (A). □

We note that the terms depending on ϵ: $\mathcal{E}_x^\epsilon(\theta, t)$, $\mathcal{F}_x^\epsilon(\theta, t)$ and $\mathcal{W}_x^\epsilon(\theta, t)$ (resp. $\mathcal{E}_y^\epsilon(\theta, t)$, $\mathcal{F}_y^\epsilon(\theta, t)$ and $\mathcal{W}_y^\epsilon(\theta, t)$) are the corrective terms arising from the bias of the parameters estimates.

Besides, we can see in Eq.(5) that at time t, the propagation of the state error e_t depends on e_{t-1} but also on the true state variable x_{t-1}. Therefore, the variance of the error e_t depends on the variance of e_{t-1} but also on the covariance between e_{t-1} and x_{t-1}.

Theorem 2.1 gives an expression of the error a posteriori e_t and of the residues ξ_t which can be rewritten as follows:

$$
\begin{aligned}
e_t &= x_t - \mathbb{E}_\theta[x_t|y_{1:t}] \\
&= \underbrace{(x_t - \mathbb{E}_{\theta_0}[x_t|y_{1:t}])}_{\text{(1) error of estimation}} + \underbrace{(\mathbb{E}_{\theta_0}[x_t|y_{1:t}] - \mathbb{E}_\theta[x_t|y_{1:t}])}_{\text{(2) correctives terms arising from the bias of parameters.}}
\end{aligned}
$$

Additionally,

$$
\begin{aligned}
\xi_t &= x_t - \mathbb{E}_\theta[\xi_t|y_{1:t}] \\
&= \underbrace{(\xi_t - \mathbb{E}_{\theta_0}[\xi_t|y_{1:t}])}_{\text{(1) true residues}} + \underbrace{(\mathbb{E}_{\theta_0}[\xi_t|y_{1:t}] - \mathbb{E}_\theta[\xi_t|y_{1:t}])}_{\text{(2) correctives terms arising from the bias of parameters.}}
\end{aligned}
$$

The expression of the correctives terms (2) are given in Corollary 1.

Corollary 1. *Let e_t given in Eq.(5), the mean of the error a posteriori is given by:*

$$
\begin{aligned}
\mathbb{E}_{\theta_0}[e_t|y_{1:t}] &= \mathbb{E}_{\theta_0}[x_t|y_{1:t}] - \mathbb{E}_{\theta}[x_t|y_{1:t}] \\
&= \left((I_{n_x \times n_x} - K_t C_\theta)A_\theta + \mathcal{F}_x^\epsilon(\theta,t)\right)\mathbb{E}_{\theta_0}[e_{t-1}|y_{1:t-1}] \\
&\quad + \mathcal{E}_x^\epsilon(\theta,t) + \mathcal{F}_x^\epsilon(\theta,t)\mathbb{E}_\theta[x_{t-1}|y_{1:t-1}] + o(\epsilon)
\end{aligned}
\tag{11}
$$

where $\mathcal{E}_x^\epsilon(\theta,t)$ and $\mathcal{F}_x^\epsilon(\theta,t)$ are given in Eq.(6) and Eq.(7).

Besides, let ξ_t given in Eq.(9), the mean of ξ_t is given by:

$$
\begin{aligned}
\mathbb{E}_{\theta_0}[\xi_t|y_{1:t}] &= \mathbb{E}_{\theta_0}[\xi_t|y_{1:t}] - \mathbb{E}_{\theta}[\xi_t|y_{1:t}] \\
&= \left(C_\theta + \mathcal{F}_y^\epsilon(\theta,t)\right)\mathbb{E}_{\theta_0}[e_t|y_{1:t}] + \mathcal{E}_y^\epsilon(\theta,t) + \mathcal{F}_y^\epsilon(\theta,t)\mathbb{E}_\theta[x_t|y_{1:t}] + o(\epsilon)
\end{aligned}
\tag{12}
$$

where $\mathcal{E}_y^\epsilon(\theta,t)$ and $\mathcal{F}_y^\epsilon(\theta,t)$ are given in Eq.(10).

Corollary 1 which is just a consequence of Theorem (2.1) gives a computable recursive expression of the expected error $\mathbb{E}_{\theta_0}[e_t|y_{1:t}]$. Given $\mathbb{E}_{\theta_0}[e_0]$ one can deduce all the values of this expectation for all $t = 1, \cdots, T$.

3 Illustration on the linear Gaussian AR(1) model:

3.1 The model

Let us consider the linear AR(1) model with measurement noise given by:

$$
\begin{cases}
y_t = x_t + \sigma_\varepsilon \varepsilon_t, & t = 1, \cdots, T \\
x_{t+1} = \phi_0 x_t + \sigma_\eta \eta_{t+1}.
\end{cases}
\tag{13}
$$

Since this model is linear and Gaussian we can apply Eq.(11) in Corollary 1 to recover the expectation of e_t when the state x_t is estimated with a biased vector of parameters. For this straighforward example, θ_0 is equal to ϕ_0. For the simulation, we take $\phi_0 = 0.7$, $\sigma_\eta^2 = 0.3$ and $\sigma_\varepsilon^2 = 0.5$.

3.2 Numerical result:

We run a Kalman filter by assuming that the parameter estimate ϕ is biased and we take $\phi = 0.85$, that is $\epsilon = 0.15$. For this model, the functions b and h are given by:

$$
b(\theta_0, x) = \phi_0 x \text{ and } h(\theta_0, x) = x
$$

The variable A_{θ_0} is equal to ϕ_0 and C_{θ_0} is equal to one. The control variables $u_t(\theta_0)$ and $d_t(\theta_0)$ are equal to zero.

Furthermore, the functions $\mathcal{E}_x^\epsilon(\theta,t), \mathcal{F}_x^\epsilon(\theta,t)$ are easily computable and given in the following lemma.

Lemma 1. *For the linear AR(1) model, the functions $\mathcal{E}_x^\epsilon(\theta, t)$ and $\mathcal{F}_x^\epsilon(\theta, t)$ are equal to:*

$$\mathcal{E}_x^\epsilon(\theta, t) = 0, \ \ and \ \mathcal{F}_x^\epsilon(\theta, t) = -\epsilon(1 - K_t)$$

Therefore, by using Eq.(11) of Corollary 1, the expectation $\mathbb{E}_{\theta_0}[e_t|y_{1:t}]$ is given by

$$\mathbb{E}_{\theta_0}[e_t|y_{1:t}] = \big((1 - K_t)(\phi - \epsilon)\big)\mathbb{E}_{\theta_0}[e_{t-1}|y_{1:t-1}] - \epsilon(1 - K_t)\mathbb{E}_\theta[x_{t-1}|y_{1:t-1}] + o(\epsilon) \tag{14}$$
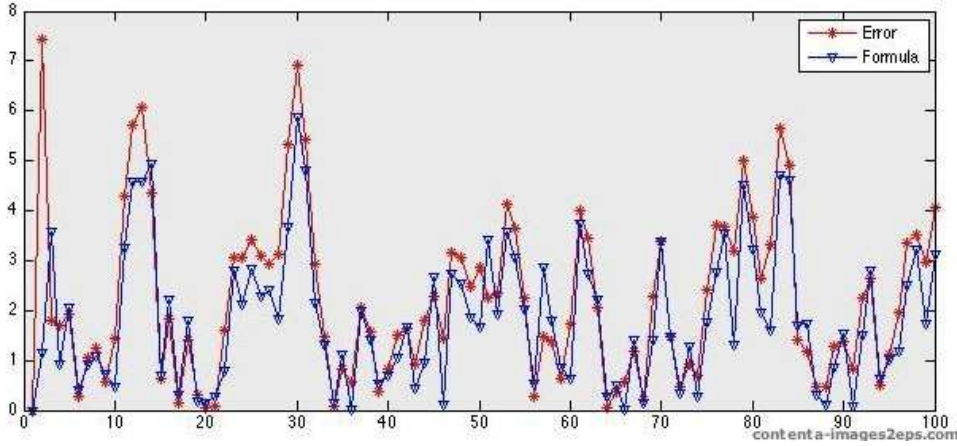

Figure 1: Red: True Error $\mathbb{E}_{\theta_0}[e_t|y_{1:t}] \times 100$. Blue: Approximation (14)$\times 100$.

This example shows that the approximation (14) explains the true error $\mathbb{E}_{\theta_0}[e_t|y_{1:t}]$ for an easy model. The term $\mathcal{F}_x^\epsilon(\theta, t)$ corresponds to the bias of x_t induced by the bias of the parameter estimate. Furthermore, we can see that the error between the true expectation $\mathbb{E}_{\theta_0}[e_t|y_{1:t}]$ and the approximation (14) corresponds to $o(\epsilon)$. A full application is given in [ElK12].

The following Theorem regards the expression of the variances matrix Σ_t^x and Σ_t^y of E_t and $\mathfrak{E}_t$ respectively.

Theorem 3.1. *The variance matrix Σ_t^x is given by:*

$$\Sigma_t^x = \begin{pmatrix} V_t^x & S_t^{'x} \\ S_t^x & P_t^x \end{pmatrix}$$

where:

$$V_t^x = (I - K_tC_\theta)A_\theta V_{t-1}^x A_\theta^{'}(I - K_tC_\theta)^{'} + \mathcal{F}_x^\epsilon(\theta)S_{t-1}(I - K_tC_\theta)A_\theta^{'}$$
$$+ (I - K_tC_\theta)A_\theta S_{t-1}^{'x}\mathcal{F}_x^{'\epsilon}(\theta) + \mathcal{F}_x^\epsilon(\theta)P_{t-1}^x\mathcal{F}_x^{'\epsilon}(\theta) + \mathbb{V}_\theta[\tilde{\mathcal{W}}_x^\epsilon(\theta)]$$
$$S_t^x = A_{\theta_0}S_{t-1}^x A_{\theta_0}^{'}(I - K_tC_\theta)^{'} + A_{\theta_0}P_{t-1}^x\mathcal{F}_x^{'\epsilon}(\theta) + Cov_\theta\left(\tilde{\mathcal{W}}_x^\epsilon(\theta), \sigma_{\theta_0}^\eta \eta_t\right)$$
$$P_t^x = A_{\theta_0}P_{t-1}^x A_{\theta_0}^{'} + Q_{\theta_0}$$

7

with:

$$\tilde{\mathcal{W}}_x^\epsilon(\theta) = \mathcal{W}_x^\epsilon(\theta) + \sigma_\theta^\eta \eta_t - K_t \sigma_\theta^\varepsilon \varepsilon_t - K_t C_\theta \sigma_\theta^\eta \eta_t \tag{15}$$

where $\mathcal{E}_x^\epsilon(\theta)$, $\mathcal{F}_x^\epsilon(\theta)$ and $\mathcal{W}_x^\epsilon(\theta)$ are given in Eq.(6), Eq.(7) and Eq.(8) in Theorem 2.1.

If $\epsilon << 1$, then $\mathbb{V}_\theta[\tilde{\mathcal{W}}_x^\epsilon(\theta)] = Q_\theta + K_t \left(C_\theta Q_\theta C'_\theta + R_\theta \right) K'_t$ and $Cov_\theta \left(\tilde{\mathcal{W}}_x^\epsilon(\theta), \sigma_{\theta_0}^\eta \eta_t \right)$ is given by:

$$Cov_\theta \left(\tilde{\mathcal{W}}_x^\epsilon(\theta), \sigma_{\theta_0}^\eta \eta_t \right) = -\epsilon \left(\frac{\partial \sigma_\theta^\eta}{\partial \theta} Q_{\theta_0} \sigma_{\theta_0}^{'\eta} - K_t \left(C_\theta \frac{\partial \sigma_\theta^\eta}{\partial \theta} + \sigma_\theta^\eta \right) Q_{\theta_0} \sigma_{\theta_0}^{'\eta} \right)$$

Additionally, the variance matrix Σ_t^y is given by

$$\Sigma_t^y = \begin{pmatrix} V_t^y & S_t^{'y} \\ S_t^y & P_t^y \end{pmatrix}$$

where:

$$V_t^y = C_\theta V_t^x C'_\theta + \mathcal{F}_y^\epsilon(\theta) S_t C'_\theta + C_\theta S'_t \mathcal{F}_y^{'\epsilon}(\theta) + \tilde{\mathcal{F}}_y^\epsilon(\theta) P_t^x \tilde{\mathcal{F}}_y^\epsilon(\theta) + \mathbb{V}_\theta[\tilde{\mathcal{W}}_y^\epsilon(\theta)]$$
$$S_t^y = C_{\theta_0} S_t^x C'_\theta + C_{\theta_0} P_t^x \mathcal{F}_y^{'\epsilon}(\theta) + Cov_\theta \left(\tilde{\mathcal{W}}_y^\epsilon(\theta), \sigma_{\theta_0}^\varepsilon \varepsilon_t \right)$$
$$P_t^y = C_{\theta_0} P_t^x C'_{\theta_0} + R_{\theta_0}$$

with:

$$\tilde{\mathcal{W}}_y^\epsilon(\theta) = \mathcal{W}_y^\epsilon(\theta) + \sigma_\theta^\varepsilon \varepsilon_t \tag{16}$$

where $\mathcal{E}_y^\epsilon(\theta)$, $\mathcal{F}_y^\epsilon(\theta)$ and $\mathcal{W}_y^\epsilon(\theta)$ are given in Eq.(10) Theorem 2.1.

If $\epsilon << 1$, then $\mathbb{V}_\theta[\tilde{\mathcal{W}}_y^\epsilon(\theta)] = R_\theta$ and $Cov_\theta \left(\tilde{\mathcal{W}}_y^\epsilon(\theta), \sigma_{\theta_0}^\varepsilon \varepsilon_t \right)$ is given by:

$$Cov_\theta \left(\tilde{\mathcal{W}}_y^\epsilon(\theta), \sigma_{\theta_0}^\varepsilon \varepsilon_t \right) = -\epsilon \left(\frac{\partial \sigma_\theta^\varepsilon}{\partial \theta} R_{\theta_0} \sigma_{\theta_0}^{'\varepsilon} \right)$$

Proof. See Appendix (B). $\qquad\qquad\square$

The quantities $\mathcal{F}_x^\epsilon(\theta, t)$ and $\mathcal{W}_x^\epsilon(\theta, t)$ (resp. $\mathcal{F}_y^\epsilon(\theta, t)$ and $\mathcal{W}_y^\epsilon(\theta, t)$) correspond to the correctives terms arising from the bias of the parameters and in particular from the correlation between e_t and the true state x_t (see Eq.(5)). This correlation induces a new Lyapunov dynamic equation for the variance matrix V_t^x. For unbiased parameters estimates, these terms are dropped and a CLT is given in [dNCdL94].

4 Concluding remarks and discussion

In this paper we provide an expression of the propagation errors on the hidden state for an initial and fixed error on the vector of parameters.

We showed that the hidden state x_t appaered in the propagation equation inducing a correlation between e_t and the true state x_t and most importantly a new Lyapunov dynamic equation for the variance matrix. By using the same assumptions than in [dNCdL94] and adding smoothness assumptions on the functions b and h and on their derivatives, one can again obtain a CLT for e_t. Nevertheless, it is not the subject of this paper.

Another remark concerns the case where ϵ is not fixed and is supposed to be a random variable. This particular case refers to the approach proposed in [HK01] for which the parameters are supposed time varying. A dynamical artificial evolution is assumed for θ such that $\theta_t = \theta_{t-1} + \sigma^Z Z$ where Z is a centered and standard gaussian random variable. To the best of our knowledge, there does not exist results about the convergence of this approach. This method fails in practice when the variance σ^Z is not small. Some authors use σ^Z decreasing with time. Hence, at each step of the filter, a small perturbation is added to the parameters. This can be seen as a small bias ϵ introduced at the first step of the filter.

A Proof of Theorem 2.1:

The proof is essentially based on a one order Taylor expansion of the functions b and h with respect to θ.

$$
\begin{aligned}
e_t = x_t - \mathbb{E}_\theta[x_t|y_{1:t}] &= x_t - \mathbb{E}_\theta[x_t|y_{1:t-1}] - K_t(y_t - \hat{y}_t^-) \\
&= u_t(\theta_0) + A_{\theta_0}x_{t-1} + \sigma_{\theta_0}^\eta \eta_t - \mathbb{E}_\theta[u_t(\theta) + A_\theta x_{t-1} + \sigma_\theta^\eta \eta_t|y_{1:t-1}] - K_t(y_t - \hat{y}_t^-) \\
&= u_t(\theta_0) + A_{\theta_0}x_{t-1} + \sigma_{\theta_0}^\eta \eta_t - \mathbb{E}_\theta[u_t(\theta) + A_\theta x_{t-1}|y_{1:t-1}] - K_t(y_t - \hat{y}_t^-) \\
&= u_t(\theta_0) + A_{\theta_0}x_{t-1} + \sigma_{\theta_0}^\eta \eta_t - u_t(\theta) - A_\theta \mathbb{E}_\theta[x_{t-1}|y_{1:t-1}] - K_t(y_t - \hat{y}_t^-) \\
&= u_t(\theta_0) + A_{\theta_0}x_{t-1} + \sigma_{\theta_0}^\eta \eta_t - u_t(\theta) - A_\theta \hat{x}_{t-1} - K_t(y_t - \hat{y}_t^-) \quad (17)
\end{aligned}
$$

Note that one can write:

$$
u_t(\theta_0) = u_t(\theta) - \epsilon \frac{\partial u}{\partial \theta}(\theta) + o(\epsilon), \quad A_{\theta_0} = A_\theta - \epsilon \frac{\partial A_\theta}{\partial \theta} + o(\epsilon), \quad \sigma_{\theta_0}^\eta = \sigma_{\theta_0}^\eta - \epsilon \frac{\partial \sigma_\theta^\eta}{\partial \theta} + o(\epsilon)
$$

Pluging into (17), one gets:

$$
e_t = A_\theta e_{t-1} - \epsilon \frac{\partial u_t}{\partial \theta}(\theta) - \epsilon \frac{\partial A_\theta}{\partial \theta} x_{t-1} + \sigma_\theta^\eta \eta_t - \epsilon \frac{\partial \sigma_\theta^\eta}{\partial \theta} \eta_t - K_t(y_t - \hat{y}_t^-) + o(\epsilon) \quad (18)
$$

Furthermore,

$$
\begin{aligned}
(y_t - \hat{y}_t^-) &= d_t(\theta_0) + C_{\theta_0}x_t + \sigma_{\theta_0}^\varepsilon \varepsilon_t - \mathbb{E}_\theta[y_t|y_{1:t-1}] \\
&= d_t(\theta_0) + C_{\theta_0}x_t + \sigma_{\theta_0}^\varepsilon \varepsilon_t - d_t(\theta) - C_\theta \mathbb{E}_\theta[x_t|y_{1:t-1}]
\end{aligned}
$$

and

$$
d_t(\theta_0) = d_t(\theta) - \epsilon \frac{\partial d}{\partial \theta}(\theta) + o(\epsilon), \quad C(\theta_0) = C(\theta) - \epsilon \frac{\partial C}{\partial \theta}(\theta) + o(\epsilon), \quad \sigma_{\theta_0}^\varepsilon = \sigma_{\theta_0}^\varepsilon - \epsilon \frac{\partial \sigma_\theta^\varepsilon}{\partial \theta} + o(\epsilon)
$$

So that:

$$
\begin{aligned}
(y_t - \hat{y}_t^-) = &-\epsilon \frac{\partial d_t}{\partial \theta}(\theta) + (\sigma_\theta^\varepsilon - \epsilon \frac{\partial \sigma_\theta^\varepsilon}{\partial \theta})\varepsilon_t + \left(C_\theta - \epsilon \frac{\partial C_\theta}{\partial \theta}\right)\left(u_t(\theta_0) + A_{\theta_0}x_{t-1} + \sigma_{\theta_0}^\eta \eta_t\right) \\
&-C_\theta \mathbb{E}_\theta[x_t|y_{1:t-1}] + o(\epsilon)
\end{aligned}
$$

Rewrite,

$$
\mathbb{E}_\theta[x_t|y_{1:t-1}] = A_\theta \mathbb{E}_\theta[x_{t-1}|y_{1:t-1}] + u_t(\theta)
$$

we get:

$$
\begin{aligned}
= &\ C_\theta A_\theta x_{t-1} - C_\theta A_\theta \hat{x}_{t-1} + \sigma_\theta^\varepsilon \varepsilon_t + C_\theta \sigma_\theta^\eta \eta_t \\
&- \epsilon \left(\frac{\partial d_t}{\partial \theta}(\theta) + C_\theta \frac{\partial A_\theta}{\partial \theta} x_{t-1} + \frac{\partial \sigma_\theta^\varepsilon}{\partial \theta} \varepsilon_t + C_\theta \frac{\partial u_t}{\partial \theta}(\theta) + C_\theta \frac{\partial \sigma_\theta^\eta}{\partial \theta} \eta_t \right. \\
&\left. + \frac{\partial C_\theta}{\partial \theta} u_t(\theta) + \frac{\partial C_\theta}{\partial \theta} A_\theta x_{t-1} + \frac{\partial C_\theta}{\partial \theta} \sigma_\theta^\eta \eta_t \right) \\
&+ \epsilon^2 \left(\frac{\partial C_\theta}{\partial \theta}\frac{\partial u_t}{\partial \theta}(\theta) + \frac{\partial C_\theta}{\partial \theta}\frac{\partial A_\theta}{\partial \theta} x_{t-1} + \frac{\partial C_\theta}{\partial \theta}\frac{\partial \sigma_\theta^\eta}{\partial \theta} \eta_t \right) + o(\epsilon) \quad (19)
\end{aligned}
$$

Define,

$$\mathcal{E}^{\epsilon}_{y-}(\theta) = -\epsilon \left(\frac{\partial d_t}{\partial \theta}(\theta) + C_\theta \frac{\partial u_t}{\partial \theta}(\theta) + \frac{\partial C_\theta}{\partial \theta} u_t(\theta) \right),$$

$$\mathcal{F}^{\epsilon}_{y-}(\theta) = -\epsilon \left(C_\theta \frac{\partial A_\theta}{\partial \theta} + \frac{\partial C_\theta}{\partial \theta} A_\theta \right),$$

$$\mathcal{W}^{\epsilon}_{y-}(\theta) = -\epsilon \left(\frac{\partial C_\theta}{\partial \theta} \sigma^\eta_\theta \eta_t + C_\theta \frac{\partial \sigma^\eta_\theta}{\partial \theta} \eta_t + \frac{\partial \sigma^\varepsilon_\theta}{\partial \theta} \varepsilon_t \right),$$

we obtain:

$$
\begin{aligned}
\xi^-_t &= y_t - \mathbb{E}_\theta[y_t|y_{1:t-1}] \\
&= C_\theta A_\theta e_{t-1} + \sigma^\varepsilon_\theta \varepsilon_t + C_\theta \sigma^\eta_\theta \eta_t + \mathcal{E}^\epsilon_{y-}(\theta) + \mathcal{F}^\epsilon_{y-}(\theta) x_{t-1} + \mathcal{W}^\epsilon_{y-}(\theta) + o(\epsilon)
\end{aligned}
$$

By combining Eq.(18) and Eq.(19), we have:

$$
\begin{aligned}
e_t &= x_t - \mathbb{E}_\theta[x_t|y_{1:t}] \\
&= (I_{n_x \times n_x} - K_t C_\theta) A_\theta e_{t-1} - K_t \sigma^\varepsilon_\theta \varepsilon_t - K_t C_\theta \sigma^\eta_\theta \eta_t + \sigma^\eta_\theta \eta_t + \mathcal{E}^\epsilon_x(\theta) + \mathcal{F}^\epsilon_x(\theta) x_{t-1} + \mathcal{W}^\epsilon_x(\theta) + o(\epsilon)
\end{aligned}
$$

where,

$$\mathcal{E}^\epsilon_x(\theta) = -\epsilon \left((I_{n_x \times n_x} - K_t C_\theta) \frac{\partial u_t}{\partial \theta}(\theta) - K_t \frac{\partial d_t}{\partial \theta}(\theta) - \frac{\partial C_\theta}{\partial \theta} u_t(\theta) \right),$$

$$\mathcal{F}^\epsilon_x(\theta) = -\epsilon \left((I_{n_x \times n_x} - K_t C_\theta) \frac{\partial A_\theta}{\partial \theta} - \frac{\partial C_\theta}{\partial \theta} A_\theta \right),$$

$$\mathcal{W}^\epsilon_x(\theta) = -\epsilon \left(\frac{\partial \sigma^\eta_\theta}{\partial \theta} \eta_t - K_t C_\theta \frac{\partial \sigma^\eta_\theta}{\partial \theta} \eta_t - K_t \sigma^\eta_\theta \frac{\partial C_{\theta_0}}{\partial \theta} \eta_t - K_t \frac{\partial \sigma^\varepsilon_{\theta_0}}{\partial \theta} \varepsilon_t \right),$$

One can deduce the *Propagation of the residues a posteriori:*

$$
\begin{aligned}
\xi_t &= y_t - \mathbb{E}_\theta[y_t|y_{1:t}] \\
&= d_t(\theta_0) + C_{\theta_0} x_t + \sigma^\varepsilon_{\theta_0} \varepsilon_t - \mathbb{E}_\theta[d_t(\theta) + C_\theta x_t + \sigma^\varepsilon_\theta \varepsilon_t|y_{1:t}] \\
&= d_t(\theta_0) - d_t(\theta) + (C_\theta - \epsilon \frac{\partial C_\theta}{\partial \theta}) x_t - C_\theta \mathbb{E}_\theta[x_t|y_{1:t}] + (\sigma^\varepsilon_\theta - \epsilon \frac{\partial \sigma^\varepsilon_\theta}{\partial \theta}) \varepsilon_t + o(\epsilon) \\
&= C_\theta e_t + \sigma^\varepsilon_\theta \varepsilon_t - \epsilon \left(\frac{\partial d_t}{\partial \theta}(\theta) + \frac{\partial C_\theta}{\partial \theta} x_t + \frac{\partial \sigma^\varepsilon_\theta}{\partial \theta} \varepsilon_t \right) + o(\epsilon)
\end{aligned}
$$

By defining:

$$\mathcal{E}^\epsilon_y(\theta) = -\epsilon \frac{\partial d_t(\theta)}{\partial \theta}$$

$$\mathcal{F}^\epsilon_y(\theta) = -\epsilon \frac{\partial C_\theta}{\partial \theta}$$

$$\mathcal{W}^\epsilon_y(\theta) = -\epsilon \frac{\partial \sigma^\varepsilon_\theta}{\partial \theta} \varepsilon_t$$

Eq.(9) follows. □

The proof of Corollary 1 is obtained by taking the expectations in Eq.(5) and Eq.(9).

B   Proof of Theorem 3.1:

By using the system (2) and Eq.(5)-Eq.(9) we can rewrite the model as follows:

$$E_t = \begin{pmatrix} \mathcal{E}_x^\epsilon(\theta) \\ u_t(\theta_0) \end{pmatrix} + \begin{pmatrix} (I - K_t C_\theta) A_\theta & \mathcal{F}_x^\epsilon(\theta) \\ 0 & A_{\theta_0} \end{pmatrix} E_{t-1} + \begin{pmatrix} \tilde{\mathcal{W}}_x^\epsilon(\theta) \\ \sigma_{\theta_0}^\eta \eta_t \end{pmatrix} \qquad (20)$$

and,

$$\mathfrak{E}_t = \begin{pmatrix} \mathcal{E}_y^\epsilon(\theta) \\ d_t(\theta_0) \end{pmatrix} + \begin{pmatrix} C_\theta & \mathcal{F}_y^\epsilon(\theta) \\ 0 & C_{\theta_0} \end{pmatrix} E_t + \begin{pmatrix} \tilde{\mathcal{W}}_y^\epsilon(\theta) \\ \sigma_{\theta_0}^\varepsilon \varepsilon_t \end{pmatrix} \qquad (21)$$

Hence, the variance matrix Σ_t^x is given by:

$$\Sigma_t^x = \begin{pmatrix} (I - K_t C_\theta) A_\theta & \mathcal{F}_x^\epsilon(\theta) \\ 0 & A_{\theta_0} \end{pmatrix} \Sigma_{t-1}^x \begin{pmatrix} (I - K_t C_\theta) A_\theta & \mathcal{F}_x^\epsilon(\theta) \\ 0 & A_{\theta_0} \end{pmatrix}' + \begin{pmatrix} \mathbb{V}[\tilde{\mathcal{W}}_x^\epsilon(\theta)] & \mathbb{C}ov(\tilde{\mathcal{W}}_x^\epsilon(\theta), \sigma_{\theta_0}^\eta \eta_t) \\ \mathbb{C}ov(\tilde{\mathcal{W}}_x^\epsilon(\theta), \sigma_{\theta_0}^\eta \eta_t) & \sigma_{\theta_0}^\eta \sigma_{\theta_0}^{'\eta} \end{pmatrix}$$

Additionally, the variance matrix Σ_t^y is given by:

$$\Sigma_t^y = \begin{pmatrix} C_\theta & \mathcal{F}_y^\epsilon(\theta) \\ 0 & C_{\theta_0} \end{pmatrix} \Sigma_t^x \begin{pmatrix} C_\theta & \mathcal{F}_y^\epsilon(\theta) \\ 0 & C_{\theta_0} \end{pmatrix}' + \begin{pmatrix} \mathbb{V}[\tilde{\mathcal{W}}_y^\epsilon(\theta)] & \mathbb{C}ov(\tilde{\mathcal{W}}_y^\epsilon(\theta), \sigma_{\theta_0}^\varepsilon \eta_t) \\ \mathbb{C}ov(\tilde{\mathcal{W}}_x^\epsilon(\theta), \sigma_{\theta_0}^\varepsilon \eta_t) & \sigma_{\theta_0}^\varepsilon \sigma_{\theta_0}^{'\varepsilon} \end{pmatrix}$$

Proposition 2.1 gives that:

$$\begin{cases} \mathcal{W}_x^\epsilon(\theta) = -\epsilon \left(\frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} \eta_t - K_t C_\theta \frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} \eta_t - K_t \sigma_\theta^\eta \frac{\partial C_{\theta_0}}{\partial \theta} \eta_t - K_t \frac{\partial \sigma_{\theta_0}^\varepsilon}{\partial \theta} \varepsilon_t \right) \\ \mathcal{W}_y^\epsilon(\theta) = -\epsilon \frac{\partial \sigma_{\theta_0}^\varepsilon}{\partial \theta} \varepsilon_t \end{cases}$$

Hence, if $\epsilon << 1$, then

$$\mathbb{V}[\mathcal{W}_x^\epsilon(\theta)] = Q_\theta + K_t \left(C_\theta Q_\theta C_\theta' + R_\theta \right) K_t' \text{ and } \mathbb{V}[\mathcal{W}_y^\epsilon(\theta)] = R_\theta$$

Furthermore, the covariances are given by:

$$
\begin{aligned}
Cov_\theta \left(\tilde{\mathcal{W}}_x^\epsilon(\theta), \sigma_{\theta_0}^\eta \eta_t \right) &= -\epsilon Cov_\theta \left(\frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} \eta_t, \sigma_{\theta_0}^\eta \eta_t \right) + \epsilon \left(K_t C_\theta \frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} \eta_t, \sigma_{\theta_0}^\eta \eta_t \right) + \epsilon Cov_\theta \left(K_t \sigma_\theta^\eta \eta_t, \sigma_{\theta_0}^\eta \eta_t \right) \\
&\quad + \epsilon Cov_\theta \left(K_t \frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} \varepsilon_t, \sigma_{\theta_0}^\eta \eta_t \right) \\
&= \epsilon \frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} Q_{\theta_0} \sigma_{\theta_0}^{'\eta} + \epsilon K_t C_\theta \frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} Q_{\theta_0} \sigma_{\theta_0}^{'\eta} + \epsilon K_t \sigma_\theta^\eta Q_{\theta_0} \sigma_{\theta_0}^{'\eta} \text{ by assumption } \mathbf{A2} \\
&= -\epsilon \left(\frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} Q_{\theta_0} \sigma_{\theta_0}^{'\eta} - K_t \left(C_\theta \frac{\partial \sigma_{\theta_0}^\eta}{\partial \theta} + \sigma_\theta^\eta \right) Q_{\theta_0} \sigma_{\theta_0}^{'\eta} \right)
\end{aligned}
$$

Additionally,

$$
\begin{aligned}
Cov_\theta \left(\tilde{\mathcal{W}}_y^\epsilon(\theta), \sigma_{\theta_0}^\varepsilon \varepsilon_t \right) &= -\epsilon Cov_\theta \left(\frac{\partial \sigma_{\theta_0}^\varepsilon}{\partial \theta} \varepsilon_t, \sigma_{\theta_0}^\varepsilon \varepsilon_t \right) \\
&= -\epsilon \left(\frac{\partial \sigma_{\theta_0}^\varepsilon}{\partial \theta} R_{\theta_0} \sigma_{\theta_0}^{'\varepsilon} \right)
\end{aligned}
$$

$\square$

References

[dNCdL94] B. d'Andréa Novel and M. Cohen de Lara. *Commande linéaire des systèmes dynamiques.* Modélisation. Analyse. Simulation. Commande. [Modeling. Analysis. Simulation. Control]. Masson, Paris, 1994. With a preface by A. Bensoussan.

[ElK12]   S. ElKolei. *Estimation des modèles à volatilité stochastique par filtrage et déconvolution.* PhD thesis, 2012.

[HK01]   M. Hürzeler and H. R. Künsch. Approximating and maximising the likelihood for a general state-space model. In *Sequential Monte Carlo methods in practice*, pages 159–175. Springer, New York, 2001.

[HS71]   P. J. Harrison and C. F. Stevens. A Bayesian approach to short-term forecasting. *Operational Res. Quart.*, 22:341–362, 1971.

[IBAK11] Edward L. Ionides, Anindya Bhadra, Yves Atchadé, and Aaron King. Iterated filtering. *Ann. Statist.*, 39(3):1776–1802, 2011.

[JPS09]   Michael S. Johannes, Nicholas G. Polson, and Jonathan R. Stroud. Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices. *Review of Financial Studies*, 22(7):2559–2599, July 2009.

[Kal60]   Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[PHH10]   Gareth W. Peters, Geoffrey R. Hosack, and Keith R. Hayes. Ecological non-linear state space model selection via adaptive particle Markov chain monte carlo (adpmcmc). *Preprint:arXIv-1005.2238v1*, 2010.