NONLINEAR FILTERING

IN ECONOMETRIC MODELS

Rolf Henriksen

Econometric Research Program
Research Memorandum No. 240

April 1979

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

Abstract

Equations are derived for the evolution of the conditional mean and covariance matrix, the extended Kalman filter, the truncated second-order nonlinear filter, and the iterated linear filter-smoother for nonlinear econometric models in structural form. It is shown that in order to get a computationally feasible solution for the truncated second-order filter, a stronger set of assumptions than usually occur will have to be made. The derivation of this filter also revealed an error which usually appears in the engineering literature on truncated second-order nonlinear filters.

1. INTRODUCTION

The use of Kalman filtering and related algorithms has been widely appreciated not only among engineers but also among economists in the last decade. While mostly the linear Kalman filter has been used in economics, several nonlinear filters have been developed and used in engineering. Unfortunately, none of these filters are directly applicable to nonlinear econometric models in structural form since this model structure has almost never been assumed in engineering problems. The purpose of this paper is to develop nonlinear filters for econometric models in structural form, and to show what types of filters that actually will be computationally infeasible.

The derivations in this paper are primarily based on Taylor series expansions and normal approximations of some probability densities, and follow in that sense along the same lines of attack as Jazwinski [1]. It will be demonstrated, however, that some stronger assumptions will have to be made in order to get a computationally feasible form of the truncated second-order nonlinear filter. Some errors in Jazwinski's derivation will also be pointed out and a more correct form of this filter will be given.

The paper is organized as follows. In Section 2 we give a description of the model being used and show how this model is related to other models being used in econometrics. In Section 3 we derive the exact equations for the evolution of the conditional mean and covariance matrix of the state vector. The extended Kalman filter is derived in Section 4.1, the truncated second-order nonlinear filter in Section 4.2, and the iterated linear filter-smoother in Section 4.3. Sone conclusions follow in Section 5. A rationale for the use of nonlinear filtering techniques is given in Appendix A while a detailed derivation of the iterated filter-smoother based on an innovations approach is given in Appendix B.

2. DESCRIPTION OF THE MODEL

Consider a nonlinear stochastic econometric model of the form

$$y_t = \phi(y_t, y_{t-1}, x_t) + \Psi(y_{t-1}, x_t)\varepsilon_t \tag{1}$$

where y_t is the vector of endogenous variables at time t which are explained by the model, x_t is a vector of exogenous variables which may include both control variables and variables not subject to control, while ε_t is a vector of disturbance terms assumed to be zero mean and independently distributed through time. $\phi(y_t, y_{t-1}, x_t)$ is a vector function with components $\phi^i(y_t, y_{t-1}, x_t)$ while $\Psi(y_{t-1}, x_t)$ is a matrix function with elements $\Psi_{ij}(y_{t-1}, x_t)$.

The model above is a system of implicit or simultaneous stochastic difference equations where the i'th equation in the structure is of the form

$$y_{i,t} = \phi^i(y_t, y_{t-1}, x_t) + \Psi^i(y_{t-1}, x_t)\varepsilon_t \tag{2}$$

where $\Psi^i(y_{t-1}, x_t)$ is the i'th row of $\Psi(y_{t-1}, x_t)$. The set of equations (1) will also be referred to as the system of structural equations. Although we have assumed no lags of order greater than one, there is actually no loss in generality. In fact, lagged endogenous variables dated prior to $t-1$ are assumed to be eliminated by introducing identities of the form $y_{j,t} = y_{k,t-1}$ while lagged exogenous variables have been eliminated by identities of the form $y_{\ell,t} = x_{m,t-1}$.

The form of the model is quite similar to the one used by Chow et al. [2], [3]. However, contrary to what is most commonly used in econometric models, we have allowed the disturbance terms ε_t to be multiplied by a matrix function $\Psi(y_{t-1}, x_t)$ instead of being purely additive. The reason for this is twofold. First, the model becomes more general and allows us to take into account a larger class of disturbances, e.g., heteroscedasity. Second, although seemingly a more complicated model than usually assumed, we shall see that all results in this paper can be obtained with almost no additional effort compared with the simpler model.

The system of structural equations may also contain models of time-varying parameters, e.g.,

$$\alpha_t = \alpha_{t-1} + \varepsilon_{\alpha,t} \tag{3}$$

or even constant parameters, i.e.,

$$\beta_t = \beta_{t-1} \tag{4}$$

The vector y_t , usually referred to as the state vector in control theory, may therefore include both endogenous variables, time-varying parameters, and constant parameters.

Observations of the endogenous variables y_t is assumed to be imperfect in the sense that they cannot be observed without error. We assume the observation device to be described by

$$z_t = h(y_t,t) + u_t \tag{5}$$

where h may be both nonlinear and, as indicated, time dependent. u_t is a vector of disturbance terms assumed to be zero mean and independently distributed through time. The processes $\{\varepsilon_t\}$ and $\{u_t\}$ are furthermore assumed to be independent.

The Errors-In-Variables Model (EVM) generally includes the possibility that also the exogenous variables are only measured with errors. Since we have not allowed this possibility in Equation (5), it seems like we would have to assume an observation equation of the form

$$z_t = h(y_t,x_t,t) + u_t$$

in order to allow the exogenous variables to be only partially observable. However, by including only partially observable exogenous variables in y_t, we can readily avoid this problem. This may introduce a slight definitional problem on the part of y_t since we have referred to y_t as the state vector. In order to be a state vector, y_t will, strictly speaking, have to be minimal in the sense that the system of equations (1) is invertible, i.e., solvable with respect to y_{t-1}. Of course, if some or all of the exogenous variables are independent and serially uncorrelated, this property cannot generally be achieved. However, what we are going to develop in the sequel will be independent of whether Equation (1) is a strict state representation or not. Partially observable exogenous variables are therefore included in y_t so that x_t consists solely of perfectly observable exogenous variables.

The two basic assumptions of EVM in structural form have traditionally been normality and serial uncorrelatedness of the explanatory variables. Mehra [4] has shown that if the assumption of serial correlatedness is relaxed, the EVM becomes identifiable[1)] for normal distributions. Serial uncorrelatedness simply means that the explanatory variables must be generated by some dynamical system, and a model of this system may easily be included in our model. It is therefore believed that the type of model which has been assumed will provide a fairly general framework for many applications in econometrics.

---

1) Identifiable is here used in the control theory meaning of the word, see Mehra [5] for an explanation of similarities and differences between this and other terms used in control theory and econometrics.

3. THE EXACT EQUATIONS FOR THE EVOLUTION OF THE CONDITIONAL MEAN AND COVARIANCE MATRIX

Define Z_t to be the sequence of observations $z_0, z_1, \ldots, z_t$, viz.

$$Z_t = (z_0, z_1, \ldots, z_t)$$

and define $\hat{y}_{t|\tau}$ to be the conditional mean of y_t given Z_τ, i.e.,

$$\hat{y}_{t|\tau} = E(y_t | Z_\tau) \ , \tag{6}$$

and let $Y_{t|\tau}$ denote the covariance matrix of y_t given Z_τ, i.e.

$$Y_{t|\tau} = E((y_t - \hat{y}_{t|\tau})(y_t - \hat{y}_{t|\tau})' | Z_\tau) \tag{7}$$

Finally, let

$$E(\varepsilon_t \varepsilon_\tau') = Q_t \delta_{t\tau}, \ E(u_t u_t') = R_t \delta_{t\tau}$$

where $\delta_{t\tau}$ is the Kronecker delta. We shall assume the random vectors ε_t and u_t, for all $t \geq 0$, to be independent of y_0 and the sequence of exogenous variables $\{x_0, x_1, \ldots, x_t\}$ up to time t.

Now, assume $\hat{y}_{t-1|t-1}$ and $Y_{t-1|t-1}$ to be known. We seek expressions for $\hat{y}_{t|t-1}$ and $Y_{t|t-1}$. From Equation (1) we readily obtain

$$\hat{y}_{t|t-1} = E(y_t | Z_{t-1}) = \hat{\phi}_{t-1} \tag{8}$$

where

$$\hat{\phi}_{t-1} = E(\phi(y_t, y_{t-1}, x_t) | Z_{t-1}) \tag{9}$$

Introducing the notation

$$E_t(.) \triangleq E(. | Z_t) \ , \tag{10}$$

we can also write Equation (8) as

$$\hat{y}_{t|t-1} = E_{t-1}(\phi) \quad ^{1)} \tag{11}$$

By multiplying both sides of Equation (1) with y_t' and taking expectations, we find

$$Y_{t|t-1} = E_{t-1}(\phi y_t') + E_{t-1}(\Psi \varepsilon_t y_t') - \hat{\phi}_{t-1}\hat{y}_{t|t-1}' \tag{12}$$

where we have used the same abbreviated type of notations as in Equation (11). Substituting for y_t' in Equation (12) we obtain

$$Y_{t|t-1} = E_{t-1}(\phi\phi') + E_{t-1}(\phi\varepsilon_t'\Psi') + E_{t-1}(\Psi\varepsilon_t\phi') + E_{t-1}(\Psi\Omega_t\Psi') - \hat{\phi}_{t-1}\hat{\phi}_{t-1}' \tag{13}$$

In case the model is explicit, i.e., in <u>reduced form</u> ϕ will be independent of y_t and both Equation (8) and Equation (13) will be explicit equations for, respectively, $\hat{y}_{t|t-1}$ and $Y_{t|t-1}$. However, since ϕ generally is a function of y_t , we should expect Equation (13) to be even more complicated than Equation (12), so there is actually no improvement in going from Equation (12) to Equation (13).

The models which until now have been used in engineering are almost exclusively in reduced form or, in continuous time, described by differential equations. Models in structural form have therefore attracted very little attention, which may explain why nonlinear filtering in implicit models is a seemingly unknown concept in control theory. With a model in reduced form, Equation (13) would be preferred to Equation (12) since the right-hand side of the latter is a function of known variables. Since both ϕ and Ψ in that case would be independent of ε_t , the second and third

---

1) This is a somewhat sloppy notation since we have not listed the arguments of ϕ which may be any triple $(y_\tau, y_{\tau-1}, x_\tau)$. However, the meaning of the notation should be obvious from the context.

right-hand terms would also drop out, leaving a somewhat simpler expression for $Y_{t|t-1}$. For models in structural form, however, Equation (12) will usually be preferred since it is believed to be far simpler in most cases.

In order to make the previous set of equations complete we finally need expressions for $\hat{y}_{t|t}$ and $Y_{t|t}$. Letting $p(y_t|Z_t)$ denote the (generalized) probability density function of y_t given Z_t, we have

$$\hat{y}_{t|t} = E_t(y_t|Z_t) = \int y_t p(y_t|Z_t)dy_t \tag{14}$$

From Bayes' rule

$$p(y_t|Z_t) = \frac{p(z_t|y_t,Z_{t-1})p(y_t|Z_{t-1})}{\int p(z_t|\xi,Z_{t-1})p(\xi|Z_{t-1})d\xi}$$

where

$$p(z_t|y_t,Z_{t-1}) = p(z_t|y_t)$$

by the assumptions on the processes $\{\varepsilon_t\}$ and $\{u_t\}$. Substituting into Equation (14) we end up with

$$\hat{y}_{t|t} = [E_{t-1}(p(z_t|y_t))]^{-1}E_{t-1}(y_t p(z_t|y_t)) \tag{15}$$

In a similar way, noting that

$$Y_{t|t} = E_t(y_t y_t'|Z_t) - \hat{y}_{t|t}\hat{y}_{t|t}' = \int y_t y_t' p(y_t|Z_t)dy_t - \hat{y}_{t|t}\hat{y}_{t|t}',$$

we find

$$Y_{t|t} = [E_{t-1}(p(z_t|y_t))]^{-1}E_{t-1}(y_t y_t' p(z_t|y_t)) - \hat{y}_{t|t}\hat{y}_{t|t}' \tag{16}$$

Although conceptually not so hard to derive, it should be noted that the right-hand sides of Equations (8), (12), and (13) generally involve expectations that require the whole conditional probability density function for their evaluation, i.e., all conditional moments. The first two moments of the conditional pdf therefore generally depend on all the other moments, and in order to obtain a computationally realizable filter, some approximations will have to be done. Equations (8), (12), and (13) are solvable only when Equations (1) and (5) are linear in which case we obtain the Kalman filter.

4. APPROXIMATE NONLINEAR FILTERS

4.1 Extended Kalman Filter

The extended Kalman filter can be derived by expanding the functions ϕ , Ψ , and h in Taylor series around $\hat{y}_{t|t-1}$ and $\hat{y}_{t-1|t-1}$ and neglecting second and higher order terms. We find

$$\phi(y_t,y_{t-1},x_t) \approx \phi(t) + \phi_1(t)(y_t-\hat{y}_{t|t-1}) + \phi_2(t)(y_{t-1}-\hat{y}_{t-1|t-1}) \tag{16}$$

where we have introduced the notations

$$\phi(t) = \phi(\hat{y}_{t|t-1},\hat{y}_{t-1|t-1},x_t)$$

$$\phi_1(t)= \frac{\partial\phi}{\partial y_t'}(\hat{y}_{t|t-1},\hat{y}_{t-1|t-1},x_t)$$

$$\phi_2(t)= \frac{\partial\phi}{\partial y_{t-1}'}(\hat{y}_{t|t-1},\hat{y}_{t-1|t-1},x_t)$$

Similarly,

$$\Psi(y_{t-1},x_t) \approx \Psi(t) + \Psi_y(t): (y_{t-1}-\hat{y}_{t-1|t-1}) \tag{17}$$

$$h(y_t,t) \approx h(t) + h_y(t)(y_t-\hat{y}_{t|t-1}) \tag{18}$$

where

$$\Psi(t) = \Psi(\hat{y}_{t-1|t-1},x_t) \ , \ \Psi_y(t) = \frac{\partial\Psi}{\partial y_{t-1}'}(\hat{y}_{t-1|t-1},x_t)$$

$$h(t) = h(\hat{y}_{t|t-1},t) \ , \quad h_y(t) = \frac{\partial h}{\partial y_t'}(\hat{y}_{t|t-1},t)$$

Notice that $\Psi_y(t)$ is a tensor and that the matrix $\Psi_y(t):(y_{t-1}-\hat{y}_{t-1|t-1})$ is given by

$$(\Psi_y(t):(y_{t-1}-\hat{y}_{t-1|t-1}))_{ij}=\frac{\partial\Psi_{ij}}{\partial y'_{t-1}}(\hat{y}_{t-1|t-1},x_t)(y_{t-1}-\hat{y}_{t-1|t-1})$$

Substituting these approximations into Equation (8) we find

$$\hat{y}_{t|t-1} = \phi(t) = \phi(\hat{y}_{t|t-1},\hat{y}_{t-1|t-1},x_t) \tag{19}$$

In order to derive an approximate equation for the covariance matrix $Y_{t|t-1}$ we shall assume that all third and higher order moments can be neglected in all joint and marginal probability laws of the system, e.g., the probability law of y_t or the joint probability law of y_t , y_{t-1} , and ε_t. Define

$$C_{t|t-1} = E_{t-1}((y_t-\hat{y}_{t|t-1})(y_{t-1}-\hat{y}_{t-1|t-1})') = E_{t-1}(\phi y'_{t-1})-\hat{\phi}_{t-1}\hat{y}'_{t-1|t-1} \tag{20}$$

$$D_{t|t-1} = E_{t-1}((y_t-y_{t|t-1})\varepsilon_t') = E_{t-1}(\phi\varepsilon_t') + E_{t-1}(\Psi\Omega_t) \tag{21}$$

Substituting the previous approximations into Equations (12), (20), and (21) we find the approximate covariance equations to be

$$Y_{t|t-1} = \phi_1(t)Y_{t|t-1} + \phi_2(t)C_{t|t-1} + \Psi(t)D_{t|t-1} \tag{22}$$

$$C_{t|t-1} = \phi_1(t)C_{t|t-1} + \phi_2(t)Y_{t-1|t-1} \tag{23}$$

$$D_{t|t-1} = \phi_1(t)D_{t|t-1} + \Psi(t)\Omega_t \tag{24}$$

From Equations (23) and (24)

$$C_{t|t-1} = (I - \phi_1(t))^{-1}\phi_2(t)Y_{t-1|t-1} \tag{25}$$

$$D_{t|t-1} = (I - \phi_1(t))^{-1}\Psi(t)\Omega_t \tag{26}$$

while, from Equation (22),

$$Y_{t|t-1} = (I - \phi_1(t))^{-1}[\phi_2(t)C'_{t|t-1} + \Psi(t)D_{t|t-1}] \tag{27}$$

or, in terms of $Y_{t-1|t-1}$ and Q_t,

$$Y_{t|t-1} = (I - \phi_1(t))^{-1}[\phi_2(t)Y_{t-1|t-1}\phi'_2(t)$$

$$+ \Psi(t)Q_t\Psi'(t)](I - \phi'_1(t))^{-1} \tag{28}$$

Another form of the covariance equation can be derived from Equation (13). Substituting the same approximations for ϕ and Ψ as previously we find

$$Y_{t|t-1} = \phi_1(t)Y_{t|t-1}\phi'_1(t) + \phi_2(t)Y_{t-1|t-1}\phi'_2(t)$$

$$+ \phi_1(t)C_{t|t-1}\phi'_2(t) + \phi_2(t)C'_{t|t-1}\phi'_1(t)$$

$$+ \phi_1(t)D_{t|t-1}\Psi'(t) + \Psi(t)D'_{t|t-1}\phi'_1(t)$$

$$+ \Psi(t)Q_t\Psi'(t) \tag{29}$$

where we have used the fact that the term $E_t(\Psi Q_t\Psi')$ in Equation (13) can be approximated as

$$E_{t-1}(\Psi Q_t\Psi') \approx E_{t-1}(\Psi(t)Q_t\Psi'(t))$$

$$+ E_{t-1}((\Psi_y(t):(y_{t-1}-\hat{y}_{t-1|t-1}))Q_t(\Psi_y(t):(y_{t-1}-\hat{y}_{t-1|t-1}))')$$

$$= \Psi(t)Q_t\Psi'(t)$$

$$+ E_{t-1}((\Psi_y(t):(y_{t-1}-\hat{y}_{t-1|t-1}))\varepsilon_t\varepsilon'_t(\Psi_y(t):(y_{t-1}-\hat{y}_{t-1|t-1}))')$$

$$\approx \Psi(t)Q_t\Psi'(t)$$

by the independency of y_{t-1} and ε_t, and the assumptions made prior to this derivation. It is easily verified that the solution of Equation (29) actually is given by Equation (28).

The previous equations enable us to compute the one-stage predicted estimate $\hat{y}_{t|t-1}$ and its covariance matrix $Y_{t|t-1}$. In order to obtain the filtered estimate $\hat{y}_{t|t}$ and the covariance matrix $Y_{t|t}$ we shall use a normal approximation and simply compute the first and second-order moments of the posterior pdf $p(y_t|Z_t)$. Using the same kind of approximations as previously we find

$$E_{t-1}(z_t) = h(\hat{y}_{t|t-1},t) \triangleq h(t)$$

$$E_{t-1}((z_t-h(t))(z_t-h(t))') = h_y(t)Y_{t|t-1}h_y'(t) + R_t$$

$$E_{t-1}((y_t-\hat{y}_{t|t-1})(z_t-h(t))') = Y_{t|t-1}h_y'(t)$$

By applying a well-known formula, cited in, for example, Jazwinski [1], p. 45, we find

$$\hat{y}_{t|t} = \hat{y}_{t|t-1} + K_t(z_t-h(t)) \tag{30}$$

$$Y_{t|t} = (I - K_t h_y(t))Y_{t|t-1} \tag{31}$$

where

$$K_t = Y_{t|t-1}h_y'(t)[h_y(t)Y_{t|t-1}h_y'(t) + R_t]^{-1} = Y_{t|t}h_y'(t)R_t^{-1} \tag{32}$$

Equations (19), (28), and (30) - (32) constitute the _extended Kalman filter_ for a model in structural form. The extended Kalman filter for a reduced-form model can be obtained by noting that ϕ is independent of y_t in that case. Then, simply by substituting $\phi(\hat{y}_{t-1|t-1},x_t)$ for $\phi(\hat{y}_{t|t-1},\hat{y}_{t-1|t-1})$ in Equation (19) and putting $\phi_1(t) = 0$ in the other equations, we have the extended Kalman filter for a reduced-form model.

4.2 Truncated Second-Order Filter

The so-called truncated second-order filter was developed for reduced-form models by Jazwinski [6], [7], and independently by Bass, et al. [8]. Nonlinearities in the functions ϕ , Ψ , and h are in this filter carried only to second order while third and higher order central moments of the vectors y_{t-1} and ε_{t-1} given Z_{t-1} are being neglected. With a model in structural form we generally have to take into consideration nonlinearities in ϕ with respect to y_t . As we shall see later on, third and higher order central moments involving y_t given Z_t cannot generally be ignored without making some stronger assumptions than one has to do with a reduced-form model. We shall, however, always assume fifth and higher order central moments of any kind to be negligible, and the magnitude of such terms will not be discussed in what follows.

As in the previous case, let us start by assuming that third and higher order central moments of all probability laws involving any number of the vectors $\{y_t, y_{t-1}, \varepsilon_t\}$ can be neglected. Some important consequences of this assumption will be outlined in the sequel.

Now, let ξ and η be arbitrary n-vectors (y_t is assumed to be an n-vector), and let us introduce the notations

$$(\xi\xi'\phi_{11}(t))_i = \sum_{j,\ell=1}^{n} \xi_j \xi_\ell \frac{\partial^2 \phi^i}{\partial y_{j,t} \partial y_{\ell,t}} (\hat{y}_{t|t-1}, \hat{y}_{t-1|t-1}, x_t) \tag{33}$$

$$(\xi\xi'\phi_{22}(t))_i = \sum_{j,\ell=1}^{n} \xi_j \xi_\ell \frac{\partial^2 \phi^i}{\partial y_{j,t-1} \partial y_{\ell,t-1}} (\hat{y}_{t|t-1}, \hat{y}_{t|t}, x_t) \tag{34}$$

$$(\xi\eta'\phi_{12}(t))_i = \sum_{j,\ell=1}^{n} \xi_j \eta_\ell \frac{\partial^2 \phi^i}{\partial y_{j,t} \partial y_{\ell,t-1}} (\hat{y}_{t|t-1}, \hat{y}_{t|t}, x_t) \tag{35}$$

$$(\xi\xi'\phi_{11}(t)) = ((\xi\xi'\phi_{11}(t))_1, \ldots, (\xi\xi'\phi_{11}(t))_n)' \tag{36}$$

$$(\xi\xi'\phi_{22}(t)) = ((\xi\xi'\phi_{22}(t))_1, \ldots, (\xi\xi'\phi_{22}(t))_n)' \tag{37}$$

$$(\xi\eta'\phi_{12}(t)) = ((\xi\eta'\phi_{12}(t))_1,\ldots,(\xi\eta'\phi_{12}(t))_n)' \tag{38}$$

Furthermore, if A is an arbitrary $n\times n$ - matrix with elements A_{ij}, we similarly write

$$(A\phi_{11}(t))_i = \sum_{j,\ell=1}^{n} A_{j\ell}\frac{\partial^2\phi^i}{\partial y_{j,t}\partial y_{\ell,t}}(\hat{y}_{t|t-1},\hat{y}_{t-1|t-1},x_t) \tag{39}$$

$$(A\phi_{11}(t)) = ((A\phi_{11}(t))_1,\ldots,(A\phi_{11}(t))_n)' \tag{40}$$

and so on.

By expanding ϕ in a Taylor series around $\hat{y}_{t|t-1}$ and $\hat{y}_{t|t}$, carrying the nonlinearities to second order only, we find

$$\hat{y}_{t|t-1} = \phi(t) + \tfrac{1}{2}(Y_{t|t-1}\phi_{11}(t)) + (C_{t|t-1}\phi_{12}(t)) + \tfrac{1}{2}(Y_{t-1|t-1}\phi_{22}(t)) \tag{41}$$

which is the expression for the one-stage predicted estimate of y_t in the truncated second-order filter.

In order to develop the approximate equation for $Y_{t|t-1}$, let us first use Equation (12). Again, carrying the nonlinearities only to second order, substituting $\hat{y}_{t|t-1}$ from Equation (41) for $\hat{\phi}_{t-1}$, and ignoring all third and higher order central moments, we end up with

$$Y_{t|t-1} = (I - \phi_1(t))^{-1}[\phi_2(t)Y_{t-1|t-1}\phi_2'(t) + \Psi(t)\Omega_t\Psi'(t)](I - \Omega_1'(t))^{-1} \tag{42}$$

$$C_{t|t-1} = (I - \phi_1(t))^{-1}\phi_2(t)Y_{t-1|t-1} \tag{43}$$

$$D_{t|t-1} = (I - \phi_1(t))^{-1}\Psi(t)\Omega_t \tag{44}$$

These are exactly the same equations that were derived for the extended Kalman filter, viz. Equations (28), (25), and (26) respectively.

Equation (42) can also be derived from Equation (13) using the same kind of approximations as previously. However, in order to do this we shall first have to make an observation which is closely related to our previous assumptions. Substituting $\hat{y}_{t|t-1}$ (as given by Equation (41)) for $\hat{\phi}_{t-1}$ in Equation (13), we obtain a lot of "squared" second order moments, e.g., expressions like $\frac{1}{4}(Y_{t|t-1}\phi_{11}(t))(Y_{t|t-1}\phi_{11}(t))'$, $\frac{1}{4}(Y_{t-1|t-1}\phi_{22}(t))(Y_{t-1|t-1}\phi_{22}(t))'$, or $\frac{1}{4}(Y_{t|t-1}\phi_{11}(t))(Y_{t-1|t-1}\phi_{22}(t))'$. Now, letting ξ be a normal random variable, $\xi \sim N(\bar{\xi},\sigma^2)$, we know that $E((\xi-\bar{\xi})^4) = 3\sigma^4 > (\sigma^2)^2$ so that the fourth-order central moment is larger than the square of the second-order central moment. The following proposition states that this is in fact true for any random variable.

Proposition

Let ξ be a random variable with mean $\bar{\xi}$. Then

$$E((\xi-\bar{\xi})^4) \geq (E((\xi-\bar{\xi})^2))^2 \tag{45}$$

Proof. By Jensen's Lemma (cited and proved in Sworder [9], pp. 27-28), any convex real valued function f of a random vector r satisfies the inequality $E(f(r)) \geq f(E(r))$. Taking $(.)^2$ to be f and ξ to be r , we immediately have the result. □

Having made the observation stated in the previous proposition, we easily see the consequences of assuming that all fourth-order central moments are negligible. All "squared" second-order moments will also have to be neglected including, by our previous assumptions, all products of second-order moments. Doing that, we find the proper approximation of $\hat{y}_{t|t-1}\hat{y}'_{t|t-1}$ to be

$$\hat{Y}_{t|t-1}\hat{Y}'_{t|t-1} \approx \phi(t)\phi'(t) + \tfrac{1}{2}\phi(t)[(Y_{t|t-1}\phi_{11}(t))' + 2(C_{t|t-1}\phi_{12}(t))' + (Y_{t-1|t-1}\phi_{22}(t))'] + \tfrac{1}{2}[(Y_{t|t-1}\phi_{11}(t)) + 2(C_{t|t-1}\phi_{12}(t)) + (Y_{t-1|t-1}\phi_{22}(t))]\phi'(t)$$

This point has been overlooked in many of the papers and books on nonlinear filtering which have appeared in the engineering literature, and even Jazwinsky [1], [6], [7] makes the same error by ignoring fourth-order central moments of the state while retaining the square of second-order central moments.

Substituting the Taylor series expansions for ϕ and Ψ into Equation (13) we find

$$Y_{t|t-1} = \phi_1(t)Y_{t|t-1}\phi_1'(t) + \phi_2(t)Y_{t-1|t-1}\phi_2'(t) + \phi_1(t)C_{t|t-1}\phi_2'(t) + \phi_2(t)C'_{t|t-1}\phi_1'(t) + \phi_1(t)D_{t|t-1}\Psi'(t) + \Psi(t)D'_{t|t-1}\phi_1(t) + \Psi(t)Q_t\Psi'(t) \quad , \tag{46}$$

the solution of which, by referring to the extended Kalman filter, is given by Equation (42).

Subject to the assumption that all third and higher order central moments of all the probability laws which are involved in computing the one-stage predictions can be neglected, we ended up with the prediction part of the truncated second-order filter to consist of Equations (41) - (44). The only difference from the extended Kalman filter is the equation for the conditional mean $\hat{Y}_{t|t-1}$. However, since the system of equations is coupled, we will generally not obtain the same solution in the two cases.

Derivation of the equations for the filtered estimates $\hat{y}_{t|t}$ and $Y_{t|t}$ can be carried out along several different lines of attack, e.g., by assuming that only first-order terms of z_t are needed in these equations or that first-order terms are needed in the equation for $\hat{y}_{t|t}$ while $Y_{t|t}$ can be assumed to be independent of z_t. Intuitively, the latter seems to be fairly attractive since this is the form of the Kalman filter, and it can be argued both on computational and stability grounds that the latter is superior. However, the author of this paper has shown [24] that this is the correct form of the filter, no matter which of the two forms we assume. It is easily seen that this form of the filtering equations can be derived by assuming normal distributions and simply computing the first and second-order moments of the posterior distribution, carrying the nonlinearities in h to second-order. Doing this, we end up with

$$\hat{y}_{t|t} = \hat{y}_{t|t-1} + K_t(z_t - h(t) - \frac{1}{2}(Y_{t|t-1}h_{yy}(t))) \tag{47}$$

$$Y_{t|t} = (I - K_t h_y(t))Y_{t|t-1} \tag{48}$$

where

$$\begin{aligned} K_t &= Y_{t|t-1}h'_y(t)[h_y(t)Y_{t|t-1}h'_y(t) + R_t]^{-1} \\ &= Y_{t|t}h'_y(t)R_t^{-1} \end{aligned} \tag{49}$$

A similar error as previously appears in the filtering equations in Jazwinsky [1], pp. 364-365, where the covariance matrix of $z_t - h(y_t,t)$ given Z_{t-1} is found to be

$$h_y(t)Y_{t|t-1}h'_y(t) + R_t - \frac{1}{4}(Y_{t|t-1}h_{yy}(t))(Y_{t|t-1}h_{yy}(t))'$$

However, a careful computation using the previous proposition actually shows that the proper approximation of this covariance matrix is

$$h_y(t)Y_{t|t-1}h_y'(t) + R_t$$

The form of the truncated second-order filter which we have developed is given by Equations (41) - (45) and Equations (47) - (49). It was derived by assuming that third and higher order central moments of the involved probability laws could be neglected. The filter is, except for the equations for $\hat{y}_{t|t-1}$ and $\hat{y}_{t|t}$, identical to the extended Kalman filter which was derived in the previous section. This similarity originates from our assumption about higher order central moments.

Now, before we terminate this section, let us investigate more thoroughly what the assumptions underlying the truncated second-order filter really amount to. In order to do this we shall use Equation (12) by carrying nonlinearities to second order. For the sake of simplicity, let us assume the system to be scalar since we are only going to show the order of magnitude of the neglected terms. Developing the term $E_{t-1}(\phi y_t)$ we find

$$\begin{aligned} E_{t-1}(\phi y_t) = E_{t-1}([\phi(t) &+ \phi_1(t)(y_t - \hat{y}_{t|t-1}) \\ &+ \phi_2(t)(y_{t-1} - \hat{y}_{t-1|t-1}) + \tfrac{1}{2}\phi_{11}(t)(y_t - \hat{y}_{t|t-1})^2 \\ &+ \tfrac{1}{2}\phi_{22}(t)(y_{t-1} - \hat{y}_{t-1})^2 + \phi_{12}(t)(y_{t-1} - \hat{y}_{t-1|t-1})(y_t - \hat{y}_{t|t-1})] \\ &\times [\hat{y}_{t|t-1} + (y_t - \hat{y}_{t|t-1}])) \end{aligned} \tag{50}$$

Equation (1) defines y_t implicitly as a function of y_{t-1}, x_t, and ε_t, viz.

$$y_t = \gamma(y_{t-1}, x_t, \varepsilon_t)$$

which when linearized around $\varepsilon_t = 0$ becomes of the form

$$y_t = f(y_{t-1}, x_t) + G(y_{t-1}, x_t)\varepsilon_t \tag{51}$$

where G, whether Ψ is assumed to be independent of y_{t-1} or not, generally will be a function of y_{t-1}.

First, let us investigate the order of magnitude of the term $E_{t-1}((y_t - \hat{y}_{t|t-1})^3)$. Using Equation (51) and carrying the nonlinearities in f and G to first-order only for the sake of simplicity, we find

$$E_{t-1}((y_t - \hat{y}_{t|t-1})^3) \approx 4 f_y(t) G(t) G_y(t) Y_{t-1|t-1} \Omega_t$$

by ignoring third-order central moments of y_{t-1}. Thus, generally, the magnitude of $E_{t-1}((y_t - \hat{y}_{t|t-1})^3)$ is of the order $Y_{t-1|t-1}\Omega_t$, i.e., the product of the covariances of y_{t-1} and ε_t given Z_{t-1}.

The same thing applies to another of the higher order terms in Equation (50). We find

$$E_{t-1}((y_{t-1} - \hat{y}_{t-1|t-1})(y_t - \hat{y}_{t|t-1})^2) \approx 2G(t) G_y(t) Y_{t-1|t-1} \Omega_t$$

by the same approximation as previously.

Developing the term $E_{t-1}(\Psi \varepsilon_t y_t)$ in Equation (12) we obtain

$$E_{t-1}(\Psi \varepsilon_t y_t) = E_{t-1}([\Psi(t)\varepsilon_t + \Psi_y(t)(y_{t-1} - \hat{y}_{t-1|t-1})\varepsilon_t$$
$$+ \frac{1}{2}\Psi_{yy}(t)(y_{t-1} - \hat{y}_{t-1|t-1})^2 \varepsilon_t][\hat{y}_{t|t-1} + (y_t - \hat{y}_{t|t-1})]) \tag{52}$$

Again, carrying the nonlinearities in Equation (51) to first order, we find

$$E_{t-1}((y_{t-1} - \hat{y}_{t-1|t-1})\varepsilon_t (y_t - \hat{y}_{t|t-1})) = G_y(t) Y_{t-1|t-1} \Omega_t$$

$$E_{t-1}((y_{t-1} - \hat{y}_{t-1|t-1})^2 \varepsilon_t (y_t - \hat{y}_{t|t-1})) = G(t) Y_{t-1|t-1} \Omega_t$$

We have somewhat informally shown that the assumptions underlying the previous second-order filter are generally permissible only if products of the covariances of the vectors y_{t-1} and ε_t given Z_{t-1} are neglible.

In order to show that these assumptions actually are permissible if the above mentioned products are negligible, we would have to do a more careful computation. Carrying the nonlinearities in Equation (12) to second-order we obtain a lot of third and fourth-order central moment involving $(y_t - \hat{y}_{t|t-1})$, $(y_{t-1} - \hat{y}_{t-1|t-1})$ and ε_t . The order of magnitude of each of these moments could then be expressed as a linear function of terms involving $Y^2_{t-1|t-1}$, $Y_{t-1|t-1}\Omega_t$, and Ω^2_t , and third and fourth-order moments involving $(y_t - \hat{y}_{t|t-1})$ to some power greater than 0 . Doing this with every third and fourth-order moment involving $(y_t - \hat{y}_{t|t-1})$ we would obtain a linear set of equations where terms involving $(y_t - \hat{y}_{t|t-1})$ to some power greater than 0 would be the unknowns while terms involving $Y^2_{t-1|t-1}$, $Y_{t-1|t-1}\Omega_t$, and Ω^2_t would be known. All the unknown terms can therefore be expressed as linear functions of $Y^2_{t-1|t-1}$, $Y_{t-1|t-1}\Omega_t$, and Ω^2_t if we assume third-order central moments of y_{t-1} and ε_t given Z_{t-1} to be negligible. Generally, without having said anything about the functions ϕ and Ψ , we do not know whether the terms involving, for example, $Y_{t-1|t-1}\Omega_t$ actually are negligible. But if we make the assumption that terms involving $Y^2_{t-1|t-1}$, $Y_{t-1|t-1}\Omega_t$, and Ω^2_t are negligible, it is clearly seen that the assumptions underlying the previous second-order filter are permissible.

Theorem

Third and fourth-order central moments involving $(y_t - \hat{y}_{t|t-1})$, $(y_{t-1} - \hat{y}_{t-1|t-1})$, and ε_t are negligible, given Z_{t-1} , if and only if all terms involving products of elements of the matrix $[Y_{t-1|t-1} \mid \Omega_t]$ are negligible.

The assumptions underlying the previous second-order nonlinear filter are therefore permissible if and only if this theorem is satisfied. It should be noted that these assumptions are stronger than the ones which usually are made for reduced-form models. They are, however, necessary in order to get a computationally feasible solution. If we had been forced to take third and fourth-order central moments involving $(y_t-\hat{y}_{t|t-1})$, $(y_{t-1}-y_{t-1|t-1})$, and ε_t into account, the filter would have become so complicated that it is question-able whether it could ever be used for any practical purposes.

If ϕ is independent of y_t , i.e., $\phi_1(t) = 0$, we obtain the following form of the predictor:

$$\hat{y}_{t|t-1} = \phi(t) + \frac{1}{2}(Y_{t-1|t-1}\phi_{22}(t)) \tag{53}$$

$$Y_{t|t-1} = \phi_{22}(t)Y_{t-1|t-1}\phi_{22}'(t) + \Psi(t)\Omega_t\Psi'(t) \tag{54}$$

However, in this case it is not necessary to assume products of terms involving $Y_{t-1|t-1}$ and Ω_t to be negligible in order to get a computationally feasible solution. Retaining these terms, we find the covariance equation for a nonlinear reduced-form model to be of the form [1)]

$$\begin{aligned} Y_{t|t-1} = {} & \phi_{22}(t)Y_{t-1|t-1}\phi_{22}'(t) + \Psi(t)\Omega_t\Psi'(t) \\ & + (Y_{t-1|t-1}\Omega_t\Psi_y^2(t)) + \frac{1}{2}(Y_{t-1|t-1}\Omega_t\Psi_{yy}(t)\Psi(t)) \\ & + \frac{1}{2}(Y_{t-1|t-1}\Omega_t\Psi_{yy}(t)\Psi(t))' \end{aligned} \tag{55}$$

where the elements of the matrices $(Y_{t-1|t-1}\Omega_t\Psi_y^2(t))$ and $(Y_{t-1|t-1}\Omega_t\Psi_{yy}(t)\Psi(t))$ are given by

$$(Y_{t-1|t-1}\Omega_t\Psi_y^2(t))_{iq} = \sum_{j,\ell=1}^{s} \sum_{p,k=1}^{n} Y_{t-1|t-1}^{pk}\Omega_t^{j\ell} \frac{\partial\Psi_{ij}}{\partial y_{p,t-1}}(t)\frac{\partial\Psi_{q\ell}}{\partial y_{k,t-1}}(t) \tag{56}$$

$$(Y_{t-1|t-1}\Omega_t\Psi_{yy}(t)\Psi(t))_{iq} = \sum_{j,\ell=1}^{s} \sum_{p,k=1}^{n} Y_{t-1|t-1}^{pk}\Omega_t^{j\ell} \frac{\partial^2\Psi_{ij}}{\partial y_{p,t-1}\partial y_{k,t-1}}(t)\Psi_{q\ell}(t) \tag{57}$$

---

1) Subscript 2 still means the partial derivative of ϕ with respect to y_{t-1} .

We have here assumed y_t and ε_t to be n and s vectors, respectively. Components of y_t are denoted by $y_{p,t}$ while elements of the matrices $Y_{t|t}$, Q_t , and Ψ are denoted by, respectively, $Y^{pk}_{t|t}$, $Q^{j\ell}_t$, and Ψ_{ij} .

An additional term,

$$-\frac{1}{4}(Y_{t-1|t-1}\phi_{22}(t))(Y_{t-1|t-1}\phi_{22}(t))'$$

appears on the right-hand side of Equation (55) in Jazwinski [1]. By the previous proposition, however, we note that this term should be dropped since its order of magnitude is less than another term which has already been neglected. Equation (55) is therefore a more correct form of the covariance equation in case all terms involving products of elements of the matrices $Y_{t-1|t-1}$ and Q_t cannot be neglected.

A comparison between Equations (54) and (55) clearly shows which terms we actually have neglected by making the stronger set of assumptions that underlie our truncated second-order filter. However, since third and fourth-order central moments of y_{t-1} given Z_{t-1} are assumed to be negligible, we should also expect third and fourth-order central moments of ε_t to be negligible. Moreover, this would imply that the elements of the matrices $Y^2_{t-1|t-1}$ and Q^2_t also were negligible which finally indicates that the assumption being made about terms involving products of elements of the matrix $[Y_{t-1|t-1} \vdots Q_t]$ is quite reasonable. Of course, without having said anything about the functional forms of ϕ and Ψ , we cannot generally conclude that these terms are negligible even if third and higher order moments y_{t-1} and ε_t given Z_t actually are. But it is at least a quite reasonable assumption.

The type of truncated second-order filter which has been derived here has the same form of the covariance equations as the extended Kalman filter.

However, the equations for the conditional means are different and this is believed to be the most important part since the main reason for introducing higher order terms in these equations is to remove biases. The truncated second-order filter should therefore perform significantly better in many cases, e.g., when the nonlinearities are strong or the uncertainties are large.

In the so-called Gaussian second-order filter nonlinearities are taken to second-order while third-order central moments are neglected, and fourth-order central moments are approximated by making a normal approximation of the probability laws. Having seen the kind of assumptions which had to be made in order to get a computationally feasible solution for the truncated second-order filter, it is easily understood why the Gaussian second-order filter generally will become so extremely complicated. In addition to retaining the terms which were neglected in the truncated filter, we would also have to approximate all fourth-order central moments. Needless to say, the computational effort involved would probably cause the filter to be without interest in most practical cases. We shall therefore not pursue this matter any further. The Gaussian second-order filter for reduced-form models has been derived in, say, Jazwinski [1].

Whether the truncated second-order filter actually will be good enough cannot generally be ascertained. Each case will have to be evaluated separately. The crucial point is of course whether fourth-order central moments can be neglected or not. To give the reader an idea of when considerable biases can arise, let us take a closer look at the covariance matrix of z_t given Z_{t-1}. Assuming for simplicity the system to be scalar and taking fourth-order central moments of z_t given Z_{t-1} into account, we find

$$\mathrm{Cov}(z_t|Z_{t-1}) = (h_y(t))^2 Y_{t|t-1} + R_t + \frac{1}{2}(h_{yy}(t))^2 Y_{t|t-1}^2$$

by a normal approximation of the pdf. If $R_t >> \frac{1}{2}(h_{yy}(t))^2 Y_{t|t-1}^2$, we would not expect much improvement by taking the nonlinearities into account since their effect is more or less "masked out" by the large observation uncertainty. However, if

$$R_t < \frac{1}{2}(h_{yy}(t))^2 Y_{t|t-1}^2 ,$$

it is obvious that the observation nonlinearities will become significant.

Similar observations can be made about the nonlinearities in the structural equations. If Q_t is large compared to fourth-order central moments, we would expect the covariance equations for $Y_{t|t-1}$ to be a good approximation. However, if this is not true, the usefulness of Equation (42) becomes more questionable.

4.3 Iterated Linear Filter-Smoother

In order to improve the filtered estimates of the extended Kalman filter when h is strongly nonlinear, an iterator can be derived from Equation (30). Furthermore, strong nonlinearities in ϕ can be handled by developing a one-stage backwards smoother for y_{t-1}. Iterated filter-smoothers have previously been derived for nonlinear reduced-form models and systems described by stochastic differential equations, first apparently by Wishner et al. [10]. We shall in this section derive a similar iterated filter-smoother for models in structural form.

Define

$$\eta_1 = \hat{y}_{t|t-1} , \quad \xi_1 = \hat{y}_{t-1|t-1} \tag{58}$$

where $\hat{y}_{t-1|t-1}$ is the previous value while $\hat{y}_{t|t-1}$ is obtained from the extended Kalman filter. Let ξ_i denote the smoothed estimate of y_{t-1} after the i'th iteration and let η_i denote the filtered estimate of y_t.

Developing all nonlinear terms in Taylor series expansions around ξ_i , η_i or/and η_{i+1} , and taking all nonlinearities to first-order only, we find the filtered estimate η_{i+1} to be

$$\eta_{i+1} = \hat{y}^1_{t|t-1} + K(t,\xi_i,\eta_i)[z_t - h(\eta_i,t) - h_y(\eta_i,t)(\hat{y}^1_{t|t-1} - \eta_i)] \quad (59)$$

where

$$\hat{y}^1_{t|t-1} = (I - \phi_1(\eta_i,\xi_i,x_t))^{-1}[\phi(\eta_i,\xi_i,x_t) + \phi_2(\eta_i,\xi_i,x_t)(\hat{y}_{t-1|t-1}-\xi_i) - \phi_1(\eta_i,\xi_i,x_t)\eta_i] \quad (60)$$

$$K(t;\xi_i,\eta_i) = Y^1_{t|t-1}h_y(\eta_i,t)[h_y(\eta_i,t)Y^1_{t|t-1}h'(\eta_i,t) + R_t]^{-1} \quad (61)$$

$$Y^1_{t|t-1} = (I - \phi_1(\eta_i,\xi_i,x_t))^{-1} \times [\phi_2(\eta_i,\xi_i,x_t)Y_{t-1|t-1}\phi_2'(\eta_i,\xi_i,x_t) + \Psi(\xi_i,x_t)Q_t\Psi'(\xi_i,x_t)] \times (I - \phi_1'(\eta_i,\xi_i,x_t))^{-1} \quad (62)$$

Furthermore, having computed η_{i+1} we obtain the smoothed estimate ξ_{i+1} of y_{t-1} from one of the following two equations (see Appendix B for more details):

$$\xi_{i+1} = \hat{y}_{t-1|t-1} + M(t;\xi_i,\eta_{i+1})[z_t - h(\eta_{i+1},t) - h_y(\eta_{i+1},t)(\hat{y}^2_{t|t-1} - \eta_{i+1})] \quad (63A)$$

or

$$\xi_{i+1} = \hat{y}_{t-1|t-1} + N(t;\xi_i,\eta_{i+1})[\eta_{i+1} - \hat{y}_{t|t-1}] \quad (63B)$$

where

$$\hat{y}^2_{t|t-1} = (I - \phi_1(\eta_{i+1},\xi_i,x_t))^{-1}[\phi(\eta_{i+1},\xi_i,t) + \phi_2(\eta_{i+1},\xi_i,x_t)(\hat{y}_{t-1|t-1}-\xi_i) - \phi_1(\eta_{i+1},\xi_i,x_t)\eta_{i+1}] \quad (64)$$

$$
\begin{aligned}
Y^2_{t|t-1} &= (I - \phi_1(\eta_{i+1},\xi_i,x_t))^{-1} \\
&\times [\phi_2(\eta_{i+1},\xi_i,x_t)Y_{t-1|t-1}\phi_2'(\eta_{i+1}\xi_I,x_t) \\
&+ \Psi(\xi_i,x_t)Q_t\Psi'(\xi_i,x_t)] \\
&\times (I - \phi_1'(\eta_{i+1},\xi_i,x_t))^{-1}
\end{aligned}
\tag{65}
$$

$$
C_{t|t-1} = (I - \phi_1(\eta_{i+1},\xi_i,x_t))^{-1}\phi_2(\eta_{i+1},\xi_i,x_t)Y_{t-1|t-1} \tag{66}
$$

$$
\begin{aligned}
M(t;\xi_i,\eta_{i+1}) &= C'_{t|t-1}h'_y(\eta_{i+1},t) \\
&\times [h_y(\eta_{i+1},t)Y^2_{t|t-1}h'_y(\eta_{i+1},t) + R_t]^{-1}
\end{aligned}
\tag{67A}
$$

$$
N(t;\xi_i,\eta_{i+1}) = C'_{t|t-1}(Y^2_{t|t-1})^{-1} \tag{67B}
$$

Finally, assume the iterated filter-smoother has converged to within the desired range of accuracy after the ℓ'th iteration. Then

$$
\hat{Y}_{t|t} = \eta_\ell \tag{68}
$$

$$
\begin{aligned}
Y_{t|t} &= (I - K(t;\xi_\ell,\eta_\ell)h_y(\eta_\ell,t))Y_{t|t-1} \\
&= (I - K(t;\xi_\ell,\eta_\ell)h_y(\eta_\ell,t))Y_{t|t-1} \\
&\times (I - K(t;\xi_\ell,\eta_\ell)h_y(\eta_\ell,t))' \\
&+ K(t,\xi_\ell,\eta_\ell)R_tK'(t;\xi_\ell,\eta_\ell) \\
&= [h'_y(\eta_\ell,t)R_t^{-1}h_y(\eta_\ell,t) + Y^{-1}_{t|t-1}]^{-1}
\end{aligned}
\tag{69}
$$

where $Y_{t|t-1}$ is obtained from Equation (62) with $\eta_i = \eta_\ell$ and $\xi_i = \xi_\ell$. Any of the three forms of the covariance equation, Equation (69) may be used, but the second is believed to be computationally more stable than the first and computationally more efficient than the third. Similar expressions can be used in the other equations for $Y_{t|t}$, i.e., Equations (31) and (48) .

A fairly detailed derivation of the iterated linear filter-smoother for structural models is given in Appendix B using an innovations approach.

5. CONCLUSION

We have seen some of the difficulties encountered in developing nonlinear filtering algorithms for nonlinear econometric models in structural form. Specifically, in order to get a computationally feasible form of the truncated second-order filter we had to make a stronger set of assumptions than usually appear in the engineering literature on the subject.

We have also derived the extended Kalman filter, the iterated linear filter-smoother, and the exact equations for the evolution of the conditional mean and covariance matrix. Although they all resemble the forms which have previously been developed for nonlinear reduced-form models, they still contain some new and interesting features.

Our analysis of the truncated second-order nonlinear filter has also helped us in pointing out an error which usually appears in the engineering literature on these filters. Using the somewhat weaker set of assumptions which usually are made in the engineering literature, we were able to derive a more correct form of the truncated second-order filter for nonlinear reduced-form models.

REFERENCES

[1] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, 1970.

[2] G. C. Chow, "An Approach to the Feedback Control of Nonlinear Econometric Systems," Annals of Economic and Social Measurement, Vol 5 (1976), 297-309.

[3] G. C. Chow, S. B. Megdal, "The Control of Large Scale Nonlinear Econometric Systems," Princeton University, Econometric Research Program, Research Memorandum No. 207, March 1977.

[4] R. K. Mehra, "Identification and Estimation of the Error-In-Variables Model (EVM) in Structural Form," Mathematical Programming Study, Vol. 5 (1976), 191-210.

[5] R. K. Mehra, "Identification in Control and Econometrics; Similarities and Differences," Annals of Economic and Social Measurement, Vol 3 (1974), 21-47.

[6] A. H. Jazwinski, "Filtering for Nonlinear Dynamical Systems," IEEE Trans. Automatic Control, Vol. AC-11 (1966), 765-766.

[7] A. H. Jazwinski, "Stochastic Processes with Application to Filtering Theory," Analytical Mechanics Associates, Inc., Seabrook, Maryland, Rept. No. 66-6, 1966.

[8] R. W. Bass, V. D. Norum, L. Schwartz, "Optimal Multichannel Nonlinear Filtering," J. Math. Anal. Appl., Vol. 16 (1966), 152-164.

[9] D. D. Sworder, Optimal Adaptive Control Systems, Academic Press, 1966.

[10] R. P. Wishner, J. A. Tabaczynski, M. Athans, "On the Estimation of the State of Noisy Nonlinear Multivariable Systems," IFAC Symp. Multivariable Control Systems, Düsseldorf, Fed. Rep. of Germany (October 1968).

[11] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems, Trans. ASME, Ser. D: J. Basic Eng., Vol. 82 (1960), 35-45.

[12] L. Schwartz, E. B. Stear, "A Computational Comparison of Several Nonlinear Minimal-Variance Filters," IEEE Trans. Automatic Control, Vol. AC-13 (1968), 83-86.

[13] A. H. Jazwinski, "Nonlinear Filtering - Numerical Experiments," NASA-Goddard Astrodynamics Conference, Greenbelt, Maryland (April 1966).

[14] T. M. Carney, R. M. Goldwyn, "Numerical Experiments with Various Optimal Estimators," J. Optimization Theory Appl., Vol. 1 (1967), 113-130.

[15] R. Henriksen, T. O. Olsen, "Comparison of Different Parameter Estimation Methods in Flotation Processes," 4th IFAC Symposium on Multivariable Technological Systems, Fredericton, N.B., Canada (July 1977).

[16] R. L. Kashyap, "Maximum Likelihood Identification of Stochastic Linear Systems," IEEE Trans. Automatic Control, Vol. AC-15 (1970), No. 1.

[17] R. K. Mehra, "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation," AIAA Journal, Vol. 9 (1971), No. 1.

[18] T. Kailath, "An Innovations Approach to Least-Squares Estimation - Part I: Linear Filtering in Additive White Noise," IEEE Trans. Automatic Control, Vol. AC-13 (1968), No. 6

[19] T. Kailath, P. Frost, "An Innovations Approach to Least-Squares Estimation -- Part III: Nonlinear Estimation in Additive White Noise," IEEE Trans. Automatic Control, Vol. AC-16 (1971), 217-226.

[20] T. Kailath, "A Note on Least Squares Estimation by the Innovations Method," SIAM J. Control, Vol. 10 (1972), 477-486.

[21] W. M. Wonham, "On the Separation Theorem of Stochastic Control," SIAM J. Control, Vol. 6 (1968), 312-326.

[22] E. Tse, "Adaptive Dual Control Methods," Annals of Economic and Social Measurement, Vol. 3 (1974), 65-84.

[23] M. Athans, "The Importance of Kalman Filtering Methods for Economic Systems," Annals of Economic and Social Measurement, Vol. 3 (1974), No. 1.

[24] R. Henriksen, "A Note on Truncated Second-Order Nonlinear Filters," Princeton University, Econometric Research Program, paper (April 1979).

APPENDIX A

A Rationale for Nonlinear Filtering in Econometric Models

The use of recursive filtering techniques has, since the introduction of the Kalman filter in 1960 [11], been widespread among control engineers for both control, identification, and estimation purposes. These techniques have also become widely appreciated in economics, see Athans [23], introducing new lines of attack for estimating both constant and time-varying parameters, and states which are not directly observable. We are here going to point out why nonlinear filtering techniques may be superior to the linear ones based on a linearized model, and to show some possible applications of these techniques.

The main reason for taking second and possibly higher order terms into account is to remove biases from the estimates. Whether these biases actually will be significant or not can only be determined in each specific case. Simulation or experimental results, see [1], [12], [13], or [14], clearly indicate that system nonlinearities may play an important part although some of the experimental results are inconclusive. With only weak nonlinearities it can probably be stated that the extended Kalman filter will generally be sufficient, see [15]. Strong nonlinearities may, however, generate biases which can only be removed by nonlinear techniques.

Now, considering applications of nonlinear filtering techniques, there are at least two apparent ones. First, since the introduction of the separation principle, see Wonham [21], it has become fairly common to use fixed structure controllers consisting of, among others, an estimator for the states and, possibly, parameters, see Tse [22]. Almost needless to say, the performance of such controllers is highly dependent on the performance of the estimator, the latter of which may only perform satisfactorily by the use of nonlinear filtering techniques.

Second, a more familiar problem in econometrics may be the estimation of unknown parameters in a model. Using the techniques which have been derived in this paper, unknown parameters can be estimated either by so-called state augmentation or by the maximum likelihood method.

The method of state augmentation has already been outlined in Section 2 where the possibility that some components of y_t may contain unknown constant or time-varying parameters was assumed. There is not much more to say about this. The final estimate $\hat{\beta}_t$ of an unknown parameter vector β given a set of observations Z_t will then simply be part of the estimated augmented state vector $\hat{y}_{t|t}$. This technique is frequently the only feasible solution in real-time applications, but it may not be so important in economics where sample periods are very large.

The maximum likelihood method is widely used in economics. We are going to conclude this Appendix by deriving the maximum likelihood method for general state models, and by showing its relationship to the previous filtering methods.

Let θ be the vector of unknown parameters which we want to estimate, and let $p(Z_N|\theta)$ be the conditional pdf. of Z_N given θ. The ML estimate of θ given Z_N is given by

$$\hat{\theta} = \text{Arg}\{\max_{\theta} p(Z_N|\theta)\} = \text{Arg}\{\max_{\theta} \log p(Z_N|\theta)\} \tag{A1}$$

Factorizing $p(Z_N|\theta)$, we find

$$\begin{aligned} p(Z_N|\theta) &= p(z_0, z_1, \ldots, z_N|\theta) \\ &= p(z_N|Z_{N-1},\theta)p(Z_{N-1}|\theta) \\ &= p(z_N|Z_{N-1},\theta)p(z_{N-1}|Z_{N-2},\theta)p(Z_{N-2}|\theta) \\ &\vdots \\ &= \prod_{t=0}^{N} p(z_t|Z_{t-1},\theta) \end{aligned} \tag{A2}$$

Therefore

$$\log p(Z_N|\theta) = \sum_{j=0}^{N} \log p(z_t|Z_{t-1},\theta) \qquad)A3)$$

The evaluation of the true ML estimate would require the calculation of $p(z_t|Z_{t-1},\theta)$ using an optimal nonlinear filter to compute all the moments of z_t given Z_{t-1} and θ . However, using a normal approximation of $p(z_t|Z_{t-1},\theta)$ with mean and covariance matrix given by one of the previous filters, we are able to find a computationally feasible solution.

Define

$$\hat{z}_{t|t-1} = E(z_t|Z_{t-1},\theta) \qquad (A4)$$

$$S_{t|t-1} = \mathrm{Cov}(z_t|Z_{t-1},\theta) \qquad (A5)$$

Using the normal approximation of $p(z_t|Z_{t-1},\theta)$ we find

$$\log p(Z_N|\theta) = -\frac{1}{2}\sum_{t=0}^{N} [\nu_t' S_{t|t-1}^{-1}\nu_t + \log|S_{t|t-1}|] + C \qquad (A6)$$

where ν_t , <u>the innovations process</u>, is defined as

$$\nu_t = z_t - E(z_t|Z_{t-1},\theta) = z_t - \hat{z}_{t|t-1} \qquad (A7)$$

Using the <u>extended Kalman filter</u>, we find

$$\hat{z}_{t|t-1} = h(\hat{y}_{t|t-1},t) \triangleq h(t) \qquad (A8)$$

$$S_{t|t-1} = h_y(t)Y_{t|t-1}h_y'(t) + R_t \qquad (A9)$$

$$\nu_t = z_t - h(\hat{y}_{t|t-1},t) \qquad (A10)$$

Using the <u>truncated second-order filter</u>, we obtain

$$\hat{z}_{t|t-1} = h(\hat{y}_{t|t-1}) + \frac{1}{2}(Y_{t|t-1}h_{yy}(t)) \qquad (A11)$$

$$S_{t|t-1} = h_y(t)Y_{t|t-1}h_y'(t) + R_t \qquad (A12)$$

$$\nu_t = z_t - h(\hat{y}_{t|t-1}) - \frac{1}{2}(Y_{t|t-1}h_{yy}(t)) \tag{A13}$$

In case the <u>iterated linear filter-smoother</u> is being used, we find, since approximately $\hat{y}^1_{t|t-1} = \hat{y}^2_{t|t-1} \triangleq \hat{y}_{t|t-1}$ and $Y^1_{t|t-1} = Y^2_{t|t-1} \triangleq Y_{t|t-1}$ when sufficient convergence has been obtained,

$$\hat{z}_{t|t-1} = h(\eta_\ell,t) + h_y(\eta_\ell,t)(\hat{y}_{t|t-1}-\eta_\ell) \tag{A14}$$

$$S_{t|t-1} = h_y(\eta_\ell,t)Y_{t|t-1}h_y'(\eta_\ell,t) + R_t \tag{A15}$$

$$\nu_t = z_t - h(\eta_\ell,t) - h_y(\eta_\ell,t)(\hat{y}_{t|t-1}-\eta_\ell) \tag{A16}$$

Derivation of similar results can be found in Kashyap [16] and Mehra [4], [17].

APPENDIX B

Derivation of the Iterated Linear Filter-Smoother Using an Innovations Approach

We start by deriving the smoothing equations for a linear reduced-form model. The results from that are then transferred to a linearized version of the structural model.

Let

$$y_t = A_t y_{t-1} + C_t x_t + G_t \varepsilon_t \tag{B1}$$

$$z_t = H_t y_t + u_t \tag{B2}$$

be a linear reduced-form model and let $\hat{z}_{t|t-1}$ denote the conditional mean of y_t given $Z_t = (z_0,\ldots,z_{t-1})$. The <u>innovations</u> process, ν_t, is defined as

$$\nu_t = z_t - \hat{z}_{t|t-1}\,, \tag{B3}$$

i.e., the difference between the actual observation z_t and the one-stage predicted observation $\hat{z}_{t|t-1}$ or simply the one-stage prediction error. Let $\hat{y}_{t|t-1}$ denote the conditional expectation of y_t given Z_t. We have the relation

$$\hat{z}_{t|t-1} = E(z_t|Z_{t-1}) = H_t\hat{y}_{t|t-1} \tag{B4}$$

The innovations process can be shown to be a white, zero-mean process with covariance matrix

$$E(\nu_t\nu_t') = H_tY_{t|t-1}H_t' + R_t \tag{B5}$$

where

$$Y_{t|t-1} = \operatorname{cov}(y_t|Z_{t-1}) = A_tY_{t-1|t-1}A_t' + G_tQ_tG_t' \tag{B6}$$

Letting $\hat{y}_{\tau|t}$ denote the conditional expectation of y_τ given Z_t , it can be shown to satisfy the orthogonality condition

$$E((y_\tau - \hat{y}_{\tau|t})\nu_s'|Z_t) = 0 , \quad s = 0,\ldots,t \tag{B7}$$

In the linear case, when $E((Y_\tau - \hat{y}_{\tau|t})\nu_s'|Z_t)$ actually can be shown to be independent of Z_t , Equation (B7) reduces to

$$E((y_\tau - \hat{y}_{\tau|t})\nu_s') = 0, \quad s = 0,\ldots,t \tag{B8}$$

see Kailath [18].

Similar results also hold in the nonlinear case. The process ν_t will still be white, but its covariance matrix will be different from the one given by Equation (B5). Furthermore, the orthogonality condition given in Equation (B7) will also have to be modified, see Kailath et al. [19], [20].

It can also be shown that the processes ν_t and z_t are statistically equivalent in the sense that they are both obtainable from the other by causal and causally invertible linear operators, see Kailath [18] for a proof of this. The meaning of this is simply that all conditional expectations can be carried out given $\{\nu_0,\ldots,\nu_t\}$ instead of $\{z_0,\ldots,z_t\}$ without changing the result.

Since the system described by Equations (B1) - (B2) is linear, the estimate $\hat{y}_{\tau|t}$ can be expressed in terms of the innovations process as

$$\hat{y}_{\tau|t} = \sum_{s=0}^{t} M_{\tau,s}\nu_s \tag{B9}$$

where the matrices $M_{\tau,s}$ can be found by using the orthogonality condition and the whiteness of ν_s .

Specifically, let $\tau = t-1$. Using Equation (B8) we find

$$E(y_{t-1}\nu_s') = M_{t-1,s}(H_s Y_{s|s-1} H_s' + R_s) \tag{B10}$$

i.e.

$$M_{t-1,s} = E(y_{t-1}\nu_s')(H_s Y_{s|s-1} H_s' + R_s)^{-1} \tag{B11}$$

Substituting into Equation (B9) we find

$$\hat{y}_{t-1|t} = \sum_{s=0}^{t-1} M_{t-1,s}\nu_s + M_{t-1,t}\nu_t = \hat{y}_{t-1|t-1} + M_{t-1,t}\nu_t \tag{B12}$$

Furthermore,

$$\begin{aligned} E(y_{t-1}\nu_t') &= E(y_{t-1}(z_t-\hat{z}_{t|t-1})') \\ &= E(y_{t-1}(y_t-\hat{y}_{t|t-1})'H_t') \\ &= Y_{t-1|t-1}A_t'H_t' , \end{aligned}$$

i.e.,

$$M_{t-1,t} = Y_{t-1|t-1}A_t'H_t'(H_t Y_{t|t-1} H_t' + R_t)^{-1} \tag{B13}$$

The filtered estimate $\hat{y}_{t|t}$ can be written as

$$\hat{y}_{t|t} = \hat{y}_{t|t-1} + K_t \nu_t ,$$

i.e.,

$$K_t \nu_t = \hat{y}_{t|t} - \hat{y}_{t|t-1} \tag{B14}$$

The process $K_t \nu_t$ will also be white but generally not of full rank, i.e., its covariance matrix may be singular since rank $K = \dim \nu_k = \dim z_k \leq \dim y_k$. We can, however, still express the smoothed estimate $\hat{y}_{t-1|t}$ in terms of the process $K_s \nu_s$. First, note that

$$E(K_s \nu_s \nu_s' K_s') = K_s (H_s Y_{s|s-1} H_s' + Q_s) K_s' \tag{B15}$$

Since $K_s \nu_s$ has the same rank as ν_s, we can write

$$\begin{aligned} \hat{y}_{t-1|t} &= \sum_{s=0}^{t} N_{t-1,s} K_s \nu_s = \sum_{s=0}^{t} N_{t-1,s} (\hat{y}_{s|s} - \hat{y}_{s|s-1}) \\ &= \hat{y}_{t-1|t-1} + N_{t-1,t} (\hat{y}_{t|t} - \hat{y}_{t|t-1}) \end{aligned} \tag{B16}$$

Again, using the orthogonality condition in Equation (B8) and noting that rank $K_t = \dim \nu_t$, we find

$$E(y_{t-1} \nu_t' K_t') = N_{t-1,t} [K_t (H_t Y_{t|t-1} H_t' + Q_t) K_t'] \tag{B17}$$

Furthermore,

$$E(y_{t-1} \nu_t' K_t') = Y_{t-1|t-1} A_t' H_t' K_t'$$

Substituting the expression for K_t,

$$K_t = Y_{t|t-1} H_t' (H_t Y_{t|t-1} H_t + W_t)^{-1} ,$$

we find the solution of Equation (B17) to be

$$N_{t-1,t} = Y_{t-1|t-1}A_t'Y_{t|t-1}^{-1} \tag{B18}$$

Now, returning to the iterated linear filter-smoother, assume η_i and ξ_i to be given at the i'th stage of iteration. Linearizing $h(y_t,t)$ around η_i (which is believed to be the best estimate of y_t) we find

$$h(y_t,t) = h(\eta_i,t) + h_y(\eta_i,t)(y_t - \eta_i)$$

Taking the conditional expectation given Z_{t-1} yields

$$\begin{aligned}\eta_{i+1} &= \hat{y}^1_{t|t-1} + K(t;\xi_i,\eta_i)[y_t - h(\eta_i,t) \\ &\quad - h_y(\eta_i,t)(\hat{y}^1_{t|t-1} - \eta_i)]\end{aligned} \tag{B19}$$

where $\hat{y}^1_{t|t-1}$, from the partially linearized system equation

$$\begin{aligned}y_t &= (I - \phi_1(\eta_i,\xi_i,x_t)^{-1}[\phi(\eta_i,\xi_i,x_t) \\ &\quad + \phi_2(\eta_i,\xi_i,x_t)(y_{t-1} - \xi_i) - \phi_1(\eta_i,\xi_i,x_t)\eta_i \\ &\quad + \Psi(y_{t-1},x_t)\varepsilon_t] ,\end{aligned} \tag{B20}$$

is found to be

$$\begin{aligned}\hat{y}^1_{t|t-1} &= (I - \phi_1(\eta_i,\xi_i,x_t))^{-1}[\phi(\eta_i,\xi_i,x_t) \\ &\quad + \phi_2(\eta_i,\xi_i,x_t)(\hat{y}_{t-1|t-1} - \xi_i) - \phi_1(\eta_i,\xi_i,x_t)\eta_i]\end{aligned} \tag{B21}$$

Furthermore,

$$K(t;\xi_i,\eta_i) = Y^1_{t|t-1}h_y(\eta_i,t)[h_y(\eta_i,t)Y^1_{t|t-1}h_y'(\eta_i,t) + R_t]^{-1} \tag{B22}$$

where

$$\begin{aligned}Y'_{t|t-1} &= (I - \phi_1(\eta_i,\xi_i,t))^{-1}[\phi_2(\eta_i,\xi_i,t)Y_{t-1|t-1}\phi_2'(\eta_i,\xi_i,t) \\ &\quad + \Psi(\xi_i,x_t)Q_t\Psi'(\xi_i,x_t)](I-\phi_1'(\eta_i,\xi_i,x_t))^{-1}\end{aligned} \tag{B23}$$

Now, having obtained a presumably better estimate η_{i+1} of y_t, we relinearize the functions $h(y_t, t)$ and $\phi(y_t, y_{t-1}, x_t)$ around η_{i+1}. From Equation (B12) we obtain the smoothed estimate ξ_{i+1} of y_{t-1} as

$$\xi_{i+1} = \hat{y}_{t-1|t-1} + M(t;\xi_i,\eta_{i+t})\nu_t \tag{B24}$$

where

$$\nu_t = y_t - h(\eta_{i+1},t) - h_y(\eta_{i+1},t)(\hat{y}^2_{t|t-1} - \eta_{i+1}) \tag{B25}$$

$$\hat{y}^2_{t|t-1} = (I - \phi_1(\eta_{i+1},\xi_i,x_t))^{-1}[\phi(\eta_{i+1},\xi_i,x_t)$$

$$+ \phi_1(\eta_{i+1},\xi_i,x_t)(\hat{y}_{t-1|t-1} - \xi_i) - \phi_2(\eta_{i+1},\xi_i,t)\eta_{i+1}] \tag{B26}$$

whereas, from Equation (B13),

$$M(t;\xi_i,\eta_{i+1}) = Y_{t-1|t-1}A'(t;\xi_i,\eta_{i+1})h'_y(\eta_{i+1},t)$$

$$\times [h_y(\eta_{i+1},t)Y^2_{t|t-1}h'_y(\eta_{i+1},t) + R_t]^{-1} \tag{B27}$$

where

$$A(t;\xi_i,\eta_{i+1}) = (I - \phi_1(\eta_{i+1},\xi_i,x_t))^{-1}\phi_2(\eta_{i+1},\xi_i,x_t) \tag{B28}$$

$$Y^2_{t|t-1} = (I - \phi_1(\eta_{i+1},\xi_i,t))^{-1}[\phi_2(\eta_{i+1},\xi_i,x_t)Y_{t-1|t-1}\phi'_2(\eta_{i+1},\xi_i,x_t)$$

$$+ \Psi(\xi_i,x_t)Q_t\Psi'(\xi_i,x_t)](I - \phi'_1(\eta_{i+1},\xi_i,t))^{-1} \tag{B29}$$

Since

$$C_{t|t-1} = A(t;\xi_i,\eta_{i+1})Y_{t-1|t-1}, \tag{B30}$$

we can rewrite Equation (B27) as

$$M(t;\xi_i,\eta_{i+1}) = C'_{t|t-1}h'_y(\eta_{i+1},t)[h_y(\eta_{i+1},t)Y^2_{t|t-1}h'_y(\eta_{i+1},t) + R_t]^{-1} \tag{B31}$$

Finally, using Equation (B16) we find

$$\xi_{i+1} = \hat{y}_{t-1|t-1} + N(t;\xi_i,\eta_{i+1})[\hat{y}_{t-1|t-1} - \hat{y}^2_{t|t-1}] \tag{B32}$$

where

$$N(t;\xi_i,\eta_{i+1}) = C_{t|t-1}(Y^2_{t|t-1})^{-1} \quad . \tag{B33}$$