Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

FACULTAD DE ECONOMICAS

Campus de Somosaguas

28223 MADRID

Teléfono 394 26 11 - FAX 294 26 13

Documento de trabajo

A General Fixed-Interval Smoother with Exact Initial Conditions

José Casals

Sonia Sotoca

Miguel Jerez

No.9804

Mayo 1998

Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

A GENERAL FIXED-INTERVAL SMOOTHER WITH EXACT INITIAL CONDITIONS

José Casals
Miguel Jerez
Sonia Sotoca

Universidad Complutense de Madrid
Campus de Somosaguas
28223 Madrid

ABSTRACT

In this work we derive a relationship between the exact fixed-interval smoothed moments and those obtained from an arbitrarily initialized smoother. Combining this result with a conventional smoother we obtain a new algorithm with exact initial conditions, that can be applied to stationary, nonstationary or partially nonstationary systems, with deterministic and/or stochastic inputs. Besides an easy analytical derivation, other advantages of this smoother are its computational efficiency and numerical stability.

RESUMEN

En este trabajo se deriva la relación existente entre los momentos exactos de un smoother de intervalo fijo y los momentos obtenidos de un smoother inicializado arbitrariamente. Combinando este resultado con un smoother convencional se obtiene un nuevo algoritmo con condiciones iniciales exactas, que puede ser aplicado a sistemas estacionarios, no estacionarios o parcialmente no estacionarios, con inputs deterministas y/o estocásticos. Además de su fácil derivación analítica, otras ventajas de este nuevo smoother son su eficiencia computacional y su estabilidad numérica.

Key words: State-space models; Nonstationarity; Stochastic inputs; Kalman filter.

JEL classification codes: C32; C40.

Mailing address: Departamento de Economía Cuantitativa. Universidad Complutense. Somosaguas. 28223 Madrid. Spain. E-mail: eccua01@emducms1.sis.ucm.es.

1. INTRODUCTION.

Consider the state-space model:

$$z_t = Hx_t + Du_t + Cv_t \tag{1}$$

$$x_{t+1} = \Phi x_t + \Gamma u_t + Ew_t \tag{2}$$

where the observation equation (1) generates the $(m \times 1)$ vector of measures z_t, $t = 1, 2, ..., N$, u_t is a $(r \times 1)$ vector of observable inputs and the state equation (2) describes the evolution of the $(n \times 1)$ state vector x_t.

We make the following assumptions about (1)-(2):

1) The system is gaussian, i.e.
   1.1) $w_t \sim \text{IID}N(0, Q)$, $v_t \sim \text{IID}N(0, R)$, $\text{cov}(w_t v_t) = S$, for all $t = 1, 2, ..., N$,
   and 1.2) the initial state is independent of w_t and v_t and such that $x_1 \sim N(\bar{x}_1, P_1)$.

2) The parameter matrices $H, D, C, \Phi, \Gamma, E, Q, R$ and S are known; whereas $\bar{x}_1$ and P_1 are unknown.

Denoting the information available up to $t = j$ by: $\Omega_j = \{z_1, z_2, ..., z_j, u_1, u_2, ..., u_j\}$ and the first two conditional moments of the state vector by: $x_{t|j} = E(x_t \mid \Omega_j)$ and $P_{t|j} = E[(x_t - x_{t|j})(x_t - x_{t|j})^T \mid \Omega_j]$, a fixed-interval smoother is an algorithm to obtain estimates of $x_{t|N}$ and $P_{t|N}$. The operation of most smoothers is similar. In a first (*forward*) phase data are filtered from t=1 up to t=N. In a second (*backward*) phase the filtered moments are corrected from t=N up to t=1.

Smoothing is useful to interpolate missing data (Kohn and Ansley 1986), to 'clean' signals contaminated by noise (Kohn and Ansley 1987), to obtain EM estimates of the parameter matrices in state-space models (Shumway and Stoffer 1982), to compute efficient estimates of time-varying parameters (Swamy and Tavlas 1995) and to calculate the residuals of a state-space model (Kohn and Ansley 1989).

Model (1)-(2) can be stationary, nonstationary or partially nonstationary, depending on the eigenvalues of Φ. Also, u_t may include deterministic and/or stochastic inputs. These two issues - stationarity of the system and stochastic nature of the inputs - affect crucially the values of $\bar{x}_1$ and P_1, which are the initial conditions for the Kalman filter. The combined effect of nonstationarity and stochastic inputs on filter initialization was analyzed by Casals and Sotoca (1997) in a framework of maximum-likelihood estimation.

Several authors (e.g. De Jong 1989; Kohn and Ansley 1989) emphasize the importance of filter initialization in the forward phase of a smoother and propose satisfactory solutions for stationary systems with deterministic inputs. The difficulties in the nonstationary case arise from the fact that the initial state covariance matrix P_1 is arbitrarily close to infinity. In this situation the standard Kalman filter cannot be used, as the approximation $P_1 = kI$ with an arbitrary 'big' value of k often induces a numerical degradation.

Literature suggests several solutions to the problem of smoothing of nonstationary systems. One approach uses the Information filter, which is not general - for example, it requires Φ to be nondefective - and can be computationally inefficient (Ansley and Kohn 1985). Other proposal consists of modifying the Kalman filter to allow for partially diffuse priors (Ansley and Kohn 1989) but this excludes the stationary case. Finally, De Jong (1991b) derives a diffuse version of the Kalman Filter, which is used in different frameworks by De Jong (1991a) and De Jong and Chu-Chun-Lin (1994b). None of the previously mentioned methods take into account the effect on filter initialization of the existence of inputs and their stochastic nature. This issue is very relevant for some applications, e.g. for the statistical analysis of time series.

The layout of this article is as follows. In Section 2 we derive a relationship between the exact smoothed moments and those obtained from a conventional smoother arbitrarily initialized with $\bar{x}_1 = 0$, $P_1 = 0$. Section 3 combines this result with De Jong (1989) smoother to obtain an algorithm which is exact and general, as it can be applied to stationary, nonstationary or partially nonstationary systems, with deterministic and/or stochastic inputs. Also, it inherits the computational efficiency and stability of De Jong (1989) smoother and its analytical derivation is pretty straightforward. Section 4 points out some immediate extensions of this algorithm to forecasting, fixed-point and fixed-lag smoothing problems, discusses briefly the properties of smoothed estimates and outlines a computationally efficient version of the proposed smoother for time-invariant systems and models in steady-state innovations form. The Appendix contains the proof of the theorem in Section 2.

2. AN EXACT EXPRESSION FOR THE SMOOTHED MOMENTS.

Consider the state-space model:

$$z_t^* = Hx_t^* + Du_t + Cv_t \tag{3}$$

$$x_{t+1}^* = \Phi x_t^* + \Gamma u_t + Ew_t \tag{4}$$

where the states and measures correspond to the initial conditions $x_1^* = 0$ and $P_1^* = 0$. Propagating the state equations (2) and (4), it follows that:

$$x_t = \Phi^{t-1} x_1 + x_t^* \tag{5}$$

where x_t^* is independent of x_1. Hence, the conditional expectation of (5) is:

$$x_{t|N} = \Phi^{t-1} x_{1|N} + x_{t|N}^* \tag{6}$$

Also from (1)-(2) and (3)-(4) it is easy to prove (Rosenberg 1973) that:

$$\tilde{z}_t = H\overline{\overline{\Phi}}_{t-1} x_1 + \tilde{z}_t^* \tag{7}$$

where $\tilde{z}_t = z_t - z_{t|t-1}$ is the sequence of Kalman Filter innovations corresponding to (1)-(2); $\tilde{z}_t^*$, defined accordingly by $\tilde{z}_t^* = z_t^* - z_{t|t-1}^*$, are the innovations resulting from the Kalman filter applied to (3)-(4) and initialized with $\bar{x}_1 = 0$ and $P_1 = 0$, hereafter KF(0,0); finally, the matrices $\overline{\overline{\Phi}}_t$ are given by $\overline{\overline{\Phi}}_t = (\Phi - K_t H)\overline{\overline{\Phi}}_{t-1}$ with $\overline{\overline{\Phi}}_1 = I$. Eq. (7) can be written for all the sample as:

$$\tilde{z} = X x_1 + \tilde{z}^* \tag{8}$$

where X is the block-diagonal matrix whose t-th block is $H\overline{\overline{\Phi}}_{t-1}$ and the $(m \times N) \times 1$ vectors $\tilde{z}$ and $\tilde{z}^*$ contain the KF(0,0) innovations $\tilde{z}_t$ and $\tilde{z}_t^*$. Note that $\tilde{z}^*$ is independent of x_1.

The problem consists then of obtaining the conditional expectations in the right-hand-side of (6), taking into account the relationship (8). The following theorem states the solution:

Theorem. The exact smoothed moments of the state in (1)-(2) can be expressed as:

$$x_{t|N} = \left\{\Phi^{t-1} - E[x_t^*(\tilde{z}^*)^T]B^{-1}X\right\} x_{1|N} + E[x_t^*(\tilde{z}^*)^T]B^{-1}\tilde{z} \tag{9}$$

$$P_{t|N} = \left\{\Phi^{t-1} - E[x_t^*(\tilde{z}^*)^T]B^{-1}X\right\} P_{1|N} \left\{\Phi^{t-1} - E[x_t^*(\tilde{z}^*)^T]B^{-1}X\right\}^T + P_{t|N}^* \tag{10}$$

where B is a block-diagonal matrix that contains the covariance matrices of $\tilde{z}_t^*$ and $P_{t|N}^*$ is the second-order smoothed moment of the state in (3)-(4).

Proof. See the Appendix.

Note that Eqs. (9)-(10) apply to both stationary and nonstationary systems, as the only

terms affected by P_1 are $x_{1|N}$ and $P_{1|N}$, and this dependence occurs through P_1^{-1}, which is finite. In addition, the deterministic and/or stochastic nature of the inputs also affects the smoothed moments through $x_{1|N}$ and $P_{1|N}$. Both remarks are justified by Eqs. (A.1)-(A.2).

The computation of $E[x_t^*(\tilde{z}^*)^T]B^{-1}X$ and $E[x_t^*(\tilde{z}^*)^T]B^{-1}\tilde{z}$ in (9)-(10) depends on the specific smoothing algorithm to be used. This issue if further discussed in next section.

3. AN EXACT SMOOTHING ALGORITHM.

The backward equations of De Jong (1989) smoother are:

$$x_{t|N} = x_{t|t-1} + P_{t|t-1} r_{t-1} \tag{11}$$

$$P_{t|N} = P_{t|t-1} - P_{t|t-1} R_{t-1} P_{t|t-1} \tag{12}$$

$$r_{t-1} = H^T B_t^{-1} \tilde{z}_t + \bar{\Phi}_t^T r_t \quad \text{with} \quad r_N = 0 \tag{13}$$

$$R_{t-1} = H^T B_t^{-1} H + \bar{\Phi}_t^T R_t \bar{\Phi}_t \quad \text{with} \quad R_N = 0 \tag{14}$$

$$\bar{\Phi}_t = \Phi - K_t H \tag{15}$$

where $x_{t|t-1}$ and $P_{t|t-1}$ have been computed in the forward step by a Kalman filter and B_t is the t-th diagonal block of B. Combining (11)-(15) with (9)-(10) yields the following algorithm:

Forward step: Propagate a KF(0,0) and:

$$\overline{\overline{\Phi}}_{t+1} = (\Phi - K_t H)\overline{\overline{\Phi}}_t \quad \text{with} \quad \overline{\overline{\Phi}}_1 = I \tag{16}$$

$$X_t = H\overline{\overline{\Phi}}_t \tag{17}$$

$$W_t = W_{t-1} + X_t^T B_t^{-1} X_t \quad \text{with} \quad W_0 = 0 \tag{18}$$

$$w_t = w_{t-1} + X_t^T B_t^{-1} \tilde{z}_t \quad \text{with} \quad w_0 = 0 \tag{19}$$

Backward step: Propagate (13)-(15) and:

$$V_{t|N} = (I - P_{t|t-1} R_{t-1})\overline{\overline{\Phi}}_t \tag{20}$$

$$x_{t|N} = x_{t|t-1} + P_{t|t-1}r_{t-1} + V_{t|N}x_{1|N} \tag{21}$$

$$P_{t|N} = P_{t|t-1} - P_{t|t-1}R_{t-1}P_{t|t-1} + V_{t|N}P_{1|N}V_{t|N}^{T} \tag{22}$$

starting from:

$$x_{1|N} = P_{1|N}(P_1^{-1}\overline{x}_1 + w_N) \tag{23}$$

and:

$$P_{1|N} = (P_1^{-1} + W_N)^{-1} \tag{24}$$

To see the relationships between (9)-(10) and the algorithm (16)-(24), one should take into account that:

$$E[x_t^*(\tilde{z}^*)^T]B^{-1}\tilde{z} = x_{t|t-1} + P_{t|t-1}r_{t-1} \tag{25}$$

$$P_{t|N}^{*} = P_{t|t-1} - P_{t|t-1}R_{t-1}P_{t|t-1} \tag{26}$$

$$\Phi^{t-1} - E[x_t^*(\tilde{z}^*)^T]B^{-1}X = V_{t|N} \tag{27}$$

From (25) and (26), it is immediate to see that the effect of the arbitrary initialization of the Kalman filter is corrected by the last term in the right-hand-side of (21)-(22).

4. FINAL REMARKS.

There are three issues that should be taken into account about the scope and assumptions of previous analysis:

First, the extension of previous fixed-interval smoothing results to forecasting, fixed-point and fixed-lag smoothing is immediate, as it only requires to modify the definition of the conditioning information set in Eqs. (9)-(10) and combining the results with any standard algorithm.

Second, we derived Eqs. (9)-(10) and the algorithm in section 3 under the gaussian assumption. As it is well known, in this situation smoothing provides maximum-likelihood estimates of the conditional moments. If the system is nonnormal, smoothed estimates are unbiased and efficient in the least-squares sense. These alternative properties apply to our results.

Third, for notational convenience we assumed in previous sections that system (1)-(2) is time-invariant. However, our results do not rely on this fact and, hence, the algorithm (16)-(24) can be generalized to time-varying systems just by adding a subindex to the nonconstant matrices. On the other hand, if the measures to be smoothed are indeed outputs of a constant coefficients system, computational gains can be obtained with the following implementation of the algorithm.

The basic idea consists of reproducing the analysis in section 2 using a KF$(\mathbf{0},\overline{P})$ instead of a KF(**0**,**0**), where $\overline{P}$ is the stationary solution of the algebraic Riccati equation of a Kalman Filter. In this case, the algorithm would be as follows:

Initial step: Compute $\overline{P}$ such that:

$$\overline{P} = \Phi\overline{P}\Phi^T + EQE^T + \overline{K}\overline{B}\overline{K}^T \tag{28}$$

where:

$$\overline{B} = H\overline{P}H^T + CRC^T \tag{29}$$

$$\overline{K} = (\Phi\overline{P}H^T + ESC^T)\overline{B}^{-1} \tag{30}$$

Under general conditions the solution of (28) exists, is unique and positive semi-definite (Chan, Goodwin and Sin 1984) and there are efficient algorithms to compute it (Ionescu, Oara and Weiss 1997).

Forward step: Propagate a KF$(\mathbf{0},\overline{P})$, which simplifies to:

$$x_{t+1|t} = \Phi x_{t|t-1} + \Gamma u_t + \overline{K}\tilde{z}_t \tag{31}$$

$$\tilde{z}_t = z_t - Hx_{t|t-1} - Du_t \tag{32}$$

and propagate the recursions (16)-(19) taking into account that $K_t = \overline{K}$ and $B_t^{-1} = \overline{B}^{-1}$.

Backward step: Propagate (13)-(15) and (20)-(22) taking into account that $K_t = \overline{K}$, $B_t^{-1} = \overline{B}^{-1}$ and $P_{t|t-1} = \overline{P}$.

The computational savings of this simplified recursion arise from a positive trade-off between a) the additional calculations required by the initial step and b) the reduced computation and memory requirements of the forward and backward steps due to three facts: first, the propagation of (31)-(32) requires less computation than the analogous

KF(0,0); second, in the time-invariant case $E[\tilde{z}_t^*(\tilde{z}_t^*)^T] = \overline{B}$ for all t and, hence, this matrix needs to be inverted only once; and third, the fact that that $K_t = \overline{K}$, $B_t^{-1} = \overline{B}^{-1}$ and $P_{t|t-1} = \overline{P}$ for all t reduces memory consumption in all the recursions.

If the system (1)-(2) is in steady-state innovations form, i.e. if $Cv_t = w_t$, previous algorithm can be further and drastically simplified because $\overline{P} = \mathbf{0}$.

REFERENCES.

Ansley, C. F., and Kohn, R. (1985), "Estimation, Filtering and Smoothing in State Space Models with Incompletely Specified Initial Conditions," *Annals of Statistics*, 13, 1286-1316.

Ansley, C. F., and Kohn, R. (1989), "Filtering and Smoothing in State Space Models with Partially Diffuse Initial Conditions," *Journal of Time Series Analysis*, 11, 4, 275-293.

Casals, J., and Sotoca, S. (1997), "Exact Initial Conditions for Maximum Likelihood Estimation of State Space Models with Stochastic Inputs," *Economics Letters*, 57, 261-267.

Chan, S. W., Goodwin, G. C., and Sin K. S. (1984), "Convergence Properties of the Riccati Difference Equation in Optimal Filtering of Nonstabilizable Systems," *IEEE Transactions on Automatic Control*, 29, 2, 110-118.

De Jong, P. (1988), "The Likelihood for a State Space Model," *Biometrika*, 75,1, 165-169.

De Jong, P. (1989), "Smoothing and Interpolation with the State-Space Model," *Journal of the American Statistical Association*, 84, 408, 1085-1088.

De Jong, P. (1991a), "Stable Algorithms for the State Space Model," *Journal of Time Series Analysis*, 12, 2, 143-157.

De Jong, P. (1991b), "The Diffuse Kalman Filter," *Annals of Statistics*, 19, 2, 1073-1083.

De Jong, P., and Chu-Chun-Lin, S. (1994a), "Stationary and Non-Stationary State Space Models," *Journal of Time Series Analysis*, 15, 2, 151-166.

De Jong, P., and Chu-Chun-Lin, S. (1994b), "Fast Likelihood Evaluation and Prediction for Nonstationary State Space Models," *Biometrika*, 81, 1, 133-142.

Ionescu, V., Oara, C., and Weiss, M. (1997), "General Matrix Pencil Techniques for the Solution of algebraic Riccati Equations: a Unified Approach," *IEEE Transactions on Automatic Control*, 42, 8, 1085-1097.

Kohn, R., and Ansley, C. F. (1986), "Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data," *Journal of the American Statistical Association*, 81, 751-761.

Kohn, R., and Ansley, C. F. (1987), "Signal Extraction for Finite Nonstationary Time Series," *Biometrika*, 74, 411-421.

Kohn, R., and Ansley, C. F. (1989), "A Fast Algorithm for Signal Extraction , Influence and Cross-Validation in State Space Models," *Biometrika*, 76, 65-79.

Rosenberg, B. M. (1973), "The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, 2, 4, 399-428.

Shumway, R. H., and Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3, 253-264.

Swamy, P. A. V. B., and Tavlas, G. S. (1995), "Random Coefficient Models: Theory and Applications," *Journal of Economic Surveys*, 9, 2, 165-196.

APPENDIX: PROOF OF EXPRESSIONS (9)-(10).

First, De Jong (1988) shows that the smoothed moments of the initial state are:

$$x_{1|N} = (X^T B^{-1} X + P_1^{-1})^{-1} (P_1^{-1} \bar{x}_1 + X^T B^{-1} \tilde{z}) \quad \text{(A.1)}$$

$$P_{1|N} = (X^T B^{-1} X + P_1^{-1})^{-1} \quad \text{(A.2)}$$

and the exact expressions of $\bar{x}_1$ and P_1 in (A.1)-(A.2) were derived by De Jong and Chu-Chun-Lin (1994a) assuming deterministic inputs, and by Casals and Sotoca (1997) for the general case. Hence, (A.1) provides the first conditional moment in the RHS of (6). Note also the remarkable similarity of (A.1)-(A.2) with the generalized least squares estimate of x_1 in (8).

Second the orthogonal projection lemma states that, for any random vector y_t:

$$E(y_t \mid \Omega_j) = E(y_t \Omega_j^T)[E(\Omega_j \Omega_j^T)]^{-1} \Omega_j \quad \text{(A.3)}$$

Applying this result to x_t^*, and taking into account (8) we obtain:

$$x_{t|N}^* = E(x_t^* \mid Xx_1 + \tilde{z}^*, u) = E[x_t^* (Xx_1 + \tilde{z}^*)^T] \left\{ E[(Xx_1 + \tilde{z}^*)(Xx_1 + \tilde{z}^*)^T] \right\}^{-1} \tilde{z} \quad \text{(A.4)}$$

which, by independence of x_t^* and x_1, simplifies to:

$$x_{t|N}^* = E[x_t^* (\tilde{z}^*)^T][X P_{1|N} X^T + B]^{-1} \tilde{z} \quad \text{(A.5)}$$

Applying the matrix inversion lemma to (A.5) and substituting in (6) we obtain:

$$x_{t|N} = \Phi^{t-1} x_{1|N} + E[x_t^* (\tilde{z}^*)^T] B^{-1} (\tilde{z} - X x_{1|N}) \quad \text{(A.6)}$$

which yields (9) after rearranging some terms. ■

On the other hand, taking into account Eqs. (8) and (25), we can write (9) as:

$$x_{t|N} = \left\{ \Phi^{t-1} - E[x_t^* (\tilde{z}^*)^T] B^{-1} X \right\} (x_{1|N} - x_1) + x_{t|N}^* \quad \text{(A.7)}$$

which by the independence of x_t^* and x_1 implies (10). ■