ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N 6/7 — 1987

Square Root Iterative Filter: Theory and Applications to Econometric Models

Carlo CARRARO, Domenico SARTORE *

ABSTRACT. – This paper provides a new algorithm for estimating state space dynamic models and, as an example, it considers the estimation of time-varying parameter models. The novel elements of the algorithm are: a simple, easily implementable, square root method which is shown to solve the numerical problems affecting the standard Kalman filter algorithm and the related information filter and smoothing algorithms;an iterative framework, where information and covariance filters and smoothing are sequentially run in order to estimate all the parameters of the model; four different algorithms to consistently estimate the distribution of the estimated parameters, which are described and then compared by performing appropriate Montecarlo experiments.

Racine carrée matricielle et filtrage itératif : théorie et applications à des modèles économétriques

RÉSUMÉ. – Cet article présente un nouvel algorithme pour estimer des modèles dynamiques à représentation d'états et, à titre d'exemple, il considère l'estimation de modèles dont les paramètres varient au cours du temps. L'aspect novateur de l'algorithme tient d'abord à une méthode de racine carrée simple, facile à mettre en œuvre, qui permet d'éviter les problèmes numériques que l'on rencontre avec le filtre de Kalman classique et les algorithmes de lissage ou le filtre d'information associés. Il tient aussi au cadre itératif où les filtres d'information et de covariance et le lissage sont appliqués successivement pour estimer tous les paramètres du modèle. On compare quatre versions différentes de l'algorithme à l'aide de simulations de Monte-Carlo.

* C. CARRARO: Department of Economics, University of Venice-Ca' Foscari, 30123 Venice, Italy. D. SARTORE: Faculty of Statistics, University of Padova, via del Santo, 16 Padova, Italy. A previous preliminary version of this paper was circulated under the title "Iterative Kalman Filter: Theory and Application to Regression Models", Nota di Lavoro 8408, University of Venice. The authors are very grateful to Antonella Basso and Enrico Dalla Vecchia for superb research assistance and accurate implementation of the SRIF algorithm and to Gregory Chow, Corrado Corradi, Andrew Harvey, Whitney Newey, an anonimous referee and the participants to the E. Malinvaud Seminar, INSEE, Paris, December 1985, for many helpful comments. The authors are of course responsible of any remaining error.

1 Introduction

The conventional Kalman filter, as proposed by KALMAN [1960], has been frequently applied to econometric models and time series analysis in recent years (see HARVEY [1981] for an accurate survey). However, all those applications of the Kalman filter utilize the standard Kalman algorithm, without facing the problems related to the actual implementation of the Kalman filter equations.

In particular, even if difficulties related to computer roundoff and other numerical problems appeared in the very early literature on the Kalman filter (see for example BELLANTONI-DODGE [1967]; SCHMIDT-WEINBERG-LUKESH [1968]; SCHMIDT [1972]), only in the engineering literature it is possible to find analyses and solutions of those difficulties. Amazingly enough, no application of Kalman filtering in economics discusses the numerical problems which affect the conventional Kalman algorithm and which are very likely to undermine the meaning of the obtained results.

It is indeed common experience (see KAMINSKI-BRYSON-SCHMIDT [1971]), including our own experience, that the conventional Kalman algorithm often provides covariance matrices of the estimated parameters which fail to be positive semi-definite and, consequently, estimated parameters which largely differ from the true parameters. Even worse, the values of the estimated covariance matrices may be explosive because of computer roundoff errors and ill-conditioned Kalman algorithms.

The above problems, in particular, the non-positive semi-definiteness of the parameter covariance matrices, may occur during one of the recursions of the filter, so that they affect the final estimates of the parameters even when the final covariance matrices are positive definite. This may explain why econometricians are seldom aware of the issues that we are going to discuss in this paper.

Even in engineering however, the most appropriate techniques for dealing with those numerical problems (the square root algorithms) have had difficulty gaining acceptance and, despite their demonstrated superior performance, have been received to a large extent as curiosities. This cold reception is due perhaps to the fact that too few articles on square root algorithms have appeared and that these articles have unfortunately led many to conclude that square root techniques are too complicate.

The first goals of this paper are therefore:

(i) to show the relevance of numerical problems for the conventional Kalman algorithm in its three forms (covariance, information and smoothing);

(ii) to propose a new, very simple, square root algorithm for both Kalman covariance and information filters and for the smoothing equations. Our algorithm may be more likely to be understood and accepted by econometricians;

(iii) to prove that our algorithm, named Square Root Iterative Filter, is very effective in solving the numerical problems affecting the conventional

Kalman algorithm so that good estimates and covariance matrices always positive semi-definite can be obtained.

It is important to emphasize that this paper not only provides new theoretical results on square root filtering, but it is also aimed at defeating the indifference surrounding square root techniques in econometrics. Therefore, several simulation experiments will be performed in order to show the superior performance of our square root algorithm when used for estimating time varying parameter models.

A second important problem which makes it difficult to apply Kalman filters to econometric models concerns the possibility of deriving explicit estimators for all the unknown parameters of the state space model derived from the original econometric model (transition and covariance matrices in particular) and for the initial conditions.

When applying the Kalman filter it is often assumed either that initial conditions and covariance and transition matrices of the state space model are known, or that they can be estimated by numerically optimizing the likelihood function of the innovations generated by the Kalman filter (see PAGAN [1980]; CHOW [1983]; LOS [1983]). This latter approach, which is known as Prediction Error Method in the engineering literature, is not always convergent and, like most numerical methods (see BASSO [1986]), does not provide precise estimates of the parameters of the model.

Other methods for estimating all the unknown elements of the state space model are based on:

(i) the E-M algorithm (see DEMPSTER-LAIRD-RUBIN [1977]; ENGLE-WATSON [1983]; LOS [1985]);

(ii) the Extended Kalman filter (see JAZWINSKI [1970]; LJIUNG [1979]);

(iii) the Innovation Correlation Method (see MEHRA [1970]; CARRARO [1983, 1985]).

Since none of the above methods is completely satisfactory from a theoretical viewpoint and efficient from a numerical viewpoint, the second goals of this paper are:

(i) to provide a theoretically simple algorithm for estimating all the unknown parameters of the state space model, including the initial conditions. This task will be accomplished by combining the covariance and information filters and the smoothing equations into an iterative square root filter;

(ii) to show the numerical efficiency of the above estimation method by performing a large number of Montecarlo experiments so that actual and estimated values can be compared.

The plan of the paper is therefore the following: section 2 deals with the theoretical characteristics of the Square Root Kalman filter and presents and discusses some new results on square root filtering. Sections 3 extends those results to the information filter and the smoothing equations and emphasizes the numerical drawbacks of former algorithms. Section 4 presents the Iterative Filter and the estimation equations for the transition and covariance matrices of the state space model and for the initial condi-

tions. Finally, section 5 describes the time varying parameter model which will be used in the Montecarlo experiments and analyses the results of the numerical simulations. A concluding section summarizes the achievements of this paper and explores its possible developments.

In particular, we hope that more reliable Kalman filtering techniques, based on the Square Root Iterative Filter proposed in this paper, can increase the number of econometric applications of state space modelling and estimation.

2 Square Root Kalman Filter

Even if the Kalman filter equations have been presented many times in the econometric literature, it is important to start the analysis of this section with a discussion of the conventional Kalman filter.

Let us therefore consider the following state space model:

$$y_t = H_t z_t + e_t \tag{1}$$

$$z_t = F_t z_{t-1} + G_t u_t \tag{2}$$

where y_t is the output of the system, z_t the state vector and the matrices $\{H_t, F_t, G_t\}$ are supposed to be known at any time t. The case of unknown system matrices will be later discussed.

Suppose that the random disturbances $\{e_t\}$, $\{u_t\}$ are independent, zero mean, gaussian white processes with

$$E(e_t e_s') = R\,\delta_{ts} \quad E(u_t u_s') = Q\,\delta_{ts} \tag{3}$$

where δ_{ts} is the Kronecker delta.

Suppose furthermore that the initial state z_0 is a gaussian random variable with mean z_0^* and covariance P_0 and that z_0 is independent of $\{e_t\}$ and $\{u_t\}$.

Therefore, the unknown elements of the model are the initial conditions (z_0, P_0), the covariance matrices R and Q and the state vectors $\{z_t;\ t=1, \ldots T\}$, where T is the number of available observations.

Suppose for the moment that $(z_0, P_0, R, Q) = \theta$ is known and define $Y_t = \{y_1, y_2, \ldots, y_t\}$. The Kalman filter equations determine the estimates $z_{t|t-1} = E(z_t | Y_{t-1}, \theta)$, $z_{t|t} = E(z_t | Y_t, \theta)$ and the associated error covariance matrices $P_{t|t-1}$, $P_{t|t}$.

Since the Kalman filter is a sequential procedure, the following equations hold for any $t = 1, 2, \ldots T$:

$$z_{t|t-1} = F_t z_{t-1|t-1} \tag{4}$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t' + G_t Q G_t' \tag{5}$$

$$v_t = y_t - H_t z_{t|t-1} \tag{6}$$

$$C_{0t} = H_t P_{t-1|t-1} H_t' + R \tag{7}$$

$$K_t = P_{t|t-1} H_t' C_{0t}^{-1} \tag{8}$$

$$z_{t|t} = z_{t|t-1} + K_t v_t \tag{9}$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1} \tag{10}$$

Equations (4) (5) represent respectively the one step ahead predictor and its covariance matrix, v_t is the one period prediction error for y_t and is called innovation, C_{0t} is the innovation covariance matrix at time t and K_t is the Kalman gain. The last two equations provide the updated estimates of the state vector z_t and the relative error covariance matrix.

As stated in the introduction, two problems frequently arise when using the conventional Kalman algorithm described by equations (4) to (10): first, the covariance matrices $P_{t|t}$ and $P_{t|t-1}$ may fail to be positive semi-definite at some t since the Kalman filter equations do not constrain $P_{t|t}$ and $P_{t|t-1}$ to be positive semi-definite. Secondly, explosive values of the covariance matrices $P_{t|t}$ and $P_{t|t-1}$ may be obtained. Both problems may imply estimated parameters which largely differ from their true values.

The above problems are often the effects of computer roundoff and usually occur for models with perfect or nearly perfect measurement or with singular or nearly singular error covariance matrices, that is when a linear combination of state vector components is known with great precision, while other combinations are essentially unobservable (see BIERMAN [1982]). Notice that the latter case often arises when estimating time varying parameter models where a subset of the parameters is constant.

The effects of numerical errors are generally manifested in the appearance of computed covariance matrices which fail to be positive semi-definite.

Among the methods that have been used in the engineering literature to improve accuracy and to maintain positive semi-definiteness and symmetry of the computed covariance matrices are:

(i) computation of only the upper triangular entries;

(ii) periodic testing and resetting of the diagonal elements (to retain positivity) and of the off-diagonal elements (to assure correlations that are less than one);

(iii) replacement of equation (10) with the more general, symmetric expression $P_{t|t} = (I - K_t H_t) P_{t|t-1} (I - K_t H_t)' + K_t R K_t'$;

(iv) the use of larger measurement and state covariance matrices when they are assumed to be known.

It is now established in the engineering literature that these methods are not reliable and our numerical experiments confirm this conclusion. [1]

In contrast, square root algorithms have been proposed in the engineering literature (accurate surveys are provided by KAMINSKI-BRYSON-SCHMIDT [1971] and BIERMAN [1982]); square root filtering has been shown to perform very well and to assure stability of the filter and positive semi-definiteness of the covariance matrices $P_{t|t}$ and $P_{t|t-1}$.

1. Due to space limits we do not report in this paper the results of the numerical experiments devoted to test the reliability of the four above methods.

Furthermore, square root algorithms are characterized by a reduced dynamic range of the numbers entering into the computations since they use about half the word lenght required by conventional algorithms. Therefore, the square root method can provide twice the effective precision of the conventional filter implementation in ill-conditioned problems.

As the computational burden for the square root filter is not much higher than for conventional filtering (see Table 2), we would expect to see a great amount of applications of this technique. However, with the exception of avionics applications, the square root filter has remained an obscure complicated technique.

In this paper we do not want to review and explain existing square root algorithms (accurate surveys are provided by KAMINSKI-BRYSON-SCHMIDT [1971] and BIERMAN [1982]), but we propose a new simple square root algorithm which can easily be implemented. This avoids the two main disadvantages of square root filtering without excessively increasing the computational burden (see Table 2).

The basic idea is to estimate two square root covariance matrices $S_{t|t}$ and $S_{t|t-1}$ such that:

$$P_{t|t-1}=S_{t|t-1}S'_{t|t-1} \tag{11}$$

$$P_{t|t}=S_{t|t}S'_{t|t} \tag{12}$$

Let us substitute the conventional equations (5) (7) (10) with the following equations:

$$A'_{1t}=\begin{bmatrix} S'_{t-1|t-1}F'_t \\ Q'_1 G'_t \end{bmatrix} \tag{13}$$

$$A'_{2t}=\begin{bmatrix} S'_{t|t-1}H'_t \\ R'_1 \end{bmatrix} \tag{14}$$

$$A'_{3t}=\begin{bmatrix} S'_{t|t-1}(I-K_t H_t)' \\ R'_1 K'_t \end{bmatrix} \tag{15}$$

where $Q_1 Q'_1=Q$ and $R_1 R'_1=R$.

It is easy to prove that $P_{t|t-1}=A_{1t}A'_{1t}$, $C_{0t}=A_{2t}A'_{2t}$ and $P_{t|t}=A_{3t}A'_{3t}$. Therefore, since the product $A_{it}A'_{it}$, $i=1, 2, 3$ is always positive semi-definite, the calculation of $P_{t|t-1}$, C_{0t}, $P_{t|t}$ by using equations (13) to (15) cannot lead to a matrix which fails to be positive semi-definite as a result of computational errors.

The following algorithm can therefore be proposed:

(i) let P_0, R and Q, when R and Q are known *a priori*, or P_0, R_0 and Q_0, the initial values of $P_{t|t}$, R and Q, when R and Q are unknown, be positive definite and determine, by using the Cholesky algorithm, S_0, R_1, Q_1;

(ii) determine A_{10} by using equation (13) and $P_{1|0}=A_{10}A'_{10}$ so that $P_{1|0}$ will be positive semi-definite by construction;

(iii) factorize $P_{1|0}$ via Cholesky and use $S_{1|0}$ to determine C_{01} and $P_{1|1}$ through the other filter equations (see Table 1);

(iv) since $P_{1|1}$ is positive semi-definite by construction, it can be factorized in order to obtain $S_{1|1}$. Then the previous steps can be repeated for $t=2$ and so on.

This simple algorithm therefore ensures the positive semi-definiteness of the covariance matrices $P_{t|t-1}$, C_{0t}, $P_{t|t}$ at any $t=1, \ldots T$. In order to understand the simplicity of our algorithm with respect to existing algorithms, we want to emphasize that the only computational addition with respect to conventional filtering is the Cholesky factorization of the covariance matrices $P_{t|t-1}$ and $P_{t|t}$ at each step.

Therefore, unlike other algorithms, the covariance matrices $P_{t|t-1}$ and $P_{t|t}$ are not propagated in square root form throughout all the steps and recursions of the filter, but we exploit their positive semi-definiteness at each step in order to simplify the algorithm. Furthermore, the matrices S's and P's are computed simultaneously and interdependently, while all other algorithms compute the sequences $\{S_{t|t-1}\}$, $\{S_{t|t}\}$ independently of $\{P_{t|t-1}\}$ and $\{P_{t|t}\}$.

This can be more easily understood by looking at Table 1 where our algorithm is compared with the most popular existing square root algorithm: the Potter algorithm in its two generalized versions proposed by Schmidt (in order to allow for Q different from zero) and by Andrews (in order to consider the vector measurement case).

Table 1 shows that the factorization of the covariance matrices $P_{t|t}$ and $P_{t|t-1}$ required by our algorithm is substituted by the computation, at each step of the filter, of the matrix T_t and the scalar γ_t in the Schmidt-Potter algorithm, [2] and by the computation of T_t and the factorization of C_{0t} in the Schmidt-Andrews algorithm. The definitions of T_t and γ_t can be found in KAMINSKI-BRYSON-SCHMIDT [1971]. In particular, the computation of T_t, performed either by the Householder algorithm or by the modified Gram-Schmidt algorithm, is particularly complicated.[3].

In contrast, the Schmidt-Potter algorithm, together with the Householder routine for computing T_t, results to be less time consuming than the other algorithms (see KAMINSKI-BRYSON-SCHMIDT, [1971]).

In order to provide a more precise comparison between our algorithm and previous square root algorithms, we computed the number of additions, multiplications, inversions and square roots required by our algorithm and then, using the coefficients provided by KAMINSKI-BRYSON-SCHMIDT [1971], we determined the run time required by our algorithm for one time and one measurement update.

2. Notice that, when y_t is not scalar, it is possible to process the r-th vector y_t as r scalar outputs. When R is diagonal, the components of y_t can be treated as independent outputs, whereas when R is not diagonal, R_1 can be computed by the Cholesky factorization and y_t^0, H_t^0 can be determined recursively by solving $R_1^{-1} y^0 = y_t$ and $R_1^{-1} H^0 = H_t$. This transformation yields a new measurement error e_t^0 with identity covariance matrix. Thus the components of y_t^0 can be processed independently.

3. See BIERMAN [1982] for an illustration of this algorithm.

TABLE 1

Square Root Algorithms

Squaring/Square Root:

$$z_{t|t-1}=F_t z_{t-1|t-1}, \qquad P_{t|t-1}=A_{1t}A'_{1t}$$
$$S_{t|t-1}S'_{t|t-1}=P_{t|t-1}$$
$$v_t=y_t-H_t z_{t|t-1}, \qquad C_{0t}=A_{2t}A'_{2t}$$
$$K_t=P_{t|t-1}H'_t C_{0t}^{-1}$$
$$z_{t|t}=z_{t|t-1}+K_t v_t, \qquad P_{t|t}=A_{3t}A'_{3t}$$
$$S_{t|t}S'_{t|t}=P_{t|t}$$

Schmidt-Potter (scalar measurement):

$$z_{t|t-1}=F_t z_{t-1|t-1}, \qquad \begin{bmatrix} S'_{t|t-1} \\ 0 \end{bmatrix}=T_t A'_{1t}$$
$$v_t=y_t-H_t z_{t|t-1}, \qquad C_{0t}=A_{2t}A'_{2t}$$
$$K_t=S_{t|t-1}S'_{t|t-1}H'_t C_{0t}^{-1}$$
$$z_{t|t}=z_{t|t-1}+K_t v_t, \qquad S_{t/t}=(I-\gamma_t K_t H_t)S_{t|t-1}$$
$$\gamma_t=(1+C_{0t}^{-1}R)^{-1}$$
$$P_{t|t-1}=S_{t|t-1}S'_{t|t-1}, \qquad P_{t|t}=S_{t|t}S'_{t|t}$$

Schmidt-Andrews (vector measurement):

$$z_{t|t-1}=F_t z_{t-1|t-1}, \qquad \begin{bmatrix} S'_{t|t-1} \\ 0 \end{bmatrix}=T_t A'_{1t}$$
$$v_t=y_t-H_t z_{t|t-1}, \qquad C_{0t}=A_{2t}A'_{2t}$$
$$B'_t B_t=C_{0t}$$
$$K_t=S_{t|t-1}S'_{t|t-1}H'_t(B_t^{-1})'B_t^{-1}$$
$$z_{t|t}=z_{t|t-1}+K_t v_t, \qquad S_{t|t}=(I-K_t B_t(B_t+R_1)^{-1}H'_t)S_{t|t-1}$$
$$P_{t|t-1}=S_{t|t-1}S'_{t|t-1}, \qquad P_{t|t}=S_{t|t}S'_{t|t}$$

Table 2 shows that our algorithm is in general more time consuming than other algorithms, so that its simplicity is traded off by a certain degree of inefficiency. However, unlike other algorithms, the complexity of our algorithm does not depend on the output vector dimension r, so that it becomes relatively more efficient as r increases (see the last column of Table 2). This result can be proven more rigorously by taking the first derivative of the difference between the number of operations performed by our algorithm and those performed by previous algorithms, with respect to r.

In any case, our algorithm may be preferred for its easy implementation by using standard computer routines and without modifying the basic structure of the conventional Kalman filter.

Furthermore, in the next sections it will be shown that our algorithm not only can be easily added to available Kalman filter algorithms, but it is also very effective in eliminating the numerical problems which affect the conventional Kalman algorithm.

TABLE 2

Approximate run time for one time and one measurement update

Filter implementation	Time		
	$n=10$, $k=5$, $r=1$	$n=10$, $k=10$, $r=10$	$n=10$, $k=1$, $r=10$
Conventional filter	14	28	24
Schmidt-Potter	18	42	36
Squaring/Square Root	32	53	36

3 Square Root Information Filter and Smoothing

Since the SRIF algorithm, which will be proposed in the next section for estimating the parameters of the state space model equations (1) to (3), is composed by three filters (covariance, information, smoothing) which run sequentially, we must derive the square root version of the information filter and the smoothing equations.

The information filter is important since it allows for diffuse prior on the parameter vector. It is indeed possible to initialize the information filter by assuming $P_0^{-1}=0$, so that the absence of *a priori* information can exactly be introduced into the filter equations. This is due to the fact that the information filter propagates the inverse error covariance matrices $P_{t|t-1}^{-1}$ and $P_{t|t}^{-1}$ and a transformed state vector $d_{t|t}$ such that $d_{t|t}=P_{t|t}^{-1} z_{t|t}$. The conventional formulation of the information filter is the following:

$$d_{t|t-1}=(I-L_t G_t')(F_t^{-1})' d_{t-1|t-1} \tag{14}$$

$$P_{t|t-1}^{-1}=(I-L_t G_t')(F_t^{-1})' P_{t-1|t-1}^{-1} F_t^{-1}(I-L_t G_t)'+L_t Q^{-1} L_t' \tag{15}$$

$$\begin{aligned} L_t=&(F_t^{-1})' P_{t-1|t-1}^{-1} F_t^{-1} G_t \\ &\times[Q^{-1}+G_t'(F_t^{-1})' P_{t-1|t-1}^{-1} F_t^{-1} G_t]^{-1} \end{aligned} \tag{16}$$

$$d_{t|t}=d_{t|t-1}+H_t' R^{-1} y_t \tag{17}$$

$$P_{t|t}^{-1}=P_{t|t-1}^{-1}+H_t' R^{-1} H_t \tag{18}$$

$$z_{t|t}=P_{t|t} d_{t|t}. \tag{19}$$

If no *a priori* information on the state vector is available, the obvious initial values are $P_0^{-1}=0$ and $d_0=0$.

However, these initial values imply that the square root method previously proposed cannot be directly used for deriving the square root information filter. The covariance matrices P_0^{-1} and $P_{1|0}^{-1}$ are indeed equal to zero and the error covariance matrices may not be invertible even in following iterations. The Cholesky factorization becomes therefore impossible.

In order to overcome this difficulty we propose the following square root recursive information filter (see CARRARO [1986] for a detailed derivation of the equations below):

$$d_{t|t-1}=(I-L_t G_t')(F_t^{-1})' d_{t-1|t-1} \tag{20}$$

$$A_{3t}=(F_t^{-1})' A_{2t-1} \tag{21}$$

$$L_t=A_{3t}A_{3t}'G_t(Q^{-1}+G_t'A_{3t}A_{3t}'G_t)^{-1} \tag{22}$$

$$A_{1t}=[(I-L_t G_t')(F_t^{-1})' A_{2t-1} \vdots L_t Q_1^{-1}] \tag{23}$$

$$P_{t|t-1}^{-1}=A_{1t}A_{1t}' \tag{24}$$

$$d_{t|t}=d_{t|t-1}+H_t' R^{-1} y_t \tag{25}$$

$$A_{2t}=[A_{1t} \vdots H_t' R_1^{-1}] \tag{26}$$

$$P_{t|t}^{-1}=A_{2t}A_{2t}' \tag{27}$$

$$z_{t|t}=P_{t|t}d_{t|t} \tag{28}$$

This recursive formulation of the square root information filter is simpler than previous formulations (see KAMINSKI-BRYSON SCHMIDT [1971] for a survey of previously existing square root information filters) and provides inverse error covariance matrices which are non negative definite at each step of the filter.

The only problem concerns the column dimension of the matrix A_{2t} which increases at any step of the algorithm and which may become very large if the number of observation and/or the number of system outputs is large. This problem can easily be solved by running the recursive square root information filter in the first n periods, where n is the state vector dimension, until the matrix $P_{t|t}^{-1}$ becomes invertible. [4]

Then $P_{n|n}^{-1}$ can be inverted and the square root Kalman (covariance) filter previously derived can be used to estimate the state vector and its covariance matrix in all the remaining periods.

The derivation of the square root smoothing formulas is slightly more complicated. First of all, it must be emphasized that the smoothing algorithm which is relevant in econometrics and time series analysis is not the fixed lag smoothing often considered in the econometric literature (see CHOW [1983] for example), but the fixed interval smoothing. The fixed lag smoothing provides indeed the projection of the unknown state vector z_t on the information set Y_j, where $j \geqslant t$ and $j-t$ is fixed. In contrast, the fixed interval smoothing provides the projection of z_t on Y_T for any $t=1, \ldots, T$, thus determining estimates of the parameters contained into the state vector which are based on all the available information.

The conventional fixed interval smoothing is described by the following equations (see ANDERSON-MOORE [1979]) :

$$A_{t-1} = P_{t-1|t-1} F_t' P_{t|t-1}^{-1} \tag{29}$$

$$z_{t-1|T} = z_{t-1|t-1} + A_{t-1}(z_{t|T} - z_{t|t-1}) \tag{30}$$

$$P_{t-1|T} = P_{t-1|t-1} + A_{t-1}(P_{t|T} - P_{t|t-1}) A_{t-1}'. \tag{31}$$

The first problem to be solved concerns equation (31). In order to apply our square root method, it must indeed be transformed into a quadratic equation composed by symmetric matrices. In particular the difference between the matrices $P_{t|T}$ and $P_{t|t-1}$ has to be eliminated.

After a little bit of algebra, equation (31) can be written in the following way (see CARRARO [1986]):

$$\begin{aligned} P_{t-1|T} = {} & (I - A_{t-1} F_t) P_{t-1|t-1} (I - A_{t-1} F_t)' \\ & + A_{t-1}(P_{t|T} + G_t Q G_t') A_{t-1}'. \end{aligned} \tag{32}$$

Our square root method can now be used recursively from $t=T$ to $t=0$, since the starting matrix $P_{T|T}$, which is obtained by using the square root Kalman (covariance) filter, is positive semi-definite by construction. Define

$$B_{t-1} = [(I - A_{t-1} F_t) S_{t-1|t-1} \vdots A_{t-1} S_{t|T} \vdots A_{t-1} G_t Q_1]. \tag{33}$$

Then, $S_{T|T}$ can be obtained by factorizing $P_{T|T}$, whereas $S_{T-1|T-1}$ was obtained during the Kalman filter recursions, so that B_{T-1} can be computed. It is easy to show that $P_{T-1|T} = B_{T-1} B_{T-1}'$. Being positive semi-definite, the covariance matrix $P_{T-1|T}$ can be factorized to obtain $S_{T-1|T}$ and the previous steps can be repeated at any time t, thus guaranteeing the positive semi-definiteness of all the error covariance matrices.

However, a second numerical problem affects the smoothing algorithm. Equation (30) can indeed be written as:

$$z_{t-1|T} = A_{t-1} z_{t|T} + (I - A_{t-1} F_t) z_{t-1|t-1}. \tag{34}$$

Since we suppose not to have prior information on the state vector z_t, the limit, as t goes to zero, of $P_{t-1|t-1}$ and $P_{t|t-1}^{-1}$ is, respectively, infinity and zero. Therefore, the limit of A_t, as t goes to zero, is indeterminate (it will be proven to be F_1^{-1} later on). However, any computer, given its limited numerical capacity, will compute $A_0 = 0$, if the conventional formula is used.

In order to solve this problem we re-write A_{t-1} and $I - A_{t-1} F_t$ in the following way (see Carraro [1986]):

$$A_{t-1} = F_t^{-1}(I - G_t Q G_t' P_{t|t-1}^{-1}) \tag{35}$$

$$I - A_{t-1} F_t = F_t^{-1}(G_t Q G_t' P_{t|t-1}^{-1}) F_t \tag{36}$$

4. If the state vector is identifiable, then $P_{t|t-1}^{-1}$ is invertible after n recursions of the square root information filter (see CARRARO [1986]). However, this condition is only a sufficient condition and depends on the number of system outputs and stochastic inputs. The condition becomes necessary and sufficient for systems with scalar output and deterministic variability of the state vector.

so that equations (3.3.2) and (3.6) can be written as:

$$F_t z_{t-1|T} = (I - G_t Q G_t' P_{t|t-1}^{-1}) z_{t|T} + (G_t Q G_t' P_{t|t-1}^{-1}) z_{t|t-1}. \tag{37}$$

As t goes to zero, the limit of $P_{t|t-1}^{-1}$ is zero (no prior information), so that the limit of A_t is F_1^{-1} from equation (35).

In contrast, if the conventional formula were used, the computer would have provided $A_0 = 0$. Symmetrically, the true limit, as t goes to zero, of $I - A_t F_{t+1}$ is zero, whereas the computer, using the conventional formula, would have computed the identity matrix. The correct value can instead be computed by using equation (36). Therefore, our square root smoothing will be based on equations (35), (36), (37), (32) and the factorization method previously described. In this way, a correct evaluation (conditional on all the available information) of the state vector will be provided and the relative error covariance matrix will be assured to be positive semi-definite.

It is important to emphasize that the smoothing algorithm is relevant not only because it provides estimates of the state vector which are conditional on the whole sample, but also because it can be considered a natural way of obtaining estimates of the initial conditions z_0 and P_0. Using equation (37) it is indeed possible to compute the projection $z_{0|T}$ and its covariance matrix $P_{0|T}$ which can be used as initial conditions in the next iterations of the algorithm (see ENGLE-WATSON [1983]). This is why, as later stated, a diffuse prior will be used as starting value for the estimation of the initial conditions z_0 and P_0.

The next section will consider this issue and other problems related with the estimation of the other unknown elements of the system (1) to (3).

4. SRIF: Square Root Iterative Filter

The previous sections have considered numerical problems related to the estimation of the state vector of the system (1) to (3). We derive now estimation equations for all the other unknown elements of the system, viz. the covariance matrices R and Q and the transition matrix F_t.

As stated in the introduction, several methods have been proposed for estimating the unknown elements contained into R, Q and F_t. In particular, the Prediction Error method, the E−M method, the Extended Kalman filter and the Innovation Correlation method have been considered. Applications of those methods in econometrics can be found in PAGAN [1980], ENGLE-WATSON [1983], CLARK [1985], CARRARO [1985].

As far as F_t is concerned, we propose an estimation method which is similar to the E−M method proposed by ENGLE-WATSON [1983] and applied

by CLARK [1985]. The basic result is contained in CARRARO [1985] where it is shown that

(38) $$\lim_{t \to \infty} E(z_{t|t}) = E(z_t)$$

and

(39) $$\lim_{t \to \infty} \mathrm{Cov}(z_{t|t}) = \mathrm{Cov}(z_t)$$

where $\mathrm{Cov}(z_t)$ indicates the true variance-covariance matrix of the state vector z_t.

Therefore, $z_{t|t}$ can be considered a "consistent" estimate of the mean state vector $E(z_t)$. Furthermore, the structure of the smoothing equations implies that there exists a number K such that, as T goes to infinity, $z_{t|T} - E(z_t) \leqslant \sigma$, for any $t \geqslant K$ and any small σ. Therefore $z_{t|T}$ can also be considered a "consistent" estimate of $E(z_t)$. This implies that the time invariant parameters of the transition matrix F_t can be consistently estimated by an appropriate regression of $z_{t|T}$ on $z_{t-1|T}$.

Suppose, for the sake of simplicity, that $F_t = B_1 + B_2 X_t$, so that F_t is a linear function of exogenous, fixed, variables X_t; $t = 1, 2 \ldots T$. Then, the matrices B_1 and B_2 can be consistently estimated by regressing $z_{t|T}$ on $z_{t-1|T}$ and $X_t z_{t-1|T}$.

Having estimated the unknown elements contained into F_t, we must provide estimation equations for the parameters of the covariance matrices R and Q. Most methods proposed in the literature for estimating R and Q are based on the innovation sequence $\{v_t\}$ and are reviewed in CARRARO-SARTORE [1985], where simulation experiments are also performed in order to test the performance of those methods. In particular the innovation correlation method (see CARRARO [1985]) was shown to provide good estimates of the covariance matrix Q, while only a numerical maximization of the innovation likelihood function could provide satisfactory estimates of R.

More precisely, R was estimated by solving the following maximization problem:

(40) $$\max_{R} \log L = \text{Const.} - 1/2 \sum_{t=1}^{T} \log (C_{0t}) - 1/2 \sum_{t=1}^{T} v_t^2 C_{0t}^{-1}$$

whereas Q was estimated by the following equation:

(41) $$Q^* = [W' W]^{-1} W' W^0$$

where

$$W = \begin{bmatrix} H_1 G \otimes H_1 G \\ \vdots \\ H_T G \otimes H_T G \end{bmatrix} \Gamma, \qquad W^0 = \begin{bmatrix} v_1 v_1' - H_1 F_1 P_0 F_1' H_1' - R \\ \vdots \\ v_T v_T' - H_T F_T P_{T-1|T-1} F_T' H_T' - R \end{bmatrix}$$

Q^* is a vector containing all the upper triangular elements of Q and Γ is a selection matrix such that $\mathrm{vec}(Q) = \Gamma Q$.

However, having obtained smoothed estimates of the state vector z_t, it is also possible to use the residuals $\hat{e}_t = y_t - H_t z_{t|T}$ and $\hat{u}_t = z_{t|T} - F_t z_{t-1|T}$ to obtain consistent estimates of R and Q. However, the covariance matrices of the residuals $\hat{e}_t$ and $\hat{u}_t$ depend on the covariance matrices $P_{t-1|T}$ and $P_{t|T}$, which do not converge to zero asymptotically. Therefore the sum of square of $\hat{e}_t$ and $\hat{u}_t$ has to be corrected in order to derive consistent estimates of R and Q.

From obvious definition, we have:

$$\hat{e}_t = e_t + H_t(z_t - z_{t|T}) \tag{42}$$

$$G_t\hat{u}_t = G_t u_t + F_t(z_{t-1} - z_{t-1|T}) - (z_t - z_{t|T}). \tag{43}$$

Defining $S = E[(z_t - z_{t|T})(G_t u_t)']$ and $V = E[e_t(z_t - z_{t|T})']$, the following equations can be derived:

$$E[\hat{e}_t\hat{e}_t'] = R + H_t P_{t|T} H_t' + 2VH_t' \tag{44}$$

$$E[G_t\hat{u}_t\hat{u}_t'G_t'] = G_t QG_t' + P_{t|T} - F_t P_{t-1|T} F_t' - 2S' \tag{45}$$

$$E[G_t\hat{u}_t\hat{e}_t'] = S'H_t' + F_t P_{t-1|T} F_t' H_t' - P_{t|T} H_t' - V' \tag{46}$$

$$E[\hat{e}_t\hat{e}_{t-1}'] = H_t F_t V' + H_t F_t P_{t-1|T} H_{t-1}'. \tag{47}$$

In order to solve for V', we write equation (4.4.4) as:

$$\begin{bmatrix} E[\hat{e}_2\hat{e}_1'] - H_2F_2P_{1|T}H_1' \\ E[\hat{e}_3\hat{e}_2'] - H_3F_3P_{2|T}H_2' \\ \vdots \\ E[\hat{e}_T\hat{e}_{T-1}'] - H_TF_TP_{T-1|T}H_{T-1}' \end{bmatrix} = \begin{bmatrix} H_2F_2 \\ H_3F_3 \\ \vdots \\ H_TF_T \end{bmatrix} V' \tag{48}$$

which can solved with respect to V' after having used the residuals $\{\hat{e}_t\}$ for estimating the covariances $E[\hat{e}_t\hat{e}_{t-1}']$. Defining the left hand side of (48) as D^0 and the right hand side as DV', it is easy to show that:

$$\hat{V}' = [D'D]^{-1}D'D^0. \tag{49}$$

The remaining equations (44) to (47) can now be used to solve with respect to R and Q. Substituting equation (46) into (45), we have:

$$\begin{aligned} &E[G_t\hat{u}_t\hat{u}_t'G_t']H_t' \\ &= G_tQG_t'H_t' - P_{t|T}H_t' \\ &\quad + F_tP_{t-1|T}F_t'H_t' - 2E[G_t\hat{u}_t\hat{e}_t'] - 2\hat{V}'. \end{aligned} \tag{50}$$

Assuming, as usual, $G_t = G$ for any $t = 1 \ldots T$, we re-write equation (50) as:

$$\begin{aligned} &GQG'[H_1'\,H_2' \ldots H_T'] \\ &= [E(G\hat{u}_1\hat{u}_1'G')H_1' \ldots E(G\hat{u}_T\hat{u}_T'G')H_T'] \\ &\quad + [P_1^0H_1' \ldots P_T^0H_T'] \\ &\quad + 2[E(G\hat{u}_1\hat{e}_1') \ldots E(G\hat{u}_T\hat{e}_T')] \\ &\quad + 2[V'\hat{V}' \ldots \hat{V}'] \end{aligned} \tag{51}$$

where $P_t^0 = P_{t|T} - F_tP_{t-1|T}F_t'$.

Defining the right hand side of (51) as N° and the matrix $G'[H_1' H_2' \ldots H_T']$ as N', the following estimation equation can be derived for Q:

$$G\hat{Q} = N^{0} N [N' N]^{-1} \tag{52}$$

which can be solved with respect to the unknown elements of Q whenever G is known (usually G is a known matrix of which the elements are either zero or one). Again, the residuals $\hat{e}_t$ and $\hat{u}_t$ must be used for estimating $E[G\hat{u}_t\hat{u}_t' G']$ and $E[G\hat{u}_t\hat{e}_t']$.

Finally, equation (44) can easily be solved for obtaining a consistent estimate of the output covariance matrix R. We have indeed:

$$\hat{R} = 1/T\left[\sum_{t=1}^{T} [\hat{e}_t\hat{e}_t' - H_t P_{t|T} H_t' - 2\hat{V}H_t']\right] \tag{53}$$

where $E[\hat{e}_t\hat{e}_t']$ can be estimated by using the sample residuals $\hat{e}_t$.

Therefore, we have now two methods for estimating the covariance matrices R and Q: the first is based on the innovation sequence $\{ v_t \}$ and the innovation likelihood function, the second is based on the smoothing residuals $\{ \hat{e}_t \}$ and $\{ \hat{u}_t \}$. In the next section we will perform simulation experiments by using both methods in order to compare their performance at least when used for estimating a simple time-varying parameter model.

We can now describe the whole structure of the Square Root Iterative Filter (SRIF) that we have implemented for estimating econometric and time series models in state space form.

SRIF Algorithm

Step 1: Choose initial values M_0, R_0, Q_0 for the covariance matrices R and Q and the matrix M containing the unknown elements of the transition matrix F and initialize the information filter by the diffuse prior $P_0^{-1} = 0$.

Step 2: Run the information filter for the first n recursions, where n is the state dimension, and invert $P_{n|n}^{-1}$.

Step 3: Run the Kalman filter, initialized by $P_{n|n}$ previously determined, from time $n+1$ to time T.

Step 4: Run the smoothing from time T to time 0 in order to obtain $\{ z_{t|T};\ t=1 \ldots T \}$, the residual sequences $\{ \hat{e}_t$ and $\hat{u}_t;\ t=1 \ldots T \}$ and estimates of the initial conditions z_0 and P_0.

Step 5: Regress $z_{t|T}$ on $z_{t-1|T}$ for estimating M and use equations (52) and (53) for estimating R and Q.

Step 6: Re-initialize the Kalman filter by using the estimated values of z_0, P_0, M, R and Q, and run the Kalman filter from time 1 to n. Then iterate the previous steps, from 3 to 6, until convergence.

A second version of SRIF, based on the innovation sequence $\{ v_t \}$ for estimating the covariance matrices R and Q, can be described by adding the following step between Step 3 and Step 4:

Step 3′: Use the innovation sequence $\{ v_t \}$ and equations (40) and (41) for estimating the covariance matrices R and Q. Then re-run Step 2 and Step 3 by using the estimated values of R and Q.

Consequently, Step 5 must be simplified in the following way:

Step 5′: Estimate M by regressing the smoothing estimate $z_{t|T}$ on $z_{t-1|T}$.

Notice that the Kalman filter and smoother are iterated until convergence of the algorithm. In the next section it will be shown that iterating SRIF on the same sample is very important in order to reduce the influence of the initial conditions M_0, R_0, and Q_0 on the final estimates on the system parameters.

Given its general features, the SRIF algorithm can be used for estimating any econometric or time series model that can be written in a suitable state space form. From the smoothing equations and (52), we have that the matrices of the state space form must indeed satisfy the following conditions:

$$\text{rank}\left[\sum_{t=1}^{T} G' H_t' H_t G\right] = n \tag{54}$$

$$\text{rank}\,[F_t] = n \quad \text{for any } t = 1, \ldots T. \tag{55}$$

Furthermore, in order to guarantee the "consistency" of the state vector estimates, the observability condition must be satisfied (see CARRARO [1985]), so that:

$$\text{rank}\,[H' \; F' \; H' \; F^{2\prime} \; H' \; \ldots \; F^{n-1\prime} \; H'] = n \tag{56}$$

when H and F are time invariant, or

$$\text{rank}\left[\sum_{i=0}^{T-1} \Omega(i, T)' H_i' R^{-1} H_i \Omega(i, T)\right] = n \tag{57}$$

where $\Omega(i, T) = F_{i+1}^{-1} F_{i+2}^{-1} \ldots F_T^{-1}$, if the system is time-varying.

Having defined features, properties and requirements of the SRIF algorithm, we need now to test its performance. Therefore, in the next section, we will present the results of several Montecarlo experiments that we have performed in order to analyse the actual characteristics of SRIF. The model used in the simulations is a simple time-varying parameter model where a subset of the parameters is assumed to be constant (this is important for emphasizing the numerical problems affecting the conventional Kalman algorithm). It will be shown that SRIF provides precise estimates of all the unknown parameters of the model, that the error covariance matrices are always positive semi-definite and that SRIF's sensitivity to different initializations is very low.

5 Montecarlo Experiments

In order to test the performance of the SRIF algorithm proposed in the previous section, we considered the following time-varying parameter model (see CARRARO [1985] for a detailed analysis of the model):

$$y_t = x_t \beta_t + e_t, \qquad e_t \sim \text{NID}(0, \sigma^2) \tag{58}$$

$$\beta_t = \text{M}\,\beta_{t-1} + \text{Z}_t\,\delta + \text{G}^* u_t, \qquad u_t \sim \text{NID}(0, \text{Q}) \tag{59}$$

where y_t is the scalar, dependent, variable, x_t the $(1 \times k)$ vector of fixed regressor, β_t a $(k \times 1)$ vector of time-varying and constant parameters and e_t a measurement error which is assumed normally and independently distributed. Furthermore, M is a $(k \times k)$ matrix representing the autoregressive structure of the vector β_t and $\text{Z}_t = \underset{k}{\text{diag}}(z_{1t} \ldots z_{kt})$ is a $(k \times p)$ matrix of fixed explanatory variables, where $\text{P} = \sum_{i=1}^{k} \text{P}_i^*$ and P_i^* is the dimension of the row vector z_{it}. Finally, δ is a $(p \times 1)$ vector of constant parameters and u_t defines the system noise which is also assumed normally and independently distributed. For the sake of simplicity, e_t and u_t are assumed to be independent. The matrix G* defines whether the variability of the i-th parameter is assumed stochastic or deterministic.

We performed the Montecarlo experiments by assuming $k = 3$, but only β_{1t} was assumed to be time-varying ($p_1^* = p = 2$ was chosen, which implies $n = 5$), so that we had to estimate the following state space form:

$$y_t = \text{H}_t b_t + e_t, \qquad e_t \sim \text{NID}(0, \sigma^2) \tag{60}$$

$$b_t = \text{F}_t b_{t-1} + \text{G}\,u_t, \qquad u_t \sim \text{NID}(0, \text{Q}) \tag{61}$$

where

$$b_t' = [\beta_0 \quad \beta_{1t} \quad \beta_2 \quad \delta_0 \quad \delta_1] \tag{62}$$

$$\text{H}_t = [1 \quad x_{1t} \quad x_{2t} \quad 0 \quad 0] \tag{63}$$

$$\text{F}_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & m & 0 & 1 & z_{1t} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{64}$$

$$\text{G}' = [0 \quad 1 \quad 0 \quad 0 \quad 0]. \tag{65}$$

The Montecarlo experiments were performed under the assumptions:

$$\beta_0 = 100, \qquad \beta_2 = 0.7 \tag{66}$$

$$\delta_0 = 10, \qquad \delta_1 = 3 \tag{67}$$

$$m = 0.4, \qquad \sigma^2 = 100, \qquad \text{Q} = 10 \tag{68}$$

which define the true values of the unknown parameters of the model.

The exogenous variables $\{ x_{1t}, x_{2t}, z_{1t}; t=1 \ldots T \}$ were generated by simulating a simple ARMA process for each exogenous variable. The characteristics of these variables remain unchanged in all experiments, except for the number of observations. All the variables were appropriately scaled.

The pseudo-random number disturbances, generated independently for each replication (using the IMSL routine GGNML) are independently and identically distributed with zero mean and variances as specified in (68). Furthermore, there is no correlation between the random disturbances.

The dependent variable y_t and the time-varying parameter β_{1t} were generated by using the structure (62) to (65), (66) to (68) and the previously described exogenous variables and random disturbances. Then, the time series $\{ y_t, x_{1t}, x_{2t}, z_{1t}; t=1 \ldots T \}$ were used as inputs for SRIF. [5]

Each Montecarlo experiment was replicated 30 times and double precision computations were performed.

The Montecarlo experiments were designed for analysing the stability of the square root equations, the sampling distribution of the SRIF estimates under different sample dimensions and the sensitivity of SRIF to different initializations.

In order to analyse the sampling distribution of the SRIF estimates of the time-varying parameter, we used Theil's inequality coefficient which was computed with respect to the difference between true and estimated parameters. In contrast, the sampling distribution of the constant parameters can be examined by comparing the computed values reported in the following Tables and the true values previously reported.

The stopping rule for the iterative filter was chosen to be $\theta^i - \theta^{i-1} \leqslant 0.01$, where $\theta' = [\beta_0 \quad \beta_2 \quad \delta_0 \quad \delta_1 \quad m \quad \sigma^2 \quad Q]$ and i indicates the i-th iteration of the Square Root Iterative Filter.

The results of the Montecarlo experiments are reported in Tables 3 to 7. [6]

TABLE 3

N = 100 R = 100 Q = 10 $m = 0.4$

Parameter	Mean of estimated parameter	Variance of estimated parameter
β_0	99.922	40.541
	(19.605)	
β_1	0.696	0.001
	(0.001)	
δ_0	10.020	0.182
	(0.052)	
δ_1	2.981	0.024
	(0.013)	
R	100.	
Q	10.	
m	0.4	
Theil's U	0.017	0.00002

Table 3 is aimed at distinguishing between the methods proposed for estimating R, Q and M and the square root algorithm that we propose in order to assure the positive semi-definiteness of the parameter covariance matrices. Therefore, Table 3 contains the average values, over thirty replications, of the estimated parameters and their variances when R, Q and M are assumed to be known. It is easy to see that the estimated parameters are very close to their true values and that all variances are positive. This is therefore a first important evidence of the high performance of the square root method we have proposed in this paper. In contrast, when the standard Kalman and smoothing algorithms were used, some variances failed to be positive, the estimated parameters were far from their true values and the algorithm was seldom convergent.

Important information is also provided by Table 4 which contains the average values, over thirty replications, of the estimated parameters and their variances, when the first version of SRIF was used. The smoothing residuals $\hat{e}_t$ and $\hat{u}_t$ were indeed used for estimating the unknown covariance matrices R and Q.

It is easy to see that the estimates of the constant and time-varying parameters are fairly good, but that R and Q are not well estimated. The estimated constant parameters are indeed very close to their true values and Theil's U indicates that the evolutionary structure of the time-varying parameters has been well captured by SRIF. In contrast, R is estimated to be equal to zero and all the variability is attributed to the system stochastic inputs u_t.

TABLE 4

N = 100 $R_0 = 1000$ $Q_0 = 100$ $m_0 = 0.2$		
Parameter	Mean of estimated parameter	Variance of estimated parameter
β_0	99.626 (401.272)	50.260
β_1	0.697 (0.0247)	0.003
δ_0	13.427 (1.030)	0.920
δ_1	3.059 (0.261)	0.038
R	0.0000	0.0000
Q	2679.402	7048918.
m	0.722 (0.163	0.071
Theil's U	0.023	0.001

5. The computer program implementing the SRIF algorithm has been written by the authors, Antonella Basso and Enrico Dalla Vecchia.

6. In Tables 3-7, the number between brackets represents the estimated variance of the parameter, whereas the second column reports the variance (over the thirty replications) of the estimated parameter.

This result can be explained by the special features of the smoothing equations which modify the state vector estimates in such a way to minimize the difference between actual and estimated system output. The flexibility of the time-varying parameter model enables the smoothing equations to provide a residual sum of squares which is zero in all samples, whatever the initializations of the filters. In other words, the smoothing's goal is not a precise estimate of the state vector parameters, but a perfect fitting of the output time series y_t. This fact and the degree of freedom provided by the time-varying structure of the parameters explain why all the variability is captured by the sum of squares of the residuals u_t.

Finally, the autoregressive parameter m is poorly estimated at the end of the first iteration of SRIF and no further iteration can be performed because of the zero value of the estimated R.

Table 5's results are therefore more interesting. In that Table we present the results of the Montecarlo experiments performed by using the second version of SRIF, which uses the innovation sequence v_t for estimating the covariance matrices R and Q.

Table 5 shows that almost all the estimates of the parameters are very precise. Only δ_0, the constant in the equation explaining the evolution of the time-varying parameter $\beta_{1\,t}$, is slightly underestimated. This is the effect of having obtained an estimate of the transition parameter m which is slightly above its true value. Nonetheless, the estimates of the time-varying parameters are very precise, as indicated by Theil's U.

Notice that the covariance matrices R and Q are much better estimated than in the previous case, when the first version of SRIF was used, even if the estimate of R is still unsatisfactory. In particular Q, the estimated covariance matrix of the disturbances u_t, is very close to its true value. Hence, the second version of SRIF should be preferred to the first version.

Two preliminary conclusions can therefore be derived from the previous analysis: first, the second version of SRIF provides better estimates of all the parameters of the model (in particular of the parameters m, R and Q).

TABLE 5

$N=100 \quad R_0=1000 \quad Q_0=100 \quad m_0=0.2$

Parameter	Mean of estimated parameter	Variance of estimated parameter
β_0	99.567	62.503
	(61.437)	
β_1	0.697	0.001
	(0.003)	
δ_0	9.082	3.449
	(0.064)	
δ_1	2.927	0.028
	(0.016)	
R	1043.23	1052284.
Q	10.286	5.718
m	0.447	0.008
Theil's U	0.0216	0.00002

Secondly, a precise estimate of m is crucial for obtaining very good estimates of δ_0, R and Q. The regression approach implemented in the second version of SRIF seems to be the appropriate method for obtaining a good estimate of m.

In order to support this conclusion, we implemented a third version of SRIF, which uses the innovation likelihood maximization approach for estimating m, R and Q. Using this method, the most used in engineering and economics even if computationally the most expensive, we replace Step 3′ of the second version of SRIF by:

Step 3″: Maximize the innovation likelihood (40) with respect to R, Q and M by appropriate numerical methods. Then re-run Step 2 and Step 3, by using the estimated values of R, Q and M, until convergence.

As a consequence, Step 5 and Step 5′ must be eliminated.

The results obtained by using the third version of SRIF are presented in Table 6 which shows that all the parameters, including the transition parameter m, are well estimated by the innovation likelihood maximization method. In particular, Theil's U shows that the second and third version of SRIF perform in a similar way even with respect to the time-varying parameter. Only the covariance matrix R is still poorly estimated, thus affecting the standard errors of the estimated parameters.

As a final test we therefore implemented a fourth version of SRIF, which uses the E-M method for estimating the unknown elements of F_t, R and Q (see ENGLE-WATSON [1983]). The results are reported in Table 7.

It is easy to see that the E-M method does not improve the estimates of the parameters m, R, Q, whereas the other parameters are still satisfactorily estimated.

We can therefore conclude that no clear evidence on the best method to be used for estimating the parameters contained into the matrices R, Q and M can be deduced from our Montecarlo experiments, even if the second

TABLE 6

$N = 100 \quad R_0 = 1000 \quad Q_0 = 100 \quad m_0 = 0.2$

Parameter	Mean of estimated parameter	Variance of estimated parameter
β_0	100.324	56.229
	(51.449)	
β_1	0.694	0.001
	(0.002)	
δ_0	10.508	1.916
	(0.056)	
δ_1	2.982	0.024
	(0.014)	
R	795.72	152177.
Q	9.502	3.135
m	3.775	0.005
Theil's U	0.0211	0.00004

TABLE 7

N = 100 $R_0 = 1000$ $Q_0 = 100$ $m_0 = 0.2$		
Parameter	Mean of estimated parameter	Variance of estimated parameter
β_0	101.886 (196.897)	170.511
β_1	0.709 (0.011)	0.006
δ_0	10.905 (0.088)	2.618
δ_1	2.997 (0.023)	0.029
R	7258.240	466087700.
Q	7.706	5.569
m	0.355	0.008
Theil's U	0.0291	0.00026

version of SRIF, based on the innovation correlation approach, is probably to be preferred for its computational simplicity. [7]

In contrast, a clear and important conclusion is supported by our results (see Table 3 above all). The square root algorithms implemented in SRIF were able to solve all numerical problems that normally affect the application of Kalman filters and smoother to real data. The covariance matrices of the estimated parameters are always positive semi-definite and the algorithms always convergent. In contrast, when the standard Kalman algorithm was used for estimating the simple time-varying parameter model previously presented, the covariance matrices were often ill conditioned (negative diagonal elements) and the algorithm was seldom convergent. It is true that the simulated model, by implying a state vector with a single stochastic element, emphasizes that type of numerical problems. However, our square root algorithm was able to perform very well even in this difficult case.

6 Conclusions

This paper has proposed and tested a new algorithm for estimating econometric and time series models in state space form. This algorithm, named SRIF, modifies the standard Kalman filter and smoother algorithms in order to guarantee the positive semi-definiteness of the error covariance matrices at each step of the algorithms.

The Montecarlo experiments that we performed show that SRIF actually succeeds in providing well defined covariance matrices and parameter estimates which are close to their true values.

However, the numerical experiments presented in this paper provide only preliminary evidence about the performance of our Square Root Iterative Filter. New evidence should be obtained by applying SRIF to unobserved components models, missing observation models, ARMA models, rational expectations models, etc.

We believe that more experiments, tests and applications can only confirm that SRIF is both a numerically reliable algorithm, and an efficient and flexible tool for estimating econometric and time series models.

7. We must emphasize that the poor estimates obtained, on average, for the covariance matrices R and Q are mainly explained by the presence of two abnormal samples among the thirty samples used for simulation. If we exclude those two samples, both the innovation correlation, the maximum likelihood and the E-M method provide estimates of m, R and Q very close to their true values, thus also improving the estimates of the other parameters.

• References

ANDERSON, B. D. and MOORE, J. B. (1979). – *Optimal Filtering*, Prentice Hall, Englewood Cliffs, New Jersey.

BASSO, A. (1986). – « Metodi di Stima Nonlineare in Econometria », *Tesi di Laurea*, Università di Venezia.

BELLANTONI, J. F. and DODGE, K. W. (1967). – "A Square Root Formulation of the Kalman-Schmidt Filter", *AIAA Journal*, 5, p. 1309-1314.

BIERMAN, G. J. (1982). – *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York.

CARRARO, C. (1983). – "Identification and Estimation of Regression Models with Structural Variability of the Parameters", BASAR T. and PAU L. F., Eds., *Dynamic Modelling and Control of National Economies*, Pergamon Press, Oxford.

CARRARO, C. (1985). – "Regression and Kalman Filter Methods for Time-Varying Econometric Models", Econometric Research Program, *Research Memorandum n.* 320, Princeton University.

CARRARO, C. (1986). – "Square Root Kalman Algorithms in Econometrics", forthcoming in *Computer Science for Economics and Management*.

CARRARO, C. and SARTORE, D. (1985). – "Efficient Estimation Algorithms for Time-Varying Parameter Models", *mimeo*, University of Venice.

CHOW, G. C. (1983). – *Econometrics*, McGraw Hill, New York.

CLARK, P. K. (1985). – "A State Space Model for Trend GNP", *mimeo*, Yale University.

ENGLE, R. F. and WATSON, M. W. (1983). – "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models", *Journal of Econometrics*, 23, p. 385-400.

HARVEY, A. (1981). – "The Kalman Filter and its Applications in Econometrics and Time Series Analysis", *paper presented* at the 6th Symposium on Operation Research, Augsburg.

KALMAN, R.E. 1960). – "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering*.

KAMINSKI, P. G., BRYSON, A. E. and SCHMIDT, S. F. (1971). – "Discrete Square Root Filtering: A Survey of Current Techniques", *IEEE Trans. on Automatic Control*, p. 727-736.

JAZWINSKI, A. H. (1970). – *Stochastic Processes and Filtering Theory*, Academic Press, New York.

LIJUNG, L. (1979). – "Asymptotic Behaviour of the Extented Kalman Filter as a Parameter Estimator of Linear Systems", *IEEE Trans. on Automatic Control*, p. 36-50.

LOS, C. A. (1983). – "Econometrics of Models with Evolutionary Parameter Structures, *Ph. D. Dissertation*, Columbia University.

LOS, C. A. (1985). – "Montecarlo Studies of Econometric Estimators of Evolutionary Parameter Structures", *Research Paper* 8502, FRB of New York.

MEHRA, R. K. (1970). – "On the Identification of Variances and Adaptive Filtering", *IEEE Trans. on Automatic Control*.

PAGAN, A. R. (1980). – "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients", *Journal of Econometrics*, p. 341-363.

SCHMIDT, S. F. (1972). – "Precision Navigation for Approach and Landing Operation", *Proceedings*, Joint Automatic Control Conference, Stanford University, p. 455-463.

SCHMIDT, S. F., WEINBERG, J. P. and LUKESH, J. S. (1968). – "Case Study of Kalman Filtering in the C-5 Aircraft Navigation System", *Proceedings*, Joint Automatic Control Conf., University of Michigan, p. 59-109.