Learning and filtering
Notes

Jaroslav Borovička[†]

November 23, 2017

Contents

1 Introduction

I will consider two types of filtering in continuous time. The first filter goes back to Wonham (1964) and concerns the filtering of an unobserved state state driven by a finite-dimensional Markov chain. The second filter is the continuous-time version of the Kalman filter, introduced by Kalman and Bucy (1961).

The reason why we are studying these cases is that we want to understand the implications for pricing of risk by agents who face these filtering problems.

2 Wonham filter

Let z be an n-state Markov chain in continuous time, encoded by a coordinate vector. A realized state z_t is a zero-one coordinate vector in $\mathbb{R}^n$. The transition matrix over an interval

[†]E-mail: jaroslav.borovicka@nyu.edu. All errors are mine.

ε is $\exp(\varepsilon A)$ where A is the intensity matrix and $\exp(\cdot)$ is the matrix exponential given by the Taylor series expansion. We have

$$\begin{aligned} \exp(\varepsilon A)\,\mathbf{1}_n &= \mathbf{1}_n \qquad \forall\varepsilon \\ A\mathbf{1}_n &= 0 \end{aligned}$$

Also, notice that for $i \neq j$ we have for two coordinate vectors u_j and u_i

$$(u_j)' \exp(\varepsilon A)\, u_i \geq 0 \tag{1}$$

for each $\varepsilon > 0$ and

$$\lim_{\varepsilon \to 0} (u_j)' \exp(\varepsilon A)\, u_i = 0$$

so that it better be that the derivative w.r.t. ε of (1) at $\varepsilon = 0$ is nonnegative

$$(u_j)' A u_i \geq 0$$

so that the matrix A has nonnegative off-diagonal elements. The condition $A\mathbf{1}_n = 0$ then implies that the diagonal elements are necessarily nonpositive.

Suppose that a scalar observable signal y_t evolves according to

$$dy_t = \kappa z_t dt + \sigma dW_t$$

where z_t is a coordinate vector describing an unobservable state of the economy, represented by an n-state Markov chain with transition probability matrix $\exp(\tau A)$ over an interval τ. W_t is a (multivariate) Brownian motion and κ and σ are constant row vectors. dy_t capture the evolution of the signal, but the agent does not observe z_t directly, he has to solve a filtering problem.

The joint distribution of the Brownian motion W and the path of the Markov chain z define a probability space $(\Omega, \mathcal{F}, P)$. We will distinguish two information structures.

1. $\mathcal{F}_t$ is the full information filtration, generated by the signal history (W_s, z_s), $0 \leq s \leq t$, or, equivalently, by (y_s, z_s), $0 \leq s \leq t$.
2. $\mathcal{H}_t$ is the information set generated by observations of $\{y_s : 0 \leq s \leq t\}$ only, and an imperfect signal about state z_0 which implies a distribution of $z_0 \mid \mathcal{H}_0$ given by a probability vector $\bar{z}_0 = E[z_0 \mid \mathcal{H}_0]$. The initial distribution $\bar{z}_0$ can be viewed as a prior about the true initial state of the system. Naturally, $\bar{z}_0 \in \mathbb{R}^n_+$ with $\bar{z}_0' \mathbf{1}_n = 1$.

Our task is to produce a our best forecast of the state z_t at time t, given the information from the signal:

$$\bar{z}_t = E[z_t \mid \mathcal{H}_t].$$

We provide a heuristic derivation of the filter.

2.1 A local regression

Let ε be a small time interval. Then the approximate evolution over the interval ε is given by

$$y_{t+\varepsilon} - y_t \approx \kappa z_t \varepsilon + \sigma \left[W_{t+\varepsilon} - W_t \right] \tag{2}$$

Notice that the variance over time interval ε is given by

$$\begin{aligned} E\left[\left(y_{t+\varepsilon} - y_t \right)^2 \mid z_t \right] &\approx \varepsilon^2 \left(\kappa z_t \right) + \varepsilon \left| \sigma \right|^2 \\ \lim_{\varepsilon \to 0} \frac{E\left[\left(y_{t+\varepsilon} - y_t \right)^2 \mid z_t \right]}{\varepsilon} &= \left| \sigma \right|^2 \end{aligned}$$

Observe that in this calculation, we did not have to remove the conditional trend, and that in fact we also have

$$\lim_{\varepsilon \to 0} \frac{E\left[\left(y_{t+\varepsilon} - y_t - \kappa z_t \varepsilon \right)^2 \mid z_t \right]}{\varepsilon} = \left| \sigma \right|^2$$

Further, we obtain the same result even if we subtract the conditional trend calculated under the information set generated by the observed signal

$$\lim_{\varepsilon \to 0} \frac{E\left[\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \right)^2 \mid z_t \right]}{\varepsilon} = \left| \sigma \right|^2 \tag{3}$$

So the amount of risk exposure over the infinitesimal time period seems to be the same regardless of which information set we use. The local variation is governed purely by the stochastic term in the diffusion, while the trend is locally smooth.

We want to produce an evolution equation for $\bar{z}_t$. Observe that

$$E\left[z_{t+\varepsilon} \mid z_t \right] = \exp\left(\varepsilon A \right)' z_t \approx z_t + \varepsilon A' z_t \tag{4}$$

where we use a first-order Taylor series approximation of the transition matrix $\exp(\varepsilon A)$. Thus $A' z_t$ represents the 'local trend' of the underlying state. Using iterated expectations, we further get

$$\begin{aligned} E\left[\bar{z}_{t+\varepsilon} \mid \mathcal{H}_t \right] &= E\left[E\left[z_{t+\varepsilon} \mid \mathcal{H}_{t+\varepsilon} \right] \mid \mathcal{H}_t \right] = E\left[z_{t+\varepsilon} \mid \mathcal{H}_t \right] = \\ &= E\left[E\left[z_{t+\varepsilon} \mid \mathcal{F}_t \right] \mid \mathcal{H}_t \right] = E\left[\exp\left(\varepsilon A \right)' z_t \mid \mathcal{H}_t \right] = \\ &= \exp\left(\varepsilon A \right)' \bar{z}_t \approx \bar{z}_t + \varepsilon A' \bar{z}_t \end{aligned}$$

Thus $\bar{z}_t$ is a sufficient statistics for $z_{t+\varepsilon}$ given $\mathcal{H}_t$. In the limit we get

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(E\left[z_{t+\varepsilon} \mid \mathcal{H}_t \right] - \bar{z}_t \right) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(E\left[\bar{z}_{t+\varepsilon} \mid \mathcal{H}_t \right] - \bar{z}_t \right) = A' \bar{z}_t$$

where $A' \bar{z}_t$ is the 'local trend' of the forecast of the underlying state, and we can write the evolution under the filtration $\mathcal{H}_t$ as

$$d\bar{z}_t = A' \bar{z}_t dt + \ldots dW_t$$

where $\ldots$ is a volatility matrix that needs to be determined.

Observe that we can decompose the evolution of the signal as

$$y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \approx \varepsilon \kappa \left(z_t - \bar{z}_t \right) + \sigma \left(W_{t+\varepsilon} - W_t \right) \tag{5}$$

where the left-hand side represents the newly arrived information through the signal at time $t+\varepsilon$, i.e. the difference between the observed signal $y_{t+\varepsilon}$ and theforecasted signal $y_t + \kappa \bar{z}_t \varepsilon$. The 'suprise' movement in the signal thus can be understood as being composed of the error in the forecasted trend $\varepsilon \kappa \left(z_t - \bar{z}_t \right)$ and the random evolution $\sigma \left(W_{t+\varepsilon} - W_t \right)$.

Imagine the following least squares regression of $z_{t+\varepsilon} - \left(\bar{z}_t + \varepsilon A' \bar{z}_t \right)$ on $y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon$. This regression forecasts the change in the state z_t relative to the predicted value

$$z_{t+\varepsilon} - \left(\bar{z}_t + \varepsilon A' \bar{z}_t \right)$$

using new information from the signal,

$$y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon$$

under the information set $\mathcal{H}_t$. This is just an information decomposition. Let us denote the corresponding regression coefficient $\Delta \left(\bar{z}_t \right)$ (which is an $n \times 1$ vector)

$$z_{t+\varepsilon} - \left(\bar{z}_t + \varepsilon A' \bar{z}_t \right) = \Delta \left(\bar{z}_t \right) \left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \right) + \eta_{t+\varepsilon}$$

which, after post-multiplying by $\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \right)$ and taking expectations conditional on $\mathcal{H}_t$, implies

$$E \left[z_{t+\varepsilon} \left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \right) \mid \mathcal{H}_t \right] = \Delta \left(\bar{z}_t \right) E \left[\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \right)^2 \mid \mathcal{H}_t \right]$$

because $E \left[\left(\bar{z}_t + \varepsilon A' \bar{z}_t \right) \left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \right) \mid \mathcal{H}_t \right] = 0$. The right-hand side expression is calculated in (3), and thus we get

$$\begin{aligned}
\Delta \left(\bar{z}_t \right) &= \left(|\sigma|^2 \varepsilon \right)^{-1} E \left[z_{t+\varepsilon} \left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \right) \mid \mathcal{H}_t \right] = \\
&= \left(|\sigma|^2 \varepsilon \right)^{-1} E \left[z_{t+\varepsilon} \left(\varepsilon \kappa \left(z_t - \bar{z}_t \right) + \sigma \left(W_{t+\varepsilon} - W_t \right) \right) \mid \mathcal{H}_t \right] = \\
&= \left(|\sigma|^2 \right)^{-1} E \left[z_{t+\varepsilon} \left(\left(z_t - \bar{z}_t \right)' \right) \mid \mathcal{H}_t \right] \kappa' = \\
&= \left(|\sigma|^2 \right)^{-1} E \left[z_{t+\varepsilon} z_t' - z_{t+\varepsilon} \bar{z}_t' \mid \mathcal{H}_t \right] \kappa'
\end{aligned}$$

Observe that equation (4) implies

$$\begin{aligned}
E \left[z_{t+\varepsilon} z_t' \mid \mathcal{H}_t \right] &= E \left[E \left[z_{t+\varepsilon} \mid z_t \right] z_t' \mid \mathcal{H}_t \right] \approx \\
&\approx E \left[\left(z_t + \varepsilon A' z_t \right) z_t' \mid \mathcal{H}_t \right] = \\
&= E \left[\left(I + \varepsilon A' \right) z_t z_t' \mid \mathcal{H}_t \right] = \\
&= \left(I + \varepsilon A' \right) diag \left(\bar{z}_t \right)
\end{aligned}$$

and further

$$\begin{aligned}
E \left[z_{t+\varepsilon} \bar{z}_t' \mid \mathcal{H}_t \right] &= E \left[z_{t+\varepsilon} \mid \mathcal{H}_t \right] \bar{z}_t' \approx \\
&\approx E \left[z_t + \varepsilon A' z_t \mid \mathcal{H}_t \right] \bar{z}_t' = \\
&= \left(I + \varepsilon A' \right) \bar{z}_t \bar{z}_t'
\end{aligned}$$

Thus, eliminating terms of order ε^2, we get

$$\begin{aligned}\Delta(\bar{z}_t) &= \left(|\sigma|^2\right)^{-1} E\left[z_{t+\varepsilon}z_t' - z_{t+\varepsilon}\bar{z}_t' \mid \mathcal{H}_t\right]\kappa' \approx \\ &\approx \left(|\sigma|^2\right)^{-1}\left(diag(\bar{z}_t) - \bar{z}_t\bar{z}_t'\right)\kappa'\end{aligned}$$

Notice that equation (2) implies that under the $\mathcal{H}_{t+\varepsilon}$ information set, we can write

$$y_{t+\varepsilon} - y_t \approx \kappa\bar{z}_t\varepsilon + \sigma\left[\bar{W}_{t+\varepsilon} - \bar{W}_t\right]$$

where $\bar{W}_{t+\varepsilon} - \bar{W}_t$ is orthogonal to $\mathcal{H}_t$. Thus $y_{t+\varepsilon} - y_t - \kappa\bar{z}_t\varepsilon$ is orthogonal to $\mathcal{H}_t$, and can be indeed interpreted as new information for the forecasting of $\bar{z}_{t+\varepsilon}$ not contained in $\bar{z}_t$. Thus we can write the evolution of $\bar{z}_t$ as consisting of two components

$$\begin{aligned}\bar{z}_{t+\varepsilon} &= E\left[\bar{z}_{t+\varepsilon} \mid \mathcal{H}_t\right] + \Delta(\bar{z}_t)\left(y_{t+\varepsilon} - y_t - \kappa\bar{z}_t\varepsilon\right) \approx \\ &\approx \bar{z}_t + \varepsilon A'\bar{z}_t + \Delta(\bar{z}_t)\left(y_{t+\varepsilon} - y_t - \kappa\bar{z}_t\varepsilon\right)\end{aligned}$$

Sending $\varepsilon \to dt$ finally gives us

$$d\bar{z}_t = A'\bar{z}_t dt + \Delta(\bar{z}_t)\left(dy_t - \kappa\bar{z}_t dt\right)$$

where the first term on the right-hand side represent the contribution of the information about the evolution contained in $\mathcal{H}_t$ (local time trend), and the second term represents the impact of newly arrived information in $\mathcal{H}_{t+dt}$ (locally stochastic term).

Equation (5) gives us

$$dy_t - \kappa\bar{z}_t dt = \kappa\left(z_t - \bar{z}_t\right)dt + \sigma dW_t$$

2.2 Innovations representation

There also exists an innovations representation under $\mathcal{H}_t$. We have argued above that $dy_t - \kappa\bar{z}_t dt$ is unpredictable under $\mathcal{H}_t$ information and has variance $|\sigma|^2 dt$, and thus

$$\frac{dy_t - \kappa\bar{z}_t dt}{|\sigma|} = \frac{\kappa\left(z_t - \bar{z}_t\right)}{|\sigma|}dt + \frac{\sigma}{|\sigma|}dW_t = d\bar{W}_t$$

where $\bar{W}$ is a standard univariate Brownian motion adapted to $\{\mathcal{H}_t : t \geq 0\}$. The reason is that, as we have also argued above, the first term on the right-hand side is locally smooth (its variance is of order $(dt)^2$). We can easily verify that

$$\begin{aligned}E\left[\left(\bar{W}_t - \bar{W}_{t-u}\right)\bar{W}_{t-u} \mid \mathcal{H}_{t-u}\right] &= 0 \\ E\left[\bar{W}_t^2 \mid \mathcal{H}_0\right] &= t\end{aligned}$$

so that indeed $\bar{W}$ is a standard univariate Brownian motion adapted to $\{\mathcal{H}_t : t \geq 0\}$.

Observe that **our partial information model** was

$$\begin{aligned}dy_t &= \kappa z_t dt + \sigma dW_t \qquad (6) \\ z_t &\ldots \text{ Markov state with intensity matrix } A\end{aligned}$$

where z_t was unobserved (thus partial information model). We transformed this model into a **full information model**

$$\begin{aligned} dy_t &= \kappa \bar{z}_t dt + |\sigma| \, d\bar{W}_t \\ d\bar{z}_t &= A' \bar{z}_t dt + |\sigma| \, \Delta(\bar{z}_t) \, d\bar{W}_t \end{aligned} \tag{7}$$

where we have full information about the evolution of $(y_t, \bar{z}_t)$. In doing this, we transformed the multivariate Brownian motion W_t into $\bar{W}_t$ which is a univariate Brownian motion under $\{\mathcal{H}_t : t \geq 0\}$.

Looking at evolution (7), we only need to add an initial distribution for $\bar{z}_0$. This amounts to specifying a prior distribution. Then, from the observations of dy_t and the knowledge of $\bar{z}_t$, I can perfectly infer $d\bar{W}_t$ and thus also $d\bar{z}_t$.

From the point of view of asset pricing models, we want to think about whether we price shocks to $\bar{W}$, or shocks to W and z. These are different environments—continuous process $\bar{z}$ vs. process z with jumps.

2.3 Filtering with more signals

This section repeats the previous scalar case, with algebra and notation adapted to the case with multiple signals. With more signals, there will be as many dimensions of $\bar{W}$ as signals. The interesting cases are those where W is at least the same dimension as y, i.e., y has non-degenerate local stochastic dynamics.

Suppose that an observable signal y_t evolves according to

$$dy_t = \kappa z_t dt + \sigma dW_t$$

where y_t is a vector in $\mathbb{R}^m$, κ is an $m \times n$ matrix, z represents an n-state Markov chain, σ is an $m \times k$ matrix and B is a k-dimensional Brownian motion. We will assume that $m \leq k$, otherwise two signals are perfectly correlated and then we would be able to learn the state immediately.

Let $\mathcal{H}_t$ be the information set at time t generated by the signal history $\{y_u : 0 \leq u \leq t\}$, and we are given an initial distribution $\bar{z}_0$ of z.

The variance over a small time interval ε is given by

$$\begin{aligned} \sigma\sigma' &= |\sigma|^2 = \lim_{\varepsilon \to 0} \frac{E\left[(y_{t+\varepsilon} - y_t)^2 \mid z_t\right]}{\varepsilon} = \\ &= \lim_{\varepsilon \to 0} \frac{E\left[(y_{t+\varepsilon} - y_t - \kappa z_t \varepsilon)^2 \mid z_t\right]}{\varepsilon} = \\ &= \lim_{\varepsilon \to 0} \frac{E\left[(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon)^2 \mid z_t\right]}{\varepsilon} \end{aligned}$$

We want to produce an evolution equation for $\bar{z}_t$. Observe that

$$E\left[z_{t+\varepsilon} \mid z_t\right] \approx z_t + \varepsilon A' z_t \tag{8}$$

where we use a first order approximation of the transition matrix $\exp(\varepsilon A)$. Thus $A' z_t$ represents the 'local trend' of the underlying state. Further

$$E\left[z_{t+\varepsilon} \mid \mathcal{H}_t\right] \approx \bar{z}_t + \varepsilon A' \bar{z}_t$$

and thus

$$\lim_{\varepsilon \to 0} \frac{E\left[z_{t+\varepsilon} \mid \mathcal{H}_t\right] - \bar{z}_t}{\varepsilon} = A' \bar{z}_t$$

where $A' \bar{z}_t$ is the 'local trend' of the forecast of the underlying state. Observe that we can write

$$y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon \approx \varepsilon \kappa \left(z_t - \bar{z}_t\right) + \sigma\left(W_{t+\varepsilon} - W_t\right) \tag{9}$$

where the left-hand side represents the newly arrived information.

We can run the following regression

$$z_{t+\varepsilon} = \Delta\left(\bar{z}_t\right)\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon\right) + \eta_{t+\varepsilon}$$

to find the matrix $(n \times m)$ coefficient $\Delta\left(\bar{z}_t\right)$. We have

$$\begin{aligned}
E\left[z_{t+\varepsilon}\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon\right)' \mid \mathcal{H}_t\right] &= \Delta\left(\bar{z}_t\right) E\left[\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon\right)\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon\right)' \mid \mathcal{H}_t\right] = \\
&= \Delta\left(\bar{z}_t\right) E\left[\left(\sigma\left(W_{t+\varepsilon} - W_t\right)\right)\left(\sigma\left(W_{t+\varepsilon} - W_t\right)\right)' \mid \mathcal{H}_t\right] = \\
&= \Delta\left(\bar{z}_t\right) \sigma \sigma' \frac{1}{\varepsilon}
\end{aligned}$$

where we used equation (9). Thus

$$\begin{aligned}
\Delta\left(\bar{z}_t\right) &= E\left[z_{t+\varepsilon}\left(y_{t+\varepsilon} - y_t - \kappa \bar{z}_t \varepsilon\right)' \mid \mathcal{H}_t\right]\left(\sigma \sigma'\right)^{-1} \frac{1}{\varepsilon} = \\
&= E\left[z_{t+\varepsilon}\left(\kappa\left(z_t - \bar{z}_t\right)\right)' \mid \mathcal{H}_t\right]\left(\sigma \sigma'\right)^{-1} = \\
&= E\left[z_{t+\varepsilon}\left(z_t - \bar{z}_t\right)' \mid \mathcal{H}_t\right] \kappa' \left(\sigma \sigma'\right)^{-1} = \\
&= \left(\operatorname{diag}\left(\bar{z}_t\right) - \bar{z}_t \bar{z}_t'\right) \kappa' \left(\sigma \sigma'\right)^{-1}
\end{aligned}$$

where the last line uses results from the unidimensional case. Thus

$$d\bar{z}_t = A' \bar{z}_t dt + \Delta\left(\bar{z}_t\right)\left(dy_t - \kappa \bar{z}_t dt\right)$$

and equation (9) implies, as $\varepsilon \to 0$,

$$dy_t - \kappa \bar{z}_t dt = \kappa\left(z_t - \bar{z}_t\right) dt + \sigma dW_t$$

where the LHS is unpredictable, so that

$$\bar{\sigma} d\bar{B}_t = \kappa\left(z_t - \bar{z}_t\right) dt + \sigma d\bar{W}_t$$

To calculate $\bar{\sigma}$, take variances of both sides to get

$$\bar{\sigma} \bar{\sigma}' = \sigma \sigma'$$

Any $\bar{\sigma}$ that satisfies this equation is fine. Notice that $\bar{\sigma}$ is $m \times m$, while σ is $m \times k$, where in general $m \leq k$.

Thus our system evolves as

$$\begin{aligned} dy_t &= \kappa \bar{z}_t dt + \bar{\sigma} d\bar{W}_t \\ d\bar{z}_t &= A' \bar{z}_t dt + \Delta\left(\bar{z}_t\right) \bar{\sigma} d\bar{W}_t \end{aligned} \tag{10}$$

where y is in $\mathbb{R}^m$, κ is $m \times n$, $\bar{z}$ is an n-dimensional vector with entries in $[0,1]$, $\bar{\sigma}$ is $m \times m$, $\bar{W}$ is an m-dimensional Brownian motion, A is an $n \times n$ intensity matrix, and $\Delta\left(\bar{z}_t\right)$ is an $n \times m$ matrix.

3 Kalman–Bucy filter

This is the second type of a filtering problem, where both signals and hidden state are conditionally normally distributed. We will do the continuous-time version of the Kalman (1960) filter which is due to Kalman and Bucy (1961):

$$\begin{aligned} dx_t &= Ax_t dt + BdW_t \\ dy_t &= Dx_t dt + GdW_t \end{aligned}$$

where the first equation is the state evolution, the second is the signal evolution, and W is a vector Brownian motion.

In Wonham filtering, all we had to do is keep track of the mean. Here, it is also simple, since everything is nice and normal. All we have to keep track is the mean

$$\bar{x}_t = E\left[x_t \mid \mathcal{H}_t\right]$$

and the conditional covariance matrix for x_t given y_t

$$\Sigma_t = E\left[\left(x_t - \bar{x}_t\right)\left(x_t - \bar{x}_t\right)' \mid \mathcal{H}_t\right] \tag{11}$$

Σ_t may be singular, thus accounting for some states being observable.

The derivation of the filter is postponed to a problem set. The resulting innovations representation takes the form

$$\begin{aligned} dy_t &= D\bar{x}_t + \bar{G} d\bar{W}_t \\ d\bar{x}_t &= A\bar{x}_t dt + K_t\left(dy_t - D\bar{x}_t dt\right) = A\bar{x}_t dt + K_t \bar{G} d\bar{W}_t. \end{aligned} \tag{12}$$

where $\bar{G}\bar{G}' = GG'$ with $\bar{G}$ nonsingular, and $\bar{W}$ is a Brownian motion under $\mathcal{H}_t$, which satisfies

$$GdW_t + D\left(x_t - \bar{x}_t\right) dt = \bar{G} d\bar{W}_t.$$

Further, the Kalman gain and the conditional variance matrix satisfy

$$\begin{aligned} \frac{d\Sigma_t}{dt} &= A\Sigma_t + \Sigma_t A' + BB' - K_t GG' K_t'. \\ K_t &= \left[BG' + \Sigma_t D'\right]\left(GG'\right)^{-1} \end{aligned} \tag{13}$$

4 Learning about an unknown model

Let us assume the following pair of models

$$\begin{aligned} dx_t(i) &= A(i)\, x_t(i)\, dt + B(i)\, dW_t(i) \\ dy_t &= D(i)\, x_t(i)\, dt + G(i)\, dW_t(i) \end{aligned}$$

where $i = 0, 1$ is not observed and x_t is also not observed. Also, we assume

$$G(i)\, G(i)' = \bar{G}\bar{G}'$$

where $\bar{G}$ is nonsingular, so that the variance matrices for dy_t are the same for both models. This makes the learning problem nontrivial in continuous time (otherwise we could learn about the variances from sampling infinitely fast). Observe that if W has more components than y, this effectively selects a subvector of the components of the Brownian motion.

Notice that if x was observable, then we could employ standard inference techniques from statistics to infer the parameters.

We will work in steps.

4.1 Dynamics conditional on a model

Assume that the model is known. Then we produced the Kalman–Bucy filter formulas for

$$\begin{aligned} E\left[x_t(i) \mid \mathcal{H}_t\right] &= \bar{x}_t(i) \\ E\left[\left(x_t(i) - \bar{x}_t(i)\right)\left(x_t(i) - \bar{x}_t(i)\right)' \mid \mathcal{H}_t\right] &= \Sigma_t(i) \end{aligned}$$

where $\Sigma_t(i)$ depends only on the elapsed time (usual property of the Kalman filter).

Define

$$K_t\left(\Sigma_t(i)\right) = \left[B(i)\, G'(i) + \Sigma_t(i)\, D'(i)\right]\left[G(i)\, G(i)'\right]^{-1}$$

Then the evolution is

$$d\bar{x}_t(i) = A(i)\,\bar{x}_t(i)\, dt + K_t\left(\Sigma_t(i)\right)\left[dy_t - D(i)\,\bar{x}_t(i)\, dt\right]$$

and

$$\frac{d\Sigma_t(i)}{dt} = A(i)\,\Sigma_t(i) + \Sigma_t(i)\, A(i)' + \left(B(i) - K_t\left(\Sigma_t(i)\right) G(i)\right)\left(B(i) - K_t\left(\Sigma_t(i)\right) G(i)\right)'$$

This can be done for an arbitrary number of models i.

4.2 Log-likelihood ratios in continuous time

Consider a static model

$$y \sim N\left(\mu_i, \Lambda\right)$$

against a baseline model

$$y \sim N\left(0, \Lambda\right)$$

The log-likelihood ratio of the two models

$$-\frac{1}{2}\left(y-\mu_i\right)'\Lambda^{-1}\left(y-\mu_i\right)+\frac{1}{2}y'\Lambda^{-1}y=\mu_i'\Lambda^{-1}y-\frac{1}{2}\mu_i'\Lambda^{-1}\mu_i \tag{14}$$

We want to produce continuous-time log-likelihoods for our continuous-time models introduced in Section 4.1.

The continuous-time log-likelihood of model i against a model with zero drift is, based on the intuition from formula (14),

$$l_t\left(i\right)=\int_0^t\left[D\left(i\right)\bar{x}_u\left(i\right)\right]'\left(\bar{G}\bar{G}'\right)^{-1}dy_u-\frac{1}{2}\int_0^t\left[D\left(i\right)\bar{x}_u\left(i\right)\right]'\left(\bar{G}\bar{G}'\right)^{-1}\left[D\left(i\right)\bar{x}_u\left(i\right)\right]du$$

In particular, we want to look at an infinitesimal increment

$$dl_t\left(i\right)=\left[D\left(i\right)\bar{x}_t\left(i\right)\right]'\left(\bar{G}\bar{G}'\right)^{-1}dy_t-\frac{1}{2}\left[D\left(i\right)\bar{x}_t\left(i\right)\right]'\left(\bar{G}\bar{G}'\right)^{-1}\left[D\left(i\right)\bar{x}_t\left(i\right)\right]dt \tag{15}$$

This can again be constructed for any number of models indexed by i. The interpretation of these formulas is the same as in discrete time. Our data are samples are dy_t, with conditionally normal distribution under $\mathcal{H}_t$ given by (12).

Here I construct the log-likelihood ratio against a model with the same variance, under the models dynamics implied by the Kalman filter, with the state variables $\bar{x}\left(i\right)$ and $\Sigma_t\left(i\right)$ specified above. Notice that once we have the evolution for $\bar{x}\left(i\right)$, we can construct the evolution of the log-likelihood (everything conditional on a model).

To infer the posterior, I need to specify a prior distribution for each model, given by initial conditions $\bar{x}_0\left(i\right)$ and $\Sigma_0\left(i\right)$.

4.3 Posterior model probability

We now specify the problem for two models. Denote

$$\bar{\imath}_t=E\left[i\mid\mathcal{H}_t\right]$$

the probability of model one conditional on $\mathcal{H}_t$. Then by Bayes rule

$$\bar{\imath}_t=\frac{\exp\left(l_t\left(1\right)\right)\bar{\imath}_0}{\exp\left(l_t\left(1\right)\right)\bar{\imath}_0+\exp\left(l_t\left(0\right)\right)\left(1-\bar{\imath}_0\right)}$$

where $\bar{\imath}_0$ is the prior probability of model 1 at time 0.

Our life if easy when we recognize that $\bar{\imath}_t$ is a martingale under $\mathcal{H}_t$. Simply by law of iterated expectations,

$$E\left[\bar{\imath}_{t+\tau}\mid\mathcal{H}_t\right]=E\left[E\left[i\mid\mathcal{H}_{t+\tau}\right]\mid\mathcal{H}_t\right]=E\left[i\mid\mathcal{H}_t\right]=\bar{\imath}_t$$

Notice that we take into account that the agent (naturally) believes in the prior that he imposed.

4.4 New information

The 'new information' arriving in time interval dt is

$$dy_t - [\bar{\imath}_t D(1)\bar{x}_t(1)\,dt + (1-\bar{\imath}_t) D(0)\bar{x}_t(0)\,dt]$$

We now want to compute the local evolution of the belief $\bar{\imath}_t$. Notice first that we can consider the function

$$f(r_0, r_1) = \log \bar{\imath} + r_1 - \log[\bar{\imath}\exp(r_1) + (1-\bar{\imath})\exp(r_0)]$$

and we have

$$\begin{aligned}
\frac{\partial f(r_0,r_1)}{\partial r_1} &= 1 - \frac{\bar{\imath}\exp(r_1)}{\bar{\imath}\exp(r_1) + (1-\bar{\imath})\exp(r_0)} = 1 - \exp(f(r_0,r_1)) \\
\frac{\partial f(r_0,r_1)}{\partial r_0} &= -\frac{(1-\bar{\imath})\exp(r_0)}{\bar{\imath}\exp(r_1) + (1-\bar{\imath})\exp(r_0)} = \exp(f(r_0,r_1)) - 1 = -\frac{\partial f(r_0,r_1)}{\partial r_1}
\end{aligned}$$

Therefore, the application of Itô's lemma to $\log \bar{\imath}_t$ yields

$$\begin{aligned}
d\log \bar{\imath}_t &= dl_t(1) - \frac{\exp(l_t(1))\bar{\imath}_0 dl_t(1) + \exp(l_t(0))(1-\bar{\imath}_0)\,dl_t(0)}{\exp(l_t(1))\bar{\imath}_0 + \exp(l_t(0))(1-\bar{\imath}_0)} + \text{terms in } dt = \\
&= (1-\bar{\imath}_t)(dl_t(1) - dl_t(0)) + \text{terms in } dt
\end{aligned}$$

Substituting in for $dl_t(i)$ from (15), we obtain

$$\begin{aligned}
d\log \bar{\imath}_t &= (1-\bar{\imath}_t)\left\{(D(1)\bar{x}_t(1))'\left(\bar{G}\bar{G}'\right)^{-1} - (D(0)\bar{x}_t(0))'\left(\bar{G}\bar{G}'\right)^{-1}\right\}\cdot \\
&\quad \cdot\{dy_t - [\bar{\imath}_t D(1)\bar{x}_t(1)\,dt + (1-\bar{\imath}_t)D(0)\bar{x}_t(0)\,dt]\} + \\
&\quad + \text{different terms in } dt
\end{aligned}$$

where the 'different terms in dt' also contain dt terms that have been added to the middle line. In particular, we removed the model specific drifts in $dl_t(i)$ defined in (15) and replaced them with one average drift so that the term

$$dy_t - [\bar{\imath}_t D(1)\bar{x}_t(1)\,dt + (1-\bar{\imath}_t) D(0)\bar{x}_t(0)\,dt]$$

does not have a drift under the information of the agent who does not observe the two models separately. I can do this since these are all dt operations.

Notice that log-probabilities are not martingales but probabilities are, so that we can infer that probabilities in levels evolve as

$$\begin{aligned}
d\bar{\imath}_t &= \bar{\imath}_t(1-\bar{\imath}_t)\left[(D(1)\bar{x}_t(1) - D(0)\bar{x}_t(0))'\left(\bar{G}\bar{G}'\right)^{-1}\right]\cdot \\
&\quad \cdot\{dy_t - [\bar{\imath}_t D(1)\bar{x}_t(1)\,dt + (1-\bar{\imath}_t)D(0)\bar{x}_t(0)\,dt]\}
\end{aligned}$$

It is essential to understand this formula. Think of a single signal (y is a scalar). Think about a situation where we obtained a higher signal than expected (the second line in the previous formula is positive). Then if model 1 has a larger mean than model 0, then the probability $\bar{\imath}$ of model 1 goes up. In usual situations, $\bar{\imath}$ will converge either to zero or one as we learn the model.

We can now define

$$d\widetilde{W}_t = \bar{G}^{-1}(dy_t - \bar{\imath}_t D(1)\bar{x}_t(1)\,dt - (1-\bar{\imath}_t)D(0)\bar{x}_t(0)\,dt)$$

$\tilde{W}$ is a Brownian motion relative to $\mathcal{H}$.

5 Applications to risk prices: Breeden model

See problem set.

References

Kalman, Rudolph Emil. 1960. "New Approach to Linear Filtering and Prediction Problems." *Transactions of the ASME–Journal of Basic Engineering* 82 (Series D):35–45.

Kalman, Rudolph Emil and Richard Snowden Bucy. 1961. "New Results in Linear Filtering and Prediction Theory." *Journal of Basic Engineering* 83 (1):95–108.

Wonham, W. Murray. 1964. "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering." *SIAM Journal on Control and Optimization* 2 (3):347–369. URL http://link.aip.org/link/?SJC/2/347/1.