Kalman and particle filtering

The Kalman and particle filters are algorithms that recursively update an estimate of the state and find the innovations driving a stochastic process given a sequence of observations. The Kalman filter accomplishes this goal by linear projections, while the particle filter does so by a sequential Monte Carlo method.

Since both filters start with a state-space representation of the stochastic processes of interest, Section 1 presents the state-space form of a dynamic model. Section 2 introduces the Kalman filter and Section 3 develops the particle filter. For extended expositions of this material, see Doucet, de Freitas, and Gordon (2001), Durbin and Koopman (2001), and Ljungqvist and Sargent (2004).

1. The state-space representation of a dynamic model

A large class of dynamic models can be represented by a state-space form:

$$X_{t+1} = \varphi(X_t, W_{t+1}; \gamma) \tag{1}$$

$$Y_t = g(X_t, V_t; \gamma). \tag{2}$$

This representation handles a stochastic process by finding three objects: a vector that describes the position of the system (a *state*, $X_t \in X \subset R^l$) and two functions, one mapping the state today into the state tomorrow (*the transition equation*, (1)) and one mapping the state into observables, Y_t (*the measurement equation*, (2)). An iterative application of the two functions on an initial state X_0 generates a fully specified stochastic process. The variables W_{t+1} and V_t are independent *i.i.d.* shocks. A realization of T periods of observables is denoted by $y^T \equiv \{y_t\}_{t=1}^T$ with $y^0 = \{\phi\}$. Finally, γ, which belongs to the set $\Upsilon \subset R^n$, is a vector of parameters. To avoid stochastic singularity, we assume that $\dim(W_t) + \dim(V_t) \geq \dim(Y_t)$ for all t.

This framework can accommodate cases in which the dimensionality of the shocks is zero, where the shocks have involved structures, or where some or all of the states are observed. Also, at the cost of heavier notation, we could deal with more general problems. For example, the state could be a function or a correspondence, and the transition equation a functional operator. The basic ideas are, however, identical.

The transition and measurement equations may come from a statistical description of the process or from the equilibrium dynamics of an economic model. For example, dynamic stochastic general equilibrium models can be easily written in state-space form with the transition and measurement equations formed by the policy functions that characterize the optimal behaviour of the agents of the model. This observation tightly links modern dynamic macroeconomics with the filtering tools presented in this article.

It is important to note that there are alternative timing conventions for the state-space representation of a dynamic model and that, even while the timing convention is

S. N. Durlauf et al. (eds.), *Macroeconometrics and Time Series Analysis*

kept constant, the same model can be written in different state-space forms. All of those representations are equivalent, and the researcher should select the form that best fits her needs.

2. The Kalman filter

The Kalman filter deals with state-space representations where the transition and measurement equations are linear and where the shocks to the system are Gaussian. The procedure was developed by Kalman (1960) to transform ('filter') some original observables y_t into Wold innovations a_t and estimates of the state x_t. With the innovations, we can build the likelihood function of the dynamic model. With the estimates of the states, we can forecast and smooth the stochastic process.

We begin with the state-space system defined by the transition equation:

$$x_{t+1} = Ax_t + G\omega_{t+1}, \omega_{t+1} \sim \mathcal{N}(0, Q)$$

and the measurement equation:

$$y_t = Cx_t + \upsilon_t, \upsilon_t \sim \mathcal{N}(0, R)$$

where A, G, C, Q, and R are known matrices.

There are different ways to derive and interpret the Kalman filter, including an explicitly Bayesian one. We follow a simple approach based on linear least-square projections. The reader will enhance her understanding with the more general expositions in Durbin and Koopman (2001) and Ljungqvist and Sargent (2004).

Let $x_{t|t-1} = E(x_t|y^{t-1})$ be the best linear predictor of x_t given the history of observables until $t-1$, i.e., y^{t-1}. Let $y_{t|t-1} = E(y_t|y^{t-1}) = Cx_{t|t-1}$ be the best linear predictor of y_t given y^{t-1}. Let $x_{t|t} = E(x_t|y^t)$ be the best linear predictor of x_t given the history of observables until t, i.e., y^t. Let $\Sigma_{t|t-1} \equiv E((x_t - x_{t|t-1})(x_t - x_{t|t-1})'|y^{t-1})$ be the predicting error variance-covariance matrix of x_t given y^{t-1}. Finally, let $\Sigma_{t|t} \equiv E((x_t - x_{t|t})(x_t - x_{t|t})'|y^t)$ be the predicting error variance-covariance matrix of x_t given y^t.

How does the Kalman filter work? Let's assume we have $x_{t|t-1}$ and $y_{t|t-1}$, that is, an estimate of the state and a forecast of the observable given y^{t-1}. Then, we observe y_t. Thus, we want to revise our linear predictor of the state and obtain an estimate, $x_{t|t}$, that incorporates the new information. Note that $x_{t+1|t} = Ax_{t|t}$ and $y_{t+1|t} = Cx_{t+1|t}$, so we can go back to the first step and wait for the y_{t+1} next period. Therefore, the key of the Kalman filter is to obtain $x_{t|t}$ from $x_{t|t-1}$ and y_t.

We do so with the formula:

$$x_{t|t} = x_{t|t-1} + K_t(y_t - y_{t|t-1}) = x_{t|t-1} + K_t(y_t - Cx_{t|t-1}),$$

that is, our new value $x_{t|t}$ is equal to $x_{t|t-1}$ plus the difference between the actual y_t and the forecasted $y_{t|t-1}$, times a matrix K_t, called the Kalman gain. Durbin and Koopman (2001) derive this formula from probabilistic foundations. Ljungqvist and Sargent (2004) find it through an application of a Gram–Schmidt orthogonalization procedure.

Then, if we choose K_t to minimize $\Sigma_{t|t}$, we get $K_t = \Sigma_{t|t-1}C'(C\Sigma_{t|t-1}C' + R)^{-1}$. This expression shows the determinants of K_t. If we made a big mistake forecasting $x_{t|t-1}$ using past information ($\Sigma_{t|t-1}$ large), we give a lot of weight to the new information (K_t large). Also, if the new information is noisy (R large), we give a lot of weight to the old prediction (K_t small).

Now, note that $\Sigma_{t|t} \equiv E((x_t - x_{t|t})(x_t - x_{t|t})'|y^t) = \Sigma_{t|t-1} - K_t C\Sigma_{t|t-1}$. Therefore, from $x_{t|t-1}$, $\Sigma_{t|t-1}$, and y_t, we compute $x_{t|t}$ and $\Sigma_{t|t}$ using K_t. Also, we derive $\Sigma_{t+1|t} = A\Sigma_{t|t}A' + GQG'$, $x_{t+1|t} = Ax_{t|t}$, and $y_{t+1|t} = Cx_{t+1|t}$.

We collect all the previous steps. We start with some estimates of the state $x_{t|t-1}$, the observables $y_{t|t-1}$, and the variance-covariance matrix $\Sigma_{t|t-1}$. Then we observe y_t and compute $x_{t+1|t}$, $y_{t+1|t}$, and $\Sigma_{t+1|t}$.

Thus, the Kalman filter can be recursively written as follows:

- $y_{t|t-1} = Cx_{t|t-1}$
- $K_t = \Sigma_{t|t-1}C'(C\Sigma_{t|t-1}C' + R)^{-1}$
- $\Sigma_{t|t} = \Sigma_{t|t-1} - K_t C\Sigma_{t|t-1}$
- $x_{t|t} = x_{t|t-1} + K_t(y_t - Cx_{t|t-1})$
- $\Sigma_{t+1|t} = A\Sigma_{t|t}A' + GQG'$
- $x_{t+1|t} = Ax_{t|t}$.

The differences between the observable and its forecast, $a_t = y_t - y_{t|t-1} = y_t - Cx_{t|t-1}$ are, by construction, Wold innovations. Moreover, since the system is linear and Gaussian, a_t is normally distributed with zero mean and variance $C\Sigma_{t|t-1}C' + R$. That is why the Kalman filter is a whitening filter: it takes as an input a correlated sequence y^T and it produces a sequence of white noise innovations a_t.

With this last result, we write the likelihood function of $y^T = \{y_t\}_{t=1}^T$ as:

$$\log L(y^T|A, G, C, Q, R) = \sum_{t=1}^{T} \log L(y_t|y^{t-1}A, G, C, Q, R)$$
$$= -\sum_{t=1}^{T}\left[\frac{N}{2}\log 2\pi + \frac{1}{2}\log|C\Sigma_{t|t-1}C' + R| + \frac{1}{2}\sum_{t=1}^{T} a_t'(C\Sigma_{t|t-1}C' + R)^{-1}a_t\right].$$

This likelihood is one of the most important results of the Kalman filter. With it, we can undertake statistical inference in the dynamic model, both with maximum likelihood and with Bayesian approaches.

An important step in the Kalman filter is to set the initial conditions $x_{1|0}$ and $\Sigma_{1|0}$. If we consider stationary stochastic processes, the standard approach is to set $x_{1|0} = x^*$ and $\Sigma_{1|0} = \Sigma^*$ such that $x^* = Ax^*$ and

$$\Sigma^* = A\Sigma^*A' + GQG' = [I - A \otimes A]^{-1}vec(GQG').$$

Non-stationary time series require non-informative prior conditions for $x_{1|0}$. This approach, called the diffuse initialization of the filter, begins by postulating that $x_{1|0}$ is

equal to:

$$x_{1|0} = \tau + \Phi\delta + G\omega_0, \omega_0 \sim \mathcal{N}(0, Q) \text{ and } \delta \sim \mathcal{N}(0, \kappa I_q)$$

where τ is given and Φ and G are formed by columns of the identity matrix such that $\Phi G' = 0$. This structure allows for some elements of $x_{1|0}$ to have a known joint distribution, while, by letting $\kappa \to \infty$, to formalize ignorance with respect to other elements. Clearly, $x_{1|0} = E(x_{1|0}) = \tau$. To determine the initial variance, we expand $\Sigma_{1|0} = \kappa\Phi\Phi' + GQG'$ as a power series of κ^{-1} and take $\kappa \to \infty$ to find the dominant term of the expansion. Durbin and Koopman (2001) provide details.

The Kalman filter can also be applied for smoothing, that is, to obtain $x_{t|T}$, an estimate of x_t given the whole history of observables, that is, y^T. Smoothing is of interest when the state x_t has a structural interpretation of its own. Since smoothing uses more information than filtering, the predicting error variance covariance matrix of x_t given y^T will be smaller than $\Sigma_{t|t-1}$. Finally, we note that the Kalman filtering problem is the dual of the optimal linear regulator problem.

3. The particle filter

The Kalman filter relies on the linearity and normality assumptions. However, many models in which economists are interested are nonlinear and/or non-Gaussian. How can we undertake the forecast, smoothing, and estimation of dynamic models when any of those two assumptions are relaxed?

Sequential Monte Carlo methods, in particular the particle filter, reproduce the work of the Kalman filter in those nonlinear and/or non-Gaussian environments. The key difference is that, instead of deriving analytic equations as the Kalman filter does, the particle filter uses simulation methods to generate estimates of the state and the innovations. If we apply the particle filter to a linear and Gaussian model, we will obtain the same likelihood (as the number of simulations grows) that we would if we used the Kalman filter. Since it avoids simulations, the Kalman filter is more efficient in this linear and Gaussian case.

We present here only the basic particle filter. Doucet, de Freitas and Gordon (2001) discuss improvements upon the basic filter. Fernández-Villaverde and Rubio-Ramírez (2007) show how this particle filter can be implemented to estimate dynamic stochastic general equilibrium models. Our goal is to evaluate the likelihood function of a sequence of realizations of the observable y^T implied by a stochastic process at a parameter value γ:

$$L(y^T; \gamma) = p(y^T; \gamma). \tag{3}$$

Our first step is to factor the likelihood function as:

$$\begin{aligned} p(y^T; \gamma) &= \prod_{t=1}^{T} p(y_t | y^{t-1}; \gamma) \\ &= \prod_{t=1}^{T} \iint p(y_t | W^t, X_0, y^{t-1}; \gamma) \times p(W^t, X_0 | y^{t-1}; \gamma) dW^t dX_0, \end{aligned} \tag{4}$$

where X_0 is the initial state of the model and the p's represent the relevant densities. In general, the likelihood function (4) cannot be computed analytically. The particle filter uses simulation methods to estimate it.

Before introducing the filter, we assume that, for all γ, x_0, w^t, and t, the following system of equations:

$$\begin{aligned} X_1 &= \varphi(x_0, w_1; \gamma) \\ y_m &= g(X_m, V_m; \gamma) \qquad \text{for } m = 1, 2, \ldots t \\ X_m &= \varphi(X_{m-1}, w_m; \gamma) \quad \text{for } m = 2, 3, \ldots t \end{aligned}$$

has a unique solution, (v^t, x^t), and we can evaluate $p(v^t; \gamma)$. This assumption implies that we can evaluate the conditional densities $p(y_t|w^t, x_0, y^{t-1}; \gamma)$ for all γ, x_0, w^t, and t. Then, we have:

$$p(y_t|w^t, x_0, y^{t-1}; \gamma) = |dy(v_t; \gamma)|p(v_t; \gamma)$$

for all γ, x_0, w^t, and t, where $|dy(v_t; \gamma)|$ stands for the determinant of the Jacobian of y_t with respect to V_t evaluated at v_t.

Conditional on having N draws of $\{\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N\}_{t=1}^T$ from the sequence of densities $\{p(W^t, X_0|y^{t-1}; \gamma)\}_{t=1}^T$, the law of large numbers implies that the likelihood function ((4)) can be approximated by:

$$p(y^T; \gamma) \simeq \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(y_t|w^{t|t-1,i}, x_0^{t|t-1,i}, y^{t-1}; \gamma).$$

This observation shows that the problem of evaluating the likelihood ((4)) is equivalent to the problem of drawing from $\{p(W^t, X_0|y^{t-1}; \gamma)\}_{t=1}^T$. Since the algorithm does not require any assumption about the distribution of the shocks except the ability to evaluate $p(y_t|w^t, x_0, y^{t-1}; \gamma)$, either analytically or by simulation, we can deal with models with a rich specification of non-Gaussian innovations. But, how do we sample from $\{p(W^t, X_0|y^{t-1}; \gamma)\}_{t=1}^T$?

Let $\{x_0^{t-1,i}, w^{t-1,i}\}_{i=1}^N$ be a sequence of N *i.i.d.* draws from $p(W^{t-1}, X_0|y^{t-1}; \gamma)$. Let $\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N$ be a sequence of N *i.i.d.* draws from $p(W^t, X_0|y^{t-1}; \gamma)$. We call each draw $(x_0^{t,i}, w^{t,i})$ a *particle* and the sequence $\{x_0^{t,i}, w^{t,i}\}_{i=1}^N$ a *swarm of particles*. Also, define the weights:

$$q_t^i = \frac{p(y_t|w^{t|t-1,i}, x_0^{t|t-1,i}, y^{t-1}; \gamma)}{\sum_{i=1}^N p(y_t|w^{t|t-1,i}, x_0^{t|t-1,i}, y^{t-1}; \gamma)}. \tag{5}$$

The next proposition shows how to use $p(W^t, X_0|y^{t-1}; \gamma)$, the weights $\{q_t^i\}_{i=1}^N$, and importance sampling to draw from $p(W^t, X_0|y^t; \gamma)$:

Proposition 1 *Let* $\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N$ *be a draw from* $p(W^t, X_0|y^{t-1}; \gamma)$. *Let the sequence* $\{\tilde{x}_0^i, \tilde{w}^i\}_{i=1}^N$ *be a draw with replacement from* $\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N$ *where* q_t^i *is the*

probability of $(x_0^{t|t-1,i}, w^{t|t-1,i})$ *being drawn* $\forall i$. *Then* $\{\tilde{x}_0^i, \tilde{w}^i\}_{i=1}^N$ *is a draw from* $p(W^t, X_0|y^t;\gamma)$.

Then, with a draw $\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N$ from $p(W^t, X_0|y^{t-1};\gamma)$, we get a draw $\{x_0^{t,i}, w^{t,i}\}_{i=1}^N$ from $p(W^t, X_0|y^t;\gamma)$ and we generate a sequence of particles $\{\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N\}_{t=1}^T$ from the sequence $\{p(W^t, X_0|y^{t-1};\gamma)\}_{t=1}^T$. Given some initial conditions, we can recursively apply the idea of the previous proposition as summarized by the algorithm:

Step 0, Initialization: Set $t \rightsquigarrow 1$. Initialize $p(W^{t-1}, X_0|y^{t-1};\gamma) = p(X_0;\gamma)$.
Step 1, Prediction: Sample N values $\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N$ from the conditional density $p(W_1^t, X_0|y^{t-1};\gamma) = p(W_t;\gamma)p(W^{t-1}, X_0|y^{t-1};\gamma)$.
Step 2, Filtering: Assign to each draw $(x_0^{t|t-1,i}, w^{t|t-1,i})$ the weight q_t^i as defined in (5).
Step 3, Sampling: Sample N times with replacement from $\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N$ with probabilities $\{q_t^i\}_{i=1}^N$. Call each draw $(x_0^{t,i}, w^{t,i})$. If $t < T$ set $t \rightsquigarrow t+1$ and go to Step 1. Otherwise stop.

With the algorithm's output $\{\{x_0^{t|t-1,i}, w_1^{t|t-1,i}\}_{i=1}^N\}_{t=1}^T$, we obtain the estimate of the states in each period and compute the likelihood:

$$p(y^T;\gamma) \simeq \frac{1}{N}\left(\prod_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} p(y_t|w^{t|t-1,i}, x_0^{t|t-1,i}, y^{t-1};\gamma)\right).$$

The **sampling step** is the heart of the algorithm. If we skip it and weight each draw in $\{x_0^{t|t-1,i}, w^{t|t-1,i}\}_{i=1}^N$ by $\{Nq_t^i\}_{i=1}^N$, we have a sequential importance sampling. The problem with this approach is that it diverges as t grows. The reason is that, as $t \to \infty$, all the sequences become arbitrarily far away from the true sequence of states (the true sequence being a zero measure set), and the sequence that happens to be closer dominates all the remaining sequences in weight. In practice, after a few steps only one sequence has a non-zero weight. Through resampling, we eliminate this problem as we keep (and multiply) those sequences that do not diverge from the true one.

The algorithm outlined above is not the only procedure to evaluate the likelihood of nonlinear and/or non-Gaussian dynamic models. However, the alternatives, such as the extended Kalman filter, the Gaussian sum approximations, or grid-based filters, are of limited use, and many, such as the extended Kalman filter, fail asymptotically. Consequently, the particle filter is the most efficient and robust procedure to undertake inference for nonlinear and/or non-Gaussian models, and we will witness many applications of this filter in economics in future years.

JESÚS FERNÁNDEZ-VILLAVERDE

See also **Bayesian methods in macroeconometrics; state space models.**

Bibliography

Doucet, A., de Freitas, N. and Gordon, N. 2001. *Sequential Monte Carlo Methods in Practice*. New York: Springer Verlag.

Durbin, J. and Koopman, S.J. 2001. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Fernández-Villaverde, J. and Rubio-Ramírez, J.F. 2007. Estimating macroeconomic models: a likelihood approach. *Review of Economic Studies* 74, 1059–87.

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering* D82, 35–45.

Ljungqvist, L. and Sargent, T.J. 2004. *Recursive Macroeconomic Theory*, 2nd edn. Cambridge, MA: MIT Press.