A Quadratic Kalman Filter*

Alain MONFORT[†], Jean-Paul RENNE[‡], and Guillaume ROUSSELLET[§]

December, 2013

Abstract

We propose a new filtering and smoothing technique for non-linear state-space models. Observed variables are quadratic functions of latent factors following a Gaussian VAR. Stacking the vector of factors with its vectorized outer-product, we form an augmented state vector whose first two conditional moments are known in closed-form. We also provide analytical formulae for the unconditional moments of this augmented vector. Our new quadratic Kalman filter (QKF) exploits these properties to formulate fast and simple filtering and smoothing algorithms. A first simulation study emphasizes that the QKF outperforms the extended and unscented approaches in the filtering exercise showing up to 70% RMSEs improvement of filtered values. Second, we provide evidence that QKF-based maximum-likelihood estimates of model parameters always possess lower bias or lower RMSEs that the alternative estimators.

JEL Codes: C32, C46, C53, C57

Key-words: Non-linear filtering, non-linear smoothing, quadratic model, Kalman filter, pseudo-maximum likelihood

*Functions for the Quadratic Kalman Filter are implemented with the `R`-software and are available on the `runmycode`-website.

[†] CREST, Banque de France, and Maastricht University, alain.monfort@ensae.fr

[‡] Banque de France, jeanpaul.renne@banque-france.fr

[§] Corresponding author. Banque de France, CREST and CEREMADE guillaume.roussellet@banque-france.fr

The authors thank participants to the CREST financial econometrics seminar, to the Banque de France seminar, and to the 2013 CFE-ERCIM conference. Particular thanks are addressed to Michel JUILLARD for providing access to the Banque de France computation platform Hubeco.

1 Introduction

This paper proposes a new discrete-time Kalman filter for state-space models where the transition equations are linear and the measurement equations are quadratic. We call this method the *Quadratic Kalman Filter* (QKF). While this state-space model have become increasingly popular in the applied econometrics literature, existing filters are either highly computationally intensive, or not specifically fitted to the linear-quadratic case. We begin by building the augmented vector of factors stacking together the latent vector and its vectorized outer-product. To the best of our knowledge, this paper is the first to derive analytically and provide closed-form formulae of both the conditional and the unconditional first-two moments of this augmented vector.[2] Using these moments, the transition equations of the augmented vector are expressed in an affine form. Similarly, the measurement equations are rewritten as affine functions of the augmented vector of factors. We thus obtain an *augmented state-space model* that is fully linear.

We perform the derivation of the QKF filtering and smoothing algorithms by applying the linear Kalman algorithms to the augmented state-space model. To do so, we approximate the conditional distribution of the augmented vector of factors given its own past by a multivariate Gaussian distribution. Since no adaptation of the linear algorithm is needed, the QKF combines simplicity of implementation and fast computational speed. We apply the same method for the derivation of the Quadratic Kalman Smoothing algorithm (QKS). Indeed, since the QKF and QKS requires no simulations, it represents a convenient alternative to particle filtering.

To compare our filter with the popular existing *traditional filters* (see Tanizaki (1996)), namely the first- and second-order extended and the unscented Kalman filters, we implement a Monte-Carlo experiment. In order to explore a broad range of cases, we build a benchmark state-space model with different values for *(i)* the persistence of the latent process, *(ii)* the importance of noise variance in the observable, and *(iii)* the importance of quadratic terms in the observables. RMSE measures are computed in each case. We compare the filters with respect to two different criteria: filtering, i.e. retrieving latent factors precisely from a fixed set of parameters, and parameter estimation, i.e. the capacity to estimate the state-space model parameters.

First, these computations provide evidence of the superiority of the QKF filtering over its competitors in all cases. When the measurement equations are fully quadratic, the QKF is the only filter able to capture the non-linearities and to produce time-varying evaluations of the latent factors. This results in up to 70% lower RMSEs for the QKF compared to the other filters, all cases considered. For measurement equations with both linear and quadratic terms, the QKF still results – to a smaller extent – in lower filtering RMSEs. These results are robust to the persistence degree of the latent process and the size of the measurement noise. Also, we emphasize that the

[2] Buraschi, Cieslak, and Trojani (2008) provide formulae of conditional first-two moments for the specific case of centred Wishart processes.

first-order extended Kalman filter performs particularly poorly in some cases and should therefore be discarded for filtering in the linear-quadratic model.

Second, the QKF-based maximum-likelihood estimates of model parameters always possess lower bias or lower RMSEs that the alternative estimators. We provide evidence that this superiority is robust to the degree of persistence of the latent process, to the degree of linearity of the measurement equations, and to the size of the measurement errors. We conclude that the QKF results in the best bias/variance trade-off for the pseudo-maximum likelihood estimation.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the non-linear filtering literature and its applications. Section 3 presents the state-space model and builds the QKF. Section 4 performs a comparison of the QKF with popular competitors using Monte-Carlo experiments. Section 5 concludes. Proofs are gathered in the Appendices.

2 Literature review

The existing *traditional* non-linear filters use linearization techniques to transform the state-space model. First and second-order extended Kalman filters build respectively on first and second-order Taylor expansions of transition and measurement equations. The first-order extended Kalman filter is extensively covered in Anderson and Moore (1979). To reduce the errors linked to the first-order approximations, Athans, Wishner, and Bertolini (1968) develop a second-order extended Kalman filter. This method is treated in continuous and continuous-discrete time in Gelb, Kasper, Nash, Price, and Sutherland (1974) and Maybeck (1982). Bar-Shalom, Kirubarajan, and Li (2002) or Hendeby (2008, Chapter 5.) propose a complete description of this second-order filter. In the general non-linear case, both methods require numerical approximations of gradients and Hessian matrices, potentially increasing the computational burden.[3] The unscented Kalman filter belongs more to the class of *deterministic density estimators*, and was originally implemented as an alternative to the previous techniques for applications in physics. It is a derivative-free method which is shown to be computationally close to the second-order extended Kalman filter in terms of complexity. Exploiting applications in radar-tracking and localization, the unscented filter is proved to perform at least as well as the second-order Gaussian extended filter (see Julier, Uhlmann, and Durrant-Whyte (2000) or Julier and Uhlmann (2004)).[4] Whereas many other filters exist, both the extended and unscented filters have been the most widely used in recent applied physics and econometrics.[5]

We consider here a specification in which the transition equations are affine and the measurement

[3] Gustafsson and Hendeby (2012) build a derivative-free version of the second-order extended Kalman filter which avoids issues due to numerical approximations, but shows a similar computational complexity.

[4] A complete description of the algorithm can be found in Merwe and Wan (2001), Julier and Uhlmann (2004), or Hendeby (2008). Also, a general version of the algorithm is provided in the Appendix.

[5] Other filters comprise, among others, higher order extended Kalman filters, importance resampling, particle and Monte-Carlo filters, Gaussian sum filters.

equations are quadratic. It first extends the static case used by studies dealing with quadratic regressions where explanatory variables are measured with errors (see Kuha and Temple (2003), Wolter and Fuller (1982) for an earth science application, and Barton and David (1960) for astronomy applications). Still in the static case, Kukush, Markovsky, and Huffel (2002) illustrate the use of quadratic measurement-errors filtering for image processing purpose. The quadratic framework is also particularly suited to numerous dynamic economic models. While first-order linearization is standard and largely employed in the dynamic stochastic general equilibrium (DSGE) literature, the algorithm we develop is fitted to exploit second-order approximations.[6]

As for finance, an important field of applications of our filter is the modelling of term structures of interest rates.[7] The standard and popular Gaussian affine term-structure model (GATSM) provides yields which are affine combinations of dynamic linear auto-regressive factor processes. As these models include latent factors, the linear Kalman filter[8] has gained overwhelming popularity compared to other estimation techniques (see e.g. Duan and Simonato (1999), Kim and Wright (2005) or Joslin, Singleton, and Zhu (2011)). A natural extension of the GATSM is to assume that yields are quadratic functions of factor processes. By authorizing additional degrees of freedom while maintaining closed-form pricing formulae, this quadratic class of models provides a better fit of the data than ATSM (see Ahn, Dittmar, and Gallant (2002)). The bulk of the papers using QTSMs considers the dynamics of government-bond yield curves (e.g. Leippold and Wu (2007) and Kim and Singleton (2012)). Exploiting the fact that they can generate positive-only variables, QTSMs have also been shown to be relevant to model the dynamics of risk intensities and their implied term structures: while default intensities are considered in the credit-risk literature (see e.g. Doshi, Jacobs, Ericsson, and Turnbull (2013) and Dubecq, Monfort, Renne, and Roussellet (2013)), mortality intensities have also been modelled in this framework (Gourieroux and Monfort (2008)). In order to estimate QTSMs involving latent variables, a wide range of techniques are considered in the existing literature: Brandt and Chapman (2003), Inci and Lu (2004), Li and Zhao (2006) and Kim and Singleton (2012) use the extended Kalman filter, Lund (1997) considers the iterated extended Kalman filter[9], Leippold and Wu (2007), Doshi, Jacobs, Ericsson, and Turnbull (2013) or Chen, Cheng, Fabozzi, and Liu (2008) employ the unscented Kalman filter and Andreasen and Meldrum (2011) opt for the particle filter. Baadsgaard, Nielsen, and Madsen (2000) use the truncated second-order extended filter to estimate a term structure model with CIR latent processes. Ahn, Dittmar, and Gallant (2002) resort to the efficient method of moments (EMM). However, Duffee and Stanton (2008) show that, compared to maximum likelihood approaches, EMM has poor finite sample properties when data are persistent, a typical characteristic of bond

[6] See Pelgrin and Juillard (2004) for a review of existing algorithms to construct second-order approximations of DSGE solutions. Our approach could for instance be exploited to estimate the standard asset-pricing model of Burnside (1998) considered e.g. by Collard and Juillard (2001) (taking the rate of growth in dividends as a latent factor).

[7] See Dai and Singleton (2003) for a survey of interest-rate term-structure modelling literature.

[8] see Kalman (1960) for the original linear filter derivation. Properties are developed in e.g. Harvey (1991) or Durbin and Koopman (2012).

[9] See Jazwinski (1970) for a description of the filtering technique.

yields. Moreover, while EMM is used to estimate model parameters, it does not directly provide estimates of the latent factors.[10] Finally, Dubecq, Monfort, Renne, and Roussellet (2013) use the QKF filter that is developed hereafter.

The quadratic state-space framework that we consider in the present paper is also well-suited to work with Wishart processes. These processes have been used in various empirical-finance studies. In most cases, they are employed in multivariate stochastic volatility models (see e.g. Romo (2012), Jin and Maheu (2013), Philipov and Glickman (2006), Rinnergschwentner, Tappeiner, and Walde (2011) or Branger and Muck (2012)).[11]. Wishart processes have also been exploited in several QTSMs (Filipovic and Teichmann (2002), Gourieroux, Monfort, and Sufana (2010), Gourieroux and Sufana (2011), and Buraschi, Cieslak, and Trojani (2008)).

3 The Quadratic Kalman Filter (QKF) and Smoother (QKS)

3.1 Model and notations

We are interested in a state-space model with affine transition equations and quadratic measurement equations. We consider the following model involving a latent (or state) variable X_t of size n and an observable variable Y_t of size m. X_t might be only partially latent, that is, some components of X_t might be observed.

Definition 3.1 *The linear-quadratic state-space model is defined by:*

$$X_t = \mu + \Phi X_{t-1} + \Omega \varepsilon_t \tag{1a}$$

$$Y_t = A + BX_t + \sum_{k=1}^{m} e_k X_t' C^{(k)} X_t + D\eta_t. \tag{1b}$$

where ε_t and η_t are independent Gaussian white noises with unit variance-covariance matrices, $\Omega\Omega' = \Sigma$ and $DD' = V$. e_k is the column selection vector of size m whose components are 0 except the k^{th} one, which is equal to 1. μ and Φ are respectively a n-dimensional vector and a square matrix of size n. A and B are respectively a vector of size m and a $(n \times m)$ matrix. All $C^{(k)}$'s are without loss of generality square symmetric matrices of size $m \times m$.

A component-by-component version of the measurement equations (1b) is:

$$Y_{t,k} = A_k + B_k X_t + X_t' C^{(k)} X_t + D_k \eta_t, \quad \forall k \in \{1, \ldots, m\}, \tag{2}$$

where $Y_{t,k}$, A_k, B_k, D_k are respectively the kth row of Y_t, A, B, and D. Note that μ, Φ, Σ, A, B, $C^{(k)}$, and D might be functions of $(Y_{t-1}, Y_{t-2}, \ldots)$, that are the past values of the observable variables.

[10] Gallant and Tauchen (1998) however propose a reprojection method to recover latent variables after having estimated the model parametrization by means of EMM.

[11] See Asai, McAleer, and Yu (2006) for a review of multivariate stochastic volatility models.

Our objective is twofold: *(i)* filtering and smoothing of X_t, which consist in retrieving the values of X_t conditionally on, respectively, past and present values of Y_t, and all the observed values of $(Y_t)_{t=1,\ldots,T}$; and *(ii)* estimation of the parameters appearing in μ, Φ, Ω, A, B, $C^{(k)}$, D. Note that Ω and D are defined up to the right multiplication by an orthogonal matrix. These matrices can be fixed by imposing $\Omega = \Sigma^{1/2}$ and $D = V^{1/2}$.[12]

Throughout the paper, we use the following notations. At date t, past observations of the observed vector are denoted by $\underline{Y_t} = \{Y_t, Y_{t-1}, Y_{t-2}, \ldots, Y_1\}$, and for any process W_t:

$$\begin{aligned} W_{t|t} &\equiv \mathbb{E}\left[W_t|\underline{Y_t}\right], & P^W_{t|t} &\equiv \mathbb{V}\left[W_t|\underline{Y_t}\right], \\ W_{t|t-1} &\equiv \mathbb{E}\left[W_t|\underline{Y_{t-1}}\right], & P^W_{t|t-1} &\equiv \mathbb{V}\left[W_t|\underline{Y_{t-1}}\right], \\ \mathbb{E}_{t-1}(W_t) &\equiv \mathbb{E}\left[W_t|\underline{W_{t-1}}\right]. & \mathbb{V}_{t-1}(W_t) &\equiv \mathbb{V}\left[W_t|\underline{W_{t-1}}\right]. \end{aligned}$$

We also introduce the notation $M_{t|t-1} \equiv \mathbb{V}\left[Y_t|\underline{Y_{t-1}}\right]$ and:

$$Z_t = \left(X_t', Vec(X_t X_t')'\right)'.$$

Z_t is the vector stacking the components of X_t and its vectorized outer-product. This vector Z_t, called the **augmented state vector** (see Cheng and Scaillet (2007)), will play a key role in our algorithms. We first study the conditional moments of this vector given past information.

3.2 Conditional moments of Z_t

It can be shown (see Bertholon, Monfort, and Pegoraro (2008)) that when μ, Φ and Σ do not depend on $\underline{Y_{t-1}}$, the process (Z_t) is Compound Autoregressive or order 1 –or Car(1)–, that is to say, the conditional log-Laplace transform, or cumulant generating function defined by:

$$\log \varphi_t(u) = \log \mathbb{E}\left[\exp(u'Z_t)|\underline{Z_{t-1}}\right]$$

is affine in Z_{t-1}. This implies, in particular, that the conditional expectation $\mathbb{E}_{t-1}(Z_t)$ and the conditional variance-covariance matrix $\mathbb{V}_{t-1}(Z_t)$ of Z_t given $\underline{Z_{t-1}}$ are affine functions of Z_{t-1}. Moreover, $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$ have closed-form expressions given in the following proposition.

[12] Ω and D can be rectangular when Σ or V are not of full-rank.

Proposition 3.1 *$\mathbb{E}_{t-1}(Z_t) = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1}$ and $\mathbb{V}_{t-1}(Z_t) = \widetilde{\Sigma}_{t-1}$, where:*

$$\widetilde{\mu} = \left(\begin{array}{c} \mu \\ \\ Vec(\mu\mu' + \Sigma) \end{array} \right) \quad , \quad \widetilde{\Phi} = \left(\begin{array}{c|c} \Phi & 0 \\ \hline \mu \otimes \Phi + \Phi \otimes \mu & \Phi \otimes \Phi \end{array} \right)$$

$$\widetilde{\Sigma}_{t-1} \equiv \widetilde{\Sigma}(Z_{t-1}) = \left(\begin{array}{c|c} \Sigma & \Sigma \Gamma'_{t-1} \\ \hline \Gamma_{t-1}\Sigma & \Gamma_{t-1}\Sigma\Gamma'_{t-1} + \big(I_{n^2} + \Lambda_n\big)(\Sigma \otimes \Sigma) \end{array} \right)$$

$$\Gamma_{t-1} = I_n \otimes (\mu + \Phi X_{t-1}) + (\mu + \Phi X_{t-1}) \otimes I_n$$

Λ_n being the $n^2 \times n^2$ matrix, partitioned in $(n \times n)$ blocks, such that the (i,j) block is $e_j e_i'$ (see Appendix A.2 for Λ_n properties).

Proof See Appendix A.3. ■

Note that $\widetilde{\Sigma}_{t-1}$ is a $n(n+1) \times n(n+1)$ matrix whereas $\widetilde{\Sigma}(\bullet)$ is a $\mathbb{R}^{n(n+1)} \longmapsto \mathcal{M}_{n(n+1)\times n(n+1)}$ function, $\mathcal{M}_{n(n+1)\times n(n+1)}$ being the space of symmetric positive definite matrices of size $n(n+1)$. If μ, Φ, and Σ are functions of $\underline{Y_{t-1}}$, Proposition 3.1 still holds replacing $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$ by $\mathbb{E}(Z_t|\underline{Z_{t-1}}, \underline{Y_{t-1}})$ and $\mathbb{V}(Z_t|\underline{Z_{t-1}}, \underline{Y_{t-1}})$, respectively.

$\widetilde{\Sigma}_{t-1}(\bullet)$ is clearly a quadratic function of X_{t-1} and an affine function of Z_{t-1}, denoted by $\widetilde{\Sigma}(Z_{t-1})$ (Proposition 3.1). In the filtering algorithm, we have to compute $\mathbb{E}[\widetilde{\Sigma}(Z_{t-1})|\underline{Y_{t-1}}]$. This quantity is easily computable as $\widetilde{\Sigma}(Z_{t-1|t-1})$ only once the affine form of the function $\widetilde{\Sigma}(Z)$ is explicitly available. Proposition 3.2 details this affine form.

Proposition 3.2 *We denote $\widetilde{\Sigma}^{(i,j)}_{t-1}$ for i and j being $\{1,2\}$ the (i,j) block of $\widetilde{\Sigma}_{t-1}$. Each block of $\widetilde{\Sigma}$ is affine in Z_{t-1} and we have:*

$$\begin{aligned}
Vec\left(\widetilde{\Sigma}^{(1,1)}_{t-1}\right) &= Vec(\Sigma) \\
Vec\left(\widetilde{\Sigma}^{(1,2)}_{t-1}\right) &= \left[\Sigma \otimes (I_{n^2} + \Lambda_n)\right]\left[Vec(I_n) \otimes I_n\right]\left\{\mu + \widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}^{(2,1)}_{t-1}\right) &= \left[(I_{n^2} + \Lambda_n) \otimes \Sigma\right](I_n \otimes \Lambda_n)\left[Vec(I_n) \otimes I_n\right]\left\{\mu + \widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}^{(2,2)}_{t-1}\right) &= \left[(I_{n^2} + \Lambda_n) \otimes (I_{n^2} + \Lambda_n)\right]\left[(I_n \otimes \Lambda_n \otimes I_n)(Vec(\Sigma) \otimes I_{n^2})\right]\left\{\mu \otimes \mu + \widetilde{\Phi}_2 Z_{t-1}\right\} \\
&\qquad + \left[I_{n^2} \otimes (I_{n^2} + \Lambda_n)\right] Vec(\Sigma \otimes \Sigma)
\end{aligned} \tag{3}$$

Where $\widetilde{\Phi}_1$ and $\widetilde{\Phi}_2$ are respectively the upper and lower blocks of $\widetilde{\Phi}$ and Λ_n is defined as in Proposition 3.1. This particularly implies:

$$Vec\left[\mathbb{V}_{t-1}(Z_t)\right] = Vec\left[\widetilde{\Sigma}(Z_{t-1})\right] = \nu + \Psi Z_{t-1},$$

where ν and Ψ are permutations of the multiplicative matrices in Equation 3, and are detailed in Appendix A.4.

Proof See Appendix A.4. ∎

These results extend the computations of Buraschi, Cieslak, and Trojani (2008). While these authors express the conditional first-two moments of a central Wishart autoregressive process (see Appendix C. of Buraschi, Cieslak, and Trojani (2008)), we derive the first two-conditional moments of our augmented vector Z_t in a more general case (where $\mu \neq 0$).

3.3 Unconditional moments of Z_t and stationarity conditions

The analytic derivation of the first two unconditional moments of Z_t can, in particular, be exploited to initialize the filter. In the following subsection, we consider the standard case where μ, Φ and Σ are not depending on $\underline{Y_{t-1}}$. If the eigenvalues of Φ have a modulus strictly smaller than 1, the process (X_t) is strictly and, a fortiori, weakly stationary. Since Z_t is a function of X_t the same is true for the process (Z_t). The unconditional or stationary distribution of X_t is the normal distribution $\mathcal{N}(\mu^u, \Sigma^u)$ where:

$$\mu^u = (I - \Phi)^{-1}\mu \quad \text{and} \quad \Sigma^u = \Phi\Sigma^u\Phi' + \Sigma \tag{4}$$

Equivalently, we can write $Vec(\Sigma^u) = (I - \Phi \otimes \Phi)^{-1} Vec(\Sigma)$. The stationary distribution of Z_t is the image of $\mathcal{N}(\mu^u, \Sigma^u)$ by the function f defined by $f(x) = (x', Vec(xx')')'$. In order to initialize our filter, we need the first two moments of this stationary distribution, that is to say the unconditional expectation $\mathbb{E}(Z_t)$ and the unconditional variance-covariance matrix $\mathbb{V}(Z_t)$ of Z_t.

Proposition 3.1 gives the expressions of the conditional moments of Z_t given Z_{t-1}, namely $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$. In general, the sole knowledge of these conditional moments does not allow to compute the unconditional moments $\mathbb{E}(Z_t)$ and $\mathbb{V}(Z_t)$. However, it is important to note that, here, the affine forms of $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$ make these computations feasible analytically. More precisely, starting from any value Z_0 of Z_t at $t = 0$, the sequence $[\mathbb{E}(Z_t)', Vec(\mathbb{V}(Z_t))']'$, for $t = 1, 2, \ldots$ satisfies a first-order linear difference equation defined in the following proposition.

Proposition 3.3 *We have:*

$$\begin{bmatrix} \mathbb{E}(Z_t) \\ Vec[\mathbb{V}(Z_t)] \end{bmatrix} = \begin{pmatrix} \widetilde{\mu} \\ \nu \end{pmatrix} + \Xi \begin{bmatrix} \mathbb{E}(Z_{t-1}) \\ Vec[\mathbb{V}(Z_{t-1})] \end{bmatrix} \quad \textit{where} \quad \Xi = \begin{pmatrix} \widetilde{\Phi} & 0 \\ \Psi & \widetilde{\Phi} \otimes \widetilde{\Phi} \end{pmatrix}. \tag{5}$$

where $\widetilde{\mu}$ and $\widetilde{\Phi}$ are defined in Proposition 3.1, and ν, Ψ are defined according to Proposition 3.2.

Proof See Appendix A.5. ∎

This linear difference equation is convergent since all the eigenvalues of Ξ have a modulus strictly smaller than 1. This is easily verified: Ξ is block triangular thus its eigenvalues are the eigenvalues of $\widetilde{\Phi}$ and $\widetilde{\Phi} \otimes \widetilde{\Phi}$. Using the same argument, $\widetilde{\Phi}$ has the same eigenvalues as Φ and $\Phi \otimes \Phi$ (see Proposition 3.1). Moreover the eigenvalues of the Kronecker product of two square matrices are given by all the possible products of the first and second matrices eigenvalues. Therefore, since Φ has eigenvalues inside the unit circle, so have $\widetilde{\Phi}$, $\widetilde{\Phi} \otimes \widetilde{\Phi}$, and Ξ.

We deduce that the unconditional expectation $\widetilde{\mu}^u$ and variance-covariance $\widetilde{\Sigma}^u$ (or rather $Vec(\widetilde{\Sigma}^u)$) of Z_t are the unique solutions of:

$$\begin{pmatrix} \widetilde{\mu}^u \\ Vec\left(\widetilde{\Sigma}^u\right) \end{pmatrix} = \begin{pmatrix} \widetilde{\mu} \\ \nu \end{pmatrix} + \begin{pmatrix} \widetilde{\Phi} & 0 \\ \Psi & \widetilde{\Phi} \otimes \widetilde{\Phi} \end{pmatrix} \begin{pmatrix} \widetilde{\mu}^u \\ Vec\left(\widetilde{\Sigma}^u\right) \end{pmatrix}. \tag{6}$$

We get the following corollary:

Corollary 3.3.1 *The unconditional expectation $\widetilde{\mu}^u$ and variance-covariance $\widetilde{\Sigma}^u$ of Z_t are given by:*

$$\begin{aligned} \widetilde{\mu}^u &= \left(I_{n(n+1)} - \widetilde{\Phi}\right)^{-1} \widetilde{\mu} \\ Vec\left(\widetilde{\Sigma}^u\right) &= \left(I_{n^2(n+1)^2} - \widetilde{\Phi} \otimes \widetilde{\Phi}\right)^{-1} (\nu + \Psi \widetilde{\mu}^u) \\ &= \left(I_{n^2(n+1)^2} - \widetilde{\Phi} \otimes \widetilde{\Phi}\right)^{-1} Vec\left[\widetilde{\Sigma}\left(\widetilde{\mu}^u\right)\right], \end{aligned}$$

where $\widetilde{\mu}$ and $\widetilde{\Phi}$ are defined in Proposition 3.1.

These closed-form expressions of $\widetilde{\mu}^u$ and $\widetilde{\Sigma}^u$ will make easy the initialization of our algorithms. Note that the computation of $Vec[\widetilde{\Sigma}(\widetilde{\mu}^u)]$ requires the explicit affine expression of Appendix A.4 given by $Vec[\widetilde{\Sigma}(\widetilde{\mu}^u)] = \nu + \Psi \widetilde{\mu}^u$.

3.4 Conditionally Gaussian approximation of (Z_t)

Proposition 3.3 shows that Z_t satisfies:

$$Z_t = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1} + \widetilde{\Omega}(Z_{t-1})\xi_t, \tag{7}$$

where $\widetilde{\Omega}(Z_{t-1})$ is such that $\widetilde{\Omega}(Z_{t-1})\widetilde{\Omega}(Z_{t-1})' = \widetilde{\Sigma}(Z_{t-1})$ and (ξ_t) is a martingale difference process, with a unit conditional variance-covariance matrix (i.e. $\mathbb{E}_{t-1}(\xi_t) = 0$ and $\mathbb{V}_{t-1}(\xi_t) = I_{n(n+1)}$). In the sequel, we approximate the process (ξ_t) by a Gaussian white noise. In the standard case where μ, Φ and Σ are time-invariant, the process Z_t^*, $t = 0, 1, \ldots$, defined by $Z_0^* \sim \mathcal{N}(\widetilde{\mu}^u, \widetilde{\Sigma}^u)$ and

$$Z_t^* = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1}^* + \widetilde{\Omega}(Z_{t-1}^*)\xi_t^*,$$

where (ξ_t^*) is a standard Gaussian white noise, has exactly the same second-order properties as process (Z_t). This statement is detailed in Proposition 3.4.

Proposition 3.4 *If μ, Φ and Σ are time-invariant, the processes Z_t and Z_t^* have the same second-order properties, i.e. the same means, variances, instantaneous covariances, serial correlations, and serial cross-correlations.*

Proof It is easy to check that, for both processes, the mean, variance-covariance matrix, and lag-h covariance matrix are respectively $\widetilde{\mu}^u$, $\widetilde{\Sigma}^u$ and $\widetilde{\Phi}^h\widetilde{\Sigma}^u$. ■

3.5 The filtering algorithm

Using the augmented state vector Z_t we can rewrite the state-space model of Definition 3.1 as an augmented state-space model.

Definition 3.2 *The augmented state-space model associated with the linear-quadratic state-space model is defined by:*

$$\begin{cases} Z_t &= \widetilde{\mu} + \widetilde{\Phi} Z_{t-1} + \widetilde{\Omega}_{t-1}\xi_t, \\ Y_t &= A + \widetilde{B} Z_t + D\eta_t, \end{cases} \tag{8}$$

where η_t, A, and D are defined as in Definition 3.1, $\widetilde{\Omega}_{t-1}$ is such that $\widetilde{\Omega}_{t-1}\widetilde{\Omega}'_{t-1} = \widetilde{\Sigma}_{t-1}$, and $\widetilde{\mu}$, $\widetilde{\Phi}$, are defined as in Proposition 3.1. Matrix $\widetilde{B} \in \mathbb{R}^{m\times n(n+1)}$ is:

$$\widetilde{B} = \left[\begin{array}{c|c} & Vec\left[C^{(1)}\right]' \\ B & \vdots \\ & Vec\left[C^{(m)}\right]' \end{array}\right]$$

Approximating the process (ξ_t) by a standard Gaussian white-noise and noting that the transition and measurement equations in Formula (8) are respectively linear in Z_{t-1} and Z_t, the resulting state-space model is linear Gaussian. Whereas numerous existing filters rely on an approximation of the conditional distribution of X_t given $\underline{Y_{t-1}}$ (see e.g. the EKF and UKF in the next section), the QKF builds on an approximation of the conditional distribution of Z_t given $\underline{Z_{t-1}}$ or, equivalently, of Z_t given $\underline{X_{t-1}}$. Proposition 3.4 shows that this approximation is exact up to the second order. The conditional variance-covariance matrix of the transition noise, i.e. $\widetilde{\Omega}_{t-1}\widetilde{\Omega}'_{t-1} = \widetilde{\Sigma}_{t-1}$, is a linear function of Z_{t-1} (see Proposition 3.2), which will be replaced in the standard linear Kalman filter by $\widetilde{\Sigma}(Z_{t-1|t-1})$. At each iteration, we emphasize that this computation should always be made using the formulae of Proposition 3.2 where the affine forms in Z_{t-1} are made completely explicit (see the discussion below Proposition 3.1). Finally, we get the Quadratic Kalman Filter algorithm displayed in Table 1.

Table 1: Quadratic Kalman Filter (QKF) algorithm

Initialization:		$Z_{0\|0} = \widetilde{\mu}^u$ and $P^Z_{0\|0} = \widetilde{\Sigma}^u$.
State prediction:	$Z_{t\|t-1}$	$\widetilde{\mu} + \widetilde{\Phi} Z_{t-1\|t-1}$
	$P^Z_{t\|t-1}$	$\widetilde{\Phi} P^Z_{t-1\|t-1} \widetilde{\Phi}' + \widetilde{\Sigma}(Z_{t-1\|t-1})$
Measurement prediction:	$Y_{t\|t-1}$	$A + \widetilde{B} Z_{t\|t-1}$
	$M_{t\|t-1}$	$\widetilde{B} P^Z_{t\|t-1} \widetilde{B}' + V$
Gain:	K_t	$P^Z_{t\|t-1} \widetilde{B}' M^{-1}_{t\|t-1}$
State updating:	$Z_{t\|t}$	$Z_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^Z_{t\|t}$	$P^Z_{t\|t-1} - K_t M_{t\|t-1} K_t'$

Note: $\widetilde{\mu}^u$ and $\widetilde{\Sigma}^u$ are respectively the unconditional mean and variance of process Z_t (that are given in Corollary 3.3.1). Note that the implied value of $[(XX')_{t|t} - X_{t|t}X'_{t|t}]$, that is a covariance matrix, should be a non-negative matrix. When it is not the case, we replace its negative eigenvalues by 0 and recompute $(XX')_{t|t}$ accordingly. Such a correction is not needed in the state-prediction step: indeed, using the expression of matrices $\widetilde{\mu}$ and $\widetilde{\Phi}$, we get that $[(XX')_{t+1|t} - X_{t+1|t}X'_{t+1|t}]$ $=\Phi((XX')_{t|t} - X_{t|t}X'_{t|t})\Phi + \Sigma$, which is then positive.

Starting the algorithm at $t = 1$, we need the initial values $Z_{0|0}$ and $P^Z_{0|0}$. As emphasized previously, one can take the unconditional moments $Z_{0|0} = \widetilde{\mu}^u$ and $P_{0|0} = \widetilde{\Sigma}^u$. Note that using Equations (6), we have $\widetilde{\mu}^u = \widetilde{\mu} + \widetilde{\Phi}\mu^u$ and $\widetilde{\Sigma}^u = \widetilde{\Phi}\widetilde{\Sigma}^u\widetilde{\Sigma}' + \widetilde{\Sigma}(\mu^u)$ and, therefore, $Z_{1|0} = Z_{0|0}$, $P^Z_{1|0} = P^Z_{0|0}$. In other words we can also start the algorithm by the prediction of Y_t, for $t = 1$, using the initial values $Z_{1|0} = \widetilde{\mu}^u$ and $P^Z_{1|0} = \widetilde{\Sigma}^u$. Note that, in the filtering algorithm, the $n(n+1)$-dimensional vector Z_t could be replaced by the smaller vector $[X_t', Vech(X_tX_t)']'$ of size $n(n+3)/2$. This transformation barely changes the augmented state space model, premultiplying the lower block of Z_t by a selection matrix H_n such that $Vech(X_tX_t') = H_n Vec(X_tX_t')$. The formal definition of the selection matrix is given in Appendix A.6. The computation of conditional moments using $Vech$ is thus straightforward.

3.6 The smoothing algorithm

Contrary to most existing non-linear filters, that are presented in the next section, our QKF approach has a straightforward smoothing extension. Indeed, since our basic state-space model is linear, we just have to use the standard backward fixed-interval algorithm. Note however that the variance-covariance matrices $P^Z_{t+1|t}$ computed with the filtering algorithm using $Vec(\bullet)$ are not of full-rank since at least one component of Z_t is redundant when $n \geqslant 2$. Consequently, the smoothing algorithm must be expressed with the $Vech(\bullet)$ operator. Let us introduce the following matrices:

$$\widetilde{H}_n = \begin{pmatrix} I_n & 0 \\ 0 & H_n \end{pmatrix} \quad \text{and} \quad \widetilde{G}_n = \begin{pmatrix} I_n & 0 \\ 0 & G_n \end{pmatrix},$$

that are respectively the $\frac{n(n+3)}{2} \times n(n+1)$ and $n(n+1) \times \frac{n(n+3)}{2}$ matrices using the selection and duplication matrices H_n and G_n defined in Appendix A.6. We have:

$$Vec(X_t X_t') = G_n Vech(X_t X_t') \quad \text{and} \quad Vech(X_t X_t') = H_n Vec(X_t X_t')$$

$\widetilde{H}_n$ is defined such that $\widetilde{H}_n Z_t = [X_t', Vech(X_t X_t')']'$. The sandwich multiplication $\widetilde{H}_n P^Z_{t+1|t} \widetilde{H}_n'$ drops the redundant rows and columns. We get the following smoothing algorithm:

$$\begin{aligned}
F_t &= \left(\widetilde{H}_n P^Z_{t|t} \widetilde{H}_n'\right) \left(\widetilde{H}_n \widetilde{\Phi} \widetilde{G}_n\right)' \left(\widetilde{H}_n P^Z_{t+1|t} \widetilde{H}_n'\right)^{-1} \\
\widetilde{H}_n Z_{t|T} &= \widetilde{H}_n Z_{t|t} + F_t \left(\widetilde{H}_n Z_{t+1|T} - \widetilde{H}_n Z_{t+1|t}\right) \\
\left(\widetilde{H}_n P^Z_{t|T} \widetilde{H}_n'\right) &= \left(\widetilde{H}_n P_{t|t} \widetilde{H}_n'\right) + F_t \left[\left(\widetilde{H}_n P^Z_{t+1|T} \widetilde{H}_n'\right) - \left(\widetilde{H}_n P^Z_{t+1|t} \widetilde{H}_n'\right)\right] F_t'
\end{aligned}$$

The initial values $Z_{T|T}$ and $P^Z_{T|T}$ are obtained from the filtering algorithm.

4 Performance comparisons using Monte Carlo experiments

We simulate a linear-quadratic state-space model and compare the performance of the QKF filter against other popular non-linear filters. We distinguish two exercises, namely filtering and parameter estimation.

4.1 Usual non-linear filters

Among the popular non-linear filters, two main classes of algorithms are widely used: the extended Kalman filter (EKF) and the unscented Kalman filter (UKF). Both approximate the non-linear measurement or transition equations using linearization techniques but their spirit differ radically. This section presents these algorithms applied to the linear-quadratic state-space model of Definition 3.1. They will further be used as competitors compared to the QKF in the performance assessment.

Two versions of the EKF have been used, namely the first and second order – Gaussian – filters. Their derivations are respectively based on first- and second-order Taylor expansions of the measurement equations around $X_{t|t-1}$ at each iteration. For simplicity, we use the following notations:

$$\begin{aligned}
h(X_t) &\equiv A + B X_t + \sum_{k=1}^{m} e_k X_t' C^{(k)} X_t \\
\mathcal{G}_{t|t-1} &\equiv \frac{\partial h}{\partial X_t'}(X_{t|t-1}) = B + 2 \sum_{k=1}^{m} e_k X_{t|t-1}' C^{(k)}
\end{aligned}$$

Table 2 details both EKF algorithms in the quadratic measurement case.[13] A general non-linear

[13] Another version of the second order filter called the *truncated* second-order filter is presented in Maybeck (1982). However, it makes the assumption that the third and higher-order conditional moments of X_t given $\underline{Y_{t-1}}$ are sufficiently small to be negligible and set to 0. As a consequence, the calculation of $M_{t|t-1}$ in this algorithm can yield non positive-definite matrices showing far less computational stability than the Gaussian second-order extended

version is provided in Appendix A.7 (see also Jazwinski (1970) and Anderson and Moore (1979) for the EKF1, and Athans, Wishner, and Bertolini (1968) or Maybeck (1982) for the EKF2).

Table 2: EKF algorithms in the quadratic case

		EKF1	EKF2
Initialization:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$	
State prediction:	$X_{t\|t-1}$	$\mu + \Phi X_{t-1\|t-1}$	
	$P^X_{t\|t-1}$	$\Phi P^X_{t-1\|t-1} \Phi' + \Sigma$	
Measurement prediction:	$Y_{t\|t-1}$	$h(X_{t\|t-1})$	$h(X_{t\|t-1}) + \sum_{k=1}^{m} e_k \mathrm{Tr}\left(P^X_{t\|t-1} C^{(k)}\right)$
	$M_{t\|t-1}$	$\mathcal{G}_{t\|t-1} P^X_{t\|t-1} \mathcal{G}'_{t\|t-1}$ $+V$	$\mathcal{G}_{t\|t-1} P^X_{t\|t-1} \mathcal{G}'_{t\|t-1} + V$ $+2 \sum_{k,j=1}^{m} e_k e'_j \mathrm{Tr}\left(C^{(k)} P^X_{t\|t-1} C^{(j)} P^X_{t\|t-1}\right)$
Gain:	K_t	$P^X_{t\|t-1} \mathcal{G}'_{t\|t-1} M^{-1}_{t\|t-1}$	
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t (Y_t - Y_{t\|t-1})$	
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K'_t$	

Note: See above for the definition of $\mathcal{G}_{t|t-1}$ and $h(x)$.

In the EKF1 algorithm, both $Y_{t|t-1}$ and $M_{t|t-1}$ are grossly approximated, whereas the EKF2 incorporates the so-called *bias correction terms* which are expected to reduce the error on these moments evaluation (see fourth and fifth rows of Table 2). Even if the Taylor expansion of the measurement equation is exact in the EKF2, it implicitly approximates the conditional distribution of (Y_t, X_t) given $\underline{Y_{t-1}}$ by a Gaussian distribution, which also induces errors in the recursions.

In comparison, the UKF belongs to the class of *density-based* filters and uses a set of vectors called *sigma points.*[14]

Definition 4.1 *Let $X \in \mathbb{R}^n$ a random vector and define $m = \mathbb{E}(X)$ and $P = \mathbb{V}(X)$. Let $(\sqrt{P})_i$ denote the i^{th} column of the lower-triangular Cholesky decomposition of P. The sigma set associated with X is composed of $2n+1$ sigma points $(\mathcal{X}_i(m, P))_{i=\{0,\dots,2n\}}$ and 2 sets of $2n+1$ weights $(\mathcal{W}_i)_{i=\{0,\dots,2n\}}$ and $(\mathcal{W}^{(c)}_i)_{i=\{0,\dots,2n\}}$ defined by:*

filter. We thus left it aside in our comparison exercise. Also, higher-order extended filters can be derived with statistical linearization techniques, but are rarely used in practice (see Gelb, Kasper, Nash, Price, and Sutherland (1974)).

[14] The name *density-based filter* belongs to the terminology of Tanizaki (1996).

$$\mathcal{X}_i = \begin{cases} m & \textit{for} \quad i = 0 \\ m + \left(\sqrt{(n+\lambda)P}\right)_i & \textit{for} \quad i \in [\![1, n]\!] \\ m - \left(\sqrt{(n+\lambda)P}\right)_{i-n} & \textit{for} \quad i \in [\![n+1, 2n]\!] \end{cases} \quad \begin{aligned} \mathcal{W}_i &= \begin{cases} \lambda/(\lambda+n) & \textit{for} \quad i = 0 \\ 1/[2(\lambda+n)] & \textit{for} \quad i \neq 0 \end{cases} \\ \mathcal{W}_i^{(c)} &= \begin{cases} \mathcal{W}_i + 1 - \alpha^2 + \beta & \textit{for} \quad i = 0 \\ \mathcal{W}_i & \textit{for} \quad i \neq 0, \end{cases} \end{aligned}$$

where (α, κ, β) is a vector of tuning parameters and $\lambda = \alpha^2(n+\kappa) - n$. It is easy to see that for any (α, κ, β) we have:

$$\sum_{i=0}^{2n} \mathcal{W}_i \mathcal{X}_i = m \quad \textit{and} \quad \sum_{i=0}^{2n} \mathcal{W}_i \left(\mathcal{X}_i - m\right)\left(\mathcal{X}_i - m\right)' = \sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left(\mathcal{X}_i - m\right)\left(\mathcal{X}_i - m\right)' = P$$

The sigma set of Definition 4.1 is then used to approximate the moments of the non-linear transformation $h(X)$. The algorithm in the quadratic measurement equation case is given in Table 3. A general non-linear version is also provided in Appendix A.7.[15]

Table 3: UKF algorithm in the quadratic case

Initialization:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$ and choose (α, κ, β).
State prediction:	$X_{t\|t-1}$	$\mu + \Phi X_{t-1\|t-1}$
	$P^X_{t\|t-1}$	$\Phi P^X_{t-1\|t-1} \Phi' + \Sigma$
Sigma points:		$\left\{\mathcal{X}_{i,t\|t-1}(X_{t\|t-1}, P^X_{t\|t-1})\right\}_{i=\{1,\ldots,2n\}}$ according to Definition 4.1.
Measurement prediction:	$Y_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i \, h(\mathcal{X}_{i,t\|t-1})$
	$M_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[h(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right] \left[h(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' + V$
Gain:	K_t	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[\mathcal{X}_{i,t\|t-1} - X_{t\|t-1}\right] \left[h(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' M^{-1}_{t\|t-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K_t'$

Note: Weights $\mathcal{W}_i$ and $\mathcal{W}_i^{(c)}$ are given in Definition 4.1.

The tuning parameters (α, κ, β) are set by the user and depend on the applied filtering problem specificities (dimension size n, number of periods T, and prior knowledge on distributions). Usual values when the distribution of X_t given $\underline{Y_{t-1}}$ is assumed Gaussian are $\beta = 2$, $\kappa = 3 - n$ or 0, and $\alpha = 1$ for low dimensional problems.

[15] For an extensive description of the unscented Kalman filter, see Julier, Uhlmann, and Durrant-Whyte (2000), Julier (2002), or Julier and Uhlmann (2004), and applications in Kandepu, Foss, and Imsland (2008), or Christoffersen, Dorion, Jacobs, and Karoui (2012). For the square-root version, see Merwe and Wan (2001) or Holmes, Klein, and Murray (2008) for a square-root filtering application.

4.2 A simple example

To emphasize the specificity of the QKF compared to both EKFs and UKF, let us consider a very simple state-space model where analytical computations are feasible. Assume that $X_t = \varepsilon_t \sim \mathcal{IIN}(0, \sigma_\varepsilon^2)$. The measured univariate Y_t is given by $Y_t = X_t^2$ and is perfectly measured without noise or, equivalently, the noise is infinitely small. The natural method to retrieve X_t from Y_t is straightforward inverting the previous formula. The only uncertainty remaining is the sign of $\pm\sqrt{Y_t}$ which is impossible to infer. In that model, the distribution of Y_t is a $\gamma(1/2, 2\sigma_\varepsilon^2)$ distribution, with mean and variance respectively given by σ_ε^2 and $2\sigma_\varepsilon^4$.[16]

We compute the filtering formulae of the four aforementioned filters and compare them. The results are presented in Table 4. Despite the simplicity of the model, the EKF1 is unable to reproduce the moments of Y_t (second column of Table 4). Both the QKF and the EKF2 give the exact formulation of Y_t moments, whereas the computation of $M_{t|t-1}$ for the UKF depends on the tuning parameters (α, κ, β) (see 3^{rd} and 4^{th} rows). More importantly, looking at the last-two rows of Table 4, we see that the QKF is the only filter to update the state variables correctly in the squared components, since the second component of $Z_{t|t}$ is exactly the observed Y_t. However, all filters including the QKF produce $X_{t|t} = 0$ for all periods. Therefore the QKF is the only considered filter to jointly *(i)* correctly reproduce Y_t first-two moments, and *(ii)* produce time-varying estimates of the latent factors. We systematize this comparison to different state-space models using simulations in the next section.

Table 4: Example: computation of filters' formulae

	QKF	EKF 1	EKF 2	UKF
$X_{t\|t-1}$ (or $Z_{t\|t-1}$ for the QKF)	$\begin{pmatrix} 0 \\ \sigma_\varepsilon^2 \end{pmatrix}$	0	0	0
$P^X_{t\|t-1}$ (or $P^Z_{t\|t-1}$ for the QKF)	$\begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & 2\sigma_\varepsilon^4 \end{pmatrix}$	σ_ε^2	σ_ε^2	σ_ε^2
$Y_{t\|t-1}$	σ_ε^2	0	σ_ε^2	σ_ε^2
$M_{t\|t-1}$	$2\sigma_\varepsilon^4$	0	$2\sigma_\varepsilon^4$	$(\alpha^2\kappa + \beta)\sigma_\varepsilon^4$
$X_{t\|t}$ (or $Z_{t\|t}$ for the QKF)	$\begin{pmatrix} 0 \\ Y_t \end{pmatrix}$	0	0	0
$P^X_{t\|t}$ (or $P^Z_{t\|t}$ for the QKF)	$\begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & 0 \end{pmatrix}$	σ_ε^2	σ_ε^2	σ_ε^2

Notes: The state-space model is defined by $X_t \sim \mathcal{IIN}(0, \sigma_\varepsilon^2)$ and $Y_t = X_t^2$. 'QKF' is the Quadratic Kalman filter, 'EKF 1' and 'EKF 2' are respectively the first- and second-order extended Kalman filters, 'UKF' is the unscented Kalman filter.

[16] Recall that the density of a $\gamma(k, \rho)$ is given by $f(x) = \frac{1}{\Gamma(k)\rho^k} x^{k-1} \exp(-x/\rho)$.

4.3 Comparison of filtering performance

We compare the filtering performance of the QKF against the EKF 1 and EKF 2, and the UKF in a linear-quadratic state-space model. We parameterize state-space model as follows:

$$\begin{aligned} X_t &= \Phi X_{t-1} + \varepsilon_t \\ Y_t &= \sqrt{\theta_2(1-\theta_1)}\sqrt{1-\Phi^2}\, X_t + \sqrt{(1-\theta_2)(1-\theta_1)}\frac{1-\Phi^2}{\sqrt{2}}\, X_t^2 + \sqrt{\theta_1}\eta_t \end{aligned} \tag{9}$$

where both ε_t and η_t are zero-mean normalized Gaussian white-noises, and both X_t and Y_t are scalar variables ($n = m = 1$). Comparing with Equations (1a) and (1b), we have set $\mu = 0$ and $A = 0$ for simplicity. It is straightforward to see that the unconditional variance of Y_t is equal to 1. Therefore, the weights $(\theta_1, \theta_2) \in [0,1]^2$, should be interpreted in the following way: θ_1 is the proportion of Y_t variance explained by the measurement noise, the rest (i.e. $1-\theta_1$) being explained by the state variables in the measurement equation. θ_2 is the proportion of the variance of Y_t explained by the linear term, within the part explained by the state variables.

The performance of the different filters are assessed with respect to values of Φ, θ_1 and θ_2. We successively set $\Phi = \{0.3, 0.6, 0.9, 0.95\}$ controlling from low to very high persistence of X_t process, $\theta_1 = \{0.2, 0.25, 0.3, \ldots, 0.8\}$ and $\theta_2 = \{0, 0.25, 0.5, 0.75\}$ (for a total of 208 cases). For instance, a combination of $(\theta_1, \theta_2) = (0.2, 0.25)$ should be interpreted as *20% of Y_t variance can be attributed to the measurement noise and 80% to the latent factors, of which 25% is attributed to the linear term and 75% to the quadratic term.*[17] Degenerated cases where either $\theta_1 = 0$, or $\theta_1 = 1$ are not considered (they correspond respectively to situations with no measurement noise or no explanatory variables in the measurement equation). Also, the case where $\theta_2 = 1$ is left aside as the measurement equation becomes linear, and all the considered filters boil down to the linear Kalman filter.[18] For each value of Φ, we simulate paths of the latent process X_t of $T = 1{,}000{,}000$ periods with a starting value of $X_0 = 0$. We then simulate the measurement noises η_t and compute implied observable variables Y_t for each combination of (θ_1, θ_2). The filtering exercise is performed for each filter, initial values being known.[19] For the UKF, we set $\alpha = 1$, and $\beta = 2$ as in Christoffersen, Dorion, Jacobs, and Karoui (2012). For those values of (α, β) and scalar processes, it can be shown that $\kappa = 0$ implies the exact same recursions as the EKF2.[20] We therefore set $\kappa = 3 - n = 2$.

We denote by $\widehat{X_{t|t}}$, $\widehat{X^2_{t|t}}$ and $\widehat{P_{t|t}}$ the filtered values resulting from any filtering algorithm. The different filters are compared with respect to three measures of performance. First, we compute the RMSEs of filtered values $\widehat{X_{t|t}}$ compared to X_t. Second, we calculate RMSEs of the quadratic process $\widehat{X^2_{t|t}}$. Whereas the QKF evaluates this quantity directly in the algorithm, we recompute its underlying value for the other filters with the formula $\widehat{X^2_{t|t}} = \widehat{X_{t|t}}^2 + \widehat{P_{t|t}}$. The RMSE measures

[17] Note that in the general quadratic models that we consider here, we have $Cov(X_t, Vec(X_t X_t')) = 0$.
[18] This is in fact not obvious for the UKF, and the proof is provided in Appendix A.8.
[19] Thus we set, $X_{0|0} = 0$ and $P^X_{0|0} = 0$ for the EKFs and UKF, and $Z_{0|0} = 0_{\mathbb{R}^2}$ and $P^Z_{0|0} = 0_{\mathbb{R}^{2\times 2}}$ for the QKF.
[20] See Appendix A.9 for a proof.

for any of our estimated values are normalized by the standard deviation of the simulated process:

$$\overline{RMSE}_W = \frac{RMSE_W}{\sigma_W} = \left[\frac{T^{-1}\sum_{t=1}^{T}(W_t - \widehat{W_{t|t}})^2}{\mathbb{V}(W_t)}\right]^{1/2}$$

where $W_t = X_t$ or X_t^2 and $\widehat{W_{t|t}} = \widehat{X_{t|t}}$ or $\widehat{X^2_{t|t}}$. This measure converges to 1 if the filtered values are equal to the unconditional mean of the latent process for all periods. Consequently, if any filter yields a normalized RMSE greater than 1, a better filtering result would be obtained by setting $W_{t|t} = \mathbb{E}(W_t)$, for all t. Lastly, we compare the filters capacities to discriminate between the explanatory process and the measurement noise by computing non-normalized RMSEs of implied $\widehat{\eta}_t$. The results are respectively presented on Figures 1, 2, and 3.

[Insert Figures 1, 2, and 3 about here.]

Result 1 *When the measurement equation is fully quadratic ($\theta_2 = 0$), The* QKF *is the only considered filter capable of both:*

(i) Filtering out a substantial part of the measurement noise,

(ii) Yielding accurate evaluations of $X^2_{t|t}$.

We first analyse the case where the measurement equation is only quadratic ($\theta_2 = 0$, left column of all figures). As already noted for a specific case in the previous section, all filters are "blind" on the evaluation of $X_{t|t}$ producing a flat $\widehat{X_{t|t}} = 0$, and normalized RMSEs are equal to 1 whatever the values of Φ and θ_1 (see Figure 1, left column). However, looking at Figure 2, we see that for any relative size of the measurement errors and any persistence, the QKF yields more accurate evaluations of $X^2_{t|t}$ than the other filters, showing 5% to 60% smaller RMSEs depending on the case. Two patterns can be observed here. First, the smaller the measurement errors, the stronger the outperformance of the QKF filter compared to the others. Second, the outperformance of the QKF increases with the persistence of the latent process.[21] This better performance is confirmed by looking at the evaluation of the measurement noise, where the QKF also provides the smallest RMSEs for all values of (Φ, θ_1) (see Figure 3, first column). The reduction in the measurement noise RMSEs for the QKF compared to the others can reach 70%. This result emphasizes the substantial improvement of the fitting properties of the QKF compared to those of the other filters.

Result 2 *For measurement equations where the linearity degree goes from 25% to 50%, the* QKF *beats the other filters, especially for the evaluation of $X^2_{t|t}$. Eventually, for levels of about 75% of linearity in the measurement equation, the RMSEs of all filters converge to the same values.*

[21] We see this as a pleasant feature for term-structure modelling applications where yields are typically highly persistent and measured with low errors.

We turn now to the cases where the measurement equation has from 25% to 50% of linearity degree ($\theta_2 = \{0.25, 0.5\}$, second and third columns of all figures). We first leave aside the EKF 1 (see result 3). For $\widehat{X_{t|t}}$, normalized RMSEs are more or less the same for the EKF 2 and the UKF in all cases. In comparison, the QKF is either equivalent, either showing smaller RMSEs for high-persistent cases ($\Phi = 0.9$ or $\Phi = 0.95$, third and fourth rows of Figure 1). This better performance is confirmed when looking at Figure 2. In all cases, the QKF possesses lowest RMSEs for $\widehat{X^2_{t|t}}$. For example, for $\Phi = 0.9$, $\theta_1 = 0.2$ and $\theta_2 = 0.25$, the QKF shows RMSEs slightly below 60% of X_t^2 standard deviation whereas the others are all above 70% (see Figure 2, third row of panel (b)). Unsurprisingly, this evidence places the QKF ahead of its competitors for the de-noising exercise: for panels (b) and (c) of Figure 3, RMSEs of $\widehat{\eta}_t$ are always below the others for the QKF. Looking at panel (d) where the measurement equation is 75% linear (fourth column of all figures), we see that all RMSEs eventually converge to each other for all filters. This is consistent with the fact that all filters reduce to the standard Kalman filter when the measurement equation is fully linear.

Result 3 *The* EKF 1 *should be discarded for filtering, especially when the variance of the measurement errors is low (cases where θ_1 is low).*

Looking at Figures 1 and 2, we notice a very unpleasant behaviour of the EKF 1. For low measurement errors, RMSEs of both $\widehat{X_{t|t}}$ and $\widehat{X^2_{t|t}}$ can reach values greater than 1, especially in panels (b) and (c) where the measurement equation shows medium linearity degree (see second and fourth columns of Figures 1 and 2). This catastrophic performance can be particularly observed for low persistence, low linearity degree, and low measurement errors: when $\Phi = 0.3$, $\theta_1 = 0.2$ and $\theta_2 = 0.25$, $\widehat{X_{t|t}}$ and $\widehat{X^2_{t|t}}$ show respectively 120% and 200% normalized RMSE values. That is to say filtered values yielded by the EKF 1 prove to be very poor in some cases.

This Monte-Carlo experiment provides evidence that in terms of filtering, the QKF largely dominates both EKFs and the UKF for evaluating X_t and X_t^2, as well as for de-noising the observable y_t. This is particularly the case when the degree of linearity in the measurement equation is low. Increasing the degree of linearity produces closer RMSEs for the QKF, the EKF2, and the UKF ; the EKF1 shows a very unstable behaviour. In the next section, we explore the characteristics of the different techniques in terms of parameter estimation.

4.4 Pseudo maximum likelihood parameter estimation

To compare the filters with respect to parameter estimation, we simulate the same benchmark model given in Equation (9). We estimate the vector of parameter $\beta = (A, B, C, \Phi, \sigma_\eta)$ for some specific values of (Φ, θ_1, θ_2). To explore the finite sample properties of the different estimators, we set $T = 200$ and simulate 1000 dynamics for a given set of (Φ, θ_1, θ_2). This provides us with the empirical marginal distributions of the estimators. As usual in non-linear filter estimation, the technique is only pseudo-maximum likelihood as the distribution of Y_t given $\underline{Y_{t-1}}$ is approximated as a Gaussian.[22]

To avoid local maxima, a two-step estimation is performed. First, a stochastic maximization algorithm is launched to select a potential zone for the global maximum. Second, a simplex algorithm is used to refine the estimates in the selected zone.[23] This procedure makes the results reliable at the cost of extended computational burden. This particular reason leads us to select three paradigmatic cases for the simulated processes. The first considered case is fully quadratic with high persistence and low measurement error variance ($\Phi = 0.9$, $\theta_1 = 0.05$, and $\theta_2 = 0$). In the second case, we decrease the persistence of the latent process and increase the size of measurement errors setting $\Phi = 0.6$, $\theta_1 = 0.2$, and keeping $\theta_2 = 0$. In the last case we introduce a linear component in the measurement equation, with the parametrization: $\Phi = 0.6$, $\theta_1 = 0.2$, and $\theta_2 = 0.25$. More linear cases ($\theta_2 > 0.25$) were not considered as we emphasized in Section 4.3 that the four filters yield closer results in those cases. For identification purpose, we also impose $\hat{B} > 0$. Results in terms of bias, standard errors, and RMSEs are presented in Table (8). Comparisons of filters on average across panels are provided in Table 9.

$\left[\text{Insert Table (8) about here.} \right]$

Result 4 *The* QKF *pseudo maximum likelihood estimates are either the less biased, either possess the lowest RMSEs for all parameters. In addition, on average across panels, the* QKF *is the less biased filter and possesses lowest RMSEs. This superiority is robust to the degree of persistence of the latent process, to the degree of linearity of the measurement equation, and to the size of the measurement errors.*

Over the three panels, the results of Tables 8 and 9 are in favour of our QKF maximum likelihood estimates. We first concentrate on panel (a) results. For the five estimated parameters, the QKF shows smaller bias than the other filters: for $\widehat{A}$, $\widehat{B}$, and $\widehat{\Phi}$, the bias of the QKF estimates corresponds to half the bias of the EKF 2 and the UKF. In addition, for four out of the five parameters, the QKF estimates yield smaller RMSEs even though it often entails higher standard deviation than its competitors (see Table 8, panel (a)). The same general pattern can be observed for panel (b), where persistence degree is smaller. Consistently with the intuition, the QKF always outperforms its competitors for estimating parameters B and C. This shows a better capacity to discriminate the influence of linear and quadratic terms in the observable. While panel (c) introduces some linearity in the measurement equation ($B \neq 0$), the QKF still beats the other filters for four (resp. three) out of five parameters in terms of bias (resp. RMSEs). In the end, looking at Table 9, we observe the superiority of the QKF across all cases: on 13 (resp. 11) out

[22] It should be noted here that none of the filters produce exact first-two conditional moments of Y_t given $\underline{Y_{t-1}}$. The asymptotic properties of the pseudo maximum likelihood are therefore not relevant.

[23] The stochastic algorithm used is the articificial bee colony of Karaboga and Basturk (2007). In order to limit computational time, we consider 256 people in the population, and only 10 iterations. Then, the best member of the population is selected as initial conditions for the Nelder-Mead algorithm.

of 15 parameters, the QKF estimates possess the lowest bias (resp. RMSEs) compared to the others.

[Insert Table (9) about here.]

Result 5 *On average across cases,*

- *The* QKF *never yields the worst bias or RMSEs of all filters.*
- *The* EKF 1 *estimates possess the largest RMSEs and standard deviations.*
- *The* UKF *estimates possess the lowest standard deviations, but are the most biased.*
- *The* EKF 2 *is rarely the best in terms of both bias, standard deviations and RMSEs, but is also rarely the worst.*

We turn now to comparing the average results of the different filters. Table 9 presents the number of times each filter is best and worst in terms of bias, standard deviations and RMSEs. We have already emphasized that the QKF estimates surpass the others on average in terms of bias and RMSEs. A striking feature presented in Table 9 is also that QKF estimates are never the most biased, neither possess the biggest RMSEs (see first column). Overall, these results underline a better bias/variance trade-off for the QKF compared to the other filters.

The results of Tables 8 and 9 also confirm the concerns about the EKF 1 performance: out of 15 estimates, 6 are the most biased, 10 possess the biggest standard deviations, and 9 possess the highest RMSEs. This poor performance is particularly observable for the estimation of C in panel (b) and (c) of Table 8: the standard deviations of the estimates are respectively 18 and 10, and their RMSEs are more than 10 times bigger than those of the other filters. This can be explained by the fact that the curvature of the EKF 1 log-likelihood along the C-axis is very close to zero. Hence the estimate $\widehat{C}$ can move a lot along the line with very little change in the log-likelihood. This corroborates the incapacity of the EKF 1 to deal with high non-linearities in the measurement equation, as already noted in the filtering performance comparison (see previous section).

Interestingly, the UKF also shows some concerning features for parameter estimation. It is the most biased for 8 parameters out of 15, which is the worst bias performance among all filters. However, it is also the filter that produces on average the smallest standard deviations for 9 parameters (see last column of Table 9). Looking at Table 8), we observe that those cases where the standard deviation is low tend to correspond to cases where the bias is highest. This bias/variance trade-off hands up being very poor regarding the RMSEs: the UKF is the best only once, and four times the worst out of the 15 parameters. Consequently, we argue that the use of the UKF should be made with caution in the linear-quadratic state-space model since it tends to result in parameter estimates that are "tightly" distributed around biased values.

Finally, the EKF 2 seems to yield better average results than both the EKF 1 (unsurprisingly) and the UKF: although it is never the less biased and possesses the lowest RMSE for only one estimate, it is also rarely the most biased or rarely shows the biggest RMSE (see Table 9). Still, those results are far less encouraging than those of the QKF and the latter should be preferred in linear-quadratic state-space model estimations.

On the whole, for most estimates, the QKF is less biased and possesses the lowest RMSEs. Despite a slightly poorer performance on the standard deviations, the QKF maximum likelihood estimates show a better bias/variance trade-off than its competitors. Also, the consideration of 3 different panels provide evidence that these results are neither altered by the degree of curvature in the measurement equation, nor by the persistence of the latent process or by the size of the measurement errors. These finite-sample estimation properties emphasize the superiority of the QKF for practical applications.

5 Conclusion

In this paper, we develop the quadratic Kalman filter (QKF), a fast and efficient technique for filtering and smoothing state-space models where the transition equations are linear and the measurement equations are quadratic. Building the *augmented vector of factors* stacking together the latent vector with its vectorized outerproduct, we provide analytical formulae of its first-two conditional and unconditional moments. With this new expression of the latent factors, we show that the state-space model can be expressed in a fully linear form with non-Gaussian residuals. Using this new formulation of the linear-quadratic state-space model, we adapt the linear Kalman filter to obtain the Quadratic Kalman Filter and Smoother algorithms (resp. QKF and QKS). Since no simulation is required in the computations, both QKF and QKS algorithms are computationally fast and stable. We compare performance of the QKF against the extended and unscented versions of the Kalman filter in terms of filtering and parameter estimation. Our results suggest that for both filtering and pseudo-maximum likelihood estimation, the QKF outperforms its competitors. For filtering, the higher the curvature of the measurement equation, the more effective the QKF compared to the other filters. For parameter estimation, the QKF shows either smaller bias or smaller RMSEs than its competitors.

A Appendix

A.1 Useful algebra

We detail hereby some properties of both the Kronecker product and the $Vec(\bullet)$ operator. Their proofs are available in Magnus and Neudecker (1988). These properties will be used extensively in the proofs presented in Appendices A.2, A.3 and A.4.

Proposition A.1 *Let m_1 and m_2 be two size-n vectors, M_1 and M_2 be two square matrices of size n. Let also P, Q, R, and S be four matrices with respective size $(p \times q)$, $(q \times r)$, $(r \times s)$, and $(s \times t)$. We have:*

(i) $Vec(m_1 m_2') = m_2 \otimes m_1$.

(ii) $Vec(M_1 \otimes M_2) = (I_n \otimes \Lambda_n \otimes I_n)\,[Vec(M_1) \otimes Vec(M_2)]$ *where Λ_n is defined in Lemma A.1.*

in particular: $Vec(M_1 \otimes m_1) = Vec(M_1) \otimes m_1$ *and* $Vec(M_1 \otimes m_1') = (I_n \otimes \Lambda_n)\,[Vec(M_1) \otimes m_1]$

(iii) $Vec(PQR) = (R' \otimes P)Vec(Q)$

(iv) $Vec(PQ) = (I_r \otimes P)Vec(Q) = (Q' \otimes I_p)Vec(P)$

(v) $(PQ) \otimes (RS) = (P \otimes R)(Q \otimes S)$.

A.2 Properties of the commutation matrix

Lemma A.1 *Let Λ_n be the $(n^2 \times n^2)$ commutation matrix partitioned in $(n \times n)$ blocks, whose (i,j) block is $e_j e_i'$. Let M_1 and M_2 be two square matrices of size n, and m be a vector of size $(n \times 1)$. We have:*

(i) $\Lambda_n = \sum_{i,j=1}^{n} (e_i e_j') \otimes (e_j e_i')$

(ii) *Λ_n is orthogonal and symmetric:* $\Lambda_n^{-1} = \Lambda_n' = \Lambda_n$

(iii) $\Lambda_n Vec(M_1) = Vec(M_1')$

(iv) $\Lambda_n(M_1 \otimes M_2)\Lambda_n = M_2 \otimes M_1$

(v) $\Lambda_n(M_1 \otimes m) = m \otimes M_1$.

Proof (i) Straightforward by definition.

(ii) Λ_n is symmetric:

$$\Lambda_n' = \sum_{i,j=1}^{n} (e_j e_i') \otimes (e_i e_j') = \Lambda_n.$$

Λ_n is orthogonal:

$$\Lambda_n \Lambda_n' = \sum_{i,j=1}^{n} [(e_i e_j') \otimes (e_j e_i')][(e_j e_i') \otimes (e_i e_j')] = \sum_{i,j=1}^{n} (e_i e_i') \otimes (e_j e_j') = I_{n^2}.$$

(iii)

$$\begin{aligned}
\Lambda_n Vec(M_1) &= \sum_{i,j=1}^{n} \left[(e_i e_j') \otimes (e_j e_i')\right] Vec(M_1) \\
&= \sum_{i,j=1}^{n} Vec\left[(e_j e_i') M_1 (e_i e_j')'\right] \\
&= \sum_{i,j=1}^{n} Vec\left[e_j M_1^{(i,j)} e_i'\right] = Vec(M_1').
\end{aligned}$$

(iv) By definition,

$$M_1 \otimes M_2 = \sum_{i,j=1}^{n} (M_1^{(,i)} \otimes M_2^{(,j)})(e_i' \otimes e_j'),$$

where $M_1^{(,i)}$ and $M_2^{(,j)}$ are respectively the i^{th} and j^{th} columns of matrices M_1 and M_2. Therefore we have:

$$\begin{aligned}
\Lambda_n(M_1 \otimes M_2)\Lambda_n &= \sum_{i,j=1}^{n} \Lambda_n(M_1^{(,i)} \otimes M_2^{(,j)})(e_i \otimes e_j)'\Lambda_n \\
&= \sum_{i,j=1}^{n} \left[\Lambda_n(M_1^{(,i)} \otimes M_2^{(,j)})\right] \left[\Lambda_n(e_i \otimes e_j)\right]' \\
&= \sum_{i,j=1}^{n} \left[\Lambda_n Vec(M_2^{(,j)} M_1^{(,i)'})\right] \left[\Lambda_n Vec(e_j e_i')\right]' \\
&= \sum_{i,j=1}^{n} (M_2^{(,j)} \otimes M_1^{(,i)})(e_j \otimes e_i)' \\
&= M_2 \otimes M_1.
\end{aligned}$$

(v) With the same notations,

$$\begin{aligned}
\Lambda_n(M_1 \otimes m) &= \Lambda_n \sum_{i=1}^{n} (M_1^{(,i)} e_i') \otimes m \\
&= \Lambda_n \sum_{i=1}^{n} (M_1^{(,i)} \otimes m) e_i' \\
&= \Lambda_n \sum_{i=1}^{n} Vec(m M_1^{(,i)'}) e_i' \\
&= \sum_{i=1}^{n} Vec(M_1^{(,i)} m') e_i' \\
&= \sum_{i=1}^{n} (m \otimes M_1^{(,i)}) e_i' \\
&= m \otimes M_1.
\end{aligned}$$

■

A.3 Z_t conditional moments calculation

Lemma A.2 *If $\varepsilon \sim \mathcal{N}(0, I_n)$, we have*

$$\mathbb{V}\left[Vec(\varepsilon\varepsilon')\right] = I_{n^2} + \Lambda_n,$$

where Λ_n is given in Lemma A.1.

Proof

$$Vec(\varepsilon\varepsilon') = \left[(\varepsilon\varepsilon_1)',\, (\varepsilon\varepsilon_2)',\, \ldots,\, (\varepsilon\varepsilon_n)'\right]'.$$

$\mathbb{V}\left[Vec(\varepsilon\varepsilon')\right]$ is a $(n^2 \times n^2)$ matrix, partitioned in $(n \times n)$ blocks, whose (i,j) block is $V_{i,j} = cov(\varepsilon\varepsilon_i, \varepsilon_j\varepsilon)$. The (k,ℓ) entry of $V_{i,j}$ is $cov(\varepsilon_k\varepsilon_i, \varepsilon_j\varepsilon_\ell)$. Two cases can be distinguished:

- *Case 1:* if $i \neq j$, then the only non-zero terms among the $cov(\varepsilon_k\varepsilon_i, \varepsilon_j\varepsilon_\ell)$ are obtained for $k = j$ and $i = \ell$. By the properties of standardized Gaussian distribution, we have $\mathrm{cov}(\varepsilon_i\varepsilon_j, \varepsilon_i\varepsilon_j) = \mathbb{V}(\varepsilon_i\varepsilon_j) = 1$. Finally, $V_{i,j} = e_j e_i'$.
- *Case 2:* if $i = j$, then the non-zero terms among the $cov(\varepsilon_k\varepsilon_i, \varepsilon_j\varepsilon_\ell)$ are obtained for $k = \ell = i$ and its value is 2, or for $k = \ell \neq i$, and its value is 1. Finally, $V_{i,i} = I_n + e_i e_i'$.

Putting case 1 and 2 together, we get $\mathbb{V}\left[Vec(\varepsilon\varepsilon')\right] = I_{n^2} + \Lambda_n$. ∎

Proposition 3.1 $\mathbb{E}_{t-1}(Z_t) = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1}$ and $\mathbb{V}_{t-1}(Z_t) = \widetilde{\Sigma}_{t-1}$, where:

$$\begin{aligned}
\widetilde{\mu} &= \begin{pmatrix} \mu \\ Vec(\mu\mu' + \Sigma) \end{pmatrix}, \quad \widetilde{\Phi} = \left(\begin{array}{c|c} \Phi & 0 \\ \hline \mu \otimes \Phi + \Phi \otimes \mu & \Phi \otimes \Phi \end{array}\right) \\
\widetilde{\Sigma}_{t-1} &\equiv \widetilde{\Sigma}(Z_{t-1}) = \left(\begin{array}{c|c} \Sigma & \Sigma\Gamma_{t-1}' \\ \hline \Gamma_{t-1}\Sigma & \Gamma_{t-1}\Sigma\Gamma_{t-1}' + \left[I_{n^2} + \Lambda_n\right](\Sigma \otimes \Sigma) \end{array}\right) \\
\Gamma_{t-1} &= I_n \otimes (\mu + \Phi X_{t-1}) + (\mu + \Phi X_{t-1}) \otimes I_n,
\end{aligned}$$

Λ_n being the $n^2 \times n^2$ matrix, defined in Lemma A.1.

Proof

$$\begin{aligned}
\mathbb{E}_{t-1}(X_t) &= \mu + \Phi X_{t-1} \\
\mathbb{E}_{t-1}[X_t X_t'] &= \mathbb{E}_{t-1}\left(\mu + \Phi X_{t-1} + \Omega\varepsilon_t\right)\left(\mu + \Phi X_{t-1} + \Omega\varepsilon_t\right)' \\
&= \mu\mu' + \mu X_{t-1}'\Phi' + \Phi X_{t-1}\mu' + \Phi X_{t-1} X_{t-1}'\Phi' + \Sigma
\end{aligned}$$

Using the $Vec(\bullet)$ operator properties of Proposition A.1, (iii), we obtain:

$$\mathbb{E}_{t-1}\left[Vec(X_t X_t')\right] = Vec(\mu\mu' + \Sigma) + (\mu \otimes \Phi)X_{t-1} + (\Phi \otimes \mu)X_{t-1} + (\Phi \otimes \Phi)Vec(X_{t-1}X_{t-1}')$$

Finally,

$$\boxed{\mathbb{E}_{t-1}(Z_t) = \left(\begin{array}{c} \mu \\ \\ Vec(\mu\mu' + \Sigma) \end{array}\right) + \left(\begin{array}{c|c} \Phi & 0 \\ \hline \mu \otimes \Phi + \Phi \otimes \mu & \Phi \otimes \Phi \end{array}\right) Z_{t-1}}$$

For the conditional variance-covariance matrix, we have $\mathbb{V}_{t-1}(X_t) = \Sigma$ and

$$\begin{aligned} \mathbb{V}_{t-1}\left[Vec(X_t X_t')\right] &= \mathbb{V}_{t-1}\Big[Vec(\overbrace{\mu\mu' + \mu X_{t-1}'\Phi' + \Phi X_{t-1}\mu' + \Phi X_{t-1}X_{t-1}'\Phi'}^{\text{deterministic given } \underline{Z_{t-1}}} \\ &\quad +(\mu + \Phi X_{t-1})\varepsilon_t'\Omega' + \Omega\varepsilon_t(\mu' + X_{t-1}'\Phi') + \Omega\varepsilon_t\varepsilon_t'\Omega')\Big] \end{aligned}$$

Using properties Proposition A.1, $(iii - iv)$,

$$\begin{aligned} \mathbb{V}_{t-1}\left[Vec(X_t X_t')\right] &= \mathbb{V}_{t-1}\big[(I_n \otimes \mu + \mu \otimes I_n + I_n \otimes \Phi X_{t-1} + \Phi X_{t-1} \otimes I_n)\Omega\varepsilon_t + Vec(\Omega\varepsilon_t\varepsilon_t'\Omega)\big] \\ &\equiv \mathbb{V}_{t-1}\left[\Gamma_{t-1}\Omega\varepsilon_t + Vec(\Omega\varepsilon_t\varepsilon_t'\Omega)\right] \\ &= \Gamma_{t-1}\Sigma\Gamma_{t-1}' + \mathbb{V}_{t-1}\left[(\Omega \otimes \Omega)Vec(\varepsilon_t\varepsilon_t')\right] \quad \text{as} \quad \mathrm{Cov}_{t-1}[\varepsilon_t, Vec(\varepsilon_t\varepsilon_t')] = 0 \\ &= \Gamma_{t-1}\Sigma\Gamma_{t-1}' + (\Omega \otimes \Omega)(I_{n^2} + \Lambda_n)(\Omega' \otimes \Omega') \quad \text{(using Lemma A.2.)} \end{aligned}$$

Proposition A.1, (v) implies that $(\Omega \otimes \Omega)(\Omega \otimes \Omega)' = \Sigma \otimes \Sigma$. Therefore, we have:

$$\begin{aligned} & (\Omega \otimes \Omega)(\Omega \otimes \Omega)' = (\Omega \otimes \Omega)\Lambda_n(\Omega \otimes \Omega)'\Lambda_n \quad \text{(using Lemma A.1, } (iv)) \\ \Longleftrightarrow \quad & (\Omega \otimes \Omega)\Lambda_n(\Omega \otimes \Omega)' = \Lambda_n(\Sigma \otimes \Sigma) \quad \text{since} \quad \Lambda_n = \Lambda_n^{-1} \\ \Longleftrightarrow \quad & (\Omega \otimes \Omega)(I_{n^2} + \Lambda_n)(\Omega \otimes \Omega)' = (I_{n^2} + \Lambda_n)(\Sigma \otimes \Sigma). \end{aligned}$$

Hence:

$$\mathbb{V}_{t-1}\left[Vec(X_t X_t')\right] = \Gamma_{t-1}\Sigma\Gamma_{t-1}' + (I_{n^2} + \Lambda_n)(\Sigma \otimes \Sigma).$$

Using again the fact that ε_t and $Vec(\Omega\varepsilon_t\varepsilon_t'\Omega')$ are non-correlated, we have:

$$\begin{aligned} \mathrm{cov}_{t-1}\left[Vec(X_t X_t'), X_t\right] &= \mathrm{cov}_{t-1}\left[\Gamma_{t-1}\Omega\varepsilon_t, \Omega\varepsilon_t\right] \\ &= \Gamma_{t-1}\Sigma \end{aligned}$$

Finally, the conditional variance-covariance matrix of Z_t given X_{t-1} is

$$\boxed{\widetilde{\Sigma}_{t-1} = \left(\begin{array}{c|c} \Sigma & \Sigma\Gamma_{t-1}' \\ \hline \Gamma_{t-1}\Sigma & \Gamma_{t-1}\Sigma\Gamma_{t-1}' + (I_{n^2} + \Lambda_n)(\Sigma \otimes \Sigma) \end{array}\right)}. \tag{10}$$

■

A.4 Proof of Proposition 3.2

We want to explicitly disclose the affine form of $\widetilde{\Sigma}(Z_{t-1})$. In order to achieve this, we consider the four blocks of the matrix in Equation (10) and express the vectorized form of each block. First, let us show that $Vec(\Gamma_{t-1}\Sigma)$ is affine in Z_{t-1}. We have:

$$\begin{aligned}
\Gamma_{t-1} &= I_n \otimes (\mu + \Phi X_{t-1}) + (\mu + \Phi X_{t-1}) \otimes I_n \\
&= I_n \otimes (\mu + \Phi X_{t-1}) + \Lambda_n \left[I_n \otimes (\mu + \Phi X_{t-1})\right] \quad \text{(using Lemma A.1, } (v)\text{)} \\
&= (I_{n^2} + \Lambda_n) \left[I_n \otimes (\mu + \Phi X_{t-1})\right].
\end{aligned}$$

Therefore we have:

$$\begin{aligned}
Vec(\Gamma_{t-1}\Sigma) &= Vec\left\{(I_{n^2} + \Lambda_n)\left[I_n \otimes (\mu + \Phi X_{t-1})\right]\Sigma\right\} \\
&= \left[\Sigma \otimes (I_{n^2} + \Lambda_n)\right] Vec\left\{I_n \otimes (\mu + \Phi X_{t-1})\right\} && \text{(Prop. A.1, } (iii)\text{)} \\
&= \left[\Sigma \otimes (I_{n^2} + \Lambda_n)\right]\left[Vec(I_n) \otimes (\mu + \Phi X_{t-1})\right] && \text{(Prop. A.1, } (ii)\text{)} \\
&= \left[\Sigma \otimes (I_{n^2} + \Lambda_n)\right] Vec\left[(\mu + \Phi X_{t-1})Vec(I_n)'\right] && \text{(Prop. A.1, } (i)\text{)} \\
&= \left[\Sigma \otimes (I_{n^2} + \Lambda_n)\right]\left[Vec(I_n) \otimes I_n\right](\mu + \Phi X_{t-1}) && \text{(Prop. A.1, } (iv)\text{)} \\
Vec(\Sigma\Gamma'_{t-1}) &= Vec\left\{\Sigma\left[(I_{n^2} + \Lambda_n)\left(I_n \otimes (\mu + \Phi X_{t-1})\right)\right]'\right\} \\
&= Vec\left\{\Sigma\left[I_n \otimes (\mu + \Phi X_{t-1})'\right](I_{n^2} + \Lambda_n)\right\} \\
&= \left[(I_{n^2} + \Lambda_n) \otimes \Sigma\right] Vec\left[I_n \otimes (\mu + \Phi X_{t-1})'\right] && \text{(Prop. A.1, } (iii)\text{)} \\
&= \left[(I_{n^2} + \Lambda_n) \otimes \Sigma\right](I_n \otimes \Lambda_n)\left[Vec(I_n) \otimes (\mu + \Phi X_{t-1})\right] && \text{(Prop. A.1, } (ii)\text{)} \\
&= \left[(I_{n^2} + \Lambda_n) \otimes \Sigma\right](I_n \otimes \Lambda_n)\left[Vec(I_n) \otimes I_n\right](\mu + \Phi X_{t-1}) && \text{(Prop. A.1, } (i - iv)\text{)}.
\end{aligned}$$

We turn now to the lower-right block of the conditional variance-covariance matrix of Z_t. We have:

$$
\begin{aligned}
& Vec(\Gamma_{t-1}\Sigma\Gamma'_{t-1}) \\
= \;& Vec\left\{(I_{n^2}+\Lambda_n)\left[I_n\otimes(\mu+\Phi X_{t-1})\right]\Sigma\left[I_n\otimes(\mu+\Phi X_{t-1})'\right](I_{n^2}+\Lambda_n)\right\} \\
= \;& \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad Vec\left\{\left[I_n\otimes(\mu+\Phi X_{t-1})\right]\Sigma\left[I_n\otimes(\mu+\Phi X_{t-1})'\right]\right\} && \text{(Prop. A.1, } (iii)) \\
= \;& \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad Vec\left\{\left[\Sigma\otimes(\mu+\Phi X_{t-1})\right]\left[I_n\otimes(\mu+\Phi X_{t-1})'\right]\right\} && \text{(Prop. A.1, } (v)) \\
= \;& \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad Vec\left\{\Sigma\otimes\left[(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right]\right\} && \text{(Prop. A.1, } (v)) \\
= \;& \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad \left[I_n\otimes\Lambda_n\otimes I_n\right]\left[Vec(\Sigma)\otimes Vec\left\{(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right\}\right] && \text{(Prop. A.1, } (ii)) \\
= \;& \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad \left[I_n\otimes\Lambda_n\otimes I_n\right]Vec\left[Vec\left\{(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right\}\times Vec(\Sigma)'\right] && \text{(Prop. A.1, } (i)) \\
= \;& \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad \left[I_n\otimes\Lambda_n\otimes I_n\right]\left[Vec(\Sigma)\otimes I_{n^2}\right]Vec\left\{(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right\} && \text{(Prop. A.1, } (iv))
\end{aligned}
$$

Finally we obtain the affine formulae for the four blocks of the conditional variance-covariance matrix $\widetilde{\Sigma}^{(i,j)}_{t-1}$ for $i,j=\{1,2\}$:

$$
\begin{aligned}
Vec\left(\widetilde{\Sigma}^{(1,1)}_{t-1}\right) &= Vec(\Sigma) \\
Vec\left(\widetilde{\Sigma}^{(1,2)}_{t-1}\right) &= \left[\Sigma\otimes(I_{n^2}+\Lambda_n)\right]\left[Vec(I_n)\otimes I_n\right]\left\{\mu+\widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}^{(2,1)}_{t-1}\right) &= \left[(I_{n^2}+\Lambda_n)\otimes\Sigma\right](I_n\otimes\Lambda_n)\left[Vec(I_n)\otimes I_n\right]\left\{\mu+\widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}^{(2,2)}_{t-1}\right) &= \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\left[I_n\otimes\Lambda_n\otimes I_n\right]\left[Vec(\Sigma)\otimes I_{n^2}\right]\left\{\mu\otimes\mu+\widetilde{\Phi}_2 Z_{t-1}\right\} \\
&\quad +\left[I_{n^2}\otimes(I_{n^2}+\Lambda_n)\right]Vec(\Sigma\otimes\Sigma),
\end{aligned}
$$

where $\widetilde{\Phi}_1$ and $\widetilde{\Phi}_2$ are respectively the upper and lower blocks of $\widetilde{\Phi}$, thus $\widetilde{\Phi}_1=\begin{pmatrix}\Phi & 0\end{pmatrix}$ and $\widetilde{\Phi}_2=\begin{pmatrix}\mu\otimes\Phi+\Phi\otimes\mu & \Phi\otimes\Phi\end{pmatrix}$.

It should be noted that the computation of $Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]$ – i.e. the analytical expressions of ν and Ψ – involves, in theory, a permutation of the previous vectorized-blocks formulae ; however, we describe hereafter a simple and pragmatic method to reconstruct $Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]$ in the QKF algorithm:

1. Use the formulae of Proposition 3.2 to construct the four vectorized blocks of $\widetilde{\Sigma}(Z_{t-1})$ as explicit affine functions of Z_{t-1} (or $\widetilde{\Sigma}(Z_{t-1|t-1})$ as affine functions of $Z_{t-1|t-1}$ in the QKF algorithm).

2. Reconstruct the square matrix $\widetilde{\Sigma}(Z_{t-1})$ from the previous vectorized blocks.

3. Vectorize the reconstructed matrix.

Using the aforementioned method does not require an analytical expression of ν and Ψ and is a fast technique to calculate both the conditional and unconditional variances in the algorithm.

A.5 Unconditional moments of Z_t

Proposition 3.3 We have:

$$\begin{bmatrix} \mathbb{E}(Z_t) \\ \\ Vec[\mathbb{V}(Z_t)] \end{bmatrix} = \begin{pmatrix} \widetilde{\mu} \\ \\ \nu \end{pmatrix} + \begin{pmatrix} \widetilde{\Phi} & 0 \\ \\ \Psi & \widetilde{\Phi}\otimes\widetilde{\Phi} \end{pmatrix} \begin{bmatrix} \mathbb{E}(Z_{t-1}) \\ \\ Vec[\mathbb{V}(Z_{t-1})] \end{bmatrix}$$

Proof The first set of equation is immediately obtained from the state-space representation. For the second set, the variance decomposition writes:

$$\begin{aligned} \mathbb{V}(Z_t) &= \mathbb{E}\left[\mathbb{V}(Z_t|\underline{Z_{t-1}})\right] + \mathbb{V}\left[\mathbb{E}(Z_t|\underline{Z_{t-1}})\right] \\ &= \mathbb{E}\left[\mathbb{V}(Z_t|\underline{Z_{t-1}})\right] + \mathbb{V}(\widetilde{\mu} + \widetilde{\Phi}Z_{t-1}) \\ &= \mathbb{E}\left[\mathbb{V}(Z_t|\underline{Z_{t-1}})\right] + \widetilde{\Phi}\mathbb{V}(Z_{t-1})\widetilde{\Phi}' \\ &= \mathbb{E}\left[\widetilde{\Sigma}(Z_{t-1})\right] + \widetilde{\Phi}\mathbb{V}(Z_{t-1})\widetilde{\Phi}' \\ \Longrightarrow \quad & Vec\left[\mathbb{V}(Z_t)\right] = \mathbb{E}\left\{Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]\right\} + (\widetilde{\Phi}\otimes\widetilde{\Phi})Vec\left[\mathbb{V}(Z_{t-1})\right] \end{aligned}$$

Denoting $Vec[\widetilde{\Sigma}(Z_{t-1})]$ by $\nu + \Psi Z_{t-1}$ we get:

$$\mathbb{E}\left\{Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]\right\} = \nu + \Psi\mathbb{E}(Z_{t-1})$$

and the result follows. ■

A.6 Selection and duplication matrices

Definition A.1 *Let P be a $(n \times n)$ symmetric matrix. Let us define a partition of $I_n = [u_n, U_n]$ where u_n is the first column of I_n and U_n is the $(n \times (n-1))$ other sub-matrix. Let Q_n be a $(n^2 \times n^2)$ matrix defined as $Q_n = (Q_{1,n}, Q_{2,n})$ such that:*

$$Q_{1,n} = I_n \otimes u_n \qquad \textit{and} \qquad Q_{2,n} = I_n \otimes U_n$$

A duplication matrix G_n *and a selection matrix* H_n *are such that:*

$$\begin{aligned} Vec(P) &= G_n Vech(P) \\ Vech(P) &= H_n Vec(P) \end{aligned}$$

and can be expressed recursively by:

$$G_{n+1} = Q_{n+1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & I_n & 0 \\ 0 & I_n & 0 \\ 0 & 0 & G_n \end{bmatrix} \qquad \text{and} \qquad H_{n+1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 \\ 0 & 0 & 0 & H_n \end{bmatrix} Q'_{n+1}$$

with $G_1 = H_1 = Q_1 = 1$*. These definitions can be found in Magnus and Neudecker (1980) or Harville (1997).*

A.7 EKF and UKF general algorithms

Let us consider a state-space model with non-linear transition and measurement equations.

$$X_t = f_t(X_{t-1}) + g_t(X_{t-1})\varepsilon_t \tag{11}$$

$$Y_t = h_t(Z_t) + d_t(Z_t)\eta_t \tag{12}$$

where f_t, G_t, h_t, D_t are function of $\underline{Y_{t-1}}$ and possibly a vector of exogenous variables. Also, $(\varepsilon_t', \eta_t')' \sim \mathcal{IIN}(0, I)$. We use the following notations:

$$\begin{aligned} F_t &= \frac{\partial f_t}{\partial x'_{t-1}}(\widehat{X}_{t-1|t-1}), \quad H_t = \frac{\partial h_t}{\partial x'_t}(\widehat{X}_{t|t-1}) \\ F^{(2)}_{i,t} &= \frac{\partial^2 f_{i,t}}{\partial x_{t-1}\partial x'_{t-1}}(\widehat{X}_{t-1|t-1}), \quad H^{(2)}_{i,t} = \frac{\partial^2 h_t}{\partial x_t \partial x'_t}(\widehat{X}_{t|t-1}) \\ G_t &= g_t(\widehat{X}_{t-1|t-1}), \quad \text{and} \quad D_t = d_t(\widehat{X}_{t|t-1}) \end{aligned}$$

Let us also denote by $e_i^{(k)}$ the vector of size k whose components are equal to 0 except the i^{th} one which is equal to 1. The EKF1 and EKF2 algorithms are respectively given in Tables 5 and 6. Keeping the same notations, Table 7 presents the recursions of the UKF.

Table 5: EKF1 algorithm in the general non-linear case

Initialize:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$.
State prediction:	$X_{t\|t-1}$	$f_t(X_{t-1\|t-1})$
	$P^X_{t\|t-1}$	$F_t P^X_{t-1\|t-1} F'_t + G_t G'_t$
Measurement prediction:	$Y_{t\|t-1}$	$h_t(X_t\|t-1)$
	$M_{t\|t-1}$	$H_t P^X_{t\|t-1} H'_t + D_t D'_t$
Gain:	K_t	$P^X_{t\|t-1} H'_t M^{-1}_{t\|t-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K'_t$

Note: See Jazwinski (1970), Anderson and Moore (1979), or Gelb, Kasper, Nash, Price, and Sutherland (1974) for a proof of the recursions.

Table 6: EKF2 algorithm in the general non-linear case

Initialize:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$.
State prediction:	$X_{t\|t-1}$	$f_t(X_{t-1\|t-1}) + \frac{1}{2}\sum_{i=1}^{n} e^{(n)}_i \operatorname{Tr}\left(F^{(2)}_{i,t} P^X_{t-1\|t-1}\right)$
	$P^X_{t\|t-1}$	$F_t P^X_{t-1\|t-1} F'_t + \frac{1}{2}\sum_{i,j=1}^{n} e^{(n)}_i \operatorname{Tr}\left(F^{(2)}_{i,t} P^X_{t-1\|t-1} F^{(2)}_{j,t} P^X_{t-1\|t-1}\right) e^{(n)'}_j + G_t G'_t$
Measurement prediction:	$Y_{t\|t-1}$	$h_t(X_t\|t-1) + \frac{1}{2}\sum_{k=1}^{m} e^{(m)}_k \operatorname{Tr}\left(H^{(2)}_{k,t} P^X_{t\|t-1}\right)$
	$M_{t\|t-1}$	$H_t P^X_{t\|t-1} H'_t + \frac{1}{2}\sum_{k,l=1}^{n} e^{(n)}_k \operatorname{Tr}\left(H^{(2)}_{k,t} P^X_{t\|t-1} H^{(2)}_{l,t} P^X_{t\|t-1}\right) e^{(n)'}_l + D_t D'_t$
Gain:	K_t	$P^X_{t\|t-1} H'_t M^{-1}_{t\|t-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K'_t$

Note: See Athans, Wishner, and Bertolini (1968) or Maybeck (1982) for a proof of the recursions.

Table 7: UKF algorithm in the general non-linear case

Initialize:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$ and choose (α, κ, β).
Sigma points:		$\left\{\mathcal{X}_{i,t-1\|t-1}(X_{t-1\|t-1}, P^X_{t-1\|t-1})\right\}_{i=\{1,\ldots,2n\}}$ according to Definition 4.1.
State prediction:	$X_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i \, f_t(\mathcal{X}_{i,t-1\|t-1})$
	$P^X_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[f_t(\mathcal{X}_{i,t-1\|t-1}) - X_{t\|t-1}\right] \left[f_t(\mathcal{X}_{i,t-1\|t-1}) - X_{t\|t-1}\right]' + G_t G_t'$
Sigma points:		$\left\{\mathcal{X}_{i,t\|t-1}(X_{t\|t-1}, P^X_{t\|t-1})\right\}_{i=\{1,\ldots,2n\}}$ according to Definition 4.1.
Measurement prediction:	$Y_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i \, h_t(\mathcal{X}_{i,t\|t-1})$
	$M_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[h_t(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right] \left[h_t(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' + D_t D_t'$
Gain:	K_t	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[\mathcal{X}_{i,t\|t-1} - X_{t\|t-1}\right] \left[h_t(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' M^{-1}_{t\|t-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K_t'$

Note: See Julier, Uhlmann, and Durrant-Whyte (2000), Julier (2002), Julier (2003), or Julier and Uhlmann (2004) for proofs of the recursions.

A.8 The UKF in a linear state-space model

Let us consider a linear state-space model which is given by Equations (1a) and (1b) putting all the $C^{(k)}$s to 0. Taking the notations of Table 7, we have $f_t(X) = \mu + \Phi X$ and $h_t(X) = A + BX$. As for $i = \{1, \ldots, 2n\}$, all the weights are equal, the sigma points are symmetrical and $\sum_{i=0}^{2n} \mathcal{W}_i = 1$, we have:

$$
\begin{aligned}
X_{t|t-1} &= \sum_{i=0}^{2n} \mathcal{W}_i \left(\mu + \Phi \mathcal{X}_{i,t-1|t-1}\right) \\
&= \left(\sum_{i=0}^{2n} \mathcal{W}_i\right) \mu + \Phi \left(\sum_{i=0}^{2n} \mathcal{W}_i \mathcal{X}_{i,t-1|t-1}\right) \\
&= \mu + \Phi X_{t-1|t-1} \\
P_{t|t-1} &= \sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[\Phi(\mathcal{X}_{i,t-1|t-1} - X_{t-1|t-1})\right] \left[\Phi(\mathcal{X}_{i,t-1|t-1} - X_{t-1|t-1})\right]' + G_t G_t' \\
&= \sum_{i=1}^{n} \mathcal{W}_i^{(c)} \Phi \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_i \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_i' \Phi' \\
&\quad + \sum_{i=n+1}^{2n} \mathcal{W}_i^{(c)} \Phi \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_{i-n} \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_{i-n}' \Phi' + G_t G_t' \\
&= 2 \sum_{i=1}^{n} \frac{(n+\lambda)}{2(\lambda+n)} \Phi \left(\sqrt{P^X_{t-1|t-1}}\right)_i \left(\sqrt{P^X_{t-1|t-1}}\right)_i' \Phi' + G_t G_t' \\
&= \Phi P_{t-1|t-1} \Phi' + G_t G_t'
\end{aligned}
$$

which proves the exact matching of the UKF and the Kalman filter for the state prediction phase. The same argument holds by linearity for the measurement prediction phase. In the linear case, the UKF shows exactly the same recursions that the linear Kalman filter, whatever the values of (α, κ, β).

A.9 The UKF in a quadratic state-space model: scalar case

Let us consider the quadratic state-space model given by Equations (1a) and (1b). Let us set the vector of tuning parameters $(\alpha, \kappa, \beta) = (1, 0, 2)$ and $n = m = 1$. From Appendix A.8, we know that the state prediction phase is exactly the same as in the linear Kalman filter, and is a fortiori the same as in the EKF2. Let us prove that the measurement prediction phase is the same for both filters for those values of (α, κ, β).

First, those tuning parameters imply $\lambda = 0$, thus:

$$\mathcal{X}_{i,t|t-1} = \begin{cases} X_{t|t-1} & \text{for} \quad i = 0 \\ X_{t|t-1} + \sqrt{P^X_{t|t-1}} & \text{for} \quad i = 1 \\ X_{t|t-1} - \sqrt{P^X_{t|t-1}} & \text{for} \quad i = 2, \end{cases} \qquad \begin{aligned} \mathcal{W}_i &= \begin{cases} 0 & \text{for} \quad i = 0 \\ 1/2 & \text{for} \quad i \neq 0 \end{cases} \\ \mathcal{W}_i^{(c)} &= \begin{cases} 2 & \text{for} \quad i = 0 \\ 1/2 & \text{for} \quad i \neq 0 \end{cases} \end{aligned}$$

Then, using the recursion of the UKF algorithm, we obtain:

$$\begin{aligned}
Y_{t|t-1} &= \frac{1}{2}\left[h(X_{t|t-1} + \sqrt{P^X_{t|t-1}}) + h(X_{t|t-1} - \sqrt{P^X_{t|t-1}})\right] \\
&= \frac{1}{2}\left\{2A + B\left(2X_{t|t-1}\right) + C\left[\left(X_{t|t-1} + \sqrt{P^X_{t|t-1}}\right)^2 + \left(X_{t|t-1} - \sqrt{P^X_{t|t-1}}\right)^2\right]\right\} \\
&= A + BX_{t|t-1} + CX^2_{t|t-1} + CP^X_{t|t-1} \\
&= h(X_{t|t-1}) + CP^X_{t|t-1} \\
M_{t|t-1} &= 2\left[h(X_{t|t-1}) - Y_{t|t-1}\right]^2 \\
&\quad + \frac{1}{2}\left\{\left[h(X_{t|t-1} + \sqrt{P^X_{t|t-1}}) - Y_{t|t-1}\right]^2 + \left[h(X_{t|t-1} - \sqrt{P^X_{t|t-1}}) - Y_{t|t-1}\right]^2\right\} + V \\
&= 2C^2(P^X_{t|t-1})^2 + \frac{1}{2}\left\{\left[B\sqrt{P^X_{t|t-1}} + C\left(P^X_{t|t-1} + 2X_{t|t-1}\sqrt{P^X_{t|t-1}} - P^X_{t|t-1}\right)\right]^2\right. \\
&\quad \left. + \left[-B\sqrt{P^X_{t|t-1}} + C\left(P^X_{t|t-1} - 2X_{t|t-1}\sqrt{P^X_{t|t-1}} - P^X_{t|t-1}\right)\right]^2\right\} + V \\
&= 2C^2(P^X_{t|t-1})^2 + V + \frac{1}{2}\left\{\vphantom{\left(\sqrt{P^X_{t|t-1}}\right)^2}\right. \\
&\qquad B^2P^X_{t|t-1} + C^2\left(2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right)^2 + 2CB\sqrt{P^X_{t|t-1}}\left(2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right) \\
&\qquad \left. + B^2P^X_{t|t-1} + C^2\left(-2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right)^2 + 2CB\sqrt{P^X_{t|t-1}}\left(2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right)\right\} \\
&= 2C^2(P^X_{t|t-1})^2 + V + B^2P^X_{t|t-1} + 4C^2X^2_{t|t-1}P^X_{t|t-1} + 2BCX_{t|t-1}P^X_{t|t-1} \\
&= \mathcal{G}^2_{t|t-1}P^X_{t|t-1} + 2C^2(P^X_{t|t-1})^2 + V
\end{aligned}$$

Both $Y_{t|t-1}$ and $M_{t|t-1}$ yield the same result as in the EKF2 recursions. Let us now turn to the Kalman gain computation.

$$\begin{aligned} K_t &= \frac{1}{2}\left\{\left[\sqrt{P^X_{t|t-1}}\sqrt{P^X_{t|t-1}}\left(B+2CX_{t|t-1}\right)-\sqrt{P^X_{t|t-1}}\sqrt{P^X_{t|t-1}}\left(-B-2CX_{t|t-1}\right)\right]\right\}M^{-1}_{t|t-1} \\ &= P^X_{t|t-1}\left(B+2CX_{t|t-1}\right)M^{-1}_{t|t-1} \\ &= P^X_{t|t-1}\mathcal{G}_{t|t-1}M^{-1}_{t|t-1} \end{aligned}$$

which is also the same gain as in the EKF2. Therefore, for $(\alpha, \kappa, \beta) = (1, 0, 2)$ and scalar transition and measurement equations, The UKF and the EKF2 possess exactly the same recursions.

Figure 1: RMSE of $\widehat{X_{t|t}}$

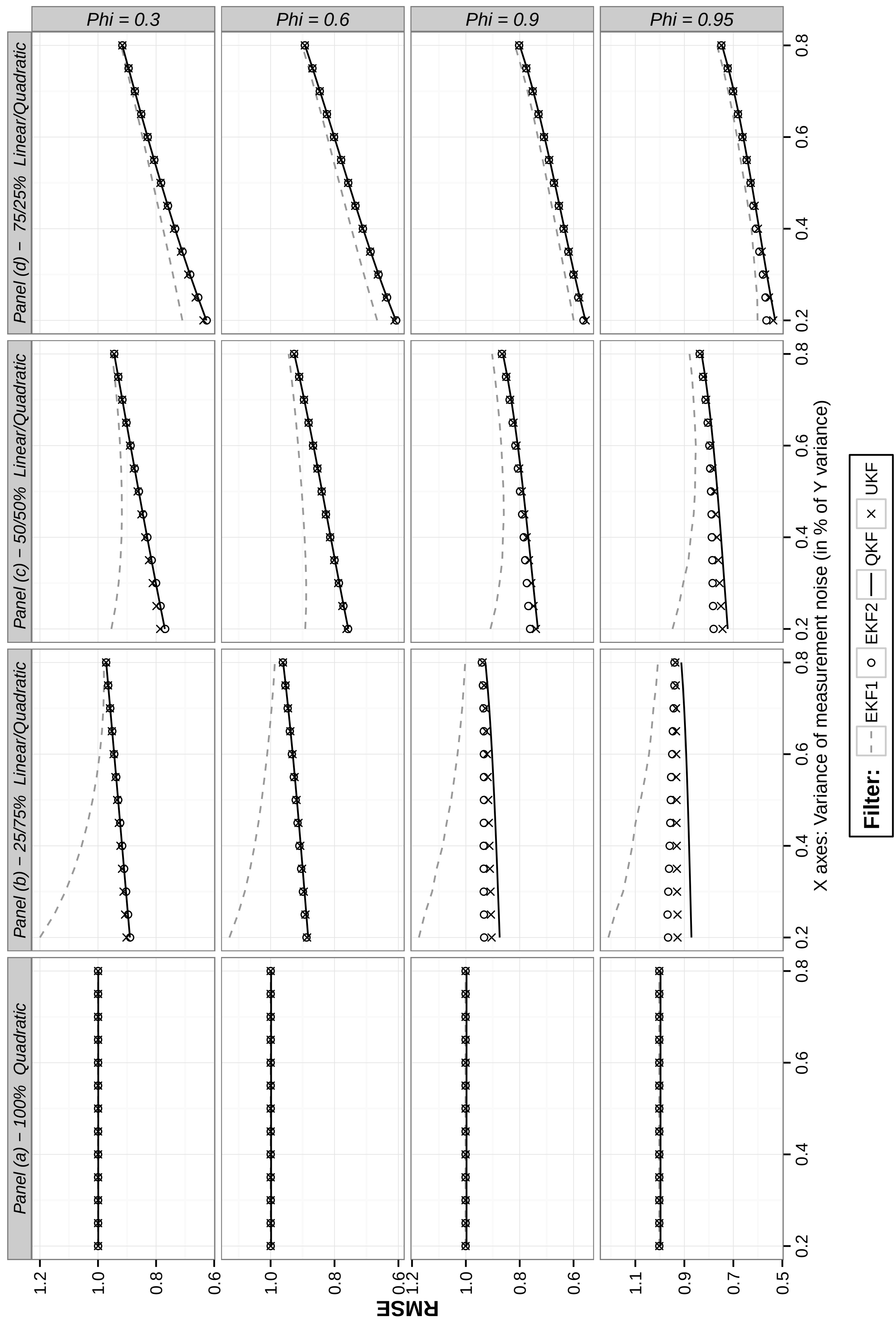

Figure 2: RMSE of $\widehat{X^2_{t|t}}$

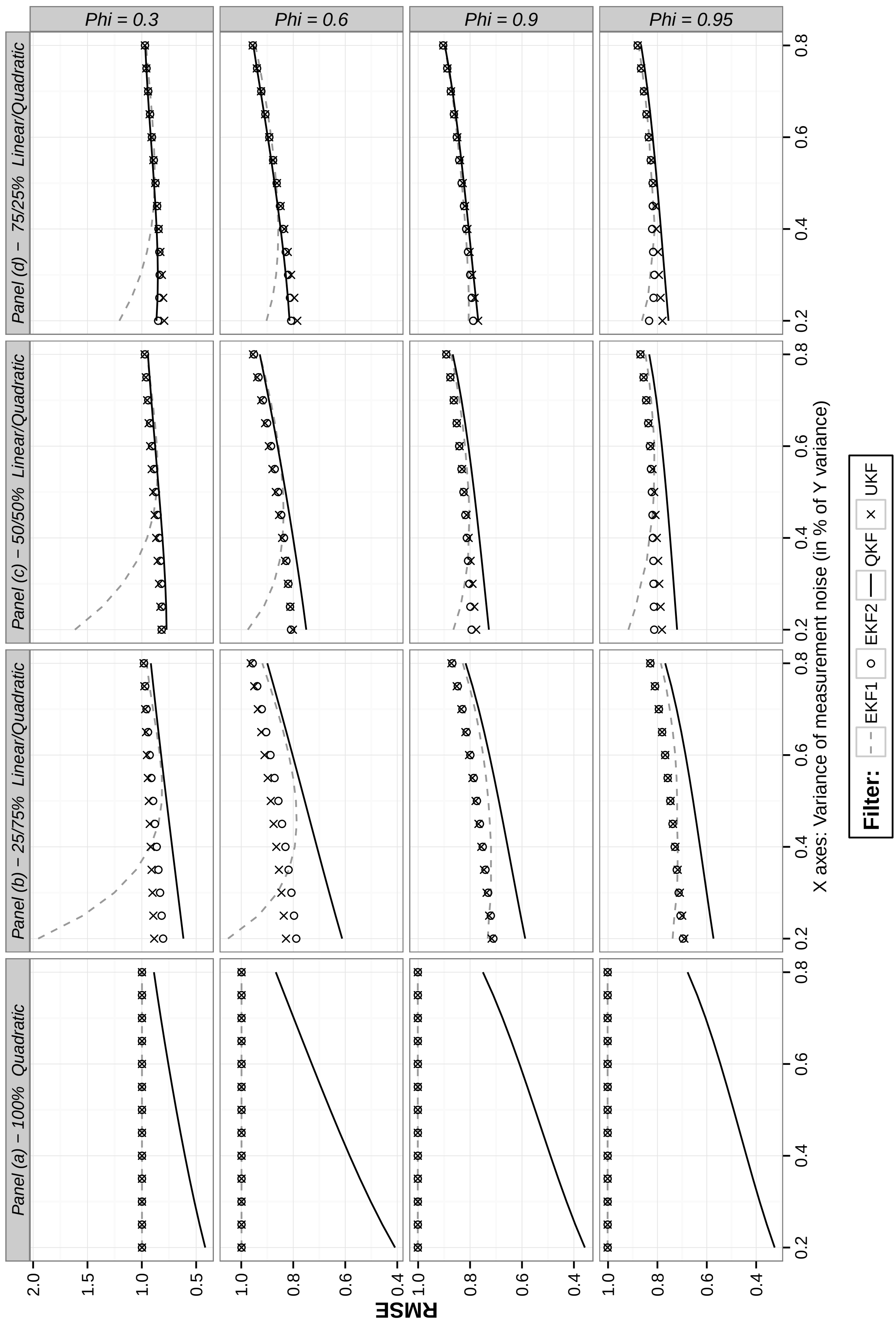

Figure 3: RMSE of $\widehat{\eta}_t$

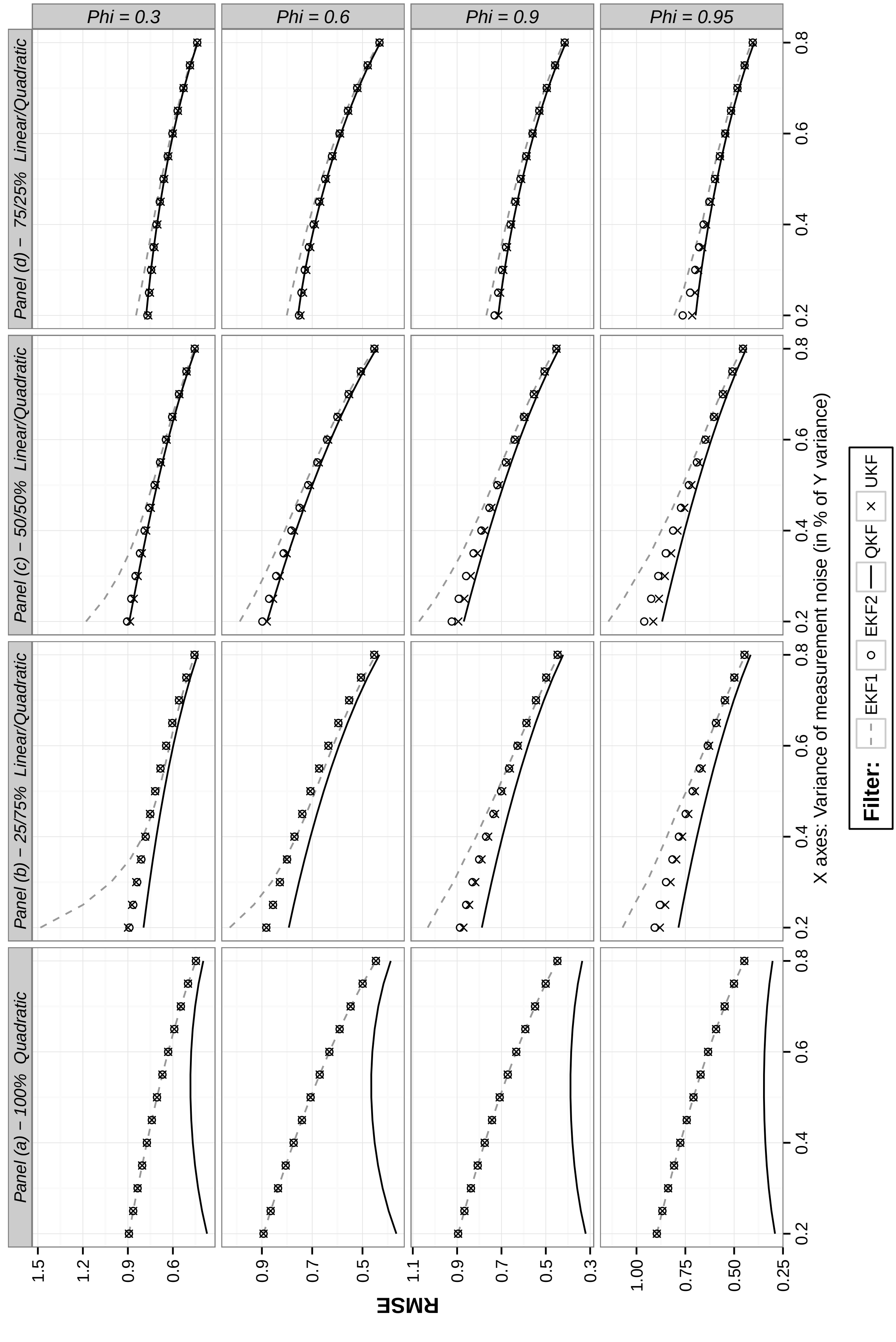

Table 8: Estimation results

		Panel (a)					Panel (b)					Panel (c)				
	Model	A	B	C	Φ	$\sqrt{\theta_1}$	A	B	C	Φ	$\sqrt{\theta_1}$	A	B	C	Φ	$\sqrt{\theta_1}$
		0	0	0.131	0.9	0.224	0	0	0.405	0.6	0.447	0	0.358	0.351	0.6	0.447
Bias	QKF	0.177*	0.188*	−0.003*	−0.063*	−0.037*	0.147*	0.269*	−0.076*	−0.074*	−0.151	0.065*	−0.089*	−0.048*	−0.029*	−0.132
	EKF 1	0.350	0.256	0.016	−0.067	0.076	0.566†	0.547	1.318†	−0.140†	0.052*	0.460†	0.158	0.309†	−0.094†	0.055*
	EKF 2	0.363	0.412	−0.022	−0.122	−0.107	0.260	0.580†	−0.143	−0.098	−0.276	0.165	0.194†	−0.107	−0.058	−0.244
	UKF	0.414†	0.415†	−0.038†	−0.124†	−0.170†	0.336	0.561	−0.200	−0.084	−0.321†	0.246	0.194†	−0.157	−0.052	−0.304†
Std.	QKF	0.319	0.202†	0.038	0.087	0.091	0.210†	0.316	0.164	0.176*	0.290	0.202†	0.299	0.144	0.163*	0.275
	EKF 1	0.211*	0.192	0.104†	0.132†	0.082*	0.115*	0.321†	18.218†	0.251†	0.309†	0.114*	0.304†	10.725†	0.236†	0.277†
	EKF 2	0.389†	0.114	0.030	0.095	0.105†	0.204	0.224	0.137	0.213	0.252	0.189	0.213	0.122	0.201	0.259
	UKF	0.277	0.108*	0.019*	0.077*	0.086	0.167	0.215*	0.100*	0.208	0.229*	0.173	0.194*	0.092*	0.186	0.233*
RMSE	QKF	0.365*	0.276*	0.038	0.108*	0.098*	0.257*	0.415*	0.181*	0.191*	0.327	0.212*	0.312	0.152*	0.166*	0.305
	EKF 1	0.409	0.320	0.105†	0.148	0.112	0.577†	0.635†	3.656†	0.287†	0.314*	0.474†	0.343†	2.042†	0.254†	0.282*
	EKF 2	0.532†	0.427	0.037*	0.155†	0.150	0.331	0.621	0.198	0.235	0.374	0.250	0.288	0.162	0.209	0.356
	UKF	0.498	0.429†	0.042	0.146	0.190†	0.375	0.601	0.224	0.225	0.395†	0.301	0.275*	0.182	0.193	0.383†

Notes: For each set of parameters, estimations were performed simulating 1,000 different dynamics. All computed values in the table are averages across simulations.EKF 1 and EKF 2 stand respectively for the first and second order extended Kalman filters. Panel (a) to (c) vary with respect to the parameters' value provided on the first row. For each simulation, the bias is calculated as $\hat{\beta} - \beta$, hence a positive value indicates an average overestimation. '*' indicates the best value among filters, '†' indicates the worst value among filters.

Table 9: Maximum likelihood performance over the three panels

	QKF	EKF 1	EKF 2	UKF
Number of times being less biased	13	2	0	0
Number of times being most biased	0	6	2	8
Number of times having smallest std.	2	4	0	9
Number of times having biggest std.	3	10	2	0
Number of times having smallest RMSEs	11	2	1	1
Number of times having smallest RMSEs	0	9	2	4

Notes: Cases are taken from Table 8 estimates. Total number of estimated parameters are 15. Note that the sum of the second row however yields a result of 16 due to equality of the EKF 2 and the UKF possessing the worst bias.

References

Ahn, D.-H., R. F. Dittmar, and A. R. Gallant (2002, March). Quadratic term structure models: Theory and evidence. *Review of Financial Studies 15*(1), 243–288.

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering.* Prentice-Hall.

Andreasen, M. and A. Meldrum (2011, January). Likelihood inference in non-linear term structure models: The importance of the zero lower bound. Boe working papers, Bank of England.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: A review. *Econometric Reviews 25*(2-3), 145–175.

Athans, M., R. Wishner, and A. Bertolini (1968). Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements. *Automatic Control, IEEE Transactions on 13*(5), 504–514.

Baadsgaard, M., J. N. Nielsen, and H. Madsen (2000). Estimating multivariate exponential-affine term structure models from coupon bond prices using nonlinear filtering. Technical report.

Bar-Shalom, Y., T. Kirubarajan, and X.-R. Li (2002). *Estimation with Applications to Tracking and Navigation.* John Wiley & Sons, Inc.

Barton, D. and F. David (1960). Models of functional relationship illustrated on astronomical data . *Bulletin of the International Statistical Institute 37*, 9–33.

Bertholon, H., A. Monfort, and F. Pegoraro (2008, Fall). Econometric asset pricing modelling. *Journal of Financial Econometrics 6*(4), 407–458.

Brandt, M. and D. Chapman (2003). Comparing multifactor models of the term structure. Wharton school working paper.

Branger, N. and M. Muck (2012). Keep on smiling? the pricing of quanto options when all covariances are stochastic. *Journal of Banking & Finance 36*(6), 1577–1591.

Buraschi, A., A. Cieslak, and F. Trojani (2008). Correlation risk and the term structure of interest rates. *Working Paper*.

Burnside, C. (1998, March). Solving asset pricing models with gaussian shocks. *Journal of Economic Dynamics and Control 22*(3), 329–340.

Chen, R.-R., X. Cheng, F. J. Fabozzi, and B. Liu (2008, March). An explicit, multi-factor credit default swap pricing model with correlated factors. *Journal of Financial and Quantitative Analysis 43*(01), 123–160.

Cheng, P. and O. Scaillet (2007). Linear-quadratic jump-diffusion modeling. *Mathematical Finance 17*(4), 575–598.

Christoffersen, P., C. Dorion, K. Jacobs, and L. Karoui (2012, May). Nonlinear kalman filtering in affine term structure models. CREATES Research Papers 2012-49, School of Economics and Management, University of Aarhus.

Collard, F. and M. Juillard (2001, June). Accuracy of stochastic perturbation methods: The case of asset pricing models. *Journal of Economic Dynamics and Control 25*(6-7), 979–999.

Dai, Q. and K. J. Singleton (2003, July). Term structure dynamics in theory and reality. *Review of Financial Studies 16*(3), 631–678.

Doshi, H., K. Jacobs, J. Ericsson, and S. M. Turnbull (2013). Pricing credit default swaps with observable covariates. *Review of Financial Studies*.

Duan, J.-C. and J.-G. Simonato (1999, September). Estimating and testing exponential-affine term structure models by kalman filter. *Review of Quantitative Finance and Accounting 13*(2), 111–35.

Dubecq, S., A. Monfort, J.-P. Renne, and G. Roussellet (2013). Credit and liquidity in interbank rates: A quadratic approach. Banque de France Working Paper Series forthcoming, Banque de France.

Duffee, G. and R. Stanton (2008). Evidence on simulation inference for near unit-root processes with implications for term structure estimation. *Journal of Financial Econometrics*.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods: Second Edition*. Number 9780199641178 in OUP Catalogue. Oxford University Press.

Filipovic, D. and J. Teichmann (2002). On finite-dimensional term structure models. Working paper, Princeton University.

Gallant, A. R. and G. Tauchen (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association 93*, 10–24.

Gelb, A., J. F. Kasper, R. A. Nash, C. F. Price, and A. A. Sutherland (Eds.) (1974). *Applied Optimal Estimation*. Cambridge, MA: MIT Press.

Gourieroux, C. and A. Monfort (2008, August). Quadratic stochastic intensity and prospective mortality tables. *Insurance: Mathematics and Economics 43*(1), 174–184.

Gourieroux, C., A. Monfort, and R. Sufana (2010, December). International money and stock market contingent claims. *Journal of International Money and Finance 29*(8), 1727–1751.

Gourieroux, C. and R. Sufana (2011, June). Discrete time wishart term structure models. *Journal of Economic Dynamics and Control 35*(6), 815–824.

Gustafsson, F. and G. Hendeby (2012). Some Relations Between Extended and Unscented Kalman Filters. *IEEE Transactions on Signal Processing 60*(2), 545–555.

Harvey, A. C. (1991, November). *Forecasting, Structural Time Series Models and the Kalman Filter*. Number 9780521405737 in Cambridge Books. Cambridge University Press.

Harville, D. (1997). *Matrix Algebra from a statistician's perspective*. Springer.

Hendeby, G. (2008). *Performance and implementation of non-linear filtering*. Department of electrical engineering, Linkopings University.

Holmes, S., G. Klein, and D. M. Murray (2008). A square root unscented kalman filter for visual monoslam. Technical report, Active Vision Laboratory, Department of Engineering Science of Oxford.

Inci, A. C. and B. Lu (2004, June). Exchange rates and interest rates: can term structure models explain currency movements? *Journal of Economic Dynamics and Control 28*(8), 1595–1624.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*.

Jin, X. and J. M. Maheu (2013, March). Modeling realized covariances and returns. *Journal of Financial Econometrics 11*(2), 335–369.

Joslin, S., K. J. Singleton, and H. Zhu (2011). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies 24*(3), 926–970.

Julier, S. J. (2002). The scaled unscented transformation. In *in Proc. IEEE Amer. Control Conf*, pp. 4555–4559.

Julier, S. J. (2003). The spherical simplex unscented transformation. In *Proceedings of the American Control Conference*.

Julier, S. J. and J. K. Uhlmann (2004). Unscented filtering and nonlinear estimation. In *Proceedings of the IEEE*, pp. 401–422.

Julier, S. J., J. K. Uhlmann, and H. F. Durrant-Whyte (2000). A new approach for filtering nonlinear systems. Technical report, Robotic Research Group.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering 82*(Series D), 35–45.

Kandepu, R., B. Foss, and L. Imsland (2008). Applying the unscented kalman filter for nonlinear state estimation. *Journal of Process Control*.

Karaboga, D. and B. Basturk (2007, November). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *J. of Global Optimization 39*(3), 459–471.

Kim, D. H. and K. J. Singleton (2012). Term structure models and the zero-lower bound: A empirical investigation of japanese yields. Working paper, Stanford University.

Kim, D. H. and J. H. Wright (2005). An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. Technical report.

Kuha, J. and J. Temple (2003, April). Covariate measurement error in quadratic regression. *International Statistical Review 71*(1), 131–150.

Kukush, A., I. Markovsky, and S. V. Huffel (2002, November). Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. *Computational Statistics & Data Analysis 41*(1), 3–18.

Leippold, M. and L. Wu (2007). Design and estimation of multi-currency quadratic models. *Review of Finance*.

Li, H. and F. Zhao (2006, 02). Unspanned stochastic volatility: Evidence from hedging interest rate derivatives. *Journal of Finance 61*(1), 341–378.

Lund, J. (1997). Non-linear kalman filtering techniques for term-structure models. Technical report.

Magnus, J. and H. Neudecker (1980). *The Elimination Matrix: Some Lemmas and Applications*. Society for Industrial and Applied Mathematics. Journal on Algebraic and Discrete Methods.

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

Maybeck, P. S. (1982). *Stochastic Models, Estimation, and Control - Vol.2*.

Merwe, R. V. D. and E. A. Wan (2001). The square-root unscented kalman filter for state and parameter-estimation. In *in International Conference on Acoustics, Speech, and Signal Processing*, pp. 3461–3464.

Pelgrin, F. and M. Juillard (2004, August). Which order is too much? an application to a model with staggered price and wage contratcs. Computing in Economics and Finance 2004 58, Society for Computational Economics.

Philipov, A. and M. Glickman (2006). Factor multivariate stochastic volatility via wishart processes. *Econometric Reviews 25*(2-3), 311–334.

Rinnergschwentner, W., G. Tappeiner, and J. Walde (2011, September). Multivariate stochastic volatility via wishart processes: A comment. *Journal of Business & Economic Statistics 30*(1), 164–164.

Romo, J. M. (2012). Worst-of options and correlation skew under a stochastic correlation framework. *International Journal of Theoretical and Applied Finance (IJTAF) 15*(07), 1250051–1–1.

Tanizaki, H. (1996). *Nonlinear Filters: Estimation and Applications*.

Wolter, K. M. and W. A. Fuller (1982). Estimation of the quadratic errors-in-variables model. *Biometrika 69*, 175–182.