4
Restricted Kalman filtering: methodological issues

This chapter is concerned with some presumed new methods about imposing linear restrictions in state space modeling. The plan I will follow is this. In section 4.1, I propose an alternative restricted Kalman filtering that is indicated to situations in which the linear restrictions are time-invariant and the state vector follows a general random walk. In section 4.2, I present another alternative restricted Kalman filtering under a reduced linear state space model, which will be confronted with the previous augmented restricted Kalman filtering from several standpoints. At last, section 4.3 deals with the imposition of linear restrictions in the prediction of the state vector.

4.1
Random walk state vectors under time-invariant restrictions

In this section, the paradigm of augmenting the measurement equation, in order to accomplish linear restrictions in state vector estimation, changes. Actually, this brief change in course deserves some attention because it may highlight a potential framework in restricted Kalman filtering.

The result of this section, the proof of which is still carried out by elementary Hilbert space theory, is the following:

Theorem 5 *If the linear state space model in (2-1) is such that* $c_t = 0$ *and* $T_t = R_t = I$*, then (i)*$A\alpha_1 = q$ *(with* q *deterministic) and (ii)*$AQ_tA' = 0$ *for all t=1,2,... are sufficient to*

$$Aa_{t|j} = q \quad for\ all \quad t, j = 1, 2, ... \tag{4-1}$$

Proof: Fix t and j. Once again, denote by $\pi_{S'}$ the linear orthogonal projection onto S'. Now observe that, from a trivial recursion on the state equation,

$$\alpha_t = \alpha_1 + \sum_{j=1}^{t-1} \eta_{t-j}. \tag{4-2}$$

Pre-multiplying both sides of (4-2) by A implies

$$A\alpha_t = A\alpha_1 + \sum_{j=1}^{t-1} A\eta_{t-j} = q + 0 = q, \tag{4-3}$$

where the second equality comes from hypotheses (i) and (ii). Denoting the i^{th} row from A by $A_i = [c_{i1} \dots c_{im}]$, it follows that

$$\begin{aligned} A_i a_{t|j} &= c_{i1} a_{t1|j} + \cdots + c_{im} a_{tm|j} = c_{i1} \pi_{S'}(\alpha_{t1}) + \cdots + c_{im} \pi_{S'}(\alpha_{tm}) \\ &= \pi_{S'}(c_{i1}\alpha_{t1} + \cdots + c_{im}\alpha_{tm}) = \pi_{S'}(A_i \alpha_t) = \pi_{S'}(q_i) \\ &= q_i, \end{aligned}$$

where the third, fifth and sixth equalities come respectively from the linearity of $\pi_{S'}$, from (4-3), and from the fact that $q_i \in \mathcal{R}(\pi_{S'}) = S'$. Since t, j and i were taken arbitrarily, the theorem is proved. □

I should make explicit some practical gains from this last proposition, applicable to models in which the state vector evolves as (possibly heteroscedastic) random walks. The first bonus is that there is no need to increase the dimension of the measurement equation any longer. The second is that, by imposing the enunciated restrictions on the initial state vector and on the covariance matrices of the error terms from the state equation, maximum likelihood estimation can be sharply enhanced whenever some of the unknown parameters belong to those matrices. The third advantage is that the restrictions are satisfied by any type of state estimation, whether it is a prediction, updating or smoothing.

4.2 Reduced restricted Kalman filtering

4.2.1 Motivation

In dealing with a linear regression model under linear restrictions, there are two ways of estimation. Actually, both prove to be numerically equivalent and are known by the name of restricted least squares. The first way was already revisited in section 3.3 (cf. expressions in (3-15)), while the second is implemented by rewriting a reduced model with transformed data and then applying usual OLS estimation to the transformed data (cf. Davidson and MacKinnon, 1993).

My aim in this section is to propose a restricted Kalman filtering under a reduced modeling framework. While the usual restricted Kalman filtering by augmentation discussed so far can be viewed as a generalization of the first way to impose linear restrictions in a static linear regression model, the new approach to be now developed, in turn, resembles the second. One feature to be listed among others is that, even though both approaches of restricted least

squares produce exactly the same result, the two restricted Kalman filtering (the augmented and the reduced) *do not* always result in the same estimated state vectors.

4.2.2 The method

In the remaining of this section 4.2, consider the measurement equation in (2-1), the restrictions in (2-5) and the following

Assumption 7 *The (possibly random) vector* $q_t = (q_{t1}, \ldots, q_{tk})'$ *is linearly predeterminate to* $1, Y_1, \ldots Y_t$, *for all* $t = 1, 2, \ldots$ *; that is,* $q_{ti} \in span\{1, Y_{11}, \ldots, Y_{1p}, \ldots, Y_{t1}, \ldots, Y_{tp}\}$, *for all* $i = 1, \ldots, k$ *and* $t = 1, 2, \ldots$.

The basic perspective behind the alternative restricted Kalman filtering is much the same as that of the reduced modeling in linear regression under linear restrictions: some state coordinates are rewritten as an affine function of the others and the result is appropriately placed in the measurement equation.

The *method*:
Let t be an arbitrary time index.

1. Without any loss of generality write the linear restrictions in (2-5) as

$$A_{t,1}\alpha_{t,1} + A_{t,2}\alpha_{t,2} = [A_{t,1}\ A_{t,2}]\left(\alpha'_{t,1}, \alpha'_{t,2}\right)' = q_t, \tag{4-4}$$

   where $A_{t,1}$ is a $k \times k$ full rank matrix.

2. Solve (4-4) for $\alpha_{t,1}$ which should result in

$$\alpha_{t,1} = A_{t,1}^{-1}q_t - A_{t,1}^{-1}A_{t,2}\alpha_{t,2}. \tag{4-5}$$

3. Take (4-5) and put it in the measurement equation of the model in (2-1) - from which we drop d_t without loosing generality at all - aiming to obtain

$$\begin{aligned} Y_t &= Z_{t,1}\alpha_{t,1} + Z_{t,2}\alpha_{t,2} + \varepsilon_t \\ &= Z_{t,1}\left(A_{t,1}^{-1}q_t - A_{t,1}^{-1}A_{t,2}\alpha_{t,2}\right) + Z_{t,2}\alpha_{t,2} + \varepsilon_t \\ &= Z_{t,1}A_{t,1}^{-1}q_t - Z_{t,1}A_{t,1}^{-1}A_{t,2}\alpha_{t,2} + Z_{t,2}\alpha_{t,2} + \varepsilon_t \\ \Rightarrow Y_t^* &\equiv Y_t - Z_{t,1}A_{t,1}^{-1}q_t = \left(Z_{t,2} - Z_{t,1}A_{t,1}^{-1}A_{t,2}\right)\alpha_{t,2} + \varepsilon_t \\ &\equiv Z_{t,1}^*\alpha_{t,2} + \varepsilon_t. \end{aligned}$$

4. Now, postulate a transition equation for the unrestricted state vector $\alpha_{t,2}$. This equation leads to the following *reduced* linear state space model

when it is put together with the measurement equation derived in the last step:

$$\begin{aligned} Y_t^* &= Z_{t,2}^* \alpha_{t,2} + \varepsilon_t \quad , \quad \varepsilon_t \sim (0, H_t) \\ \alpha_{t+1,2} &= T_{t,2}\alpha_{t,2} + c_{t,2} + R_{t,2}\eta_{t,2} \quad , \quad \eta_{t,2} \sim (0, Q_{t,2}) \\ \alpha_{1,2} &\sim (a_{1,2}, P_{1,2}). \end{aligned} \tag{4-6}$$

5. For the reduced model in (4-6), apply the usual Kalman filtering to obtain $a_{t,2|j}$ and $P_{t,2|j}$, for all $j \geq t$.

6. Reconstitute the estimate $a_{t,1|j}$ and its mean square error matrix $P_{t,1|j}$ by means of the affine relation given in (4-5):

$$\begin{aligned} a_{t,1|j} &= A_{t,1}^{-1} q_t - A_{t,1}^{-1} A_{t,2} a_{t,2|j} \\ P_{t,1|j} &= (A_{t,1}^{-1} A_{t,2}) P_{t,2|j} (A_{t,1}^{-1} A_{t,2})'. \end{aligned} \tag{4-7}$$

The above algorithm deserves some qualification. First, I must say that the approach is not completely new, since a particular case was conveniently used in Doran and Rambaldi (1997); what I am doing here is to put it in a more general framework. In addition, observe that j *does have to be greater than or equal to* t due to steps 5 and 6 (cf. Assumption 7). Another aspect is that the specification for the state equation in step 4 could be extracted from the *complete* state equation in (2-1), but if one does not want to think or worry about a full transition system, then one could concentrate only in modeling the block $\alpha_{t,2}$.

4.2.3 Reducing versus augmenting

Among the advantages of the reduced model approach over the augmented model, I cite:

– *Mathematical consistency*: Once the state equation is chosen after the reducing task, the method avoids any risk of obtaining measurement and state equations theoretically inconsistent with each other.

– *Computational efficiency*: While the augmenting approach increases the dimension of the practical problem (indeed, the length of the measurement vectors increases from p to $p + k$!), the reduced model approach goes in an opposite direction by not altering the size of the measurement equation and shortening the size of the state equation (from m to $m-k$). In other words, the augmenting approach "augments" the dimensions of the practical problem while the reduced model approach "reduces" them.

– *Model selection*: The reduced model approach enables one to investigate the plausibility of the assumed linear restrictions by using information criteria (*e.g. AIC* and *BIC*). The competing model would be the unrestricted one as given by (2-1), the (*quasi*) likelihood function of which is surely comparable with that one from the restricted model in equations (4-6).

Stepping further towards the comparison between the reducing and the augmenting approaches, I present two results. Both are related to the augmented model suggested in Theorem 1, and reveal that, for certain types of state restrictions, it is much less flexible. The first proposition concerns limitations on the state equation. Note that the first three conditions listed below are quite general, since they are verified for several state space specifications (*e.g.* zero-mean initial state vectors, whatever diffuse or non-diffuse) and for many types of linear restrictions (*e.g.*, all the deterministic ones):

Proposition 1 *Suppose the partition in (4-4) is such that* $A_{t,1} \equiv A_1$. *Also, admit the following conditions:*

(i) $T_t = diag\,(T_{t,1}, T_{t,2})$, *where* $T_{t,1}$ *is* $k \times k$.

(ii) $\left(A_1^{-1}A_{t,2}T_{t,2} - T_{t,1}A_1^{-1}A_{t,2}\right) E\,(\alpha_{t,2}) = 0$.

(iii) $E\,(q_t) = E\,(q_{t+1}) = \bar{q}$.

Then, (i), (ii) *and* (iii) *are sufficient for* $T_{t,1} = I_{k\times k}$. *Now, suppose* (i), (ii) *and* (iii) *valid for all* $t \geq 1$ *and consider the additional conditions:*

(iv) $\forall t \geq 1 : A_{t,2} \equiv A_2$, *such that* A_2 *has null kernel.*

(v) $\forall t \geq 1 : q_t \equiv q$ *(possibly random).*

Now, (i) *to* (v) *are sufficient for* $T_t = I_{m\times m}$.

Proof: For ease of notation, set $c_t = 0$ and $R_t = I$ in the augmented version of model (2-1). From (4-5) and from condition (i), I have

$$A_1^{-1}q_{t+1} - A_1^{-1}A_{t,2}T_{t,2}\alpha_{t,2} - A_1^{-1}A_{t,2}\eta_{t,2} = T_{t,1}A_1^{-1}q_t - T_{t,1}A_1^{-1}A_{t,2}\alpha_{t,2} + \eta_{t,1},$$

which is equivalent to

$$A_1^{-1}q_{t+1} - T_{t,1}A_1^{-1}q_t = \left(A_1^{-1}A_{t,2}T_{t,2} - T_{t,1}A_1^{-1}A_{t,2}\right)\alpha_{t,2} + \eta_{t,1} + A_1^{-1}A_{t,2}\eta_{t,2}. \quad (4\text{-}8)$$

Taking expectations on both sides of (4-8) and using conditions (ii) and (iii), I arrive at

$$\left(I - T_{t,1}\right) A_1^{-1}\bar{q} = 0. \quad (4\text{-}9)$$

From (4-9), I necessarily have $T_{t,1} = I_{k\times k}$. Finally, under (i), (ii) and (iii) valid for all $t \geq 1$, the conditions (iv) and (v) imply $T_{t,2} = I_{(m-k)\times(m-k)}$ (indeed: get $A_2\alpha_{t,2} = q - A_1\alpha_{t,1}$ from (4-4), pre-multiply this latter identity by a left inverse of A_2, and recall that affine functions of random walks are also random walks). □

It becomes clear from Proposition 1 that, if one chooses the augmenting approach for dealing with important types of restrictions, there would be no possibility left but a random walk evolution for at least a block of the state vector.

The second proposition is stated below. Its condition (vii), as one can directly see, is a quite natural set-up, since this avoids some pathological behaviors from the measurement equation, such as non-ergodic stationarity:

Proposition 2 *Suppose conditions* (i), (ii) *and* (iii) *of Proposition 1 are valid for all* $t \geq 1$, *as well as* (iv) *and* (v), *with* q *degenerated. Also assume that:*

(vi) $\forall t \geq 1 : Q_t \equiv diag\left(\sigma_{t1}^2, \ldots, \sigma_{tm}^2\right)$.

(vii)$\forall t \geq 1$ *and* $\forall i = 1, \ldots, m : \sigma_{1i}^2 = \cdots = \sigma_{ti}^2 = 0 \Rightarrow Var\left(\alpha_{1i}\right) = 0$.

Then, $Q_t = O_{m \times m}$ *for all* $t \geq 1$.

Proof: Take an arbitrary $t \geq 1$. From (i) to (v), the decomposition in (4-4) collapses to $A_1\alpha_{t+1,1} + A_2\alpha_{t+1,2} = q$. This implies $rank\left(V\left(\alpha_{t+1}\right)\right) \leq m - k$. But, as $T_s = I$ for all $s = 1 \ldots t$ (cf. Proposition 1), I must have $max\left\{rank\left(P_1\right), rank\left(Q_1\right), \ldots, rank\left(Q_t\right)\right\} \leq m - k$. Then, using (vi), there exist $i_1, \ldots, i_k \in \{1, \ldots, m\}$ such $\sigma_{s\,i_j}^2 = 0$ for all $s = 1, \ldots, t$ and $j = 1, \ldots, k$. Conveniently rearranging α_{t+1}, I get a partition $\left(\alpha_{t+1,1}^{*\prime}, \alpha_{t+1,2}^{*\prime}\right)'$ such that $Var\left(\alpha_{t+1,1}^*\right) = O_{k \times k}$ (cf. Proposition 1 again and condition (vii)). Than,

$$O_{k \times k} = Var\left(q^*\right) = Var\left(A_1^*\alpha_{t+1,1}^* + A_2^*\alpha_{t+1,2}^*\right) = A_2^* Var\left(\alpha_{t+1,2}^*\right) A_2^{*\prime}. \quad (4\text{-}10)$$

From (4-10) I finally obtain $Q_{s,2}^* = O_{(m-k)\times(m-k)}$ for all $s = 1, \ldots t$. □

This last result rules out any possibility of non-degenerated state vectors under contemporaneously uncorrelated errors $\eta_{t1}, \ldots, \eta_{tm}$. This limitation, as the previously raised from Proposition 1, surely does not arise under the reducing approach.

4.2.4 Geometrical considerations

Let me now grasp some intuitive insight from the described method and therefore geometrically understand what this alternative restricted Kalman filtering, as well as the previous one by augmentation, is in fact "doing" to the state vector.

I in the first place defend that one could work in an equivalent way with the model

$$\begin{aligned} Y_t &= Z^*_{t,2}\alpha_{t,2} + d^*_t + \varepsilon_t \ \ , \ \ \varepsilon_t \sim (0, H_t) \\ \alpha_{t+1,2} &= T_{t,2}\alpha_{t,2} + c_{t,2} + R_{t,2}\eta_{t,2} \ \ , \ \ \eta_{t,2} \sim (0, Q_{t,2}) \\ \alpha_{1,2} &\sim (a_{1,2}, P_{1,2}), \end{aligned} \tag{4-11}$$

where $d^*_t \equiv Z_{t,1}A^{-1}_{t,1}q_t$, since (4-6) and (4-11) are equivalent linear state space representations. Indeed, by Assumption 7, d^*_t just defined is necessarily deterministic - or linearly predeterminate to $Y_1, \ldots, Y_t$. So the sets $\{1, Y_{11}, \ldots, Y_{1p} \ldots, Y_{j1}, \ldots, Y_{jp}\}$ and $\{1, Y^*_{11}, \ldots, Y^*_{1p} \ldots, Y^*_{j1}, \ldots, Y^*_{jp}\}$ produce the same univariate innovations in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, which implies they span the same subspace. Therefore, the Kalman filtering (updating or smoothing equations) applied to (4-6) - or, as already argued, to (4-11) - is projecting each α_{tj}, $j = k+1, \ldots, m$ onto $span\{1, Y_{11}, \ldots, Y_{1p}, \ldots, Y_{j1}, \ldots, Y_{jp}\}$, $j \geq t$, as it is done in a regular state space estimation. *But (4-6) - or equivalently (4-11) - together with (4-5) make explicit the fact that* $\alpha_{t,1} = (\alpha_{t1}, \ldots, \alpha_{tk})'$ *can be affinely extracted from* $\alpha_{t,2} = (\alpha_{t,k+1}, \ldots, \alpha_{tm})'$. Then, it becomes possible to project each coordinate of $\alpha_{t,2}$ first and subsequently obtain the projections of each coordinate of $\alpha_{t,1}$ using (4-7). In light of such considerations, one could consider the reduced model approach as some kind of a "two-stage" state estimation.

Within the previous augmenting procedure, one in turn has to project directly (by means of the Kalman equations applied to an augmented model) each coordinate of the entire $\alpha_t = (\alpha_{t1}, \ldots, \alpha_{tm})'$ onto the *bigger* subspace $span\{1, Y_{11}, \ldots, Y_{1p}, q_{11}, \ldots, q_{1k}, \ldots, Y_{j1}, \ldots, Y_{jp}, q_{j1}, \ldots, q_{jk}\}$ - strictly bigger if at least one of the $q_1, \ldots, q_t$ is not linearly predeterminate to the measurements. Figures 4.1 and 4.2 illustrate these highlighted geometrical differences.

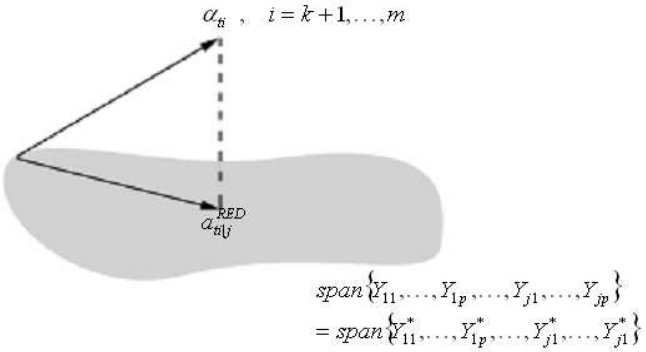

Figura 4.1: Geometrical meaning of this section's reduced model approach: Here only part of the coordinates of the state vector are directly projected onto the original spanned subspace. The other coordinates projections are obtained by formula (4-7).

4.3

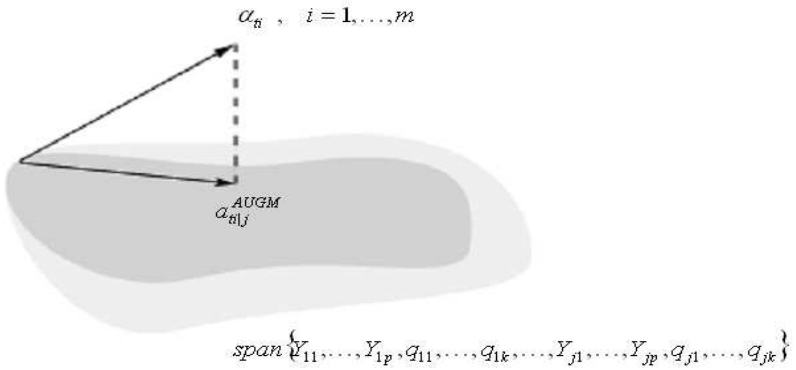

Figura 4.2: Geometrical meaning of the previous augmenting approach: Here each coordinate of the state vector is directly projected onto an augmented subspace.

Predictions from a restricted state space model

The original proposal for adopting an augmented model, which I re-evoke again in this section, does not, in general, guarantee that linear restrictions on the state vector are carried over to the Kalman prediction equations (*immediate example*: except for the local level model, there is no extension of Corollary 1 for the prediction equations when one is dealing with any of the structural models - cf. Harvey, 1989 - put in their respective state space forms). However, there is one exception. This particular case is described by a state vector that follows a possibly heteroscedastic random walk, and is considered in the following

Corollary 2 *Under the conditions presented in Theorem 1, in addition to* (i) $c_t = 0$ *and* (ii) $T_t = R_t = I$ *, it follows that*

$$A_t a_{t+1|t} = q_t.$$

In this section I propose a simple strategy to further extend (namely, for *any* type of linear state space model) the restricted Kalman filtering and smoothing up to schemes aimed at prediction. As a matter of fact, what I seek certainly differs from the method by Pandher (2002). In turn, the grounds of my proposal are built up on the ideas of missing values state space treatment and of the decomposition used in the second proof of Theorem 1 and in the proof of Theorem 3.

Consider that one is willing to extrapolate the state vector and/or the measurements up to h steps ahead in the future; that is, one wants to obtain $a_{n+1|n},\ldots,a_{n+h|n}$ and/or $\hat{Y}_{n+1|n},\ldots,\hat{Y}_{n+h|n}$. But, similarly to everything that has been done so far in this Thesis, it is known *a priori* that, for all $j = 1,\ldots,h$, $A_{n+j}\alpha_{n+j} = q_{n+j}$, where A_{n+j} is a $k \times m$ matrix and q_{t+j} is

a $k \times 1$ (possibly random) vector; this knowledge is nothing more than the confirmation that Assumption 1 is not confined to a particular time series of size n. So the question is how to make $a_{n+1|n}, \ldots, a_{n+h|n}$ satisfy those same theoretical constraints.

The proposed answer starts again from adopting an augmented model. Then, the augmented version of (2-1) is rewritten in order to accomplish (or *recall* to better say) the enunciated "future" restrictions:

$$\begin{pmatrix} Y_t \\ q_t \end{pmatrix} = \begin{pmatrix} Z_t \\ A_t \end{pmatrix} \alpha_t + \begin{pmatrix} d_t \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & 0 \end{pmatrix} \right) \tag{4-12}$$

$$\begin{aligned} \alpha_{t+1} &= T_t\alpha_t + c_t + R_t\eta_t \ , \ \eta_t \sim (0, Q_t) \\ t &= 1, \ldots, n, n+1, \ldots, n+h. \end{aligned}$$

Stepping ahead, observe now that the model in (4-12) can be decomposed in a way that stresses that the researcher is actually dealing with the (possibly multivariate) series

$$Y_1, q_1, Y_2, q_2, \ldots, Y_n, q_n, Y_{n+1}, q_{n+1}, Y_{n+2}, q_{n+2}, \ldots, Y_{n+h}, q_{n+h}, \tag{4-13}$$

where there are missing measurements; $Y_{n+1}, Y_{n+2}, \ldots, Y_{n+h}$ are obviously absent up to time n. So the series in (4-13) presents blanks and should be appropriately recast as

$$Y_1, q_1, Y_2, q_2, \ldots, Y_n, q_n, \quad , q_{n+1}, \quad , q_{n+2}, \ldots, \quad , q_{n+h}. \tag{4-14}$$

The obtention of $a_{n+1|n}, a_{n+2|n}, \ldots, a_{n+h|n}$ is then almost equivalent to the application of the Kalman smoothing to the "incomplete" series in (4-14), using the following equivalent version of (4-12). The measurement equation is defined by

$$Y_{t,i} = Z_{t,i}\alpha_{t,i} + d_{t,i} + \varepsilon_{t,i} \ , \ \varepsilon_{t,i} \sim (0, H_{t,i}).$$

When $i = 1$, nothing is changed from the measurement equation in (2-1) of section 2.1. But for $i = 2$ we must have

$$Y_{t,2} = q_t, \ \ Z_{t,2} = A_t, \ \ d_{t,2} = 0 \text{ and } H_{t,2} = 0.$$

Regarding the state equation, just notice that, for all t, $\alpha_{t,2} = \alpha_{t,1}$ and $\alpha_{t+1,1} = T_t\alpha_{t,2} + c_t + R_t\eta_t$, $\eta_t \sim (0, Q_t)$. The use of the word "almost" is justified by the fact that the information from $q_{n+1}, \ldots, q_{n+h}$ does enter in the Kalman estimation; so the result does not necessarily equal to $a_{n+1|n}, a_{n+2|n}, \ldots, a_{n+h|n}$, which theoretically use the information only up to q_n. The modifications in

the original Kalman equations due to missing observations are discussed in Durbin and Koopman (2001), section 4.8. If the researcher wants to treat the series in (4-14) under a univariate framework, he or she shall refer to section 6.4 of that same book - this univariate choice brings some gains exclusively on the computational side since the whole vectors $Y_{n+1}, Y_{n+2}, \ldots, Y_{n+k}$ are lacking. Finally, notice that Theorem 1 is doing its job by guaranteeing that, for all $j = 1, \ldots, h$, $A_{n+j}a_{n+j|n} = q_{n+j}$. For the reasons just explained in this paragraph, "$n + j|n$" is an abuse of notation.

It is now time to harvest the resuming computational algorithm:

1. Decompose the model in (4-12) striving to get the series in (4-14).

2. Store the "new" observations while respecting the missing values positions.

3. Apply the Kalman smoothing equation to the stored observations, appropriately modified to account for the missing values.

4. Take the smoothed states corresponding to the missing values positions as the *predicted state vectors under linear restrictions*.