Nota Técnica N°4 / 2021
Representación de estado-espacio, filtrado y suavizado de Kalman: Nociones básicas y ejemplos de aplicaciones macro-financieras

Eduardo A. Corso y Constanza Matarrelli

Febrero 2021

BANCO CENTRAL DE LA REPÚBLICA ARGENTINA

ie | BCRA
INVESTIGACIONES ECONÓMICAS

Nota técnica Nº: 4

Representación de estado-espacio, filtrado y suavizado de Kalman: Nociones básicas y ejemplos de aplicaciones macro-financieras

Eduardo Ariel Corso[1] y Constanza Matarrelli[2]

Febrero 2021

Resumen: En este documento se presenta una introducción a las representaciones de estado-espacio y a las nociones de filtrado y suavizado de Kalman. Adicionalmente se describen algunos de sus usos en aplicaciones macro-financieras. El objetivo es facilitar a los lectores no especializados la interpretación y utilización de tales herramientas.

Introducción

En sus distintas variantes, los modelos macro-financieros buscan representar la interrelación dinámica entre la estructura temporal de la curva de rendimientos (en algunos casos la estructura temporal de las expectativas de inflación) y un conjunto de variables macroeconómicas, tales como el producto, la tasa de inflación y la tasa de depreciación.

Numerosos modelos macro-financieros como también modelos específicos de estructura temporal de tasas de curva de rendimientos y de expectativas de inflación, incluyen como parte de su representación la dinámica de variables no observables. Tomemos como ejemplo el ampliamente utilizado enfoque dinámico de curva de rendimientos de Nelson y Siegel, al que nos referiremos en la sección de aplicaciones. Este modelo explica la evolución de la estructura temporal de tasas de interés a partir de la dinámica de tres factores no observables, denominados nivel, pendiente y curvatura.

La principal característica de un modelo de estado-espacio —y de su creciente utilización en los modelos macro financieros— es que permite representar la evolución temporal de un conjunto de variables de estado (que pueden ser o no observables) y de un conjunto de variables observables que aportan información sobre las primeras. Esta representación nos permite inferir los momentos de las distribuciones condicionales de las variables de estado, a medida que las nuevas observaciones de las señales —variables observables— se van incorporando al conjunto de información.

En general, estaremos interesados en dos tipos de inferencia: filtrado, y suavizado. Como explicaremos más adelante, la diferencia entre ambos procesos radica en la característica del conjunto de información utilizado[3].

El resto del documento se estructura de la siguiente manera. En la sección I describiremos las principales características de un modelo de estado-espacio. En la sección II y III analizaremos los conceptos de filtrado y suavizado de Kalman, respectivamente. En la sección IV describiremos el rol del filtro de Kalman en la construcción de la función de verosimilitud, con el objeto de estimar los parámetros de una representación de estado-espacio. En la sección V abordamos el problema de observaciones faltantes en el conjunto de información de las variables observables. Por último, en la sección VI nos referiremos a algunas aplicaciones específicas en modelos macro financieros.

---

[1] Doctor en Economía. Subgerente de modelos Macro Financieros. Gerencia de Modelos y Pronósticos. Banco Central de la República Argentina. Mail: eduardo.corso@bcra.gob.ar

[2] Mag. En Finanzas. Analista de modelos Macro Financieros. Gerencia de Modelos y Pronósticos. Banco Central de la República Argentina. Mail: constanza.matarrelli@bcra.gob.ar

[3] Para un análisis exhaustivo de estos temas puede consultarse en Durbin y Koopman (2012) y Shumway y Stoffer (2006).

I. *Modelos de estado-espacio*

Representación básica

Los modelos de estado-espacio permiten modelar la secuencia temporal de un vector de variables observables $\{Y_\tau\}_{\tau=1}^{T}$ como dependientes de la evolución temporal de un vector de variables de estado que pueden ser o no observables $\{S_\tau\}_{\tau=1}^{T}$. A lo largo de este trabajo, asumiremos que el vector S_t está compuesto por variables no observables. La forma de representación básica de un modelo de estado-espacio Gaussiano puede escribirse como:

$$Y_t = Z_t \cdot S_t + e_t \qquad (1)$$

$$S_t = B_t \cdot S_{t-1} + \varepsilon_t \qquad (2)$$

donde

$$\begin{pmatrix} \varepsilon_t \\ e_t \end{pmatrix} \sim iid\ N\left(0, \begin{bmatrix} Q_t & 0 \\ 0 & R_t \end{bmatrix}\right) \qquad (3)$$

$$S_0 \sim N\left(S_{0|0}, P_{0|0}\right)$$

La primera ecuación, denominada de observación, describe la relación entre las variables observables Y_t y las variables de estado no observables S_t. Las variables observables pueden interpretarse como señales ruidosas de las variables de estado, siendo e_t el vector de errores de medición u observación Gaussianos. La segunda ecuación, denominada ecuación matricial de transición, describe el proceso estocástico que determina la dinámica de las variables de estado. La especificación de la matriz B_t y Q_t, determina la interrelación dinámica de las variables de estado.

Los modelos de estado-espacio son una herramienta notablemente flexible. El subíndice t en las matrices de parámetros de (1)-(3) indican que pueden ser cambiantes en el tiempo. También podemos incluir variables exógenas d_t y c_t y constantes a_t y u_t en las ecuaciones de observación y transición, respectivamente. De esta manera, una representación más amplia podría ser:

$$Y_t = a_t + Z_t \cdot S_t + D_t \cdot d_t + e_t$$

$$S_t = u_t + B_t \cdot S_{t-1} + C_t \cdot c_t + \varepsilon_t$$

Para simplificar la notación de nuestros desarrollos, a lo largo del trabajo utilizaremos la forma básica (1)-(2).

Estimación

En las aplicaciones prácticas de los modelos de estado-espacio, en general los valores paramétricos de las matrices Z_t, B_t ,R_t y Q_t son desconocidos, y deben ser estimados. Para inferir los valores de estos parámetros utilizamos la función de verosimilitud —o distribución muestral. Podemos escribir la función de verosimilitud conjunta del sistema (1)-(2) como el producto de las densidades condicionales $f(Y_t|\ Y^{t-1})$:

$$\begin{aligned} f(Y_1, Y_2, \ldots, Y_t;\ \delta) &= f(Y_1|Y_0, \delta) \cdot f(Y_2|Y_1, \delta)\ \ldots \\ &\quad \cdot f(Y_3|Y_2, Y_1, \delta) \cdot\ \ldots \\ &\quad f(Y_T|Y_{T-1}, \ldots, Y_1, \delta) \\ &= \textstyle\prod_{t=1}^{T} f(Y_t|Y^{t-1}, \delta) \qquad (4) \end{aligned}$$

Donde $y^0 = \phi$ e $Y^{t+1} = (Y_1, Y_2, \ldots, Y_{t-1})$ para $t \geq 2$. El logaritmo de la función de verosimilitud resulta entonces:

$$\ln L(Y^T, \delta) = \textstyle\sum_{t=1}^{T} \ln f(Y_t|Y^{t-1}, \delta) \qquad (5)$$

Donde $\delta = \{B^*, Z^*, Q^*, R^*\}$ es el conjunto de valores paramétricos del modelo.

Para construir la función de verosimilitud, necesitamos derivar las funciones de densidad condicional $f(Y_t|Y^{t-1}, \delta)$, para $t = 1, 2, \ldots, T$. Como veremos en la sección V, el filtro de Kalman nos permitirá obtener recursivamente tales distribuciones, y de esa manera construir la función de verosimilitud asociada con un modelo de estado-espacio. Además de su relevancia en el proceso de estimación de la estructura paramétrica, existe otra función del filtro de Kalman de notable utilidad, y a la que nos referiremos a continuación: la inferencia de los momentos condicionales de las variables del sistema.

Inferencia sobre las variables no observables

Una de las razones principales por la que estaremos interesados en trabajar con modelos de estado-espacio, es en realizar inferencia sobre los momentos de las variables no observables, en base a las realizaciones de las variables observables $\{Y_t\}$. Este cómputo puede realizarse con un procedimiento *forward* o *backward*. Por procedimiento *forward* nos referimos a que la información disponible hasta el momento t consiste en una secuencia de observaciones para $t = 1{:}\,t; \{Y_1, Y_2, \ldots, Y_t\} = Y^t$. En el caso del procedimiento *backward*, la información utilizada también incluye las observaciones entre $t = t{:}\,T$, donde T es el momento terminal de la muestra. En este caso, el conjunto de información resulta Y^T.

Nos referimos al cómputo *foward* como filtrado, y al *backward* como suavizado. Ambos constituyen dos criterios alternativos de inferencia sobre los momentos de las variables no observables. En resumen, lo que distingue al filtrado y al suavizado de Kalman es el conjunto de información utilizado, y la característica de la recursión. En el filtrado de Kalman el conjunto de información utilizada es Y^t, y el cómputo de los momentos de las distribuciones condicionales de las variables de estado se realiza recursivamente hacia adelante. En el suavizado de Kalman se utiliza como conjunto de información toda la muestra —es decir Y^T—, y el proceso de cómputo tendrá componentes recursivos *backward* y *forward*. Veremos ambos casos en detalle en las secciones III y IV.

En este trabajo utilizamos una representación básica de un modelo de estado-espacio Gaussiano, donde e_t y ε_t se encuentran normalmente distribuidos. La derivación del filtrado y suavizado de Kalman que presentamos a continuación se basa en ese supuesto. Sin embargo, el supuesto de normalidad sobre la distribución conjunta de Y_t y S_t no es necesario para tal derivación. Los procesos de filtrado y suavizado nos permiten inferir los momentos condicionales de las variables de estado aún en ausencia de este supuesto.

En Durbin y Koopmans (2012) se muestra que la derivación del filtrado y suavizado de Kalman sólo requiere del uso de propiedades elementales de la teoría de regresiones multivariadas. Bajo supuestos alternativos, los autores plantean cuatro lemas que representan en un sentido apropiado la regresión de S_t sobre Y_t, punto de partida para la derivación del filtro y suavizado de Kalman.

Los esquemas alternativos contemplados por los autores se resumen en un conjunto de cuatro lemas:

En su primer lema, asumiendo que S_t e Y_t se encuentran distribuidos normalmente de forma conjunta, se muestra que la distribución de S_t condicionada a Y_t es normal.

En el segundo lema, abandonando el supuesto de normalidad conjunta, los autores derivan la estimación lineal insesgada de mínima varianza de S_t dado Y_t.

En el tercer lema, asumiendo un enfoque bayesiano y nuevamente normalidad conjunta, derivan la distribución a posteriori de S_t dados los valores observados de Y_t

Finalmente, en el cuarto lema, manteniendo el enfoque bayesiano, los autores nuevamente eliminan el supuesto de normalidad conjunta y derivan una densidad cuasi posterior de S_t dado Y_t, cuyo vector de medias es lineal en Y_t y posee matriz de mínima varianza.

Debido a que el vector de medias y la matriz de covarianzas de la estimación de S_t dadas las realizaciones de Y_t son las mismas bajo las diversas estructuras de supuestos subyacentes tras cada uno de los cuatro lemas, Durbin y Koopmans muestran que la derivación del filtro es independiente no sólo del supuesto de normalidad conjunta entre las variables observables y de estado, sino también del enfoque empleado: clásico o bayesiano.

La derivación del filtro y suavizado de Kalman que presentamos a continuación es consistente con el conjunto de supuestos del lema 1 de Durbin y Koopmans. Una variante al lema 2 que también se encuentra frecuentemente en la literatura, se formula en términos de la minimización del error cuadrático medio –ver apéndice—.

II. Filtrado de Kalman

Bajo el supuesto de que S_0, e_t y ε_t se distribuyen normalmente, también lo hacen S_t e Y_t, dado que resultan de combinaciones lineales de variables normalmente distribuidas.

En el caso de la inferencia de las distribuciones condicionales mediante el proceso de filtrado, nuestro objetivo final será obtener la siguiente distribución:

$$S_t|Y^t \sim N\left(S_{t|t}, P_{t|t}\right) \tag{6}$$

Es decir, la distribución de S_t condicionada a Y^t.

Para ello requeriremos primero hallar las siguientes distribuciones condicionales:

$$S_t|Y^{t-1} \sim N\left(S_{t|t-1}, P_{t|t-1}\right) \tag{7}$$

$$Y_t|Y^{t-1} \sim N\left(Y_{t-1}, \Omega_{t|t-1}\right) \tag{8}$$

Adicionalmente debemos obtener la distribución conjunta de S_t e Y_t, condicional a Y^{t-1}.

Por la ecuación (2), los momentos condicionales de la distribución (7) resultan:

$$S_{s|t-1} = E[S_t|Y^{t-1}] = B_t \cdot S_{t-1|t-1} \tag{9}$$

$$\begin{aligned} P_{t|t-1} &= var(S_t|Y^{t-1}) \\ &= E\left[\left(S_t - S_{t|t-1}\right)\left(S_t - S_{t|t-1}\right)'|Y^{t-1}\right] \end{aligned}$$

$$= E\left[\begin{matrix}(B_t S_{t-1} + \varepsilon_t - B_t S_{t-1|t-1}) \\ \times (B_t S_{t-1} + \varepsilon_t - B_t S_{t-1|t-1})'\end{matrix} |Y^{t-1}\right]$$

$$= E\left[\begin{matrix}B_t(S_{t-1} - S_{t-1|t-1}) \\ \times (S_{t-1} - S_{t-1|t-1})' B_t{}'\end{matrix} |Y^{t-1}\right]$$

$$+ E[\varepsilon_t \varepsilon_t{}' |Y^{t-1}]$$

De manera que:

$$P_{t|t-1} = B_t P_{t-1|t-1} B_t' + Q_t \qquad (10)$$

De la ecuación (1), los momentos condicionales de la distribución (8) resultan:

$$Y_{t|t-1} = E[Y_t|Y^{t-1}]$$

$$= Z_t \cdot S_{t|t-1} \qquad (11)$$

$$\Omega_{t|t-1} = var(Y_t|Y^{t-1})$$

$$= E\left[(Y_t - Y_{t|t-1})(Y_t - Y_{t|t-1})' |Y^{t-1}\right]$$

$$= E\left[\begin{matrix}(Z_t S_t + e_t - Z_t S_{t|t-1}) \\ \times (Z_t S_t + e_t - Z_t S_{t|t-1})'\end{matrix} |Y^{t-1}\right]$$

$$= E\left[Z_t(S_t - S_{t|t-1})(S_t - S_{t|t-1})' Z_t{}' |Y^{t-1}\right]$$

$$+ E[e_t e_t |Y^{t-1}]$$

De manera que:

$$\Omega_{t|t-1} = Z_t P_{t|t-1} Z_t' + R_t \qquad (12)$$

Respecto a los momentos de la distribución conjunta condicionada a Y^{t-1}, sólo nos resta obtener la covarianza condicional:

$$cov(S_t, Y_t|Y^{t-1}) =$$

$$= E\left[(S_t - S_{t|t-1})(Y_t - Y_{t|t-1})' |Y^{t-1}\right]$$

$$= E\left[\begin{matrix}(S_t - S_{t|t-1}) \\ \times (Z_t S_t + e_t - Y_{t|t-1})'\end{matrix} |Y^{t-1}\right]$$

$$= E\left[\begin{pmatrix} S_t S_t' Z_t' + S_t e_t' - S_t Y_{t|t-1}' \\ -S_{t|t-1} S_t' Z_t{}' - S_{t|t-1} e_t' \\ +S_{t|t-1} Y_{t|t-1}{}' \end{pmatrix} |Y^{t-1}\right]$$

$$= E[S_t S_t{}' |Y^{t-1}] Z_t' - E\left[S_t Y_{t|t-1} |Y^{t-1}\right] - E\left[S_{t|t-1} S_t' |Y^{t-1}\right] Z_t{}' + E\left[S_{t|t-1} Y_{t|t-1}' |Y^{t-1}\right]$$

$$= (E[S_t S_t{}' |Y^{t-1}] Z_t{}' - E[S_t|Y^{t-1}] E[S_t{}' |Y^{t-1}]) Z_t{}'$$

$$= var(S_t|Y^{t-1}) Z_t'$$

De manera que[4]:

$$cov(S_t, Y_t|Y^{t-1}) = P_{t|t-1} Z_t{}' \qquad (13)$$

Utilizando las ecuaciones (11)-(13), podemos escribir la distribución condicional conjunta de S_t e Y_t respecto de Y^{t-1} como:

$$\begin{pmatrix} S_t \\ Y_t \end{pmatrix} |Y^{t-1} \sim N\left(\begin{bmatrix} S_{t|t-1} \\ Y_{t|t-1} \end{bmatrix} \begin{bmatrix} P_{t|t-1} & P_{t|t-1} Z_t{}' \\ Z_t P_{t|t-1} & \Omega_{t|t-1} \end{bmatrix} \right) \qquad (14)$$

Nótese que a partir de (14) estamos en condiciones de obtener la distribución de S_t condicional a Y_t e Y^{t-1}, lo que equivale a obtener la distribución de S_t condicionada a Y^t —objetivo del filtrado de Kalman— especificada en (6). Para ello, aplicamos el *teorema de la distribución condicional* para el caso de la distribución normal multivariada (14), de manera que:

$$S_{t|t} = E[S_t|Y^t]$$

$$= S_{t|t-1} + P_{t|t-1} Z_t{}' \Omega_{t|t-1}^{-1} (Y_t - Y_{t|t-1})$$

y

$$P_{t|t} = var(S_t|Y^t)$$

$$= P_{t|t-1} - P_{t|t-1} Z_t{}' \Omega_{t|t-1}^{-1} Z_t P_{t|t-1}$$

Definiendo a la *ganancia de Kalman* como:

$$K_t = P_{t|t-1} Z_t{}' \Omega_{t|t-1}^{-1}$$

resulta:

$$S_t|Y^t \sim$$

$$N\left(\begin{matrix} S_{t|t} = \\ S_{t|t-1} + K_{t|t-1}(Y_t - Y_{t|t-1}) \end{matrix} ; \begin{matrix} P_{t|t} = \\ (I_t - K_t Z_t) P_{t|t-1} \end{matrix} \right)$$

Que es la distribución que estábamos buscando.

En resumen, el procedimiento de filtrado para obtener recursivamente la distribución de S_t condicionada a Y^t puede separarse en dos pasos —o fases. En el primero, denominado de *predicción*, se obtienen los momentos de las distribuciones (7) y (8):

[4] Dado que: $var(x) = E[x^2] + E[x]^2$.

$$S_{t|t-1} = B_t S_{t-1|t-1} \quad (15)$$

$$P_{t|t-1} = B_t P_{t-1|t-1} B_t' + Q_t \quad (16)$$

$$Y_{t|t-1} = Z_t S_{t|t-1} \quad (17)$$

$$\Omega_{t|t-1} = Z_t P_{t|t-1} Z_t' + R_t \quad (18)$$

Nótese que los momentos obtenidos en (15)-(18) dependen de los valores pasados de los momentos de la distribución (6). Es decir, los valores pasados de los momentos de la distribución de la variable de estado en t una vez revelado Y_t. Este paso se realiza en la segunda fase del filtrado, denominado de *actualización*, que consiste en actualizar la *ganancia de Kalman* y los momentos de la distribución de S_t condicionado al conjunto de información Y^t:

$$K_t = P_{t|t-1} Z_t{}' \Omega_{t|t-1}^{-1} \quad (19)$$

$$S_{t|t} = S_{t|t-1} + K_t\left(Y_t - Y_{t|t-1}\right) \quad (20)$$

$$P_{t|t} = (I - K_t Z_t) P_{t|t-1} \quad (21)$$

De esta manera, partiendo de las condiciones iniciales $S_{0|0}$ y $P_{0|0}$, las ecuaciones (14)-(21) permiten obtener recursivamente las distribuciones (6)-(8) para cada momento *t*.

Detengámonos en el rol que cumple la ganancia de Kalman en este proceso. Para ello, veamos la intuición detrás de la matriz K_t. La misma puede pensarse como un cociente entre el "ruido" generado por la variable de estado —$P_{t|t-1}Z_t{}'$— y el ruido del conjunto de señales en el período t —$\Omega_{t|t-1}$—. Cuanto mas ruidosas sean las señales en t, mayores las varianzas de los errores de medición. Las mayores varianzas de R, se traducirán en mayores varianzas de Ω, y en una menor ganancia de Kalman. En efecto, cuanto más ruidosas la señales, menor corrección se realizará sobre la media y la varianza de S_t, una vez realizada la observación Y_t.

Por otro lado, cuanto más ruidoso sea el proceso de las variables de estado, mayor valor informativo tendrá la incorporación de señales para la inferencia de los momentos de S_t. Esto se reflejará en una mayor ganancia de Kalman.

III. Suavizado de Kalman

Como vimos en la sección previa, el filtrado de Kalman nos permite obtener los momentos de los estados latentes utilizando las realizaciones de las variables observables en el período $1\colon t$. Si bien el filtro de Kalman es suficiente para computar la verosimilitud de los datos, es subóptimo para estimar los momentos de las variables de estado. Para su estimación es deseable utilizar toda la información de la muestra $1\colon T$. Es decir, es deseable obtener el valor esperado del vector de estados, condicionado a la información Y^T. El suavizado de Kalman nos permite obtener tales momentos. A diferencia del filtrado de Kalman, en el que la recursión es hacia adelante, el suavizado de Kalman combina un proceso de recursión *hacia atrás* —partiendo de los valores terminales de las variables de estado—, con un proceso *hacia adelante*, en el que se utilizan las distribuciones del filtrado de Kalman. Es decir, el filtrado de Kalman constituye un insumo para el suavizado de Kalman. En este sentido se trata de un proceso que combina ambos tipos de recursiones.

Específicamente, a lo que nos referiremos como *suavizado de Kalman* en este trabajo es al algoritmo de suavizado de Rauch, Tung y Striebel (RTS)[5]. Como veremos, se trata de un algoritmo que itera directamente sobre las estimaciones realizadas por filtrado y suavizado.

Para su derivación, volveremos a utilizar el *teorema de la distribución condicional para una distribución normal multivariada*. Nuestro punto de partida es la distribución condicional conjunta de S_t y S_{t+1} respecto de Y^t. Ya hemos derivado anteriormente sus varianzas —ver ecuaciones (6) y (7) adaptando t—. Sólo nos resta obtener la expresión de la covarianza entre S_t y S_{t+1}, condicional a Y^t:

$$cov(S_t, S_{t-1}|Y^t) =$$

$$= cov(S_t, B_{t+1}S_t|Y^t) + cov(S_t, \varepsilon_{t+1}|Y^t)$$

$$= cov(S_t, B_{t+1}S_t|Y^t)$$

$$= E\left[\left(S_t - S_{t|t}\right)\left(B_{t+1}S_t - B_{t+1}S_{t|t}\right)'\lfloor Y^t\right]$$

$$= E\left[\left(S_t - S_{t|t}\right)\left(S_t - S_{t|t}\right)' B_{t+1}{}'\lfloor Y^t\right]$$

De manera que:

$$cov(S_t, S_{t-1}|Y^t) = P_{t|t} B_t{}'$$

La distribución condicional conjunta de S_t y S_{t+1} respecto de Y^t resulta:

[5] Rauch et al. (1965).

$$\begin{bmatrix} S_t \\ S_{t+1} \end{bmatrix} |Y^t \sim$$

$$N\left(\begin{bmatrix} S_{t|t} \\ S_{t+1|t} \end{bmatrix} \quad \begin{bmatrix} P_{t|t} & P_{t|t}B_{t+1}' \\ B_{t+1}P_{t|t} & P_{t+1|t} \end{bmatrix}\right) \quad (22)$$

Como nuestro objetivo es un algoritmo que realice una recursión *hacia atrás* de las variables de estado, estamos interesados en obtener la distribución de S_t, condicionada a S_{t+1} e Y^t. Utilizando el teorema de distribución condicional en (22) para una distribución normal multivariada, resulta[6]:

$$E[S_t|S_{t+1}, Y^t] =$$

$$S_{t|t} + P_{t|t}B_{t+1}'P_{t+1|t}^{-1}(S_{t+1} - S_{t+1|t}) \quad (23)$$

$$var(S_t|S_{t+1}, Y^t) =$$

$$P_{t|t} - P_{t|t}B_{t+1}'P_{t+1|t}^{-1}B_{t+1}P_{t|t} \quad (24)$$

Ya hemos obtenido la distribución de los estados condicionales a S_{t+1} e Y^t. Ahora debemos extender el conjunto de información incorporando las observaciones entre $t{:}T$. Para ello utilizaremos las siguientes propiedades de Markov:

$$E[S_t|S_{t+1}, Y^T] = E[S_t|S_{t+1}, Y^t]$$

y

$$var[S_t|S_{t+1}, Y^T] = var[S_t|S_{t+1}, Y^t]$$

Aplicadas en (23) y (24), obtenemos:

$$E[S_t|S_{t+1}, Y^T] =$$

$$S_{t|t} + P_{t|t}B_{t+1}'P_{t+1|t}^{-1}(S_{t+1} - S_{t+1|t}) \quad (25)$$

$$var(S_t|S_{t+1}, Y^T) =$$

$$P_{t|t} - P_{t|t}B_{t+1}'P_{t+1|t}^{-1}B_{t+1}P_{t|t} \quad (26)$$

Por último, aplicando propiedades de esperanza y varianza condicional[7], resulta:

$$\begin{aligned} S_{t|T} &= E[S_t|Y^T] \\ &= E[E[S_t|S_{t+1}, Y^T]|Y^T] \\ &= E\left[S_{t|t} + P_{t|t}B_{t+1}'P_{t+1|t}^{-1}(S_{t+1} - S_{t+1|t})|Y^T\right] \\ &= S_{t|t} + L_t(S_{t+1|T} - S_{t+1|t}) \end{aligned}$$

Donde:

$$L_t = P_{t|t}B_{t+1}'P_{t+1|t}^{-1}$$

Para la varianza condicional obtenemos:

$$\begin{aligned} P_{t|T} &= Var[S_t|Y^T] \\ &= Var[E[S_t|S_{t+1}, Y^T]|Y^T] \\ &\quad + E[Var[S_t|S_{t+1}, Y^T]|Y^T] \\ &= Var\left[S_{t|t} + L_t(S_{t+1|T} - S_{t+1|t})|Y^T\right] \\ &\quad + E\left[P_{t|t} - L_tP_{t+1|t}L_t'|Y^T\right] \\ &= L_tVar\left[S_{t|t}|Y^T\right]L_t' + P_{t|t} - L_tP_{t+1|t}L_t' \\ &= L_tP_{t+1|T}L_t' + P_{t|t} - L_tP_{t+1|t}L_t' \\ &= P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t' \end{aligned}$$

En resumen, el algoritmo RTS puede escribirse como:

$$S_{t|T} = S_{t|t} + L_t(S_{t+1|T} - S_{t+1|t}) \quad (27)$$

$$P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t' \quad (28)$$

Donde:

$$L_t = P_{t|t}B_{t+1}'P_{t+1|t}^{-1} \quad (29)$$

Con las condiciones iniciales dados por:

$$S_{T|T} = S_T$$

$$P_{T|T} = P_T$$

El algoritmo de suavizado puede interpretarse como una mejora retrospectiva de la estimación de la variable de estado S_t obtenida con el filtro de Kalman, utilizando los datos adicionales observados entre $t{:}T$.

En la recursión (27)-(28), las series $\{S_{t|T}\}$ y $\{P_{t|T}\}$ son el resultado del proceso de suavizado, mientras que $\{S_{t+1|t}\}$, $\{S_{t|t}\}$, $\{P_{t+1|t}\}$ y $\{P_{t|t}\}$ surgen del filtro de Kalman. En ese sentido el algoritmo combina una recursión *backward* con una *forward*.

Veamos la intuición de la ecuación (27). El vector de variables filtradas $S_{t|T}$ surge de corregir el valor actualizado a t del filtro de Kalman $S_{t|t}$, por el

[6] Puede resultar extraño expresar la distribución condicional a S_{t+1}, pero lo expresamos de esta forma por conveniencia. Posteriormente lo re expresaremos, utilizando la ley de las expectativas iteradas.

[7] Específicamente, para tres variables X, Y,Z:
$$E[X|Z] = E[E[X|Y,Z]|Z]$$
$$Var[X|Z] = Var[E[X|Y,Z]|Z] + E[Var[X|Y,Z]|Z]$$

término $L_t(S_{t+1|T} - S_{t+1|t})$. El paréntesis muestra la diferencia entre el valor del vector de estado obtenido en $t+1$ utilizando la información posterior al momento t —$S_{t+1|T}$—, y el valor del vector de estado pronosticado para $t+1$ en base a la información disponible hasta t —$S_{t+1|t}$—. La diferencia entre $S_{t+1|T} - S_{t+1|t}$ se debe a que el primer término consta de información adicional, entre $t{:}T$. Cuanto de esta información adicional se incorpore al valor actualizado del filtro de Kalman $S_{t|t}$ depende de la matriz L_t, que determinará la ganancia informativa.

La matriz L_t cumple un rol equivalente al de K_t, pero refiriéndose a la ganancia de información por disponer de la inferencia del vector de estado S_{t+1}, utilizando información hasta T —a diferencia de K_t, que refería a la ganancia informativa aportada por la señal Y_t—. En efecto, L_t se define como el ratio entre la matriz de covarianza entre S_{t+1} y S_t —$P_{t|t}B_{t+1}{}'$—, que puede interpretarse como la información que en la recursión *backward* aporta para S_t la inferencia de S_{t+1} y la matriz de varianzas y covarianzas de S_{t+1} —$P_{t+1|t}$—, que representa al ruido q de estado—. Este cociente constituye una medida de cuan informativa resulta el proceso recursivo S_{t+1} para la dinámica de S_t. Cuanto más informativa resulte la inferencia de la variable de estado en $t+1$ para la inferencia de sus momentos en t, mayor el valor de la ganancia L_t, y por ende mayor la corrección de $S_{t|t}$ para obtener $S_{t|T}$.

Los mismos razonamientos esbozados anteriormente se hacen extensivos para el caso de la matriz de covarianzas (28).

IV. Filtrado de Kalman y función de verosimilitud

Como hemos mencionado en la sección II, además de la inferencia, el filtro de Kalman nos permite evaluar la función de verosimilitud, y de esta manera brindarnos la herramienta necesaria para estimar la estructura paramétrica de la representación de estado-espacio (1)-(2). Para ello, partamos de un momento discrecional t. Podemos utilizar $S_{t-1|t-1}$ y $P_{t-1|t-1}$ junto con la ecuación de transición para obtener $S_{t|t-1}$ y $P_{t|t-1}$. En otras palabras, podemos utilizar la fase de predicción del filtrado de Kalman. Con $S_{t|t-1}$, calculamos el pronóstico $Y_{t|t-1} = Z_t S_{t|t-1}$. Una vez realizada la observación Y_t, estamos en condiciones de obtener el error de pronóstico:

$$u_t = Y_t - Y_{t|t-1}$$

$$= Y_t - Z_t S_{t|t-1}$$

$$= e_t + Z(S_t - S_{t|t-1})$$

Debido a que los errores son Gaussianos, resulta $u_t \sim N(0, Z_t P_{t|t-1} Z_t + R)$. Adicionalmente, como $Y_t = Y_{t|t-1} + u_t$, tenemos que:

$$f(Y_t|Y^{t-1}, \delta) = f(u_t, t)$$

Con el objeto de construir la función de verosimilitud, necesitamos obtener las distribuciones $f(Y_t|Y^{t-1}, \delta)$, para todo t. Dado $S_{t-1|t-1}$ y $P_{t-1|t-1}$, computamos $f(Y_t|Y^{t-1}, \delta)$ utilizando la función de densidad normal multivariada:

$$f(Y_t|Y^{t-1}, \delta) = \frac{1}{\sqrt{(2\pi)^t |Z_t P_{t|t-1} Z_t' + R|}} \times$$

$$exp\left(-\frac{u_t'(Z_t P_{t|t-1} Z_t' + R)^{-1} u_t}{2}\right)$$

Posteriormente, para computar $f(Y_{t+1}|Y^t, \delta)$ requerimos de $S_{t|t}$ y $P_{t|t}$. En otras palabras, necesitamos utilizar el paso de actualización del filtro de Kalman para incorporar la nueva información disponible en t. De esta manera, utilizando de forma iterativa las dos fases del filtro de Kalman podemos obtener sucesivamente todas las distribuciones condicionales, y de esa manera computar la función de verosimilitud $f(Y_1, Y_2, \dots, Y_t;\ \delta) = \prod_{t=1}^{T} f(Y_t|Y^{t-1}, \delta)$.

V. Observaciones faltantes

Por último, cabe señalar que una característica adicional de los modelos de estado-espacio es que permiten abordar el problema de series de tiempo observadas con irregularidad (es decir, con observaciones faltantes). Para un determinado t, puede suceder que para alguna o bien todas las variables del vector de observación no se disponga de valor observable. Abordaremos de manera breve ambos casos a continuación.

Supongamos que la secuencia temporal de las variables observables Y_t de dimensión $q_t \times 1$ es irregular y que, en cada t, podemos definir un vector de variables observadas Y_t^1 de dimensión $q_t^1 \times 1$ y un vector Y_t^2 de variables faltantes de dimensión $q_t^2 \times 1$, con $q_t^1 + q_t^2 = q$.

La ecuación de observación puede reescribirse según:

$$\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} \times S_t + \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix}$$

Donde $cov \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix} = \begin{bmatrix} R_t^{11} & R_t^{12} \\ R_t^{21} & R_t^{22} \end{bmatrix}$ corresponde a la matriz de varianzas y covarianzas de los errores de medición del grupo de variables observadas y el grupo de variables faltantes.

Si consideramos sólo el vector de variables observadas, podemos reemplazar el modelo estado espacio en (1)-(2) por:

$$Y_t^1 = Z_t^1 \cdot S_t + e_t^1$$

$$S_t = B_t \cdot S_{t-1} + \varepsilon_t$$

Donde ahora la ecuación de observación Y_t^1 tiene dimensión $q_t^1 \times 1$ en el momento t.

Con respecto al cambio en la dimensión del vector de observación, en la sección II puede observarse que la derivación del filtro de Kalman no presenta ninguna limitación que excluya dichos casos. Es decir, aunque la dimensión de Y_t varíe, las ecuaciones (16), (19), (20) y (21) que surgen de la derivación del filtro de Kalman se cumplen realizando las transformaciones o sustituciones correspondientes[8].

Sin embargo, puede resultar computacionalmente más fácil mantener la dimensión de la ecuación de observación e imputar valor 0 a los componentes faltantes. Por ejemplo, si en t no hay datos, las ecuaciones de actualización (19)-(21) se modifican tal que:

$$Y_t = \begin{bmatrix} Y_t^1 \\ 0 \end{bmatrix}$$

$$Z_t = \begin{bmatrix} Z_t^1 \\ 0 \end{bmatrix}$$

$$R_t = \begin{bmatrix} R_t^{11} & 0 \\ 0 & I_t^{22} \end{bmatrix}$$

Siendo I_t^{22} la matriz identidad cuadrada de dimensión igual a la del vector Y_t^2 de observaciones faltantes. Con dichos reemplazos, la estimación por máxima verosimilitud se lleva a cabo igual que en el caso de una serie de datos completa.

Si, en cambio, todo vector de observación Y_t contiene observaciones faltantes, entonces los componentes de la matriz K de la ganancia de Kalman adoptarán un valor 0. Entonces, para el conjunto Y_t de variables faltantes, las ecuaciones (20) y (21) devienen:

$$S_{t|t} = S_{t|t-1}$$

$$P_{t|t} = P_{t|t-1}$$

Por lo tanto, es posible estimar un modelo estado espacio tanto en el caso de observaciones faltantes para algunas de las variables del vector de observación, como en el caso de observaciones faltantes para todo el vector de variables Y_t.

VI. Ejemplos de aplicaciones macro-financieras

En esta sección describiremos dos ejemplos de representaciones de estado-espacio de modelos utilizados en la literatura macro-financiera. El primero, es el modelo dinámico de Nelson y Siegel, para la representación de la estructura temporal de curvas de rendimientos. El segundo, es el modelo de Arouba (2018) para la representación de la estructura temporal de expectativas de inflación.

Modelo dinámico de Nelson y Siegel de estructura temporal de tasas de interés

El modelo de Nelson y Siegel (1987) representa la estructura temporal de la curva cupón cero como la suma entre una constante y una función de Laguerre, que consiste en un polinomio multiplicado por un término de decaimiento exponencial. Siguiendo esta estructura, en la representación dinámica del modelo NS propuesta por Diebold y Li (2006), la evolución temporal del rendimiento de un bono con madurez τ, $y(\tau)$ está dada por la suma ponderada de tres factores no observables, denominados nivel L_t, pendiente S_t y curvatura C_t, de acuerdo con la siguiente expresión:

$$\begin{aligned} y_t(\tau) = \; & L_t - \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) S_t \\ & + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) C_t + e_t \end{aligned} \qquad (30)$$

El factor nivel se encuentra multiplicado por 1, mientras que los factores pendiente y curvatura se multiplican por dos funciones que describen el comportamiento de mediano y largo plazo de la estructura temporal, respectivamente.

El parámetro λ caracteriza la tasa de decaimiento exponencial de ambas funciones. Valores pequeños

[8] Lo mismo sucede si alguno o todos los parámetros de la ecuación de estado o de la ecuación de transición $(B_t,\ Z_t, Q_r,\ R_t)$ cambia dentro del período muestral.

de λ producen una caída suave, y permite ajustar mejor la curva a plazos largos, mientras que valores mayores producen una caída más pronunciada, y permiten ajustar mejor la curva en el corto plazo. Este parámetro también caracteriza el horizonte temporal en la que el factor curvatura alcanza su mayor ponderación. Nótese que tanto la función que multiplica al factor pendiente como al factor curvatura tienden a cero con $\tau \to \infty$, de manera que $\lim\ y_t(\tau \to \infty) \ \to L_t$. El término e_t constituye un término de error.

Existe una amplia literatura que respalda el uso empírico del modelo NS para representar estática y dinámicamente la curva de rendimientos —ver Diebold y Rudebusch (2013)—. En su versión dinámica, el modelo NS se especifica como una representación de estado-espacio, en el que la expresión (30) constituye la ecuación matricial de observación para τ variables observables correspondientes a los retornos cupón cero de títulos con vencimiento τ. El modelo se completa con la ecuación matricial de transición, que caracteriza la interacción dinámica entre los factores no observables L_t, S_t y C_t.

A modo de ejemplo, consideremos el caso de tres ecuaciones de observación, correspondientes a un horizonte temporal de 1, 6 y 12 meses vista. Las ecuaciones matriciales de observación y transición (1) y (2) resultan respectivamente:

$$\begin{bmatrix} \pi_{t\to t+1} \\ \pi_{t\to t+6} \\ \pi_{t\to t+12} \end{bmatrix} = \qquad (31)$$

$$\begin{bmatrix} 1 & \dfrac{1-e^{-\lambda}}{\lambda} & \dfrac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} \\ 1 & \dfrac{1-e^{-6\lambda}}{6\lambda} & \dfrac{1-e^{-6\lambda}}{6\lambda} - e^{-6\tau} \\ 1 & \dfrac{1-e^{-12\lambda}}{12\lambda} & \dfrac{1-e^{-12\tau}}{12\lambda} - e^{-12\tau} \end{bmatrix} \cdot \begin{bmatrix} L_t \\ S_t \\ C_t \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \\ e_t^3 \end{bmatrix}$$

$$\begin{bmatrix} L_{t+1} \\ S_{t+1} \\ C_{t+1} \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} \cdot \begin{bmatrix} L_t \\ S_t \\ C_t \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \varepsilon_t^3 \end{bmatrix} \qquad (32)$$

Con:

$$R = \begin{bmatrix} \sigma_{e_1}^2 & 0 & 0 \\ 0 & \sigma_{e_2}^2 & 0 \\ 0 & 0 & \sigma_{e_3}^2 \end{bmatrix} \quad \text{y} \quad Q = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_{12}}^2 & \sigma_{\varepsilon_{13}}^2 \\ \sigma_{\varepsilon_{21}}^2 & \sigma_{\varepsilon_2}^2 & \sigma_{\varepsilon_{23}}^2 \\ \sigma_{\varepsilon_{31}}^2 & \sigma_{\varepsilon_{32}}^2 & \sigma_{\varepsilon_3}^2 \end{bmatrix}$$

En la ecuación matricial (32), se representan los factores nivel, pendiente y curvatura como procesos AR(1), con shocks correlacionados (Matriz Q no restringida).

Modelo de Arouba de estructura temporal de expectativas de inflación

En su trabajo “Term Structures of Inflation Expectations and Real Interest Rates” de 2016, Borağan Arouba presenta una metodología de estimación de la estructura temporal de las expectativas de inflación en los Estados Unidos, en base a información de encuestas[9]. Su modelo se caracteriza principalmente por dos elementos. El primero, es utilizar el modelo dinámico de Nelson y Siegel (31) y (32) para representar las expectativas de inflación. Definiendo $\pi_t(\tau)$ a la expectativa de inflación entre t y $t+\tau$ —a las que denominaremos expectativas *spot*—, resulta:

$$\pi_{t\to t+\tau}(\tau) = \ L_t - \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) S_t + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) C_t + e_t \qquad (33)$$

De acuerdo con esta especificación, L_t captura la inflación esperada de largo plazo, S_t la diferencia entre la inflación de corto y de largo plazo y C_t la inflación esperada relativa entre el mediano y el corto y largo plazo.

El segundo elemento que caracteriza su modelo es la ampliación de las ecuaciones de observación para incorporar el relevamiento de expectativas de inflación en períodos futuros, es decir entre $t+\tau_1$ y $t+\tau_2$ —a las que denominaremos expectativas *forward*—. Para la derivación de este conjunto adicional de ecuaciones de observación, comenzaremos denotando $\Pi_{t\to t+s}$ a la tasa de inflación bruta anualizada entre t y $t+s$ como:

$$\Pi_{t\to t+s} = (\Pi_{t\to t+r})^{\frac{r}{s}} \times (\Pi_{t+r\to t+s})^{\frac{(r-s)}{s}}$$

Aplicando logaritmos, se obtiene:

[9] La Reserva Federal de Filadelfia publica mensualmente la evolución de la curva de expectativas de inflación calculada con la metodología de Arouba (2016). https://www.philadelphiafed.org/research-and-data/real-time-center/atsix

$$\Pi_{t\to t+s} = \frac{r}{s}\times(\pi_{t\to t+r}) + \frac{(r-s)}{s}\times(\pi_{t+r\to t+s})$$

Donde:

$$\pi_{t\to t+s} = \log\Pi_{t\to t+s} = \frac{12}{s-t}[\log P_{t+s} - \log P_t]$$

Es el logaritmo de la tasa de inflación anualizada entre el mes t y el mes s, expresado en tanto por uno. Reordenando términos, resulta:

$$\pi_{t+r\to t+s} = \frac{s}{r-s}\times\pi_{t\to t+s} - \frac{r}{s-r}\times\pi_{t\to t+r} \qquad (34)$$

Nótese que la ecuación (34) expresa la inflación forward $\pi_{t+r\to t+s}$ en términos de las inflaciones spot $\pi_{t\to t+s}$ y $\frac{r}{s-r}\times\pi_{t\to t+r}$. De esta manera, aplicando la ecuación (33) en (34) para los correspondientes horizontes temporales, obtenemos:

$$\pi_{t+\tau_1\to t+\tau_2} = L_t + \frac{e^{-\lambda\tau_1} - e^{-\lambda\tau_2}}{\lambda(\tau_2-\tau_1)}(C_t - S_t)$$
$$+\left(\frac{\tau_1 e^{-\lambda\tau_1} - \tau_2 e^{-\lambda\tau_2}}{\tau_2-\tau_1}\right)C_t + e_t \qquad (35)$$

La ecuación matricial (35) nos permite ampliar el conjunto de ecuaciones de observación (33) para incorporar relevamientos de tasas de inflación esperada en períodos futuros.

En el ejemplo que se plantea a continuación, se parte de la estructura básica del modelo de Arouba, y se amplía la representación de estado-espacio incorporando los efectos de las variaciones cambiarias realizadas sobre la estructura temporal de expectativas de inflación. Para ello incorporamos dos factores no observables adicionales. El primero, que represente los efectos de los shocks cambiarios mensuales sobre las expectativas de inflación. El segundo, que capte los efectos de la depreciación acumulada a lo largo del último año sobre las expectativas inflacionarias. Como señales ruidosas de estos factores no observables utilizaremos la tasa de depreciación mensual, y la tasa de depreciación acumulada en los últimos 12 meses, respectivamente. De esta manera, a las ecuaciones de observación (33) y (34) adicionaremos las siguientes:

$$\xi_{t-1\to t} = ED1_t + e_t \qquad (36)$$

$$\xi_{t-2\to t} = ED12_t + e_t \qquad (37)$$

Donde $\xi_{t-1\to t}$ es la tasa de depreciación realizada durante el último mes y $\xi_{t-12\to t}$ es la tasa de depreciación observada en los últimos 12 meses. Las variables $ED1_t$ y $ED12_t$ corresponden a los factores no observables, asociados con los impactos cambiarios de corto y mediano plazo.

Para el caso argentino, una fuente de información válida para las ecuaciones de observación (33) y (35) la constituye el relevamiento de expectativas de mercado (REM) que lleva a cabo el BCRA. Tomando como referencia esta encuesta, podríamos utilizar por ejemplo dos tasas de inflación entre t y $t+\tau$ —*i.e. spot*— y seis tasas esperadas futuras entre $t+\tau$ y $t+\tau+\theta$ —*i.e. forward*—. En el primer caso podría utilizarse la inflación esperada un mes y 12 meses vista —$\pi_{t\to t+1}$ y $\pi_{t\to t+12}$, mientras que las tasas futuras podrían corresponderse con valores de $\tau = 1$, 2, 3, 4 y 5 meses y $\theta = 1$ —$\pi_{t+1\to t+2}$, $\pi_{t+2\to t+3}$, $\pi_{t+3\to t+4}$, $\pi_{t+4\to t+5}$, $\pi_{t+5\to t+6}$— y un caso para $\tau = 12$ y $\theta = 12$, —$\pi_{t+12\to t+24}$.

El Cuadro 1 muestra la representación de estado-espacio resultante bajo estos supuestos.

Cuadro 1
Representación de estado-espacio

Ecuación matricial de observación

$$
\begin{bmatrix} \pi_{t\to t+1} \\ \pi_{t+1\to t+2} \\ \pi_{t+2\to t+3} \\ \pi_{t+3\to t+4} \\ \pi_{t+4\to t+5} \\ \pi_{t+5\to t+6} \\ \pi_{t\to t+12} \\ \pi_{t+12\to t+24} \\ \xi_{t-1\to t} \\ \xi_{t-12\to t} \end{bmatrix}
=
\begin{bmatrix}
1 & \frac{1-e^{-\lambda}}{\lambda} & \frac{1-e^{-\lambda}}{\lambda} - e^{-\lambda} & 0 & 0 \\
1 & \frac{e^{-\lambda}-e^{-2\lambda}}{\lambda} & \frac{e^{-\lambda}-e^{-2\lambda}}{\lambda} - e^{-\lambda} - e^{-2\lambda} & 0 & 0 \\
1 & \frac{e^{-2\lambda}-e^{-3\lambda}}{\lambda} & \frac{e^{-2\lambda}-e^{-3\lambda}}{\lambda} - e^{-2\lambda} - e^{-3\lambda} & 0 & 0 \\
1 & \frac{e^{-3\lambda}-e^{-4\lambda}}{\lambda} & \frac{e^{-3\lambda}-e^{-4\lambda}}{\lambda} - e^{-3\lambda} - e^{-4\lambda} & 0 & 0 \\
1 & \frac{e^{-4\lambda}-e^{-5\lambda}}{\lambda} & \frac{e^{-4\lambda}-e^{-5\lambda}}{\lambda} - e^{-4\lambda} - e^{-5\lambda} & 0 & 0 \\
1 & \frac{e^{-5\lambda}-e^{-6\lambda}}{\lambda} & \frac{e^{-5\lambda}-e^{-6\lambda}}{\lambda} - e^{-5\lambda} - e^{-6\lambda} & 0 & 0 \\
1 & \frac{1-e^{-12\lambda}}{12\lambda} & \frac{1-e^{-12\lambda}}{12\lambda} - e^{-12\lambda} & 0 & 0 \\
1 & \frac{e^{-12\lambda}-e^{-24\lambda}}{12\lambda} & \frac{e^{-12\lambda}-e^{-24\lambda}}{12\lambda} - e^{-12\lambda} - e^{-24\lambda} & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\cdot
\begin{bmatrix} L_t \\ S_t \\ C_t \\ Dc_t \\ Dl_t \end{bmatrix}
+
\begin{bmatrix} e_1 \\ e_2 \\ e_2 \\ e_2 \\ e_2 \\ e_2 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}
$$

Ecuación matricial de transición

$$
\begin{bmatrix} L_t \\ S_t \\ C_t \\ Dc_t \\ Dl_t \end{bmatrix}
=
\begin{bmatrix}
b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\
b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\
b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\
b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\
b_{51} & b_{52} & b_{53} & b_{54} & b_{55}
\end{bmatrix}
\cdot
\begin{bmatrix} L_{t-1} \\ S_{t-1} \\ C_{t-1} \\ Dc_{t-1} \\ Dl_{t-1} \end{bmatrix}
+
\begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \varepsilon_t^3 \\ \varepsilon_t^4 \\ \varepsilon_t^5 \end{bmatrix}
$$

Con

$$
R = \begin{bmatrix}
\sigma_{e1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma_{e2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma_{e2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma_{e2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_{e2}^2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma_{e2}^2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e1}^2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e2}^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e3}^2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e4}^2
\end{bmatrix}
\qquad
Q = \begin{bmatrix}
\sigma_{\varepsilon 1}^2 & 0 & 0 & 0 & 0 \\
0 & \sigma_{\varepsilon 2}^2 & 0 & 0 & 0 \\
0 & 0 & \sigma_{\varepsilon 3}^2 & 0 & 0 \\
0 & 0 & 0 & \sigma_{\varepsilon 4}^2 & 0 \\
0 & 0 & 0 & 0 & \sigma_{\varepsilon 5}^2
\end{bmatrix}
$$

Nótese que, en este caso, a diferencia del ejemplo del modelo de Nelson y Siegel, se ha especificado una matriz de transición no restringida, con shocks independientes sobre las variables de estado (matriz Q diagonal). Alternativamente, podrían especificarse procesos AR(1) para los factores de observables (matriz B diagonal) con shocks correlacionados (matriz Q no restringida). Como puede también observarse, se ha sumido que todas las expectativas de inflación spot comparten la misma varianza en su error de observación (matriz R). El mismo supuesto se asume para el caso de las expectativas forward.

VII. Conclusión

En esta nota técnica se presentó de manera resumida una descripción de la representación de estado-espacio y de los procedimientos de filtrado y suavizado de Kalman. Numerosos modelos macro-financieros incluyen variables no observables y la representación de estado-espacio permite inferir sus momentos a partir de la información que aportan variables que actúan como señales ruidosas. Con respecto a los procedimientos para obtener tales inferencias, hemos descrito el proceso de filtrado de Kalman (proceso forward) y de suavizado (que emplea una recursión que combina componentes backward y forward) dependiendo del conjunto de información utilizado. El filtro de Kalman utiliza sólo la información disponible hasta el momento t para obtener una inferencia de las *distribuciones condicionales* de la media y de las matrices de varianzas y covarianzas de las variables no observables. Por su parte, el suavizado de Kalman amplía el conjunto de información, utilizando toda la muestra disponible, hasta el momento terminal T.

En el presente trabajo detallamos las ecuaciones que caracterizan ambos procedimientos con el objetivo de facilitar su interpretación y su aplicación en modelos macro-financieros que busquen representar la interrelación dinámica entre variables observables y no observables.

Adicionalmente, hemos referido al uso del filtro de Kalman para la estimación de la estructura paramétrica de las representaciones de estado-espacio a través de la evaluación de la función de verosimilitud.

Por último, a modo de ejemplo, propusimos dos aplicaciones de representaciones de estado-espacio en modelos macro-financieros. La primera, el modelo dinámico de Nelson y Siegel para la representación de la estructura temporal de curva de rendimientos. En segundo lugar, una ampliación del modelo básico de expectativas de inflación de Arouba, al que le incorporamos dos factores no observables que conceptualmente buscan captar los efectos de los shocks cambiarios realizados en el corto y mediano plazo, sobre las expectativas de inflación.

VIII. Referencias

Aruoba, S.B. (2016), “Term Structures of Inflation Expectations and Real Interest Rates”, Federal Reserve Bank of Philadelphia Working Paper 16-09/R.

Aruoba, S.B., F.X. Diebold, and C. Scotti (2009), “Real-Time Measurement of Business Conditions”, Journal of Business and Economic Statistics, 27(4), 417-427.

Diebold, F.X., and C. Li (2006), “Forecasting the Term Structure of Government Bond Yields”, Journal of Econometrics, 130(2), 337-364.

Diebold, F.X., and G.D. Rudebusch (2013), Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach, Princeton University Press.

Diebold, F.X., G.D. Rudebusch, and S.B. Aruoba (2006), “The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach”, Journal of Econometrics, 131(1-2), 309-338.

Durbin and Koopman (2012). “Times Series Analysis by State Space Methods”. Second Edition. Oxford University Press.

Hamilton, J.D (1994). “Time Series Analysis”, Princeton University Press.

Kollo, Tonu and Dietrich von Rosen (2005), “Advanced Multivariate Statistics with Matrices”, *Mathematics and Its Applications*. Springer.

Nelson, C.R., and A.F. Siegel (1987), “Parsimonious Modeling of Yield Curves”, Journal of Business, 60(4), 473-489.

Shumway, Robert H. and David S. Stoffer (2006). “Time Series Analysis and Its Applications: With R Examples”. Springer Texts in Statistics. Second Edition.

XIX. Anexo

Como se ha mencionado en la sección I, la derivación del filtro de Kalman que se presenta en este trabajo asume que $S_o, e_t\ y\ \varepsilon_t$ se distribuyen normalmente. En ese caso, para el cómputo de las medias y matrices de varianzas y covarianzas condicionales de las variables de estado, se utiliza el *teorema de la distribución condicional para una distribución normal multivariada*. En términos de los casos alternativos propuestos por Durbin y

Koopmans, tal derivación se encuentra en línea con los supuestos del Lema 1.

En el presente anexo se deriva el filtro de Kalman prescindiendo del supuesto de normalidad, lo que se corresponde con el Lema 2 de Durbin y Koopmans. En este caso, el filtro de Kalman puede ser pensado como un algoritmo que computa pronósticos de mínimos cuadrados lineales del vector de estado, sobre la base de la información observada en t —Ver Hamilton (1994).

Desde esta perspectiva, el filtro óptimo surge de la minimización del error cuadrático medio. En un modelo estado espacio del tipo:

$$Y_t = ZS_t + e_t \tag{38}$$

$$S_t = B_t S_{t-1} + \varepsilon_t \tag{39}$$

Donde B_t es la matriz de transición y se asume estacionaria, e_t es ruido blanco y no está correlacionado con ε_t, $E[e_t \varepsilon_t] = 0$. Las covarianzas de los errores se asumen estacionarias y están dadas por:

$$E[e_t e_t'] = R \tag{40}$$

$$E[\varepsilon_t \varepsilon_t'] = Q \tag{41}$$

La diferencia entre S_t y su valor estimado $\hat{S}_{t|t-1}$ es el error de pronóstico:

$$w_t = S_t - \hat{S}_{t|t}$$

Con cada error de pronóstico se asocia una matriz de varianzas y covarianzas:

$$\begin{aligned} P_{t|t} &= E[w_t w_t'] \\ &= E\left[(S_t - \hat{S}_{t|t}) \times (S_t - \hat{S}_{t|t})'\right] \end{aligned} \tag{42}$$

Como $\hat{S}_{t|t-1}$ es el estimador de la variable de estado con la información disponible hasta t-1, una vez revelada la información en t podemos escribir la ecuación de actualización de S_t como:

$$\hat{S}_{t|t} = \hat{S}_{t|t-1} + K_t(y_t - Z\hat{S}_{t|t-1}) \tag{43}$$

Sustituyendo (38) en (43):

$$\hat{S}_{t|t} = \hat{S}_{t|t-1} + K_t(ZS_t + e_t - Z\hat{S}_{t|t-1}) \tag{44}$$

Luego, sustituyendo el valor de $\hat{S}_{t|t}$ de (44) en la ecuación (42):

$$\begin{aligned} P_{t|t} = E\big[&(S_t - \hat{S}_{t|t-1} + K_t(ZS_t + e_t \\ &- Z\hat{S}_{t|t-1})) \\ &\times (S_t - \hat{S}_{t|t-1} + K_t(ZS_t + e_t \\ &- Z\hat{S}_{t|t-1}))'\big] \end{aligned}$$

Reordenando:

$$\begin{aligned} P_{t|t} = E\big[&\big(S_t - \hat{S}_{t|t-1} - K_t Z(S_t - \hat{S}_{t|t-1}) \\ &- K_t e_t\big) \\ &\times (S_t - \hat{S}_{t|t-1} - K_t Z(S_t - \hat{S}_{t|t-1}) \\ &- K_t e_t)'\big] \end{aligned}$$

$$\begin{aligned} P_{t|t} = E\big[&\big(S_t - \hat{S}_{t|t-1} - K_t Z(S_t - \hat{S}_{t|t-1}) \\ &- K_t e_t\big) \\ &\times (S_t - \hat{S}_{t|t-1} - K_t Z(S_t - \hat{S}_{t|t-1}) \\ &- K_t e_t)'\big] \end{aligned}$$

$$\begin{aligned} P_{t|t} = E\big[&\big((I - K_t Z)(S_t - \hat{S}_{t|t-1}) \\ &- K_t e_t\big) \\ &\times \big((I - K_t Z)(S_t - \hat{S}_{t|t-1}) \\ &- K_t e_t)'\big] \end{aligned}$$

Donde $(S_t - \hat{S}_{t|t-1})$ es el error de predicción en t-1 y no está correlacionado con e_t, de manera que:

$$\begin{aligned} P_{t|t} = (I - K_t Z)\, E\big[(S_t - \hat{S}_{t|t-1})(S_t - \hat{S}_{t|t-1})'\big](I \\ - K_t Z)' + K_t E[e_t e_t'] K_t' \end{aligned} \tag{45}$$

Sustituyendo (40) en (45) se obtiene la siguiente expresión para la matriz de covarianzas de los errores:

$$P_{t|t} = (I - K_t Z) P_{t|t-1} (I - K_t Z)' + K_t R K_t' \tag{46}$$

La traza de la matriz $P_{t|t}$ es el error cuadrático medio (ECM). La minimización del ECM resulta de minimizar la traza de $P_{t|t}$. Para ello se diferencia primero con respecto a K_t y se iguala a cero para encontrar la condición de minimización.

Comenzamos distribuyendo los términos en (46) y re-expresamos:

$$\begin{aligned} P_{t|t} = P_{t|t-1} - K_t Z P_{t|t-1} - P_{t|t-1} Z' K_t' + \\ K_t\big[Z P_{t|t-1} Z' + R\big] K_t' \end{aligned} \tag{47}$$

Como la traza de una matriz es igual a la traza de su traspuesta, entonces:

$$\begin{aligned} Tr\big[P_{t|t}\big] = Tr\big[P_{t|t-1}\big] - 2Tr\big[K_t Z P_{t|t-1}\big] + \\ Tr\big[K_t (Z P_{t|t-1} Z') K_t'\big] \end{aligned} \tag{48}$$

Derivamos con respecto a K_t e igualamos a cero:

$$\frac{\partial Tr[P_{t|t}]}{\partial K_t} = -2(ZP_{t|t-1})' + 2K_t(ZP_{t|t-1}Z' + R) = 0$$

Despejando K_t encontramos la expresión de la ganancia de Kalman asociada a la condición de minimización del error cuadrático medio:

$$K_t = P_{t|t-1}Z' \times (ZP_{t|t-1}Z' + R)^{-1} \quad (49)$$

Sustituyendo (49) en (47):

$$P_{t|t} = P_{t|t-1} - K_t Z P_{t|t-1}$$

$$= (I - K_t Z)P_{t|t-1} \quad (50)$$

Las ecuaciones (43), (49) y (50) son respectivamente análogas a las ecuaciones (20), (19) y (21) obtenidas a partir del supuesto de normalidad en el presente trabajo, y permiten obtener a una estimación de S_t a partir de condiciones iniciales.

Luego, de manera recursiva se puede obtener la variable de estado y la matriz de covarianzas del error de pronóstico para los períodos siguientes. Adaptando los subíndices temporales según corresponda, obtenemos:

$$\hat{S}_{t|t-1} = BS_{t-1|t-1} \quad (51)$$

$$P_{t|t-1} = E[w_{t|t-1}w'_{t|t-1}] \quad (52)$$

Siendo:

$$w_{t|t-1} = S_t - \hat{S}_{t|t-1} \quad (53)$$

Reemplazamos (39) y (51) en (53) y obtenemos una expresión para $w_{t|t-1}$:

$$w_{t|t-1} = BS_{t-1} + \varepsilon_t - BS_{t-1|t-1}$$

$$= Bw_{t-1} + \varepsilon_t \quad (54)$$

Sustituyendo la expresión (54) en (52) y recordando que el error de predicción en t-1 no está correlacionado con ε_t :

$$P_{t|t-1} = E[(Bw_{t-1} + \varepsilon_t)(Bw_{t-1} + \varepsilon_t)']$$

$$= E[(Bw_{t-1})(Bw_{t-1})'] + E[\varepsilon_t {\varepsilon_t}']$$

$$= BP_{t-1|t-1}B' + E[\varepsilon_t {\varepsilon_t}'] \quad (55)$$

Por último, reemplazamos (41) en (55) y obtenemos la expresión que estábamos buscando:

$$P_{t|t-1} = BP_{t-1|t-1}B' + Q \quad (56)$$