Filtro de Kalman: teoria y aplicaciones*

Alfonso Novales
Departamento de Economía Cuantitativa
Universidad Complutense

Diciembre 2017
Versión preliminar
No citar sin permiso del autor

Contents

1 Filtro de Kalman

El filtro de Kalman es un algoritmo para actualizar, observación a observación, la proyección lineal de un sistema de variables sobre el conjunto de información disponible, según se va disponiendo de nueva información. Para ello, es preciso representar el modelo en la formulación conocida como espacio de los estados. El filtro de Kalman permite calcular de modo sencillo la verosimilitud de un

*Estas notas se basan en el Capitulo 13 de Hamilton: "Time Series Analysis" modelo dinámico lineal, uniecuacional o multiecuacional, lo que permite estimar los parámetros de dicho modelo, así como obtener predicciones de dicho tipo de modelos.

La representación en forma de espacio de los estados es:

$$\underset{(rx1)}{\xi_{t+1}} = \underset{(rxr)}{F}\xi_t + \underset{(rx1)}{v_{t+1}}, \ \textit{ecuación de estado} \tag{1}$$

$$\underset{(nx1)}{y_t} = \underset{(nxk)}{A'}\underset{(kx1)}{x_t} + \underset{(nxr)}{H'}\xi_t + \underset{(nx1)}{w_t}, \ \textit{ecuación de observación} \tag{2}$$

con:

$$\begin{aligned}
E(v_t v_t') &= Q, \text{ matriz } rxr; \ E(v_t v_s') = 0_{rxr}, \text{ si } t \neq s \\
E(w_t w_t') &= R, \text{ matriz } nxn; \ E(w_t w_s') = 0_{nxn}, \text{ si } t \neq s \\
E(v_t w_s') &= 0, \text{ matriz } rxn \ \forall t, s \\
E(\xi_t w_s') &= 0, \text{ matriz } rxn \ \forall t, s
\end{aligned}$$

El vector ξ_t de variables de estado es generalmente no observable, mientras que y_t y x_t son observables. El vector x_t contiene variables exógenas. Ello significa que x_t no contiene información sobre ξ_{t+s} o w_{t+s} que no esté ya contenida en $(y_{t-1}, y_{t-2}, ..., y_1)$. El vector x_t puede contener retardos de y_t o variables que estén incorrelacionadas con ξ_τ y w_τ para todo τ.

Nuestro objetivo es predecir ξ_{t+1} e y_{t+1} a partir de información disponible en t. Denotamos la predicción un periodo hacia adelante como:

$$\hat{\xi}_{t+1|t} = E(\xi_{t+1} \mid Y_t);$$

donde Y_t denota el conjunto de información disponible para el analista al hacer las predicciones para el periodo $t+1 : Y_t = \left(x_t', x_{t-1}', .., x_1', y_t', y_{t-1}', ..., y_1'\right)$. En otras ocasiones calcularemos predicciones condicionales en el conjunto Y_{t-1} o en el conjunto (x_t', Y_{t-1}). Tambien analizaremos predicciones a horizontes más largos.

Que las variables x_t sean exógenas o predeterminadas implica que:

$$E(\xi_t \mid x_t, Y_{t-1}) = E(\xi_t \mid Y_{t-1}) = \hat{\xi}_{t|t-1}$$

Asociadas a cada una de estas predicciones tendremos una matriz de Error Cuadratico Medio:

$$P_{t|t-1} = E\left[\left(\xi_t - \hat{\xi}_{t|t-1}\right)\left(\xi_t - \hat{\xi}_{t|t-1}\right)'\right]$$

Suponiendo que tuviéramos condiciones iniciales $\hat{\xi}_{1|0}, P_{1|0}$, querríamos encontrar expresiones recursivas que nos permitan pasar de $\hat{\xi}_{t|t-1}, P_{t|t-1}$ a $\hat{\xi}_{t+1|t}, P_{t+1|t}$. El procedimiento recursivo comienza con $\hat{\xi}_{1|0}$, que denota la predicción de ξ_1

sin utilizar información sobre y o x. Dicha predicción es la media incondicional de ξ_1 :

$$\hat{\xi}_{1|0} = E(\xi_1)$$

con matriz de ECM:

$$P_{1|0} = E\left[(\xi_1 - E\xi_1)(\xi_1 - E\xi_1)'\right]$$

Si los autovalores de F están todos dentro del circulo unidad, entondes el proceso propuesto para ξ_t es estacionario en covarianza. En ese caso, la media incondicional de ξ_t puede obtenerse tomando esperanza matemática en (1) : $E(\xi_{t+1}) = F.E(\xi_t)$, por lo que si es estacionario en covarianza, tendremos: $(I_r - F)E(\xi_t) = 0$, y si ningun autovalor es igual a cero, entonces: $E(\xi_t) = 0$.

La varianza incondicional es:

$$\begin{aligned}\Sigma &= E(\xi_{t+1}\xi_{t+1}') = E\left[(F\xi_t + v_{t+1})(F\xi_t + v_{t+1})'\right] = F\left[E(\xi_t\xi_t')\right]F' + E(v_{t+1}v_{t+1}') = \\ &= F\Sigma F' + Q\end{aligned}$$

ecuación que tiene por solución:[1]

$$vec(\Sigma) = \left[I_{r^2} - (F \otimes F)\right]^{-1} vec(Q)$$

1.1 Comienza el procedimiento recursivo

Por tanto, bajo el supuesto mencionado para los autovalores de F, comenzaremos la recursión con $\hat{\xi}_{1|0} = 0$, y con $P_{1|0}$ igual a la matriz rxr Σ que se obtiene de la ecuación anterior. En cualquier otra circunstancia, puede tomarse otro valor de $\hat{\xi}_{1|0}$, mientras que $P_{1|0}$ puede escogerse de modo que refleje la incertidumbre existente acerca de dichos valores iniciales. Valores numericos más elevados a lo largo de la diagonal de $P_{1|0}$ reflejarán mayor incertidumbre acerca del verdadero valor de la variable de estado correspondiente.

Nuestro objetivo es predecir el valor de y_t :

$$\hat{y}_{t|t-1} \equiv E\left(y_t \mid x_t, Y_{t-1}\right)$$

Ahora bien, como:

$$E\left(y_t \mid x_t, \xi_t\right) = A'x_t + H'\xi_t$$

aplicando la ley de proyecciones iteradas, tenemos:

$$\hat{y}_{t|t-1} = A'x_t + H'E(\xi_t \mid x_t, Y_{t-1}) = A'x_t + H'\hat{\xi}_{t|t-1}$$

[1] $\Sigma = F\Sigma F' + Q \Rightarrow vec(\Sigma) = vec(F\Sigma F') + vec(Q) \Rightarrow$
$vec(\Sigma) = (F \otimes F)vec(\Sigma) + vec(Q) \Rightarrow vec(\Sigma) = \left[I_{r^2} - (F \otimes F)\right]^{-1} vec(Q)$

por lo que el error de predicción es:

$$\begin{aligned} y_t - \hat{y}_{t|t-1} &= (A'x_t + H'\xi_t + w_t) - (A'x_t + H'\hat{\xi}_{t|t-1}) = \\ &= H'\left(\xi_t - \hat{\xi}_{t|t-1}\right) + w_t \end{aligned}$$

con matriz de error cuadratico medio

$$V_t = E\left[\left(y_t - \hat{y}_{t|t-1}\right)\left(y_t - \hat{y}_{t|t-1}\right)'\right] = H'P_{t|t-1}H + R$$

donde $P_{t|t-1}$ denota la matriz de error cuadrático medio de $\xi_t - \hat{\xi}_{t|t-1}$, y donde hemos utilizado que: $E\left[w_t\left(\xi_t - \hat{\xi}_{t|t-1}\right)'\right] = 0$. Esta igualdad se justifica porque w_t está incorrelacionado con ξ_t y porque como $\hat{\xi}_{t|t-1}$ es una combinación lineal de variables en Y_{t-1}, tambien ha de estar incorrelacionada con w_t.

1.2 Actualización de inferencias acerca de ξ_t

Una vez que observamos y_t, actualizamos nuestra estimación de ξ_t para obtener: $\hat{\xi}_{t|t} = E(\xi_t \mid x_t, y_t, Y_{t-1}) = E(\xi_t \mid Y_t)$, mediante:[2]

$$\begin{aligned} \hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + \left\{E\left[\left(\xi_t - \hat{\xi}_{t|t-1}\right)\left(y_t - \hat{y}_{t|t-1}\right)'\right]\right\} \times \\ &\quad \left\{E\left[\left(y_t - \hat{y}_{t|t-1}\right)\left(y_t - \hat{y}_{t|t-1}\right)'\right]\right\}^{-1}\left(y_t - \hat{y}_{t|t-1}\right) \end{aligned}$$

Pero:

$$E\left[\left(\xi_t - \hat{\xi}_{t|t-1}\right)\left(y_t - \hat{y}_{t|t-1}\right)'\right] = E\left\{\left(\xi_t - \hat{\xi}_{t|t-1}\right)\left[H'\left(\xi_t - \hat{\xi}_{t|t-1}\right) + w_t\right]'\right\} = P_{t|t-1}H$$

por lo que:

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + P_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1}\left(y_t - A'x_t - H'\hat{\xi}_{t|t-1}\right)$$

[2] Utilizamos la expresión:

$$P(A_3 \mid A_2, A_1) = \Omega_{31}\Omega_{11}^{-1}A_1 + H_{32}H_{22}^{-1}\left(A_2 - \Omega_{21}\Omega_{11}^{-1}A_1\right) =$$

con:

$$\begin{aligned} H_{22} &= E\left[(A_2 - P(A_2 \mid A_1))(A_2 - P(A_2 \mid A_1))'\right] \\ H_{32} &= E\left[(A_3 - P(A_3 \mid A_1))(A_2 - P(A_2 \mid A_1))'\right] \end{aligned}$$

para $A_3 \equiv \xi_t, A_2 \equiv y_t, A_1 \equiv (x_t, Y_{t-1})$. Su matriz de error cuadrático medio es:

$$E\left[(A_3 - P(A_3 \mid A_2, A_1))(A_3 - P(A_3 \mid A_2, A_1))'\right] = H_{33} - H_{32}H_{22}^{-1}H_{23}$$

con error cuadrático medio (ver pie de página anterior):

$$\begin{aligned} P_{t|t} &= E\left[\left(\xi_t - \hat{\xi}_{t|t}\right)\left(\xi_t - \hat{\xi}_{t|t}\right)'\right] = \\ &= P_{t|t-1} - \left[P_{t|t-1} H_t \left(H_t' P_{t|t-1} H_t + R\right)^{-1} H_t' P_{t|t-1}\right] \end{aligned}$$

Como hemos visto, el filtro de Kalman se inicializa con:

$$\begin{aligned} \hat{\xi}_{1|0} &= E(\xi) \\ P_{1|0} &= E\left[(\xi - E(\xi))(\xi - E(\xi))\right] = \left[I_{r^2} - (F \otimes F)\right]^{-1} vec(Q) \end{aligned}$$

A partir de la ecuación de estado (1) obtenemos las ecuaciones de *predicción*:

$$\begin{aligned} \hat{\xi}_{t+1|t} &= E\left(\xi_{t+1} \mid Y_t\right) = F.E\left(\xi_t \mid Y_t\right) + E\left(v_{t+1} \mid Y_t\right) = F\hat{\xi}_{t|t} = \\ &= F\hat{\xi}_{t|t-1} + FP_{t|t-1} H \left(H' P_{t|t-1} H + R\right)^{-1} \left(y_t - A' x_t - H'\hat{\xi}_{t|t-1}\right) = \\ & F\hat{\xi}_{t|t-1} + K_t \left(y_t - \hat{y}_{t|t-1}\right) \end{aligned}$$

donde K_t denota la *ganancia* del filtro de Kalman:

$$K_t = FP_{t|t-1} H_t (H_t' P_{t|t-1} H_t + R)^{-1}$$

La matriz de Error Cuadrático Medio es:

$$\begin{aligned} P_{t+1|t} &= E\left[\left(\xi_{t+1} - \hat{\xi}_{t+1|t}\right)\left(\xi_{t+1} - \hat{\xi}_{t+1|t}\right)'\right] = \\ &= E\left[\left(F\xi_t + v_{t+1} - F\hat{\xi}_{t|t}\right)\left(F\xi_t + v_{t+1} - F\hat{\xi}_{t|t}\right)'\right] = \\ &= FP_{t|t}F' + Q = \left(FP_{t|t-1} - K_t H' P_{t|t-1}\right) F' + Q \end{aligned}$$

puesto que los productos cruzados tiene esperanza matemática nula.

1.2.1 Resumen: Filtro de Kalman

$$\begin{aligned} y_t - \hat{y}_{t|t-1} &= y_t - A' x_t - H'\hat{\xi}_{t|t-1} \\ ECM(y_t) &= V_t = H' P_{t|t-1} H + R \\ \hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + P_{t|t-1} H V_t^{-1} \left(y_t - \hat{y}_{t|t-1}\right) \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1} H V_t^{-1} H' P_{t|t-1} \\ K_t &= FP_{t|t-1} H V_t^{-1} \\ \hat{\xi}_{t+1|t} &= F\hat{\xi}_{t|t} \\ P_{t+1|t} &= FP_{t|t}F' + Q \end{aligned}$$

Nótese que si $F = I_n$, entonces $\hat{\xi}_{t+1|t} = \hat{\xi}_{t|t}$.

Por último, para formar la función log-verosimilitud bajo el supuesto de Normalidad, necesitamos:

$$E(y_t \mid x_t, Y_{t-1}) = A'x_t + H_t'\hat{\xi}_{t|t-1}$$

$$E\left[\left(y_t - A'x_t - H_t'\hat{\xi}_{t|t-1}\right)\left(y_t - A'x_t - H_t'\hat{\xi}_{t|t-1}\right)' \mid x_t, Y_{t-1}\right] = H_t'P_{t|t-1}H_t + R$$

y finalmente, la función de densidad del vector y_t cada periodo y la función de verosimilitud, en logaritmos, son:

$$\begin{aligned}\ln f(y_t \quad \mid \quad & x_t, Y_{t-1}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\left|H_t'P_{t|t-1}H_t + R\right|\right) - \\ & -\frac{1}{2}\left(y_t - A'x_t - H_t'\hat{\xi}_{t|t-1}\right)'\left(H_t'P_{t|t-1}H_t + R\right)^{-1}\left(y_t - A'x_t - H_t'\hat{\xi}_{t|t-1}\right)\end{aligned}$$

$$\log Lik = \sum_{t=1}^{T} \ln f(y_t \mid x_t, Y_{t-1})$$

Que puede maximizarse respecto de los parámetros que aparecen en las matrices F, Q, A, H y R.

1.3 Predicción s períodos hacia adelante

A partir de la ecuación de estado, y suponiendo que las variables x son deterministas, tenemos:

$$\begin{aligned}
\xi_{t+s} &= F^s\xi_t + F^{s-1}v_{t+1} + F^{s-2}v_{t+2} + ... + Fv_{t+s-1} + v_{t+s}, \;\; s = 1, 2, ... \\
E(\xi_{t+s} &\mid \xi_t, Y_t) = F^s\xi_t \Rightarrow \hat{\xi}_{t+s|t} = E(\xi_{t+s} \mid Y_t) = F^s\hat{\xi}_{t|t} \\
\xi_{t+s} - \hat{\xi}_{t+s|t} &= F^s\left(\xi_t - \hat{\xi}_{t|t}\right) + F^{s-1}v_{t+1} + F^{s-2}v_{t+2} + ... + Fv_{t+s-1} + v_{t+s} \\
&\Rightarrow P_{t+s|t} = F^sP_{t|t}(F')^s + F^{s-1}Q(F')^{s-1} + F^{s-2}Q(F')^{s-2} + ... + FQF' + Q \\
\hat{y}_{t+s|t} &= E(y_{t+s} \mid Y_t) = A'x_{t+s} + H'\hat{\xi}_{t+s|t} \Rightarrow y_{t+s} - \hat{y}_{t+s|t} = H'\left(\xi_{t+s} - \hat{\xi}_{t+s|t}\right) + w_{t+s} \\
MSE &: E\left[\left(y_{t+s} - \hat{y}_{t+s|t}\right)\left(y_{t+s} - \hat{y}_{t+s|t}\right)'\right] = H'P_{t+s|t}H + R
\end{aligned}$$

1.4 Smoothing (suavizado)

La estimación suavizada de las variables de estado es:

$$\xi_{t|T} = E\left(\xi_t \mid Y_t\right)$$

Supongamos que hemos estimado ξ_t con datos hasta t , $\xi_{t|t}$ y que posteriormente observamos el verdadero valor de ξ_{t+1}. Podríamos utilizar la expresion de actualización de una proyección lineal que vimos en el pie de página anterior para obtener:

$$\begin{aligned} E\left(\xi_t \mid \xi_{t+1}, Y_t\right) &= \xi_{t|t} + E\left[\left(\xi_t - \xi_{t|t}\right)\left(\xi_{t+1} - \xi_{t+1|t}\right)'\right]. \\ &.E\left[\left(\xi_{t+1} - \xi_{t+1|t}\right)\left(\xi_{t+1} - \xi_{t+1|t}\right)'\right]^{-1} . \left(\xi_{t+1} - \xi_{t+1|t}\right) \end{aligned}$$

El primer término de la derecha puede escribirse:

$$E\left[\left(\xi_t - \xi_{t|t}\right)\left(\xi_{t+1} - \xi_{t+1|t}\right)'\right] = E\left[\left(\xi_t - \xi_{t|t}\right)\left(F\xi_t + v_{t+1} - F\xi_{t|t}\right)'\right]$$

y como v_{t+1} está incorrelacionado con ξ_t y con $\xi_{t|t}$, entonces:

$$E\left[\left(\xi_t - \xi_{t|t}\right)\left(\xi_{t+1} - \xi_{t+1|t}\right)'\right] = E\left[\left(\xi_t - \xi_{t|t}\right)\left(\xi_t - \xi_{t|t}\right)' F'\right] = P_{t|t}F'$$

Sustituyendo esta igualdad y la definición de $P_{t+1|t}$ en la primera expresion, tenemos:

$$E\left(\xi_t \mid \xi_{t+1}, Y_t\right) = \xi_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}\left(\xi_{t+1} - \xi_{t+1|t}\right)$$

y si definimos:

$$J_t = P_{t|t}F'P_{t+1|t}^{-1}$$

llegamos a:

$$E\left(\xi_t \mid \xi_{t+1}, Y_t\right) = \xi_{t|t} + J_t\left(\xi_{t+1} - \xi_{t+1|t}\right)$$

Pero esta proyección es la misma que $E\left(\xi_t \mid \xi_{t+1}, Y_T\right)$ puesto que conocer los valores futuros y_{t+j}, x_{t+j}, $j > 0$, no añade información una vez que conocemos ξ_{t+1}. Esto se debe a que el error de proyección $\xi_t - E\left(\xi_t \mid \xi_{t+1}, Y_t\right)$ esta incorrelacionado con ξ_{t+1} por las propiedades de una proyección lineal, y tambien con $x_{t+j}, w_{t+j}, v_{t+j}, v_{t+j-1}, ..., v_{t+2}$ bajo las hipótesis que expusimos al comienzo. Por tanto, el error de proyección está incorrelacionado con y_{t+j}, x_{t+j}, $j > 0$. En consecuencia:

$$E\left(\xi_t \mid \xi_{t+1}, Y_T\right) = \xi_{t|t} + J_t\left(\xi_{t+1} - \xi_{t+1|t}\right)$$

Por la ley de las proyecciones iteradas, la proyección $E\left(\xi_t \mid Y_T\right)$ puede obtenerse proyectando $E\left(\xi_t \mid \xi_{t+1}, Y_T\right)$ sobre Y_T. Pero $\xi_{t|t}$ es una función lineal exacta de Y_t. Los coeficientes de dicha funcion son momentos poblacionales, y son por tanto constantes deterministas a efectos de una nueva proyección. Luego la proyección de $\xi_{t|t}$ sobre Y_T es $\xi_{t|t}$. Lo mismo sucede con $\xi_{t+1|t}$. El término J_t también es una función de momentos poblacionales y su proyección es él mismo. Por tanto:

$$E\left(\xi_t \mid Y_T\right) = \xi_{t|t} + J_t \left[E(\xi_{t+1} \mid Y_T) - \xi_{t+1|t}\right]$$

o:

$$\xi_{t|T} = \xi_{t|t} + J_t \left(\xi_{t+1|T} - \xi_{t+1|t}\right) \tag{3}$$

Procedemos de la siguiente manera: aplicamos el filtro de Kalman para obtener las sucesiones: $\left\{\xi_{t|t}\right\}_{t=1}^{T}, \left\{\xi_{t+1|t}\right\}_{t=1}^{T-1}, \left\{P_{t|t}\right\}_{t=1}^{T}, \left\{P_{t+1|t}\right\}_{t=1}^{T-1}$. La estimacion suavizada de la ultima observacion muestral, $\xi_{T|T}$ es la ultima observacion en la sucesión $\left\{\xi_{t|t}\right\}_{t=1}^{T}$. Luego, generamos $\{J_t\}_{t=1}^{T-1}$, y utilizamos (3) con $t = T-1$ para obtener $\xi_{T-1|T}$:

$$\xi_{T-1|T} = \xi_{T-1|T-1} + J_{T-1} \left(\xi_{T|T} - \xi_{T|T-1}\right)$$

A continuación, utilizamos la misma expresión con $t = T-2$:

$$\xi_{T-2|T} = \xi_{T-2|T-2} + J_{T-2} \left(\xi_{T-1|T} - \xi_{T-1|T-2}\right)$$

y así sucesivamente. Para obtener el error cuadratico medio, a partir de (3) tenemos:

$$\begin{aligned} \xi_t - \xi_{t|T} &= \xi_t - \xi_{t|t} - J_t\xi_{t+1|T} + J_t\xi_{t+1|t} \Rightarrow \\ \xi_t - \xi_{t|T} + J_t\xi_{t+1|T} &= \xi_t - \xi_{t|t} + J_t\xi_{t+1|t} \end{aligned}$$

Multiplicando por su traspuesta y tomando esperanzas:

$$\begin{aligned} & E\left[\left(\xi_t - \xi_{t|T}\right)\left(\xi_t - \xi_{t|T}\right)'\right] + J_t E\left[\xi_{t+1|T}\xi'_{t+1|T}\right] J'_t \\ = \quad & E\left[\left(\xi_t - \xi_{t|t}\right)\left(\xi_t - \xi_{t|t}\right)'\right] + J_t E\left[\xi_{t+1|t}\xi'_{t+1|t}\right] J'_t \end{aligned}$$

y tras varios desarrollos se llega a:

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

una expresion que puede utilizarse de nuevo hacia atrás en el tiempo, comenzando con $t = T-1$.

2 Aplicaciones

2.1 Estimación de la tendencia subyacente

2.1.1 Modelo de camino aleatorio

Consideremos un modelo de series temporales:

$$\begin{aligned} y_t &= \mu_t + e_t,\ e_t \sim N(0, \sigma_e^2) \\ \mu_{t+1} &= \beta\mu_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2) \end{aligned}$$

con $E(e_t.\eta_s) = 0, \forall t, s.$

En términos de la notación de Hamilton:

$$\begin{aligned} \xi_t &= \mu_t;\ F \equiv \beta;\ x_t = 0 \forall t;\ H_t' = 1;\ A = 0; \\ w_t &\equiv e_t; v_t = \eta_t;\ Q = \sigma_\eta^2;\ R = \sigma_e^2 \end{aligned}$$

y, teniendo en cuenta que tenemos una única variable de estado y una única variable de observación, el fitro de Kalman resulta:

$$\begin{aligned} y_t - \hat{y}_{t|t-1} &= y_t - \mu_{t|t-1} \\ V_t &= P_{t|t-1} + \sigma_e^2 \\ K_t &= \beta P_{t|t-1}/V_t \\ \hat{\mu}_{t|t} &= \hat{\mu}_{t|t-1} + P_{t|t-1}\left(y_t - \hat{y}_{t|t-1}\right)/V_t \\ P_{t|t} &= P_{t|t-1} - \frac{P_{t|t-1}^2}{V_t} = P_{t|t-1}\left(1 - \frac{P_{t|t-1}}{V_t}\right) \\ \hat{\mu}_{t+1|t} &= \beta\hat{\mu}_{t|t} = \beta\hat{\mu}_{t|t-1} + K_t\left(y_t - \hat{y}_{t|t-1}\right) \\ P_{t+1|t} &= \beta^2 P_{t|t} + \sigma_\eta^2 \end{aligned}$$

en el que todas las variables, vectores y matrices son escalares, de dimensión 1x1.

Analicemos en este ejemplo la elección de condiciones iniciales. Tenemos que:

$$\begin{aligned} y_1 - y_{1|0} &= y_1 - \mu_{1|0};\ V_1 = P_{1|0} + \sigma_e^2 \\ \mu_{2|1} &= \mu_{1|0} + \frac{P_{1|0}}{V_1} w_1 = \mu_{1|0} + \frac{P_{1|0}}{P_{1|0} + \sigma_e^2}\left(y_1 - \mu_{1|0}\right) \\ P_{2|1} &= P_{1|0}\left(1 - \frac{P_{1|0}}{P_{1|0} + \sigma_e^2}\right) + \sigma_e^2 = \frac{P_{1|0}}{P_{1|0} + \sigma_e^2}\sigma_e^2 + \sigma_e^2 \end{aligned}$$

Por tanto, si dejamos que $P_{1|0}$ tienda a infinito, tendremos: $\mu_{2|1} = y_1$, y $P_{2|1} = \sigma_e^2 + \sigma_w^2$. Esto equivale a tratar y_1 como un parámetro fijo y suponer:

$\mu_1 \sim N(y_1, \sigma_e^2)$, lo cual se conoce como una inicialización difusa, porque tomar $P_{1|0}$ muy grande significa que se tiene una importante incertidumbre acerca de la condición inicial.

2.1.2 Modelo de doble camino aleatorio

Consideremos un modelo de series temporales:

$$\begin{aligned}
y_t &= \mu_t + x_t \\
\mu_t &= g_{t-1} + \mu_{t-1} + v_t, v_t \sim i., i.d.N(0, \sigma_v^2) \\
g_t &= g_{t-1} + w_t, w_t \sim i., i.d.N(0, \sigma_w^2) \\
x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + e_t, e_t \sim i., i.d.N(0, \sigma_e^2)
\end{aligned}$$

con $E(e_t.\eta_s) = E(e_t.w_s) = E(w_t.\eta_s) = 0, \forall t, s.$

En términos de la notación de Hamilton:

$$\begin{aligned}
\xi_t &= (\mu_t, x_t, x_{t-1}, g_t); \;\; H_t' = (1, 1, 0, 0); \;\; A = 0; \\
F &\equiv \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & \phi_1 & \phi_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \\
w_t &\equiv 0 \forall t; \;\; v_t = (v_{t,} e_t, 0, w_t, \eta_t); \;\; R = 0; \;\; Q = diag(\sigma_v^2, \sigma_e^2, 0, \sigma_w^2)
\end{aligned}$$

y, teniendo en cuenta que tenemos una única variable de estado y una única variable de observación, el fitro de Kalman resulta:

$$\begin{aligned}
y_t - \hat{y}_{t|t-1} &= y_t - H'\mu_{t|t-1} \\
V_t &= H' P_{_{t|t-1}} H \\
K_t &= F P_{t|t-1} H V_t H^{-1} \\
\hat{\mu}_{t|t} &= \hat{\mu}_{t|t-1} + P_{t|t-1} H V_t^{-1} \left(y_t - \hat{y}_{t|t-1} \right) \\
P_{t|t} &= P_{t|t-1} - P_{t|t-1} H V_t^{-1} H' P_{t|t-1} \\
\hat{\mu}_{t+1|t} &= F \hat{\mu}_{t|t} = \beta \hat{\mu}_{t|t-1} + K_t \left(y_t - \hat{y}_{t|t-1} \right) \\
P_{t+1|t} &= \left(F - K_t H' \right) P_{t|t} \left(F - K_t H' \right)' + Q
\end{aligned}$$

2.1.3 Descomposición de una variable en tendencia y ciclo

Modelo:

$$\begin{aligned}
y_t &= T_t + c_t \\
T_{t+1} &= T_t + v_{t+1}, v_t \sim i, id.N(0, \sigma_v^2) \\
c_{t+1} &= \phi c_t + e_t, e_t \sim i, id.N(0, \sigma_e^2)
\end{aligned}$$

donde y_t puede ser el logaritmo del PIB, o de un indice de bolsa, por ejemplo.
Representación en espacio de los estados:
Ecuación de observación:

$$y_t = \underbrace{(1\ 1)}_{G} \underbrace{\begin{pmatrix} T_t \\ c_t \end{pmatrix}}_{xi_t}$$

Ecuación de estado:

$$\underbrace{\begin{pmatrix} T_{t+1} \\ c_{t+1} \end{pmatrix}}_{xi_{t+1}} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix}}_{F} \underbrace{\begin{pmatrix} T_t \\ c_t \end{pmatrix}}_{xi_t} + \underbrace{\begin{pmatrix} v_t \\ e_t \end{pmatrix}}_{\eta_t};\ \eta_t \sim N\left(0_2, \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_e^2 \end{pmatrix}\right)$$

Condiciones iniciales razonables serían:

$$\begin{aligned} \hat{x}i_{1|0} &= (\ln y_1;\ 0), \\ vec\left(\Sigma_{1|0}\right) &= vec\,(x_1 x_1') = \left[I_4 - F \otimes F\right]^{-1} vec(Q)\ \text{[Matriz singular, habria que buscar alternativas]} \end{aligned}$$

2.2 El filtro de Kalman con parámetros cambiantes

En muchas situaciones, la representación en espacio de los estados se obtiene con parámetros cambiantes en el tiempo:

$$\begin{aligned} \xi_{t+1} &= F(x_t)\xi_t + v_{t+1} \\ y_t &= a(x_t) + H(x_t)'\xi_t + w_t \end{aligned}$$

En esta situación, el supuesto de Normalidad condicional para las perturbaciones, que en el caso general no es necesaria, resulta crucial. Si suponemos:

$$\begin{pmatrix} v_{t+1} & \\ w_t & \end{pmatrix} | x_t, Y_{t-1} \Bigg) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q(x_t) & 0 \\ 0 & R(x_t) \end{bmatrix}\right)$$

entonces la expresion del filtro de Kalman es similar a la del caso mas restringido que contempla parametros constantes, con las variaciones que naturalmente provienen del hecho de que las matrices y vectores F, A, H son ahora cambiantes en el tiempo.

Puede probarse (Hamilton, sección 13.8) que:

$$\begin{aligned} \hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + \left\{P_{t|t-1}H(x_t)\left[H(x_t)'P_{t|t-1}H(x_t) + R(x_t)\right]^{-1}\left(y_t - a(x_t) - H(x_t)'\hat{\xi}_{t|t-1}\right)\right\} \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}H(x_t)\left[H(x_t)'P_{t|t-1}H(x_t) + R(x_t)\right]^{-1} H(x_t)'P_{t|t-1} \end{aligned}$$

de lo que se obtiene:

$$
\begin{aligned}
V_t &= H_t' P_{t|t-1} H_t + R \\
K_t &= F_t P_{t|t-1} H_t V_t^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\xi}_{t+1|t} &= F(x_t)\hat{\xi}_{t|t} \\
P_{t+1|t} &= F(x_t) P_{t|t} F(x_t)' + Q(x_t)
\end{aligned}
$$

2.3 Aplicacion: Regresión con parámetros cambiantes en el tiempo

Consideremos el modelo de regresión:

$$y_t = x_t' \beta_t + w_t$$

en el que suponemos que las variables en x_t son retardos de y_t o variables que son independientes de w_τ para todo τ. Suponemos que los parámetros evolucionan de acuerdo con:

$$\beta_{t+1} - \beta = F(\beta_t - \beta) + v_{t+1}$$

lo que implica que si los autovalores de F son inferiores a la unidad, entonces $E(\beta_t) = \beta$.

Suponemos tambien:

$$\begin{pmatrix} v_{t+1} & \\ w_t & \end{pmatrix} \mid x_t, Y_{t-1} \Bigg) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q & 0 \\ 0 & \sigma^2 \end{bmatrix} \right)$$

La regresión puede escribirse:

$$y_t = x_t' \beta + x_t' \xi_t + w_t$$

donde el vector de estado es el vector de coeficientes de la regresión en diferencias respecto de su media:

$$\xi_t = \beta_t - \beta$$

por lo que la predicción un periodo hacia adelante es:

$$E(y_t \mid x_t, Y_{t-1}) = x_t' \beta + x_t' \hat{\xi}_{t|t-1}$$

donde $\hat{\xi}_{t|t-1}$ se obtiene a partir de:

$$
\begin{aligned}
\hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + P_{t|t-1} x_t \left(x_t{}' P_{t|t-1} x_t + \sigma^2 \right)^{-1} \left(y_t - x_t' \beta - x_t{}' \hat{\xi}_{t|t-1} \right) \\
\hat{\xi}_{t+1|t} &= F \hat{\xi}_{t|t} \\
P_{t|t} &= P_{t|t-1} - P_{t|t-1} x_t \left(x_t{}' P_{t|t-1} x_t + \sigma^2 \right)^{-1} x_t{}' P_{t|t-1} \\
P_{t+1|t} &= F P_{t|t} F' + Q(x_t)
\end{aligned}
$$

comenzando con las condiciones iniciales habituales. En este caso, la matriz $x_t{}'P_{t|t-1}x_t+\sigma^2$ es escalar, por lo que multiplicar por su inversa equivale a dividir por ella. La matriz de Error Cuadrático medio es:

$$E\left[\left(y_t - x_t'\beta - x_t'\hat{\xi}_{t|t-1}\right)^2 \mid x_t, Y_{t-1}\right] = x_t{}'P_{t|t-1}x_t + \sigma^2$$

y la función de verosimilitud:

$$\ln f(y_t \quad | \quad x_t, Y_{t-1}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\left(x_t{}'P_{t|t-1}x_t + \sigma^2\right) - \\ -\frac{1}{2}\sum_{t=1}^{T}\frac{\left(y_t - x_t'\beta - x_t'\hat{\xi}_{t|t-1}\right)^2}{x_t{}'P_{t|t-1}x_t + \sigma^2}$$

En el caso de que supongamos que el vector de coeficientes β_t sigue un proceso AR(p), tendríamos:

$$\xi_{t+1} = \begin{pmatrix} \phi_1 & \phi_2 & ... & \phi_{p-1} & \phi_p \\ I_k & 0 & ... & 0 & 0 \\ 0 & I_k & ... & 0 & 0 \\ . & . & & . & . \\ . & . & ... & . & . \\ . & . & & . & . \\ 0 & 0 & ... & I_k & 0 \end{pmatrix} \xi_t + \begin{pmatrix} v_{+1} \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}$$

2.4 Aplicación: CAPM con coeficientes cambiantes en el tiempo

Consideremos el modelo en excesos de rendimiento respecto del activo sin riesgo:

$$\begin{aligned} r_t &= \alpha_t + \beta_t r_{M,t} + w_t,\ w_t \sim N(0, \sigma_w^2) \\ \alpha_{t+1} &= \alpha_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2) \\ \beta_{t+1} &= \beta_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \end{aligned}$$

con $E(e_t.\eta_s) = E(e_t.\varepsilon_s) = E(\varepsilon_t.\eta_s) = 0, \forall t, s$. El supuesto de que la evolución de cada uno de los dos coeficientes el modelo sigue un camino aleatorio es una restricción que imponemos en el modelo.

El modelo puede escribirse:

$$r_t = \bar{\alpha} + (\alpha_t - \bar{\alpha}) + (\beta_t - \bar{\beta})r_{M,t} + \bar{\beta}r_{M,t} + w_t$$

y las variables de estado van a ser los dos coeficientes, en diferencias respecto a su media. Dichas medias son dos de los parametros a estimar. Los restantes parámetros son las varianzas de los 3 términos aleatorios: w_t, η_t, ε_t.

En este caso, tenemos la representación:

$$\begin{aligned}
\begin{pmatrix} \tilde{\alpha}_{t+1} \\ \tilde{\beta}_{t+1} \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\alpha}_t \\ \tilde{\beta}_t \end{pmatrix} + \begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \\
r_t &= \begin{pmatrix} 1 & r_{M,t} \end{pmatrix} \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix} + \begin{pmatrix} 1 & r_{M,t} \end{pmatrix} \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + w_t
\end{aligned}$$

por lo que en términos de la notación de Hamilton, tenemos:

$$\begin{aligned}
\xi_t &= \left(\tilde{\alpha}_t, \tilde{\beta}_t\right)'; \; F \equiv I_2; \; A' = (\bar{\alpha}, \bar{\beta}); \; x_t' = H_t' = (1, r_{M,t}); \\
v_t &\equiv (\eta_t, \varepsilon_t)'; \; Q = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}; \; R = \sigma_w^2
\end{aligned}$$

y el filtro de Kalman resulta:

$$\begin{aligned}
y_t - \hat{y}_{t|t-1} &= r_t - A'x_t - H_t'\xi_t \\
V_t &= H_t' P_{t|t-1} H_t + \sigma_w^2, \; escalar \; 1x1 \\
K_t &= \frac{P_{t|t-1} H_t}{V_t}, \; vector \; 2x1 \\
\begin{pmatrix} \hat{\alpha}_{t|t} \\ \hat{\beta}_{t|t} \end{pmatrix} &= \begin{pmatrix} \hat{\alpha}_{t|t-1} \\ \hat{\beta}_{t|t-1} \end{pmatrix} + K_t \left(y_t - \hat{y}_{t|t-1}\right) \\
P_{t|t} &= P_{t|t-1} - K_t H_t' P_{t|t-1} \\
\begin{pmatrix} \hat{\alpha}_{t+1|t} \\ \hat{\beta}_{t+1|t} \end{pmatrix} &= \begin{pmatrix} \hat{\alpha}_{t|t} \\ \hat{\beta}_{t|t} \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_{t|t-1} \\ \hat{\beta}_{t|t-1} \end{pmatrix} + K_t \left(y_t - \hat{y}_{t|t-1}\right) \\
P_{t+1|t} &= P_{t|t} + Q
\end{aligned}$$

Nótese que pr ser $F = I_2$, entonces $\xi_{t+1|t} = \xi_{t|t}$

2.5 Aplicación: Ratio de cobertura de mínimos cuadrados

Este ratio de cobertura no es sino la estimacion por minimos cuadrados ordinarios de la pendiente en una regresion lineal simple de las variaciones en el precio del contado sobre las variaciones en el precio del futuro. Vamos a permitir que dicho ratio varíe en el tiempo:

$$\Delta S_t = \alpha + \beta_t \Delta F_t + w_t$$

denotemos por $s_t = \Delta S_t - \overline{\Delta S}, \; f_t = \Delta F_t - \overline{\Delta F}$, y el modelo se transforma en:

$$s_t = \beta_t f_t + w_t \tag{4}$$

Suponemos que la evolución del ratio β_t está bien representada por un proceso AR(1):

$$\beta_t = \beta(1-\varphi) + \varphi\beta_{t-1} + v_t$$

o, lo que es equivalente, si denotamos $\tilde{\beta} = \beta_t - \beta$:

$$\tilde{\beta}_t = \varphi\tilde{\beta}_{t-1} + v_t \tag{5}$$

Si escribimos (4) y (5) en la forma de espacio de los estados, tenemos:

$$\begin{aligned}
\tilde{\beta}_{t+1} &= \varphi\tilde{\beta}_t + v_{t+1} \\
s_t &= \beta f_t + f_t\tilde{\beta}_t + w_t
\end{aligned}$$

siendo la primera la ecuacion de estado y la segunda la ecuación de observación.

En términos de la notación de Hamilton:

$$\begin{aligned}
\xi_t &= \tilde{\beta}_t;\ F \equiv \varphi;\ A' = \beta;\ x_t = f_t;\ H_t' = f_t; \\
Q &= \sigma_v^2;\ R = \sigma_w^2
\end{aligned}$$

El filtro de Kalman se inicializa con:

$$\begin{aligned}
\hat{\xi}_{1|0} &= E(\xi) \Rightarrow \hat{\xi}_{1|0} = E\left(\tilde{\beta}\right) = 0 \\
P_{1|0} &= E\left[(\xi - E(\xi))(\xi - E(\xi))\right] = \left[I_{r^2} - (F \otimes F)\right]^{-1} vec(Q) \Rightarrow \\
P_{1|0} &= var\left(\tilde{\beta}\right) = \frac{\sigma_v^2}{1-\varphi^2}
\end{aligned}$$

y a partir de las ecuaciones que describen la recursividad del filtro:

$$\begin{aligned}
\hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + \left[P_{t|t-1}H_t\left(H_t'P_{t|t-1}H_t + R\right)^{-1}\left(y_t - A'x_t - H_t'\hat{\xi}_{t|t-1}\right)\right] \\
P_{t|t} &= P_{t|t-1} - \left[P_{t|t-1}H_t\left(H_t'P_{t|t-1}H_t + R\right)^{-1}H_t'P_{t|t-1}\right]
\end{aligned}$$

aplicadas a este ejemplo, vamos obteniendo las ecuaciones de *filtrado*:

$$\begin{aligned}
\hat{\tilde{\beta}}_{t|t} &= \hat{\tilde{\beta}}_{t|t-1} + \left[P_{t|t-1}f_t\frac{1}{f_t^2P_{t|t-1} + \sigma_w^2}\left(s_t - \beta f_t - f_t\hat{\tilde{\beta}}_{t|t-1}\right)\right] \\
P_{t|t} &= P_{t|t-1} - \frac{P_{t|t-1}^2 f_t^2}{f_t^2P_{t|t-1} + \sigma_w^2}
\end{aligned}$$

que, denotando $\Omega_t = f_t^2P_{t|t-1} + \sigma_w^2$, pueden escribirse:

$$
\begin{aligned}
\widehat{\widetilde{\beta}}_{t|t} &= \widehat{\widetilde{\beta}}_{t|t-1} + \frac{P_{t|t-1} f_t}{\Omega_t}\left(s_t - \beta f_t - f_t \widehat{\widetilde{\beta}}_{t|t-1}\right) \\
P_{t|t} &= P_{t|t-1} - \frac{P_{t|t-1}^2 f_t^2}{\Omega_t} = P_{t|t-1}\left(1 - \frac{f_t^2}{\Omega_t} P_{t|t-1}\right)
\end{aligned}
$$

La ganancia del filtro de Kalman es en este caso:

$$
K_t = F P_{t|t-1} H_t (H_t' P_{t|t-1} H_t + R)^{-1} = \varphi P_{t|t-1} f_t \frac{1}{f_t^2 P_{t|t-1} + \sigma_w^2} = \frac{\varphi P_{t|t-1} f_t}{\Omega_t}
$$

mientras que las habituales ecuaciones de *predicción*:

$$
\begin{aligned}
\hat{\xi}_{t+1|t} &= F \hat{\xi}_{t|t} \\
P_{t+1|t} &= F P_{t|t-1} F' + \sigma_v^2
\end{aligned}
$$

se convierten en este ejemplo en:

$$
\begin{aligned}
\widehat{\widetilde{\beta}}_{t+1|t} &= \varphi \widehat{\widetilde{\beta}}_{t|t} = \varphi \widehat{\widetilde{\beta}}_{t|t-1} + \varphi \frac{P_{t|t-1} f_t}{\Omega_t} w_t = \varphi \widehat{\widetilde{\beta}}_{t|t-1} + K_t \left(s_t - \beta f_t - f_t \widehat{\widetilde{\beta}}_{t|t-1}\right) \\
P_{t+1|t} &= \varphi^2 P_{t|t} + \sigma_v^2 = \varphi^2 P_{t|t-1}\left(1 - \frac{f_t^2}{\Omega_t} P_{t|t-1}\right) + \sigma_v^2
\end{aligned}
$$

Por último, para formar la función log-verosimilitud bajo el supuesto de Normalidad, necesitamos:

$$
E(y_t \mid x_t, Y_{t-1}) = A' x_t + H_t' \hat{\xi}_{t|t-1}
$$

$$
E\left[\left(y_t - A' x_t - H_t' \hat{\xi}_{t|t-1}\right)\left(y_t - A' x_t - H_t' \hat{\xi}_{t|t-1}\right)' \mid x_t, Y_{t-1}\right] = H_t P_{t|t-1} H_t' + R
$$

que, denotando $S_{t-1} = \{s_{t-1}, s_{t-2}, ...\}$ se convierten en:

$$
\begin{aligned}
E(s_t \quad &| \quad f_t, S_{t-1}) = \beta f_t + f_t \widehat{\widetilde{\beta}}_{t|t-1} \\
E\left[\left(s_t - \beta f_t - f_t \widehat{\widetilde{\beta}}_{t|t-1}\right)^2 \mid f_t, S_{t-1}\right] &= f_t^2 P_{t|t-1} + \sigma_w^2
\end{aligned}
$$

y finalmente:

$$
\log Lik = \sum_{t=1}^{T} \ln f(s_t \mid f_t, S_{t-1}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \ln\left(f_t^2 P_{t|t-1} + \sigma_w^2\right) - \frac{1}{2}\sum_{t=1}^{T} \frac{\left(s_t - \beta f_t - f_t \widehat{\widetilde{\beta}}_{t|t-1}\right)^2}{f_t^2 P_{t|t-1} + \sigma_w^2}
$$