A LOW DIMENSIONAL KALMAN FILTER FOR SYSTEMS WITH LAGGED STATES IN THE MEASUREMENT EQUATION

KRISTOFFER P. NIMARK

Abstract. This note describes how the Kalman filter can be modified to allow for the vector of observables to be a function of lagged variables without increasing the dimension of the state vector in the filter. The modified filter, which nests the standard filter, can be used to compute (i) the steady state Kalman filter (ii) the log likelihood of a parameterized state space model conditional on a history of observables (iii) a smoothed estimate of latent state variables and (iv) a draw from the distribution of latent states conditional on a history of observables.
Keywords: Kalman filter, lagged observables, Kalman smoother, simulation smoother

This note describes how the Kalman filter can be modified to allow for the vector of observable variables in the measurement equation to be a function of lagged variables. The standard approach, which is to augment the state vector of the filter to include also lagged variables, works well in most applications. However, it also doubles the dimension of the state vector which is undesirable in some applications. The modified filter presented here avoids increasing the dimension of the state by exploiting that the innovation representation can be modified so as to make it unnecessary to augment the state vector with lagged variables.

The derivation of the modified filter, which nests the standard filter as a special case, is presented in the next section. This is followed by a brief description of how to use the modified filter together with standard algorithms for the Kalman Smoother and Kalman Simulation Smoother. The last section concludes and references existing work where the modified filter has proved to be useful.

1. A filtering problem

Consider a standard state space system augmented to allow the measurement equation to depend on lagged states

$$X_t = AX_{t-1} + Cu_t : u_t \sim N(0, I) \tag{1.1}$$

$$Z_t = D_1 X_t + D_2 X_{t-1} + Ru_t \tag{1.2}$$

where X_t is the $n \times 1$ dimensional state vector, A is an $n \times n$ matrix, C is and $n \times m$ matrix. Z_t is a $p \times 1$ vector of observable variables and D_1 and D_2 are both $p \times n$ matrices and R is a $p \times m$ matrix.

Date: October 6, 2014. The author thanks Francisco Barillas, Leonardo Melosi and Stefan Pitschner for useful comments. *Address:* Economics Department, Cornell University, Ithaca, 14850, New York. *e-mail*: pkn8@cornell.edu *webpage*: www.kris-nimark.net .

Define the notation

$$X_{t|t-s} \equiv E\left[X_t \mid Z^{t-s}, X_{0|0}\right] \tag{1.3}$$

$$P_{t|t-s} \equiv E\left[\left(X_t - X_{t|t-s}\right)\left(X_t - X_{t|t-s}\right)'\right] \tag{1.4}$$

where $X_{0|0}$ is the mean of the exogenously given prior distribution of X_0 given by

$$X_0 \sim N\left(X_{0|0}, P_{0|0}\right) \tag{1.5}$$

We want to find the Kalman gain K_t in the recursive updating

$$X_{t|t} = AX_{t-1|t-1} + K_t\left[Z_t - (D_1A + D_2)X_{t-1|t-1}\right] \tag{1.6}$$

so that $X_{t|t}$ is the conditional minimum variance estimate of X_t.

1.1. **The standard approach.** The state space system (1.1) - (1.2) is standard apart from the fact that the vector of observables Z_t depends on both the current and the lagged state. A straightforward and common way to get around this problem is to redefine the state so as to include also lagged X_t to get

$$\overline{X}_t = \overline{AX}_{t-1} + \overline{C}u_t \tag{1.7}$$

$$Z_t = \overline{DX}_t + Ru_t \tag{1.8}$$

where

$$\overline{X}_t = \begin{bmatrix} X'_t & X'_{t-1} \end{bmatrix}', \quad \overline{A} = \begin{bmatrix} A & \mathbf{0} \\ I & \mathbf{0} \end{bmatrix}$$

$$\overline{C} = \begin{bmatrix} C \\ \mathbf{0} \end{bmatrix}, \quad \overline{D} = \begin{bmatrix} D_1 & D_2 \end{bmatrix}$$

The standard filter can then be applied to the augmented system (1.7) - (1.8). In most application, this does not cause any complications. However, in some cases it is desirable to have a state of low dimensionality and redefining the state as above doubles the dimension of the state, i.e. $\overline{X}_t$ is a $2n \times 1$ vector. Below, a new filter is derived that solves the filtering problem while maintaining an n-dimensional state vector.

2. A modified filter

In this section, the modified filter is derived. The system is linear with Gaussian disturbances and the minimum variance estimate of the latent state then coincides with the orthogonal projection onto the set of conditioning variables (or signals). The filter is derived using the Gram-Schmidt approach of recursively orthogonalizing the time series of observable variables.[1] This approach exploits that the projection of a random variable onto a set of mutually orthogonal signals is equivalent to adding up the projections of the variable onto the individual signals. That is,

$$E(x \mid z, y) = E(x \mid z) + E(x \mid y) \tag{2.1}$$

if

$$E(zy') = 0 \tag{2.2}$$

[1]A derivation of the standard filter along similar lines can be found in Anderson and Moore (1979).

and x, y and z are zero-mean Gaussian random variables.

It is the property (2.1) - (2.2) that will allow us to write down a recursive update equation for $X_{t|t}$. The first step is to find the projection of X_t onto the component of the period t signals that is orthogonal to information known in period $t-1$. This projection can then be added to the prior estimate $X_{t|t-1}$, i.e. the projection of X_t onto period $t-1$ information, to form a posterior estimate $X_{t|t}$. To this end, define the innovation $\widetilde{Z}_t$ as the component of Z_t that is orthogonal to period $t-1$ information

$$\widetilde{Z}_t \equiv Z_t - Z_{t|t-1} \tag{2.3}$$

so that the posterior estimate $X_{t|t}$ will be given by

$$X_{t|t} = X_{t|t-1} + E\left(X_t \mid \widetilde{Z}_t\right). \tag{2.4}$$

To solve the filtering problem we thus need to find an expression for $E\left(X_t \mid \widetilde{Z}_t\right)$. We will start by solving this problem for period 1. The resulting expressions are then straightforward to generalize to period t.

2.1. **Projecting the state onto the innovation in the observable vector.** In the initial period there are two pieces of information available: the exogenously given prior distribution (1.5) and the initial signal Z_1. By (2.4) the prior and the signal can be combined as

$$X_{1|1} = X_{1|0} + K_1 \widetilde{Z}_1 \tag{2.5}$$

to form the posterior mean $X_{1|1}$ if $K_1 \widetilde{Z}_1 = E\left(X_1 \mid \widetilde{Z}_1\right)$. From the projection theorem (e.g. Brockwell and Davis 2006), the appropriate K_1 is given by the standard projection formula

$$K_1 = E\left(X_1 \widetilde{Z}_1'\right)\left[E\left(\widetilde{Z}_1 \widetilde{Z}_1'\right)\right]^{-1}. \tag{2.6}$$

To compute the Kalman gain K_1 we thus need to derive operational expressions for $E\left(X_1 \widetilde{Z}_1'\right)$ and $E\left(\widetilde{Z}_1 \widetilde{Z}_1'\right)$.

2.2. **The covariance of the state and the innovation vector.** To find the covariance $E\left(X_1 \widetilde{Z}_1'\right)$, start by using the identities implied by (1.1) - (1.2) to rewrite the innovation as

$$\widetilde{Z}_t = \left(D_1 A + D_2\right)\left(X_0 - X_{0|0}\right) + \left(D_1 C + R\right) u_1. \tag{2.7}$$

It is helpful to define the posterior state estimation error $\widetilde{X}_t$ as

$$\widetilde{X}_t \equiv X_t - X_{t|t} \tag{2.8}$$

and use this together with (2.7) to express the covariance of the state and the innovation as

$$\begin{aligned} E\left(X_1 \widetilde{Z}_1'\right) &= E\left[\left(A\left(\widetilde{X}_0 + X_{0|0}\right) + C u_1\right)\right. \\ &\quad \left.\times\left(\left(D_1 A + D_2\right)\widetilde{X}_0 + D_1 C u_1 + R u_1\right)'\right]. \end{aligned} \tag{2.9}$$

Since $E\left(X_{0|0}\widetilde{X}_0'\right) = 0$ and $P_{0|0} \equiv E\left(\widetilde{X}_0\widetilde{X}_0'\right)$ equation (2.9) can be simplified to

$$E\left(X_1\widetilde{Z}_1'\right) = AP_{0|0}\left(D_1A + D_2\right)' + CC'D_1' + CR'. \tag{2.10}$$

We thus have the first term in the Kalman gain (2.6).

2.3. **The covariance of the innovation vector.** To find the covariance of the innovation vector $\widetilde{Z}_1$, simply use that (2.7) implies that

$$\begin{aligned} E\left(\widetilde{Z}_1\widetilde{Z}_1'\right) &= \left(D_1A + D_2\right)P_{0|0}\left(D_1A + D_2\right)' \\ &\quad + \left(D_1C + R\right)\left(D_1C + R\right)' \end{aligned} \tag{2.11}$$

yielding the second term in the Kalman gain (2.6).

2.4. **The Kalman gain.** Plugging in (2.10) and (2.11) into the (2.12) then yields the Kalman gain for the first period

$$\begin{aligned} K_1 &= \left(AP_{0|0}\left(D_1A + D_2\right)' + CC'D_1' + CR'\right) \\ &\quad \times \left[\left(D_1A + D_2\right)P_{0|0}\left(D_1A + D_2\right)' + \left(D_1C + R\right)\left(D_1C + R\right)'\right]^{-1} \end{aligned} \tag{2.12}$$

2.5. **The posterior covariance.** To find the general expressions for the Kalman filter we can simply apply the same steps (2.7) - (2.12) with each variable and matrix replaced by their period t counterparts. However, the covariance $P_{0|0}$ of the initial period estimate X_0 was given exogenously. To find a general expression for the Kalman gain we thus first need to find an expression for the posterior covariance matrix $P_{1|1}$.

First, take the expression for the period 1 estimate of X_1

$$X_{1|1} = X_{1|0} + K_1\widetilde{Z}_1 \tag{2.13}$$

and add X_1 to each side. Use the definition (2.8) and rearrange to get

$$\widetilde{X}_1 + K_1\widetilde{Z}_1 = X_1 - X_{1|0}. \tag{2.14}$$

By optimality of the filter, the posterior error $\widetilde{X}_1$ is orthogonal to the innovation $\widetilde{Z}_1$. The variance of the left hand side of (2.14) is thus simply the sum of the covariance of $\widetilde{X}_1$ and the covariance of $K_1\widetilde{Z}_1$. Using (1.4) and (2.11) we thus have

$$P_{1|1} + K_1\left[\left(D_1A + D_2\right)P_{0|0}\left(D_1A + D_2\right)' + \left(D_1C + R\right)\left(D_1C + R\right)'\right]K_1' = P_{1|0}. \tag{2.15}$$

Rearranging gives an operational expression for the posterior covariance

$$P_{1|1} = P_{1|0} - K_1\left[\left(D_1A + D_2\right)P_{0|0}\left(D_1A + D_2\right)' + \left(D_1C + R\right)\left(D_1C + R\right)'\right]K_1' \tag{2.16}$$

as desired.

2.6. **The Kalman recursions.** The general expressions for the Kalman filter are then given by

$$\begin{aligned} K_t &= \left[AP_{t-1|t-1}\left(D_1A+D_2\right)'+CC'D_1'+CR'\right] && (2.17)\\ &\quad\times\left[\left(D_1A+D_2\right)P_{t-1|t-1}\left(D_1A+D_2\right)'+\left(D_1C+R\right)\left(D_1C+R\right)'\right]^{-1} \\ P_{t|t} &= P_{t|t-1} && (2.18)\\ &\quad-K_t\left[\left(D_1A+D_2\right)P_{t-1|t-1}\left(D_1A+D_2\right)'+\left(D_1C+R\right)\left(D_1C+R\right)'\right]K_t' \\ P_{t+1|t} &= AP_{t|t}A'+CC' && (2.19)\end{aligned}$$

where the last line used that

$$X_{t+1|t}-X_t=A\left(X_{t|t}-X_t\right)+Cu_t. \tag{2.20}$$

The steady state Kalman gain K_∞ can as usual be found by iterating on (2.17) - (2.19) until convergence.

3. Computing the log likelihood

As with the standard filter, the fact that the innovations $\widetilde{Z}_t$ are i.i.d. Gaussian vectors can be used to recursively evaluate the log likelihood $\mathcal{L}$ of the data conditional on a parameterized state space system with Gaussian disturbances. It is given by

$$\mathcal{L}\left(Z \mid A,C,D1,D2,R\right)=-\frac{1}{2}\sum_{t=1}^{T}\left(p\ln\pi+\ln\left|\Omega_t\right|+\widetilde{Z}_t'\Omega_t^{-1}\widetilde{Z}_t\right) \tag{3.1}$$

where Ω_t is the covariance of the innovation vector given by the period t equivalent of (2.11).

4. The Kalman smoother for the modified system

The smoothed estimate of X_t is defined as the linear minimum variance estimate of X_t conditional on the complete history of observables, i.e.

$$X_{t|T}\equiv E\left(X_t \mid Z^T, X_{0|0}\right) \tag{4.1}$$

As shown in Hamilton (1994), the smoothed estimate of X_t can be computed by using as input the time series of $X_{t|t}$ without the need to again condition on the observables. This makes it possible to derive the smoother without an explicit role for Z_t, once we have the recursions (2.17)-(2.19) and have computed the time series for $X_{t|t}$. The smoothed estimates of X_t are then given by

$$X_{t|T}=X_{t|t}+J_{t-1}\left(X_{t+1|T}-X_{t+1|t}\right) \tag{4.2}$$

where

$$J_t=P_{t|t}A'P_{t+1|t}^{-1} \tag{4.3}$$

The covariances of the smoothed state estimation errors can be computed as

$$P_{t|T}=P_{t|t}+J_t\left(P_{t+1|T}-P_{t+1|t}\right)J_t'.$$

A smoothed estimate of X_t can then be found using the following algorithm..

4.1. A Kalman smoother algorithm.

(1) Compute the sequence $X_{t|t} : t = 1, 2, .., T$ using the forward recursions (1.6) and (2.17) - (2.19). Store $X_{t|t}, P_{t|t}$ and $P_{t+1|t}$.
(2) Compute the smoothed estimates $X_{t|T} : t = T-1, T-2, ..., 1$ using the backward recursions (4.2) - (4.3).

5. The Kalman simulation smoother for the modified system

As described in Durbin and Koopman (2002), a draw from $p(X^T|Z^T)$ can be generated by the following algorithm

5.1. A Kalman simulation smoother algorithm.

(1) Construct a draw Z^{+T} from $p(Z^T)$ using the system (1.1) (1.2) and save the draw of the state X^{+T}.
(2) Construct $Z^{*T} = Z^T - Z^{+T}$.
(3) $\widetilde{X}^T = \widehat{X}^{*T} + X^{+T}$ is then a draw from $p(X^T|Z^T)$ where $\widehat{X}^{*T} = E\left(X|^T Z^{*T}\right)$ (i.e. $\widehat{X}^{*T}$ is the output of running Z^{*T} through the smoothing algorithm above).

This algorithm involves drawing only from the i.i.d. vectors of u_t rather than from conditional distributions of the state x_t (with the exception of generating the draw from the distribution of the initial state $p(X_0)$). The latter is often singular in interesting economic applications since the state dimension is often larger than the stochastic dimension in models with endogenous state variables. A singular covariance matrix requires additional computational steps which are avoided in Durbin and Koopman's algorithm.

6. Conclusions

Above it was demonstrated how the Kalman filter can be modified to allow for lagged observables without increasing the dimension of the state vector in the filter. While the standard approach of augmenting the state vector with lagged variables works well in many applications, it also introduces additional computational burdens that in some applications have significant costs. Examples of when the modified filter has been useful include Melosi (2014) and Nimark (2012) who study models with heterogeneously informed agents. The former paper estimates a structural business cycle model and the latter a model of the term structure of interest rates. These models naturally have high-dimensional state vectors and the standard approach of augmenting the state vector with lagged values is costly, both in terms of tractability and the computational burden of solving these models.

Researchers facing similar computational challenges can download Matlab code for the modified Kalman filter, smoother and simulation smoother from the author's web page *www.krisnimark.net* .

References

[1] Anderson, B.D.O., and J.B. Moore, 1979, *Optimal filtering*, Dover Publications, New York, (2005 edition).
[2] Brockwell, P.J. and R.A. Davis, 2006, *Time Series: Theory and Methods*, Springer-Verlag.
[3] Durbin, J. and S.J. Koopman, 2002,"A simple and efficient simulation smoother for state space time series analysis", Biometrica, 89, pp. 603-615.
[4] Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press.

[5] Melosi, Leonardo, 2014, "Signaling Effects of Monetary Policy", Working Paper, Federal Reserve Bank of Chicago.

[6] Nimark, K, 2012, "Speculative Dynamics in the Term Structure of Interest Rates", Working Paper, Cornell University.