Memo: **Notes on the Disturbance Smoother by Durbin and Koopman (2002)**
Date: August 30, 2013
From: Elmar Mertens
em@elmarmertens.com

Durbin and Koopman (2002) derive an efficient smoothing algorithm for the Kalman filter as well as a neat sampling scheme for drawing from the posterior density of the states, conditional on all available observations. The sampling scheme is described in their Section 2.1 and the smoothing algorithm, called "disturbance smoother" is derived in their Section 2.2. Notice, that this simulation smoothing algorithm is more efficient than the algorithms covered in their book (Durbin and Koopman, 2001).

In this note, I restate key elements of both, using a slightly different notation for the state space system, than in the original paper.

Setup

Let an observed state vector x_t evolve according to the following transition equation:

$$x_t = A_t x_{t-1} + B_t w_t \tag{1}$$

The state is (imperfectly) measured by the vector y_t

$$y_t = C_t x_t \tag{2}$$

where A_t, B_t and C_t are assumed to be known, deterministic matrices and $w_t \sim N(0, I)$ is a white noise vector. It is assumed that the history of measurements y^t does not span the history of the state vector x^t. For notational convenience I have omitted explicit measurement errors (they are folded into x_t).[1]

Projections of the state vector onto the the set of observables will be denoted $x_{t|t} = E(x_t|y^t)$ and "innovations" with respect to the observed information are defined by $\tilde{x}_t = x_t - x_{t|t-1}$ and $\tilde{y}_t = y_t - y_{t|t-1}$. The usual Kalman filtering formulas yield

$$x_{t|t} = x_{t|t-1} + K_{t-1}\tilde{y}_t \qquad K_t = \text{Cov}\left(\tilde{x}_t, \tilde{y}_t\right) \text{Var}\left(\tilde{y}_t\right)^{-1} = \Sigma_t C_t' \left(C_t \Sigma_t C_t'\right)^{-1} \tag{3}$$

where $\text{Var}\left(\tilde{x}_t\right) \equiv \Sigma_t$ and

$$\tilde{x}_t = A_t(I - K_{t-1}C_{t-1})\tilde{x}_{t-1} + B_t w_t \tag{4}$$

$$\Rightarrow \Sigma_t = B_t B_t' + A_t \Sigma_{t-1}^* A_t' \tag{5}$$

$$\Sigma_t^* = \text{Var}\left(\tilde{x}_t - \tilde{x}_{t|t}\right) = (I - K_t C_t)\Sigma_t = \Sigma_t(I - K_t C_t)' \tag{6}$$

In the present case, without measurment error, we also have $(I - K_t C_t)\Sigma_t = (I - K_t C_t)\Sigma_t(I - K_t C_t)'$ and the Riccati equation can further be simplified into

$$\Sigma_t = B_t B_t' + A_t \Sigma_{t-1} A_t' - A_t \Sigma_{t-1} C_{t-1}' \left(C_{t-1} \Sigma_{t-1} C_{t-1}'\right)^{-1} C_{t-1} \Sigma_{t-1} A_t'$$

However, as will be seen below, equations (5) and (6) are particularly convenient for extending the model to include measurment error.

Disturbance smoother

The standard textbook derivation for smoothed projections, $x_{t|T}$ relies on the Markov transition of the state vector and starts with $E(x_t|y^t, x_{t+1})$ which involves a projection of $x_t - x_{t|t}$ onto $\tilde{x}_{t+1}$ and thus inversion of Σ_{t+1} — see for example Hamilton (1994).[2] Henceforth, I will refer to this

[1] An alternative measurement equation would be $y_t = C_t x_t + e_t$ where $e_t \sim N(0, R_t)$. As will be seen below, it is straighforward to augment the exposition below to this case.

[2] Given $E(x_t|y^t, x_{t+1})$ — which is linear in x_{t+1} — the derivation then proceeds by condition down onto y^T and thus replacing x_{t+1} by $x_{t+1|T}$. Starting with $x_{T|T}$ the algorithm then iterates backwards through time.

standard approach also as the "state smoother". If the number of states is large this might be costly to compute, and linear dependence between some of the states (or their innovations) complicates the inversion further.[3]

Typically, the number of states (including measurement errors) is larger than the number of observable variables and by projecting directly on leads of the innovations in the observables, $\tilde{y}_{t+k}$, the disturbance smoother is considerably more efficient than the standard approach. By shunning the Markov structure of the problem, it might initially appear that the disturbance smoother forgoes the neat recursive structure of the state smoother. However, as will be seen shortly, the disturbance smoother has a recursive representation and can be computed by a simple iteration backwards through time, just as the "state smoother". At each step in the iteration, the computation needs to invert however a smaller variance covariance matrix, namely $\text{Var}\left(\tilde{y}_{t+j}\right)$ instead of $\text{Var}\left(\tilde{x}_{t+j}\right)$. In fact, since $\text{Var}\left(\tilde{y}_{t+j}\right)^{-1}\tilde{y}_{t+j}$ is already computed during the forward recursion of the Kalman filter, it is sufficient to store these "scaled" innovations, without need for further matrix inversions.

The disturbance smoother for $x_{t|T}$ is based on an orthogonal partitioning of the information set y^T into y^{t-1} and future innovations $\tilde{y}_{t+j}$ ($\forall\ 0 \leq j \leq T$):[4]

$$x_{t|T} = x_{t|t-1} + E(x_t|\tilde{y}_t) + E(x_t|\tilde{y}_{t+1}) + \ldots + E(x_t|\tilde{y}_T) \tag{7}$$

$$\text{with} \quad E(x_t|\tilde{y}_{t+j}) = \text{Cov}\left(\tilde{x}_t, \tilde{y}_{t+j}\right) \text{Var}\left(\tilde{y}_{t+j}\right)^{-1} \tilde{y}_{t+j} \tag{8}$$

$\text{Var}\left(\tilde{y}_{t+j}\right)$ has already been derived and computed for the forward recursion of the Kalman filter and the covariance term equals:

$$\text{Cov}\left(\tilde{x}_t, \tilde{y}_{t+j}\right) = \text{Cov}\left(\tilde{x}_t, \tilde{x}_{t+j}\right)C'_{t+j} = \Sigma_t \left(\prod_{k=1}^{j} \tilde{A}_{t+k}\right)' C'_{t+j} = \text{Cov}\left(\tilde{x}_t, \tilde{x}_{t+j-1}\right)\tilde{A}'_{t+j}C'_{t+j} \tag{9}$$

[3]This can easily be the case when the state vector contains lagged variables, for example when the transition equation (1) corresponds to the companion form of a higher-order VAR process, or when x_t corresponds to the state vector of a DSGE model, which contains lagged endogenous variables. Formally, when $|\Sigma_{t+1}| = 0$ one could always use a pseudo-inverse computed via the singular value decomposition, and when the nature of the singularity is obvious — like a block of zeros inherited from the design of the state space — a direct computation could be feasible. See also the discussion in Kim and Nelson (1999, Chapter 8).

[4]Of course, one could also partition into y^t and leading innovations. For the recursion described below, the partitioning described above is however slightly more convenient.

where $\tilde{A}_{t+j} \equiv A_{t+j}(I - K_{t+j-1}C_{t+j-1})$.

Closer inspection of (7) and (9) reveals that the disturbance smoother can also be expressed in a recursive form, in analogous manner to equations (3) and (5) of Durbin and Koopman (2002).

The recursion relies on an auxiliary variable, $s_{t,T}$ which tracks the information contained in the innovations $\tilde{y}_t, \tilde{y}_{t+1}, \ldots \tilde{y}_T$ for constructing $x_{t|T}$. $s_{t,T}$ follows a backwards recursion, starting with $s_{T,T} = C_T' \operatorname{Var}(\tilde{y}_T)^{-1} \tilde{y}_T$ and then iterating backwards ($t = T-1, T-2, \ldots$) over

$$s_{t|T} = A_t' \, s_{t+1|T} + C_t' \operatorname{Var}(\tilde{y}_t)^{-1} \, \tilde{y}_t \tag{10}$$

$$x_{t|T} = x_{t|t-1} + \Sigma_t \, s_{t|T} \tag{11}$$

Again, the only matrix to be inverted is $\operatorname{Var}(\tilde{y}_t)$ which is of full rank and typically of smaller size than Σ_t.[5] Additional efficiency is gained by constructing and storing "scaled" observables $\tilde{z}_t \equiv \operatorname{Var}(\tilde{y}_t)^{-1}\tilde{y}_t$ during the forward recursion of the Kalman filter, which can then be recycled during the backwards recursion without further need of any matrix inversions.

Measurement error The model can easily be augmented with measurment error in the observer equation, $y_t = C_t x_t + e_t$ with $e_T \sim N(0, R_t)$ such that $\operatorname{Var}(\tilde{y}_t) = C_t \Sigma_t C_t' + R_t$. Using the Kalman gain $K_t = \operatorname{Cov}(\tilde{x}_t, \tilde{y}_t) \operatorname{Var}(\tilde{y}_t)^{-1}$, the variance of state innovations Σ_t can then still be computed from (5) and (6).[6]

Sampling from the posterior

Instead of iteratively drawing the states, as in Carter and Kohn (1994), Durbin and Koopman (2002) use a more generic approach to sampling the states, which does not explicitly rely on the dynamic nature of the state space.

The goal is to sample a vector X, consisting of the stacked elements x_1, x_2 etc., from the posterior density $p(X|Y)$,where the vector Y consists of the stacked history y^t. Assuming conjugate

[5]Notice that a singular $\operatorname{Var}(\tilde{y}_t)$ is indicative of collinear measurements in the original setup of the problem.

[6]To see this, note that $\Sigma_t^* = (I - K_t C_t)\Sigma_t(I - K_t C_t)' + K_t R_t K_t' = \Sigma_t - \Sigma_t C_t' \operatorname{Var}(\tilde{y}_t)^{-1} C_t \Sigma_t$.

normal prior, this posterior is also a normal distribution, characterized by its mean $\hat{X} = E(X|Y)$ — which can be computed from the disturbance smoother as described above (or the conventional state smoother) – and its variance covariance matrix, which shall be denoted W.

Consider jointly simulating new vectors X^+ and Y^+ based on the prior distribution of the variables. Again, using the Kalman smoother, the conditional mean $\hat{X}^+ = E(X|Y)$ is straightforward to compute; and the posterior variance equals $E\left((X^+ - \hat{X}^+)^2\right) = W$.[7]

Draws from $p(X|Y)$, denoted $\tilde{X}$, can then simply be constructed from

$$\tilde{X} = X^+ - \hat{X}^+ + \hat{X} \tag{12}$$

It is straightforward to see that $\tilde{X}$ has the correct mean $\hat{X}$. Furthermore, since X^+ has been drawn independently from Y, it follows that

$$E\left((\tilde{X} - \hat{X})^2|Y\right) = E\left((X^+ - \hat{X}^+)^2|Y\right) = E\left((X^+ - \hat{X}^+)^2\right) = W \tag{13}$$

The draws $\tilde{X}$ are thus jointly normal with appropriate mean and variance equal, $\tilde{X}$ are thus draws from the posterior density $p(X|Y)$. Given efficient routines for simulating and smoothing the model, this sampling scheme is straightforward to implement.

References

C. K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*, volume 24 of *Oxford Statistical Science Series*. Oxford University Press, 2001.

J. Durbin and S. J. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–615, September 2002.

James D. Hamilton. *Time-Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

Chang-Jin Kim and Charles R. Nelson. *State-Space Models with Regime Switching*. The MIT Press, Cambridge, MA, 1999.

[7]Recall that the second moments of the Kalman filter are independent from the realizations of the observables.