3
Restricted Kalman filtering: theoretical issues

During all this chapter, I will discuss several topics concerning the theory of imposing the linear restrictions enunciated under a quite general form in (2-5) from Assumption 1. In section 3.1, I will present and compare three different derivations of the restricted Kalman updating and smoothing equations under an augmented modeling approach. In section 3.2, I prove the statistical efficiency due to the imposition of restrictions, and this shall be done using a geometrical framework. Stepping forward, I try in section 3.3 to establish the equivalence between the restricted Kalman filtering and something that could be termed *recursive restricted least squares* estimator. Finally, in section 3.4, I investigate how initial diffuse state vectors affect the use of the Kalman smoother under linear restrictions.

3.1
Augmented restricted Kalman filtering: alternative proofs

3.1.1
Geometrical proof

When estimating state space models under linear restrictions as given in eq. (2-5), the natural task is to impose these very restrictions on the state estimators given by the Kalman equations in order to obtain a more meaningful result. The following theorem guarantees that such task is possible for the updating and smoothing equations, whenever one adopts an augmented measurement equation:

Theorem 1 *If the measurement vectors Y_t are replaced by $Y_t^* = (Y_t', q_t')'$, the matrices Z_t are replaced by $Z_t^* = [Z_t' \; A_t']'$, the vectors d_t are replaced by $d_t^* = (d_t', 0')'$, and the measurement equation error vectors ε_t are replaced by $\varepsilon_t^* = (\varepsilon_t', 0')'$, then the Kalman updating and smoothing equations applied to this new linear state space models satisfy the same linear restrictions given in (2-5), that is,*

$$A_t a_{t|t} = q_t \tag{3-1}$$

$$A_t a_{t|n} = q_t \tag{3-2}$$

First proof of Theorem 1: Denote the subspace generated by the augmented measurements up to time j, where $j \in \{t, t+1, \ldots, n\}$, by $S'' = span\{1, Y_{11}, \ldots, Y_{1p}, q_{11}, \ldots, q_{1k}, \ldots, Y_{j1}, \ldots, Y_{jp}, q_{j1}, \ldots, q_{jk}\}$, the unique linear orthogonal projection onto S'' by $\pi_{S''}$ and the i^{th} row from A_t by $A_{ti} = [c_{ti1} \ldots c_{tim}]$. Then, by making use of linearity of $\pi_{S''}$ and of the linear restrictions established in Assumption 1, it follows that

$$\begin{aligned} A_{ti} a_{t|j} &= c_{ti1} a_{t1|j} + \cdots + c_{tim} a_{tm|j} = c_{ti1} \pi_{S''}(\alpha_{t1}) + \cdots + c_{tim} \pi_{S''}(\alpha_{tm}) \\ &= \pi_{S''}(c_{ti1}\alpha_{t1} + \cdots + c_{tim}\alpha_{tm}) = \pi_{S''}(A_{ti}\alpha_t) = \pi_{S''}(q_{ti}) \\ &= q_{ti} \ , \end{aligned}$$

where the last equality comes from the fact that q_{ti} belongs to $\mathcal{R}(\pi_{S''}) = S''$. Since i is arbitrary, the theorem is proved. □

Theorem 1 was originally due to Doran (1992, p.570, 571), but our proof also reveals some gains:

1. It does not presume that F_t is invertible for all t.
2. It unifies in a single argument both updating and (*any type of*) smoothing equations.
3. It does not make any explicit use of Kalman updating or smoothing equations.
4. It is a shorter and consequently more elegant proof.

The first methodological contribution of this proof, namely the guarantee that the augmented measurement procedure is able to deal with *any* type of linear restriction, is directly related to item 1. Many examples of restrictions that would decrease the rank of F_t are of deterministic nature, whether they originate from economic theories or not (to be even more specific: consider for instance the *portfolio* accounting restriction in time-varying extensions of the asset class factor models due to Sharpe, 1992). The second contribution, related to item 2, is that *any* set of state smoothing - from which we mention the traditional *fixed-interval*, *fixed-point* and *fixed-lag* estimators (cf. Anderson and Moore, 1979) - must yield restricted estimated states.

The following consequence from Theorem 1 has already proved to be useful, once it was conveniently used by Doran (1996) in a state space estimation of population totals.

Corollary 1 *If some univariate equations of the measurement vector Y_t have errors with zero variance, then*

$$Z_{t2}a_{t|t} = Y_{t2} \ \ and \ \ Z_{t2}a_{t|n} = Y_{t2} \tag{3-3}$$

where Z_{t2} is the block from Z_t that corresponds to the block Y_{t2} from Y_t whose coordinates have null variance errors.

Proof: It is enough to see that Y_t can be written as $[Y_{t1}' \ Y_{t2}']'$ and that Z_t, in turn, can be written as $[Z_{t1}' \ Z_{t2}']'$. Establishing that $A_t = Z_{t2}$ and $q_t = Y_{t2}$, Theorem 1 guarantees the desired result. □

3.1.2
Computational proof

During this subsection, I shall consider the following additional structure to the restrictions in (2-5):

Assumption 2 *The linear restrictions in (2-5) are such that the coordinates of q_t are linearly independent in L_2 and from $1, Y_{11}, \ldots, Y_{1p}, \ldots, Y_{t1}, \ldots, Y_{tp}$. Also, suppose that $F_t > 0$ for all t.*

For the Kalman updating and smoothing equations, it is in fact an attainable task, as Theorem 1 says, to carry out Kalman filtering estimations under the above these linear restrictions. Here, this is now proved by explicitly using updating and smoothing equations, even though under strategies somewhat different from those tackled by Doran (1992).

Second proof of Theorem 1: Uncouple the augmented model by recognizing that, for all t, q_t is a "new" measurement vector that is observed "after" Y_t and "before" Y_{t+1}. This recognition leads to a new linear state space representation entirely equivalent to the augmented model. The measurement equation for this representation is defined by

$$Y_{t,j} = Z_{t,j}\alpha_{t,j} + d_{t,j} + \varepsilon_{t,j} \ , \ \varepsilon_{t,j} \sim (0, H_{t,j}) \tag{3-4}$$

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

$$Y_{t,2} = q_t, \ \ Z_{t,2} = A_t, \ \ d_{t,2} = 0 \text{ and } H_{t,2} = 0. \tag{3-5}$$

Regarding the state equation, just notice that, for all t, $\alpha_{t,2} = \alpha_{t,1}$ and $\alpha_{t+1,1} = T_t\alpha_{t,2} + c_t + R_t\eta_t$, $\eta_t \sim (0, Q_t)$. Within this equivalent framework, it becomes possible to treat the imposing of the linear restriction in time t as a new update of the state vector. Implementing: consider the state updating equation given in (2-3), already applied to the above equivalent model for t fixed and $j = 2$:

$$\begin{aligned} a_{t,2|t,2} &= a_{t|t-1,2} + P_{t|t-1,2} Z'_{t,2} F_{t,2}^{-1} (Y_{t,2} - Z_{t,2} a_{t|t-1,2}) \\ &= a_{t|t-1,2} + P_{t|t-1,2} Z'_{t,2} (Z_{t,2} P_{t|t-1,2} Z'_{t,2} + H_{t,2})^{-1} (Y_{t,2} - Z_{t,2} a_{t|t-1,2}) \\ &= a_{t|t-1,2} + P_{t|t-1,2} A'_t (A_t P_{t|t-1,2} A'_t)^{-1} (q_t - A_t a_{t|t-1,2}), \end{aligned}$$

where the second equality comes from the very expression of F_t (cf. the established notation in section 2.2) and the third comes from (3-5). Now, since $(A_t P_{t|t-1,2} A'_t)^{-1}$ is a genuine inverse (cf. Assumption 2) and $a_{t,2|t,2} = a_{t|t}$, this last updated state vector being the one associated with the augmented model, pre-multiply both sides of the last identity by A_t in order to get (3-1).
Now, rephrase the state smoothing equations in (2-4) for the augmented model as follows:

$$\begin{aligned} &a_{t|n} = a_{t|t-1} + P_{t|t-1} r_{t-1} \\ &r_{t-1} = Z_t^{*'} F_t^{-1} \upsilon_t + \left(T_t - T_t P_{t|t-1} Z_t^{*'} F_t^{-1} Z_t^{*}\right)' r_t, \quad \text{where} \quad Z_t^* = \begin{bmatrix} Z_t \\ A_t \end{bmatrix}. \end{aligned}$$

Of course, other quantities would also have deserved asterisks, but they are suppressed for ease of notation. Placing the expression of r_t in $a_{t|n}$, it follows that

$$\begin{aligned} a_{t|n} &= a_{t|t-1} + P_{t|t-1}(Z_t^{*'} F_t^{-1} \upsilon_t + (T_t - T_t P_{t|t-1} Z_t^{*'} F_t^{-1} Z_t^{*})' r_t) \\ &= a_{t|t-1} + P_{t|t-1} Z_t^{*'} F_t^{-1} \upsilon_t + P_{t|t-1}(T_t - T_t P_{t|t-1} Z_t^{*'} F_t^{-1} Z_t^{*})' r_t \\ &= a_{t|t} + (P_{t|t-1} T_t' - P_{t|t-1} Z_t^{*'} F_t^{-1} Z_t^{*} P_{t|t-1} T_t') r_t, \end{aligned}$$

where the last equality follows from the Kalman updating equation in (2-3). Pre-multiplying both sides by A_t, it follows that

$$A_t a_{t|n} = A_t a_{t|t} + (A_t P_{t|t-1} T_t' - A_t P_{t|t-1} Z_t^{*'} F_t^{-1} Z_t^{*} P_{t|t-1} T_t') r_t.$$

According to Doran (1992), eq. (22) (from Assumption 2, F_t from the augmented model is invertible), it follows that $A_t P_{t|t-1} Z_t^{*'} F_t^{-1} = \begin{bmatrix} \underset{k \times p}{0} & \underset{k \times k}{I} \end{bmatrix}$. Use this together with (3-1) already proved to obtain

$$\begin{aligned} A_t a_{t|n} &= q_t + \left(A_t P_{t|t-1} T_t' - [0 \;\; I] \begin{bmatrix} Z_t \\ A_t \end{bmatrix} P_{t|t-1} T_t' \right) r_t \\ &= q_t + \left(A_t P_{t|t-1} T_t' - A_t P_{t|t-1} T_t' \right) r_t = q_t, \end{aligned}$$

which gives identity (3-2) $\square$

It is worth to be noticed that there is no methodological novelty here. In turn, the contribution offered here comes from this second proof, which certainly deserves some qualification. Even though not encompassing significant generalizations like those verified in the first proof, and besides being considerably longer, this second proof makes use of simple matrix operations, which illustrate potentially useful strategies that can be evoked more times in future research. Indeed, one should recall that quite the same decomposition used in the part of the proof related to the updating equations has been the great responsible for the well-known treatment of multivariate state space models under an univariate framework; cf. Durbin and Koopman (2001), section 6.4. On the other hand, the part related to the smoothing equation is entirely based on de Jong (1989)'s smoothing recursions, which are mathematically transparent and computationally efficient.

3.1.3
Conditional expectation proof

The main goal of this subsection is to give a third and last proof for the augmented restricted Kalman filtering. For such, I must add other structure (quite "traditional", we would say) to the linear state space model in (2-1).

Assumption 3 *ε_t and η_t are independent (in time, between each other and of α_1) Gaussian stochastic processes. Also, α_1 is a Gaussian random vector.*

Besides considering this new "parametric"framework, denote by $\mathcal{F}_j$ the σ-field generated by the measurement vectors up to time j; that is $\mathcal{F}_j \equiv \sigma\left(Y_1, \ldots, Y_j\right)$. Also set $\hat{a}_{t|j} \equiv E\left(\alpha_t | \mathcal{F}_j\right)$ and $\hat{P}_{t|j} \equiv V\left(\alpha_t | \mathcal{F}_j\right)$. Under Assumption 3, the Kalman recursions are versions of these conditional moments when $j = t-1$, $j = t$ and $j = n$; see Anderson and Moore (1979), Harvey (1989), Harvey (1993), Tanizaki (1996) and Durbin and Koopman (2001). Consequently, all the properties of the conditional expectation can be conveniently used in order to allow a very quick proof for Theorem 1:

Third proof of Theorem 1: Let t be an arbitrary time instant. Define $\mathcal{F}_j^* \equiv \sigma\left(Y_1, q_1, \ldots, Y_j, q_j\right)$. Fixing j in $\{t, t+1, \ldots, n\}$, it follows with probability 1 that

$$A_t \hat{a}_{t|j} = A_t E\left(\alpha_t | \mathcal{F}_j^*\right) = E\left(A_t \alpha_t | \mathcal{F}_j^*\right) = E\left(q_t | \mathcal{F}_j^*\right) = q_t, \qquad (3\text{-}6)$$

where the third equality is due to the restrictions in (2-5) and the fourth equality naturally comes from the very $\mathcal{F}_j^*$-measurability of q_t. Finally make $j = t$ and $j = n$. $\square$

The most evident comparison between this third proof and the previous proofs is concerned with length and elegance. Besides, it maintains the same generality in terms of linear restrictions and state smoothing, which has been guaranteed already by the first proof given in subsection 3.1.1.

Now, on the potentially usefulness from this third proof:

1. The additional normality and independence assumptions, although restricting a little the scope of Theorem 1, can be considered an asset because these are straightforwardly generalizable to other types of state space models - the non-Gaussian and/or nonlinear state space models. The only drawback is that most of statistical techniques designed to handle more general state space models do require the existence of an expression for the conditional laws $p\left(y_t|\alpha_t\right)$, which are obscured by the "singularity" incurred in the augmenting procedure.

2. Finally, the third proof plainly reveals that the *Bayesian approach* for state space modeling (cf. West and Harisson, 1997; Durbin and Koopman, 2001; and Shumway and Stoffer, 2006) can also deal with linear restrictions by adopting augmented measurement equations as well. Yet, one in such case shall be aware of some unavoidable singularities.

3.2 Statistical efficiency

In this section I demonstrate the statistical efficiency - in terms of mean square estimation error - of the restricted Kalman filtering discussed so far. For this, I shall make use of a geometrical perspective, something that might be general enough, while still grasping at intuition and simplicity. For what follows, it is important to bear in mind that the Kalman recursions, in addition to being recursive computational formulae from an operational standpoint, give linear orthogonal projections evaluations onto some specific subspaces spanned by the model measurements.

I begin by quoting the well-established and useful fact:

Lemma 1 *Take a Hilbert space $\mathcal{H}$, two subspaces $\mathcal{M}, \mathcal{N}$ of $\mathcal{H}$ and the linear orthogonal projections $\pi_{\mathcal{M}}$ and $\pi_{\mathcal{N}}$. If $\mathcal{M} \subseteq \mathcal{N}$, then, for each $x \in \mathcal{H}$, $\pi_{\mathcal{M}}\left(\pi_{\mathcal{N}}\left(x\right)\right) = \pi_{\mathcal{M}}\left(x\right)$.*

Proof: $\mathcal{N}$ is, by its own, a Hilbert space (because it is closed) and $\mathcal{M}$ is a closed subspace of $\mathcal{N}$. Then, using the Orthogonal Projection Theorem (cf. Theorem 5.20 of Kubrusly, 2001), we get $\mathcal{N} = \mathcal{M} + \left(\mathcal{M}^{\perp} \cap \mathcal{N}\right)$. So, from Proposition

5.58 of Kubrusly (2001), it follows that $\pi_{\mathcal{N}} = \pi_{\mathcal{M}} + \pi_{\mathcal{M}^{\perp}\cap\mathcal{N}}$; tautologically, for each $x \in \mathcal{H}$,

$$\pi_{\mathcal{N}}(x) = \pi_{\mathcal{M}}(x) + \pi_{\mathcal{M}^{\perp}\cap\mathcal{N}}(x). \tag{3-7}$$

Now, apply $\pi_{\mathcal{M}}$ on both sides of (3-7). □

Some additional notation must also be set:

- $a_{t|j}$, $P_{t|j}$ and S' are defined as previously and relate to the *standard* state space model;
- $a^{*}_{t|j}$, $P^{*}_{t|j}$ and S'' are obtained with the *augmented* state space model, associated with Theroem 1

Now, everything needed for formally guaranteeing the statistical efficiency has been gathered. Two demonstrations are given. Both are based on a strong geometrical appeal and have an inductive style, in the sense that, firstly, individual coordinates of the state vector are tackled and then, in a second moment, the strategy is generalized for arbitrary linear combinations of these coordinates. But they do differ in some aspects. The first proof concentrates on the optimality of the linear orthogonal projection that comes directly from first principles, while the second proof is rather "constructive", uses Lemma 1 and focuses on a standard decomposition.

Theorem 2 *$P^{*}_{t|j} \leq P_{t|j}$ in the usual ordering of symmetric matrices.*

First proof of Theorem 2: Let $i = 1, \ldots, m$. Since the set containing the original model measurements until time j is contained in the correspondent set the from augmented model, it follows that $S' \subseteq S'' \equiv$ *span*$\{1, Y_{11}, \ldots, Y_{1p}, q_{11}, \ldots, q_{1k}, \ldots, Y_{j1}, \ldots, Y_{jp}, q_{j1}, \ldots, q_{jk}\}$. Therefore, from Theorem 5.53 of Kubrusly (2001),

$$E\left[\left(\alpha_{ti} - a^{*}_{ti|j}\right)^2\right] = \inf_{Y \in S''} E\left[(\alpha_{ti} - Y)^2\right] \leq E\left[\left(\alpha_{ti} - a_{ti|j}\right)^2\right].$$

Generalizing: take $x = (x_1, \ldots, x_m)' \in R^m$. Using linearity, the linear orthogonal projections onto S' and onto S'', both evaluated in $x'\alpha_t = x_1\alpha_{t1} + \cdots + x_m\alpha_{tm}$, are given by

$$\begin{aligned} x'a_{t|j} &= x_1 a_{t1|j} + \cdots + x_m a_{tm|j} \\ &\text{and} \\ x'a^{*}_{t|j} &= x_1 a^{*}_{t1|j} + \cdots + x_m a^{*}_{tm|j}\,. \end{aligned}$$

Observing that $x'a_{t|j} \in S''$ (because $a_{ti|j} \in S' \subseteq S'' \forall i = 1, \ldots, m$ and S'' is a linear manifold), it follows that

$$\begin{aligned}
x'P^*_{t|j}x &= x'E\left[(\alpha_t - a^*_{t|j})(\alpha_t - a^*_{t|j})'\right]x = E\left[x'(\alpha_t - a^*_{t|j})(\alpha_t - a^*_{t|j})'x\right] \\
&= E\left[(x'\alpha_t - x'a^*_{t|j})(x'\alpha_t - x'a^*_{t|j})'\right] = E\left[\left(x'\alpha_t - x'a^*_{t|j}\right)^2\right] \\
&= \inf_{Y \in S''} E\left[\left(x'\alpha_t - Y\right)^2\right] \leq E\left[\left(x'\alpha_t - x'a_{t|j}\right)^2\right] \\
&= E\left[(x'\alpha_t - x'a_{t|j})(x'\alpha_t - x'a_{t|j})'\right] = E\left[x'(\alpha_t - a_{t|j})(\alpha_t - a_{t|j})'x\right] \\
&= x'E\left[(\alpha_t - a_{t|j})(\alpha_t - a_{t|j})'\right]x = x'P_{t|j}x.
\end{aligned}$$

Since x is arbitrary, the conclusion is that $P^*_{t|j}$ is, in fact, "less than or equal to" $P_{t|j}$. □

Second proof of Theorem 2: Consider again an arbitrary $i = 1, \ldots, m$. Recall once more that S' and S'' already defined are subspaces (closed linear manifolds) of L_2 and that $S' \subseteq S''$. Theorem 5.20 of Kubrusly (2001) asserts the existence of $\xi \in S''^{\perp}$ such that

$$\alpha_{ti} = a^*_{ti|j} + \xi. \tag{3-8}$$

Also, Theorem 5.20 of Kubrusly (2001) and Lemma 1 (make $\mathcal{H} = L_2$, $\mathcal{M} = S'$, $\mathcal{N} = S''$ and $x = \alpha_{ti}$) assure the existence of $\nu \in S'' \cap S'^{\perp}$ such that

$$a^*_{ti|j} = a_{ti|j} + \nu. \tag{3-9}$$

From the decomposition (3-8), obtain

$$E\left[\left(\alpha_{ti} - a^*_{ti|j}\right)^2\right] = E\left(\xi^2\right). \tag{3-10}$$

Now, add both decompositions (3-8) and (3-9) in order to get $\alpha_{ti} - a_{ti|j} = \xi + \nu$, and evoke Pythagorean theorem - which is licit since ξ and ν are orthogonal - to have

$$E\left[\left(\alpha_{ti} - a_{ti|j}\right)^2\right] = E\left(\xi^2\right) + E\left(\nu^2\right). \tag{3-11}$$

Identities (3-10) and (3-11) assert the claimed efficiency for each coordinate estimation of the state vector. The case of an arbitrary linear combinations $x'\alpha_t$ is dealt with in a similar fashion. □

Looking at cases in which $j \geq t$, the last theorem shows that Kalman updating and smoothing equations, when used with the augmented model, besides respecting the linear restrictions from equation (3-1), give more accurate estimators.

3.3 Restricted Kalman filtering versus restricted recursive least squares

Consider the following univariate special case of model (2-1) where the state vector is time invariant and $Z_t \equiv x_t'$ is a row vector of exogenous explanatory variables:

$$\begin{aligned} Y_t &= x_t'\beta_t + \varepsilon_t \quad , \quad \varepsilon_t \sim (0, \sigma^2) \\ \beta_{t+1} &= \beta_t. \end{aligned} \tag{3-12}$$

This model can and should be viewed as a linear regression model written in a linear state space representation. It is known (see Harvey, 1981 and 1993) that application of the Kalman state updating equation to (3-12) numerically coincides with the method of recursive least squares. Back in the days when matrix inversion was a computational burden, this equivalence proved useful since it turned out to be possible to estimate a regression model with no need to invert a "big" $X'X$ matrix. In addition, this made attainable the updating of ordinary least squares (OLS) estimates whenever new observations added to the data set. Nowadays this equivalence still deserves its merits in statistics and econometrics. Firstly, depending on the ill-conditioning of the regressors, it still may be difficult to invert "big" $X'X$ matrices, a problem that justifies recursive estimation. Secondly, this equivalence is in full connection with the traditional coefficients stability test by Brown *et al.* (1975).

The purpose of this section is to generalize the above parallel in the context of linear restrictions. I shall admit that the coefficient vector of a regression model is supposed to obey certain linear restrictions which are enunciated as

$$A\beta = q, \tag{3-13}$$

where A is a known $k \times m$ matrix, $k \leq m$, and $q = (q_1, ..., q_k)'$ is a known $k \times 1$ vector. Since the main objective is to bridge the restricted recursive estimation to the restricted Kalman filtering, structures (3-12) and (3-13) are now taken together to generate the following augmented measurement equation:

$$\begin{pmatrix} Y_t \\ q \end{pmatrix} = \begin{pmatrix} x_t' \\ A \end{pmatrix} \beta_t + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix} \right). \tag{3-14}$$

From Theorem 1, the application of the Kalman updating equation to the model in (3-14) produces updated state vectors which satisfy $Ab_{t|t} = q$. But, in fact, there is more: the terms of the sequence $\left(b_{t|t}\right)$ are the output from *on line* successive applications of restricted least squares. In order to establish this link, the restricted least square (RLS) estimator and its covariance matrix for a linear regression model $Y = X\beta + \varepsilon$, $\varepsilon \sim (0, \sigma^2 I)$, where β obeys (3-13), is recalled below:

$$\hat{\beta}_{RLS} = \hat{\beta}_{OLS} + (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(q - A\hat{\beta}_{OLS}) \tag{3-15}$$
$$Var\left(\hat{\beta}_{RLS}\right) = \sigma^2(X'X)^{-1} - (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}A(X'X)^{-1}.$$

The derivation of the expression in (3-15) is presented in almost any book on econometrics. See for instance Johnston and Dinardo (1997) or Greene (2003).

This section's result:

Theorem 3 *Under the state space model in (3-14), the Kalman state updating equation is identical to a recursive application of (3-15).*

Proof: The model in (3-14) can be decomposed in a way that recognizes q as a "new" measurement vector obtained/observed just "after" Y_t and right "before" Y_{t+1}. Thus, the measurement equation is recast as

$$Y_{t,j} = Z_{t,j}\beta_{t,j} + \varepsilon_{t,j}\ ,\ \varepsilon_{t,j} \sim (0, H_{t,j}). \tag{3-16}$$

Notice that $Y_{t,1} = Y_t$, $Z_{t,1} = x_t'$ and $H_{t,1} = \sigma^2$; on the other hand $Y_{t,2} = q$, $Z_{t,2} = A$ and $H_{t,2} = 0$. The new state equation is written in the same way as before.

It now becomes possible to regard the imposing of the linear restrictions as a new updating of the state vector. In fact, for an arbitrary t, denote the output of the Kalman updating using all the measurements from eq.(3-16) up to $Y_{t,1}$ by $\hat{\beta}_{t,1|t,1}$, which, as already discussed, equals the output of the recursive least squares - and consequently the OLS estimator - applied to the "observations"$\{Y_1, q_1, ..., q_k, ..., Y_{t-1}, q_1, ..., q_k, Y_t\}$. The state equation implies that $\hat{\beta}_{t,2|t,1} = \hat{\beta}_{t,1|t,1}$. Then, as $Y_{t,2} = q$ arrives, and by noticing that $P_{t,2|t,1} = P_{t,1|t,1} = \sigma^2\left(X_t'X_t\right)^{-1}$, $Z_{t,2} = A$ and $\upsilon_{t,2} = q - A\hat{\beta}_{t,2|t,1}$ and $F_{t,2} = A\sigma^2\left(X_t'X_t\right)^{-1}A'$, the Kalman state updating equation in (2-3) becomes

$$\hat{\beta}_{t,2|t,2} = \hat{\beta}_{t,1|t,1} + \left(X_t'X_t\right)^{-1}A'\left(A\left(X_t'X_t\right)^{-1}A'\right)^{-1}\left(q - A\hat{\beta}_{t,1|t,1}\right). \tag{3-17}$$

But, as just mentioned, $\hat{\beta}_{t,1|t,1} = \hat{\beta}_{MQO}$. Therefore, the conclusion is that the Kalman updating in (3-17) is indeed an application of RLS estimator of (3-15). The equivalence between covariance matrices can be also established analogously. $\square$

Some conceptual and practical consequences follow. First of all, it now becomes clear that the restricted Kalman filtering is indeed a generalization of the RLS estimator, a statement that, albeit intuitive, was lacking a proper formalization. In addition, Theorem 3 also shows that a regression model with random walk time-varying coefficients *under restrictions* (set $\beta_{t+1} = \beta_t + \eta_t$, $\eta_t \sim (0, Q)$, as the state equation for model (3-14)) does encapsulate the regression model with static coefficients, still under the same restrictions. Then, whenever the restricted Kalman filtering is applied to the time-varying version, both models can be compared as usual - to estimate the static model, just set $Q \equiv 0$. Finally, note that the recursive residuals obtained from recursive application of (3-15) are automatically uncorrelated - indeed, Theorem 3 says they are innovations. This is a desirable property in paving the way towards the development of a generalization of the stability test by Brown *et al.* (1975).

3.4 Initialization

3.4.1 Motivation

Besides considering the linear restrictions in equation (2-5), in this section I will also admit that some coordinates of the initial state vector α_1 have infinity variances. This is the basic set-up of the so-called *diffuse initialization* of the Kalman recursions, a subject extensively studied in Ansley and Kohn (1985), de Jong (1988), Harvey (1989), de Jong (1991), Koopman (1997), Durbin and Koopman (2001), Koopman and Durbin (2003), and de Jong and Chu-Chun-Lin (2003). Under this at least partially unspecified initial conditions, a question that comes is whether the methods of imposing linear restrictions can be derived from the very beginning. Observe that, once some elements of P_1 explode, there shall be no L_2 theory available anymore, nor could even the traditional Kalman equations be tackled, at least in the period when the effect of diffuseness - which lasts for an initial portion of the data - has not vanished yet. So, the strategies used in proofs by Doran (1992) and by Pizzinga (2008) to achieve the augmented restricted Kalman filtering (cf. Theorem 1) unfortunately become useless here. The purpose of this section is to address this theoretical issue precisely, by exploring the conditions which allow one to extend the restricted estimation to diffuse initializations, and by working out appropriately the modified versions of the Kalman equations; that is, the proof shall be "computational" instead of "geometrical".

3.4.2 Reviewing the initial exact Kalman smoother

From now on, the initial state vector is modeled as

$$\alpha_1 = a + B\delta + R_0\eta_0,$$

in which a is fixed and known, $\delta \sim (0, \kappa I_q)$, $\eta_0 \sim (0, Q_0)$, and B and R_0 are $m \times q$ and $m \times (m-q)$ selection matrices respectively, such that $B'R_0 = 0$ and $B'\alpha_1 = \delta$. In general, δ consists of initial conditions for the non-stationary terms of the m-variate process α_t. Under this fix, the exact initial Kalman smoother, obtained when $\kappa \longrightarrow +\infty$, is, in Durbin and Koopman (2001)'s notation,

$$\begin{aligned} &\upsilon_t^{(0)} = Y_t - Z_t a_t^{(0)} - d_t, \qquad F_{*,t} = Z_t P_{*,t} Z_t' + H_t, \qquad F_{\infty,t} = Z_t P_{\infty,t} Z_t' \\ &L_t^{(0)} = T_t - T_t P_{\infty,t} Z_t' F_{\infty,t}^{-1} Z_t, \quad L_t^{(1)} = -T_t P_{*,t} Z_t' F_{\infty,t}^{-1} Z_t + T_t P_{\infty,t} Z_t' F_{\infty,t}^{-1} F_{*,t} F_{\infty,t}^{-1} Z_t, \end{aligned} \tag{3-18}$$

$$a_{t+1}^{(0)} = T_t a_t^{(0)} + c_t + T_t P_{\infty,t} Z_t' F_{\infty,t}^{-1} \upsilon_t^{(0)}, \quad P_{*,t+1} = T_t P_{\infty,t} L_t^{(1)'} + T_t P_{*,t} L_t^{(0)'} + R_t Q_t R_t',$$
$$P_{\infty,t+1} = T_t P_{\infty,t} L_t^{(0)'}, \quad t = 1, \ldots, n,$$

$$\begin{aligned} &r_{t-1}^{(0)} = L_t^{(0)'} r_t^{(0)}, \quad r_{t-1}^{(1)} = Z_t F_{\infty,t}^{-1} \upsilon_t^{(0)} + L_t^{(0)'} r_t^{(1)} + L_t^{(1)'} r_t^{(0)}, \\ &a_{t|n} = a_t^{(0)} + P_{*,t} r_{t-1}^{(0)} + P_{\infty,t} r_{t-1}^{(1)}, \quad r_n^{(0)} = 0, \quad r_n^{(1)} = 0, \quad t = n, \ldots, 1, \end{aligned} \tag{3-19}$$

whenever $F_{\infty,t}$ just defined above is nonsingular. Otherwise, changes must take place in the recursions (3-18) and (3-19) (cf. Koopman and Durbin, 2003). According to Koopman (1997), there exists a time instant d after which the above recursions collapse to the traditional Kalman smoother; therefore, $P_{\infty,t} = 0$ for $t > d$ necessarily.

The presented recursions constitute the paradigm proposed in Koopman (1997), Durbin and Koopman (2001, sec. 5.3), and Koopman and Durbin (2003) for the treatment of state smoothing diffuse initialization. An alternative approach, based on the augmentation of the measurement equation, is proposed in de Jong and Chu-Chun-Lin (2003)

3.4.3
Combining exact initialization with linear restrictions

Before going to the main result of the paper, some preliminary steps must me addressed. The first is to list and to discuss the conditions under which it will be possible to combine diffuse initial conditions for the Kalman smoothing equations with the imposition of linear restrictions. Let them be enunciated and, without any loss of generality, consider them valid for all $t = 1, \ldots, n$.

Assumption 4 $\{q_{ti}, \ldots, q_{tk}\}$ *is a linearly independent subset of* $L_2(\Omega, \mathcal{F}, \mathcal{P})$.

Assumption 5 $\forall i = 1, \ldots, k : q_{ti} \notin$ *span*$\{1, Y_{11}, \ldots, Y_{1p}, q_{11}, \ldots, q_{1k}, \ldots, Y_{t-1,1}, \ldots, Y_{t-1,p}, q_{t-1,1}, \ldots, q_{t-1,k}, Y_{t,1}, \ldots, Y_{t,p}\}$.

Assumption 6 *The matrices Z_t and A_t are such that the rank of $[Z_t' \; A_t']'$ is $p+k$.*

Each Assumption should be examined in terms of generality and plausibility. Assumption 4 guarantees non-redundance of the linear restrictions. Assumption 5, in turn, definitively moves the focus towards the augmented restricted Kalman filtering, since this is actually the appropriate way of handling "stochastic" restrictions; indeed, if one faces some q_{ti} fixed or perfectly predictable given the past, the corresponding linear restrictions would be much better dealt with by the reduced restricted Kalman filtering (section 4.2), under which the exact initial Kalman smoother in (3-18) and (3-19) are straightforwardly applicable. Notice also that, under any state equation chosen, Assumptions 4 and 5 necessarily imply $A_t P_{t|t-1} A_t' > 0$. Finally, Assumption 6, besides reinforcing that the linear restrictions are distinct, should be understood as an impossibility of repeated signals along the lines of the augmented measurement equation proposed in Theorem 1, something quite natural from a practical perspective.

The second preliminary step is to quote the following auxiliary result:

Lemma 2 *Consider a state space model with the augmented measurement proposed in Theorem 1. If $F_t > 0$, then $A_t P_{t|t-1} \left[Z_t' \; A_t'\right] F_t^{-1} = [0_{k\times p} \; I_{k\times k}]$.*

Notice that Lemma 2 is actually a rephrasing of eq.(22) from Doran (1992), which has already proved to be key in the second proof of Theorem 1.

Finally, here is the main result concerning initialization:

Theorem 4 *(The initial exact restricted Kalman smoother) Suppose the augmented state space model associated to Theorem 1 satisfies Assumptions 4, 5 and 6. Then the initial exact Kalman smoother in (3-18) and (3-19) yields*

$$A_t a_{t|n} = q_t. \tag{3-20}$$

Proof: Fix an arbitrary $t \in \{1, \ldots, d\}$, where d is the length of the diffuse period associated with the augmented model. Define $\tilde{Z}_t = [Z_t' \; A_t']'$. Other quantities would also have deserved tildes, but they are suppressed for conserving notation. From the Assumptions 4 and 5, $F_{\infty,t}$ cannot be a zero matrix. Supposing first that $F_{\infty,t}$ is nonsingular, take the recursive formulae of $r_{t-1}^{(0)}$ and $r_{t-1}^{(1)}$ in (3-19) and place them in the expression of $a_{t|n}$, which gives

$$a_{t|n} = a_t^{(0)} + P_{\infty,t}\tilde{Z}_t' F_{\infty,t}^{-1} v_t^{(0)} + P_{\infty,t} L_t^{(0)\prime} r_t^{(1)} + P_{\infty,t} L_t^{(1)\prime} r_t^{(0)} + P_{*,t} L_t^{(0)\prime} r_t^{(0)}. \tag{3-21}$$

From (3-21), identity (3-20) will be proved whenever the following three claims are established.

Claim 1. $A_t\left(a_t^{(0)} + P_{\infty,t}\tilde{Z}_t' F_{\infty,t}^{-1} \upsilon_t^{(0)}\right) = q_t$.
Proof: Define $a_{t|t}^{(0)} \equiv a_t^{(0)} + P_{\infty,t}\tilde{Z}_t' F_{\infty,t}^{-1}\upsilon_t^{(0)}$. Looking at the recursion in (3-18), it follows that $a_{t|t}^{(0)}$ is the Kalman updating equation given in (2-3) applied to an augmented state space model with system matrices given by $Z_t^\dagger = \tilde{Z}_t$, $d_t^\dagger = (d_t'\ 0')'$, $H_t^\dagger = 0_{(p+k)\times(p+k)}$, $T_t^\dagger = T_t$, $c_t^\dagger = c_t$, and $Q_t^\dagger = 0$; and also $a_1^\dagger = 0$ and $P_1^\dagger = P_{\infty,1} = BB'$. $\square$

Claim 2: $A_t P_{\infty,t} L_t^{(0)'} = 0$.
Proof: Still considering the auxiliary state space model from the previous claim, use the expression of $L_t^{(0)'}$ in (3-18) and Lemma 2 to get

$$\begin{aligned} A_t P_{\infty,t} L_t^{(0)'} &= A_t P_{\infty,t} T_t' - A_t P_{\infty,t}\tilde{Z}_t' F_{\infty,t}^{-1}\tilde{Z}_t P_{\infty,t} T_t' \\ &= A_t P_{\infty,t} T_t' - \left[0_{k\times p}\ I_{k\times k}\right]\tilde{Z}_t P_{\infty,t} T_t' \\ &= 0. \ \square \end{aligned}$$

Claim 3: $A_t\left(P_{\infty,t} L_t^{(1)'} + P_{*,t} L_t^{(0)'}\right) = 0$.
Proof: From the expression of $L_t^{(1)'}$ in (3-18), it follows that

$$\begin{aligned} A_t P_{\infty,t} L_t^{(1)'} &= -A_t P_{\infty,t}\tilde{Z}_t' F_{\infty,t}^{-1}\tilde{Z}_t P_{*,t} T_t' + A_t P_{\infty,t}\tilde{Z}_t' F_{\infty,t}^{-1} F_{*,t} F_{\infty,t}^{-1}\tilde{Z}_t P_{\infty,t} T_t' \\ &= -\left[0_{k\times p}\ I_{k\times k}\right]\tilde{Z}_t P_{*,t} T_t' + \left[0_{k\times p}\ I_{k\times k}\right]\left[\tilde{Z}_t P_{*,t}\tilde{Z}_t' + diag\left(H_t, 0_{k\times k}\right)\right] F_{\infty,t}^{-1}\tilde{Z}_t P_{\infty,t} T_t' \\ &= -A_t P_{*,t} T_t' + A_t P_{*,t}\tilde{Z}_t' F_{\infty,t}^{-1}\tilde{Z}_t P_{\infty,t} T_t', \end{aligned} \tag{3-22}$$

where the second equality comes from Lemma 2 combined with the auxiliary state space model firstly evoked in Claim 1, and from the expression of $F_{*,t}$ in (3-18) associated with the the augmented model.
On the other hand, the expression of $L_t^{(0)}$ implies

$$A_t P_{*,t} L_t^{(0)'} = A_t P_{*,t} T_t' - A_t P_{*,t}\tilde{Z}_t' F_{\infty,t}^{-1}\tilde{Z}_t P_{\infty,t} T_t'. \tag{3-23}$$

Add (3-22) and (3-23). $\square$

In case of $F_{\infty,t}$ being singular, uncouple the augmented measurement $(Y_t', q_t')'$ in such a way that

$$Y_{t,1}, \ldots, Y_{t,p-1}, \left(Y_{t,p}, q_t'\right)'.$$

Without losing generality, assume that $Z_{t,p} P_{\infty,t} Z_{t,p}' > 0$, where $Z_{t,p}$ is the p^{th}-row of Z_t (recall: $P_{\infty,t} > 0$ for $t \leq d$). Since $F_{\infty,t}$ associated with $(Y_{t,p}, q_t')'$ is nonsingular (cf. Assumption 6), proceed exactly as before in order to attain (3-20). This completes the proof. $\square$

From a practical perspective, a point coming from this last result, which must be reinforced, is that, under quite general conditions, it is *always possible* to yield restricted smoothed state vectors, even when the estimation lies in the "diffuse" period (that is, for $t = 1, \ldots d$, whatever d may be). Said in other words: the beginning of the series is not critical anyhow to get more interpretable results (which is certainly the case whenever estimated state vectors under meaningful restrictions are achieved).